NASA TECHNICAL MEMORANDUM

NASA TM X-72696

NASA TM X-72696

NOISE AND SPEECH INTERFERENCE

PROCEEDINGS OF MINISYMPOSIUM

Edited by

William T. Shepherd

(NASA-TM-X-72696) NOISE AND SPEECH INTERFERENCE: PROCEEDINGS OF MINISYMPOSIUM (NASA) 230 p HC $7.50 CSCL 05E

N75-31731

Unclas
G3/53 35257

This informal documentation medium is used to provide accelerated or special release of technical information to selected users. The contents may not meet NASA formal editing and publication standards, may be revised, or may be incorporated in another publication.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LANGLEY RESEARCH CENTER, HAMPTON, VIRGINIA 23665

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM X-72696		

4. Title and Subtitle Noise and Speech Interference - Proceedings of Minisymposium	5. Report Date September 1975
	6. Performing Organization Code 2630
7. Author(s) Edited by: William T. Shepherd	8. Performing Organization Report No. TM X-72696
	10. Work Unit No. 504-09-11-01
9. Performing Organization Name and Address NASA-Langley Research Center Hampton, VA 23665	11. Contract or Grant No. N/A
	13. Type of Report and Period Covered Technical Memorandum
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The following papers are included:

1. Speech Interference Assessment - An Overview and Some Suggestions for the Future
2. A Proposed Method for Measuring Annoyance Due to Speech Interference by Noise
3. Annoyance of Time-Varying Noise While Listening to Speech
4. Effects of Three Activities on Annoyance Responses to Aircraft Sounds
5. Some Aspects of Interference Between Speech and Noise
6. Units for the Assessment of Nuisance Due to Traffic Noise in a Speech Environment
7. A New Look at Multiple Word Test Items for Evaluating Talkers, Listeners and Communication Systems
8. Tri-Word Intelligibility Test for Assessing Interword Interference
9. Is Speech Intelligibility Enough
10. Objectivity - Subjectivity Continuum in Intelligibility Testing

17. Key Words (Suggested by Author(s)) (STAR category underlined)	18. Distribution Statement
Speech Interference, Noise, Aircraft Sounds, Intelligibility Testing 71	Unclassified Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	228	$7.50

*Available from { The National Technical Information Service, Springfield, Virginia 22151
STIF/NASA Scientific and Technical Information Facility, P.O. Box 33, College Park, MD 20740

EDITOR'S INTRODUCTION

The meeting represented by these proceedings was predicated on the judgment that speech interference can be a problem. A problem to many people; to a telephone engineer; to a teacher in a classroom; to an airplane pilot communicating with an air traffic controller; or to an individual trying to talk to another individual across a room. All of these people face similar difficulties when speech interference results from noise masking or some form of filtering or some other form of disturbance which affects the quality of their communication, or of the communication situation they attempt to provide for others.

Many undesirable secondary effects result from speech interference and exacerbate the simple problem implied by a report of a reduction in the number of verbal units transferred. Effects such as reduced safety, reduced amount of knowledge transmitted or an increase in the hard to define feeling of annoyance that comes from frustration of a desire to communicate effectively.

Given that speech interference problems exist in many contexts, one logically ponders their solutions. Kernel to the solution of any problem is the definition of its limits; this implies the measurement or observation of "how much" or "what kind" or other similar qualities and quantities.

The papers contained in these proceedings address such questions; questions regarding the kinds of measurement devices or techniques to use in assessing speech interference effects; questions regarding the units to observe or measure in research; or questions regarding entirely new ideas as to what are the components of speech interference.

Considerable discussion was devoted to the annoyance aspect of speech interfering noise, an area of concern to NASA-Langley researchers. Of particular interest is the question of the usefulness of existing intelligibility assessment tools such as AI or the MRT in the annoyance domain. In this case it is important to know first if such devices can be used to predict intelligibility under various conditions, and if they can, can they then be used to reliably predict annoyance for a known or predictable speech interference situation. If the existing intelligibility devices are not adequate in the annoyance context, the question remains as to what new types of assessment devices or measuring units or techniques should be used to evaluate the speech interference/annoyance situation. A number of the conference participants presented information pertinent to these areas. A very real question concerns just what are speech interference annoyance and dissatisfaction related to? Are they related simply to a reduced number of verbal units transferred or is the picture more complex, including perhaps consideration of variation of listener or speaker effort or of listener response time, variations in all of which may occur in the face of perfect intelligibility? Dr. Dave Nagel has examined these and other possibilities in his paper.

The order of papers presented in the following pages is the order in which individuals presented them at the conference. This order was based on a quasi-random selection procedure and no assertion by the editor of relative importance of papers is intended.

LIST OF PAPERS

1. "Speech Interference Assessment - An Overview and Some Suggestions for the Future" - William Shepherd, NASA-LaRC

2. "A proposed Method for Measuring Annoyance Due to Speech Interference by Noise" - John Molino, National Bureau of Standards

3. "Annoyance of Time-Varying Noise While Listening to Speech" - Karl Pearsons, B.B & N.

4. "Effects of Three Activities on Annoyance Responses to Aircraft Sounds" Walter Gunn, NASA - LaRC

5. "Some Aspects of Interference Between Speech and Noise." - John Webster, Naval Electronics Laboratory Center.

6. "Units for the Assessment of Nuisance Due to Traffic Noise in a Speech Environment" - Chris Rice, ISVR, U.K.

7. "A New Look at Multiple Word Test Items for Evaluating Talkers, Listeners and Communication Systems" - Carl Williams, James Mosko, James Greene, Naval Aerospace Medical Research Laboratory, Pensacola, FL (Paper presented by James Mosko)

8. "Tri-Word Intelligibility Test for Assessing Interword Interference" RussellSergeant, Hunter College

9. "Is Speech Intelligibility Enough?" - David Nagel, NASA -ARC

10. "Objectivity - Subjectivity Continuum in Intelligibility Testing"

G. C. Tolhurst, University of Massachusetts.

SPEECH INTERFERENCE ASSESSMENT - AN OVERVIEW AND SOME SUGGESTIONS FOR THE FUTURE

By

William T. Shepherd

NASA-Langley Research Center
Hampton, VA.

SUMMARY

This paper considers factors important to the assessment of speech interference and effects of speech interference in a number of contexts. The principal focus is on speech interference effects resulting from noise masking, particularly that engendered by aircraft flyover noise. A discussion of various speech interference assessment devices is given along with an evaluation of their limitations when used to estimate other forms of human response. A proposed new approach is suggested which embodies evaluation of other factors besides amount of information transferred and reported annoyance.

When we talk about speech interference it is possible to consider it as resulting from at least three causes, viz, filtering, articulation distortion, and noise masking. Filtering as it applies to speech interference is generally related to electrical/electronic communication devices. Filtering is really a process of bandwidth reduction resulting from attenuation of certain component frequencies of a complex signal. As a result, a reduced amount of information reaches listener's ears. This is a rather general statement and it is not meant to suggest that the information reduction necessarily results in a reduced understanding of the speech material being transmitted over the bandwidth limiting device. One of the most interesting findings from the study of effects of filtering or bandwidth modification was the indication that bandwidth and intensity are complementary. That is if speech intelligibility is reduced as a result of reducing the bandwidth of a communication device, then intelligibility may be restored to a certain degree by increasing the intensity of the signal.

Regarding articulation distortion, Harris published an interesting study some years ago relating speech interference to the articulation distortion produced in a speaker who was simultaneously eating a sandwich. We here at this meeting are most interested in speech interference resulting from noise masking. I would like to preface my discussion in this area with a few obvious points, familiar to everyone here, for the purpose of setting the stage for my later remarks. To begin with, speech energy is mostly low frequency energy as shown in the long term spectrum for adult male speech in figure 1. Consonant speech sounds are typically found at the high frequency, low intensity end of this spectrum. It has been shown by many investigators that consonants are the information bearing elements of speech due to their quantitative preponderance and dynamic nature. Consonant sounds are generally easier to mask than vowel sounds due to their lower intensity and upward spread of masking effects. This latter factor was shown by Stevens et al who found that tones in the region of 0.3 - 0.5 Khz are the most effective speech masking sounds. Noise bands centered generally at frequencies lower than 1000 Hz are more effective speech maskers than higher frequency noise bands. Figure 1 shows the sound spectrum for a typical jet aircraft at a given instant during a flyover. As shown here, aircraft noise has lots of energy in the low frequency, high speech masking region. The aircraft noise spectrum looks very much like the speech spectrum. This suggests that aircraft noise presents a definite speech masking problem.

If we consider only aircraft noise for the moment it can be said that it has become increasingly clear that new approaches are needed to answer the different kinds of questions related to human response to this noise source. Simply measuring intelligibility for some idealized laboratory situation or inferring intelligibility using Articulation Index for example, is not enough. Such procedures are fine for telling us that telephone "A" is a better speech transmission device than telephone "B", but we need to know more than this level of information. Given the interest in community response to aircraft noise, we want to know something about the annoyance that accompanies realistic exposures to speech interfering aircraft noises. This requirement clearly establishes the need for more realistic speech test conditions and for more accurate and precise means for quantifying speech interference and subjective response.

Let's look at some of the existing speech masking evaluation procedures. A lot of early speech research was concerned with phonetics, pronunciation, aural discrimination etc. Actually, this early work was more attuned to the kinds of things we want to do in assessing speech interference. Speech interference assessment is at least partly concerned with phonetics, pronunciation and discrimination too. It was a natural scientific progression to attempt to quantify these early observations of speech processes such that a given speech interference situation might be described most efficiently say by a single number such as a test score. In making these quantification attempts, more was involved than devising a laboratory curiosity. There were immediate practical advantages related to commercial, wartime, linguistic and other interests. For example, telephone and communication hardware oriented companies had a commercial interest in such procedures since they were concerned with developing more viable speech communication devices. Wartime needs made it imperative to devise vocabularies that were least sensitive to interference. Linguistic and anthropological researchers use the artifact of speech production and perception to make inferences about differences between man and other species. Closely allied here are the needs of psychologists to determine various psychophysical thresholds related to audition. Still other needs for quantifying speech processes concern the treatment of disordered speech and hearing.

At any rate over a period of fairly recent years, a number of articulation tests have been devised and used for a number of purposes. Tests such as PB word tests, rhyme tests of Fairbanks and House which have been used extensively primarily by those interested in military communication. Sentence tests have been used more often in audiometry than in assessment of noise masking effects. There have been problems with variability of performance by groups of people on sentence tests. As shown by Rogers at the University of Connecticut, people vary widely in their abilities to predict words in sentences under marginal listening conditions, and consequently there is a large range in scores on these tests under simulative noise masking conditions. Sentence test scores are fairly uniform under high signal to noise ratio conditions, but these conditions may not reflect the masking situation that frequently occurs in airport communities. Other types of tests that may be useful in the assessment of noise effects are the content report tests devised by Ullrich and Williams.

Most of these previously described tests are similar in that controlled speech material is presented to listeners who respond in some way such as c... off a word on a list or writing in the words of a sentence.

It is possible to identify at least four factors that are important in speech interference testing, viz: The people involved (speakers, listeners); test materials (words, sentences, and by inference, the mode of listener response); equipment (earphones, loudspeakers, microphones, test rooms); and the noise or distortion affecting the speech transmission (white noise, aircraft noise, filtering). This list suggests that a lot of work may be involved in speech interference evaluation. Many others thought so and looked for ways to reduce or eliminate the need for speech interference testing. The most prominent result of these searches is Articulation Index. With AI, all that is needed is to measure speech and noise levels and make some calculations and corrections to produce an index that rates telephones, radios and other communication devices with respect to one another. AI can be useful for such evaluations particularly earphone type equipment, but certain cautions are in order regarding some of the underlying assumptions of AI. These cautions relate to the assumptions of independently contributing frequency bands and single curves relating intelligibility and AI. Bowman has presented evidence in the Journal of Sound and Vibration suggesting that neither of these assumptions may be tenable and we have some experimental results from work here at Langley hinting that the latter assumption may not be tenable. I will present these data shortly. Apart from the typical communications hardware evaluation task, there have been suggestions that AI can be used to evaluate other communication situations dealing with free field cases such as loudspeaker presentation and face-to-face communication in various types of enclosures. In these cases, the room is essentially being rated as part of the communication system. This presents a more difficult experimental situation adding effects which are harder to assess and embody as corrections which can be applied uniformly in such a device as AI. Also there have been suggestions that intelligibility scores are predictable based on a knowledge of AI. These claims are usually hedged with warnings that the scores depend on the particular talker/listener crews, their training etc. Given these warnings, it is difficult to tell what the prediction claims really mean since the results obtained from one crew to another will almost certainly be different, and it is not possible to objectively assert the superiority of one or more of a number of identically trained crews composed of similar members.

The limitations of AI in the freefield situation are suggested from results of an experiment performed here at Langley. We set out to rate the speech efficacy of a classroom using AI and PB word intelligibility tests. A speaker of general American English and a five man listener crew were trained in accordance with the instructions given in U. S. Standard S3.2. The PB word lists used were also taken from this standard. The ambient masking noise in the room was provided by two window air conditioning units. The classroom layout is shown in figure 2. Three noise conditions were evaluated. These conditions corresponded to zero, one and both air conditioners operating respectively. AI was calculated for each condition at each listener location using the octave band method as specified in ANSI Standard S3.5. The ideal voice spectrum given in this standard was used and corrected for the overall speech level as measured. The calculated AI values were corrected for visual cues and room reverberation time. Speech stimuli were presented live to the listeners. The speaker monitored his voice level with a VU meter. Speech and noise levels were previously measured separately and then together so that correct speech levels could be obtained. All acoustical measuring equipment was checked and calibrated prior to the test. The results of this experiment are shown in figure 3. The noncomparability between the present data and that given in S3.5 is really expected even though all the pertinent corrections were applied. These differences do say something about "prediction" though. Of possibly greater significance is the suggestion that the data for the three AI conditions do not fall on a single curve. Rather it appears that separate curves may be drawn through the data points for each condition. It should be pointed out here that these data are much too sparse to make any definitive judgements of this nature especially given the variability that may have resulted from the live presentation of stimuli.

However, as stated earlier, Bowman found similar results in a much more detailed experiment. Our judgement is that at least a cautious approach is required to the use of AI and interpretation of results.

When it comes to evaluating typical community or home noise situations in terms of speech interference the picture becomes less clear than for the well defined laboratory situation. AI emphasizes precision as might be needed to evaluate two similar pieces of communication hardware. However it is not clear that this type of precision is needed or buys anything that is not attainable much more simply for the community or home case. Beranek has suggested large ranges of AI for rating acceptability of rooms, office spaces etc. For example anything greater than AI of 0.5 is rated as an acceptable speech situation. This means essentially that a room with an AI of 0.6 is rated about the same as one having an AI of 0.8 on this acceptability scale. This is really a process of rank ordering and as such is not especially precise. Given this lack of precision, I think a simpler approach would involve the measurement of speech interference level. SIL tacitly recognizes the difficulty in obtaining precision and perhaps the lack of importance of such precision in a community noise context. In the final analysis, SIL probably gives essentially the same information that AI gives. Furthermore SIL has been shown by many people to be a good predictor of AI, so SIL is, in my opinion, the best existing method for evaluating steady state noise effects on speech in everyday environments.

Time varying noise presents a more difficult assessment situation. In terms of effects of time varying noise effects on speech, it is important to know what are the important aspects of the noise such as peak level, overall duration, duration above certain levels etc. To illustrate this problem, Carl Williams found that time varying noise masked speech less than steady state noise with an equivalent AI. Aircraft noise is a time-varying noise that has received a lot of attention recently. We here at Langley are particularly concerned with the effects of aircraft noise including speech effects. Our approach will involve the assessment of annoyance resulting from speech interfering noise rather than simply obtaining measures of intelligibility. This approach is of course, not new. Williams looked at acceptability of aircraft noise in the presence of speech. Langdon et al have looked at acceptability of various time varying noises during TV viewing. Dr. Gunn will report later on a study we performed at Memphis State University in which annoyance judgements were obtained during three tasks, two of which were speech communication tasks.

Others besides those just mentioned have measured the annoyance and acceptability that attend speech interfering noise. We expect to study annoyance that accompanies interference with four speech communication situations; TV viewing, tleephone use, classroom lecture, and face-to-face communication. We intend however to go beyond simply measuring information transfer and simultaneously getting annoyance judgements during speech interference situations. Actually the annoyance may result from considerably more than reduction in amount of information transferred. Such behaviors as listener confidence ratings, requests for repeats or actual repeats of information, voice level required, settings of loudness levels on audio equipment, bodily gestures, such as cupping a hand to one's ear, or turning one's head and other forms of behavior may also be significantly related to annoyance, and we expect to ultimately examine these relationships. As a jumping off point we intend to look first at differences in type of verbal stimuli in their effects on reported annoyance and also differences in method of stimuli presentation such as earphones, vs free field (loudspeaker) vs live presentation. From there we intend to build our speech interference research program in a way to reflect interest in the previously described factors.

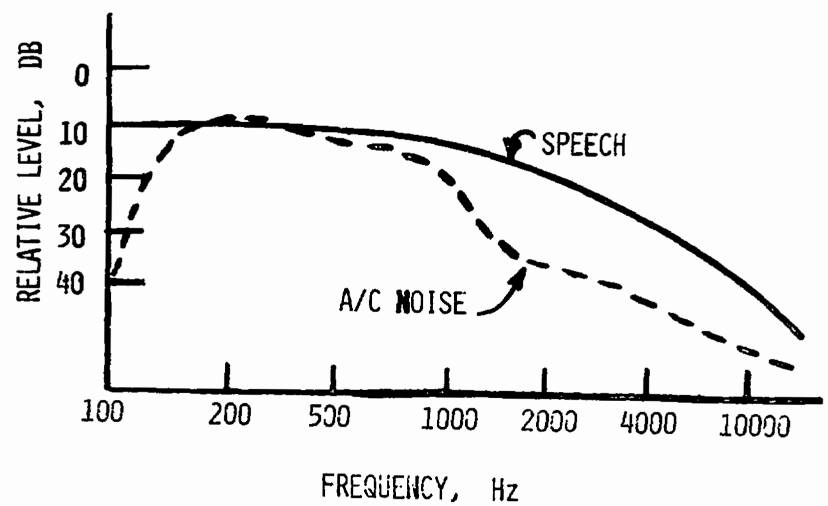
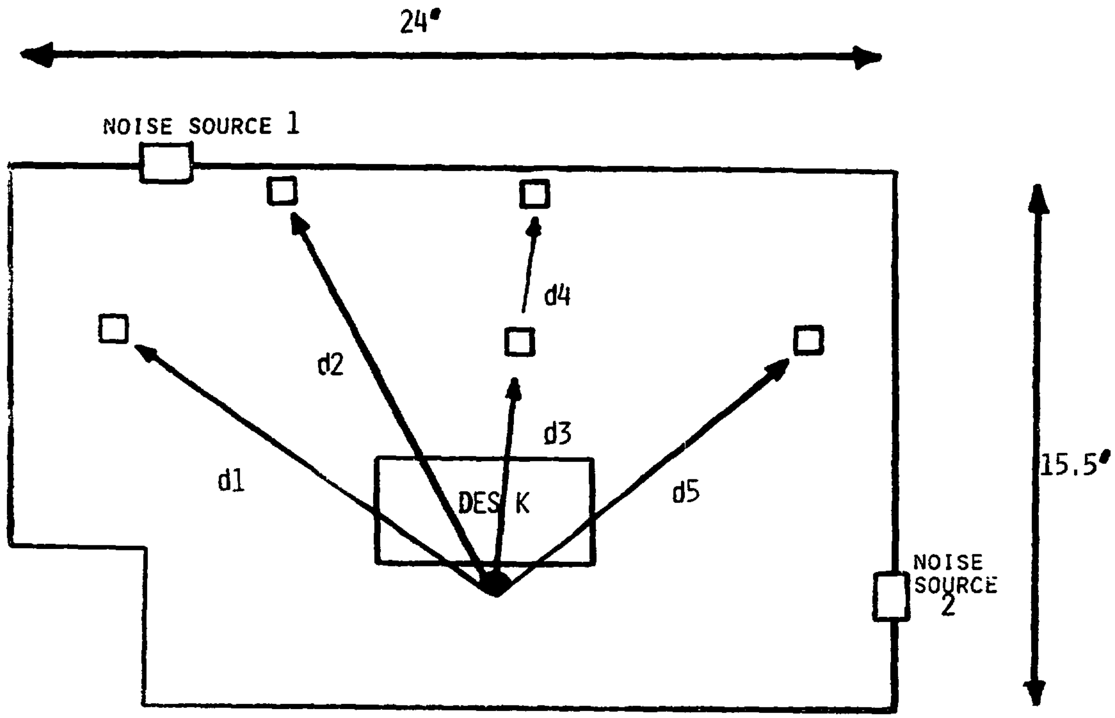


Figure 1.- Speech and Aircraft Noise Spectra



d1=13'
d2=12.5'
d3= 7'
d4=11.5'
d5=11.5'

Figure 2.- Test Room Dimensions

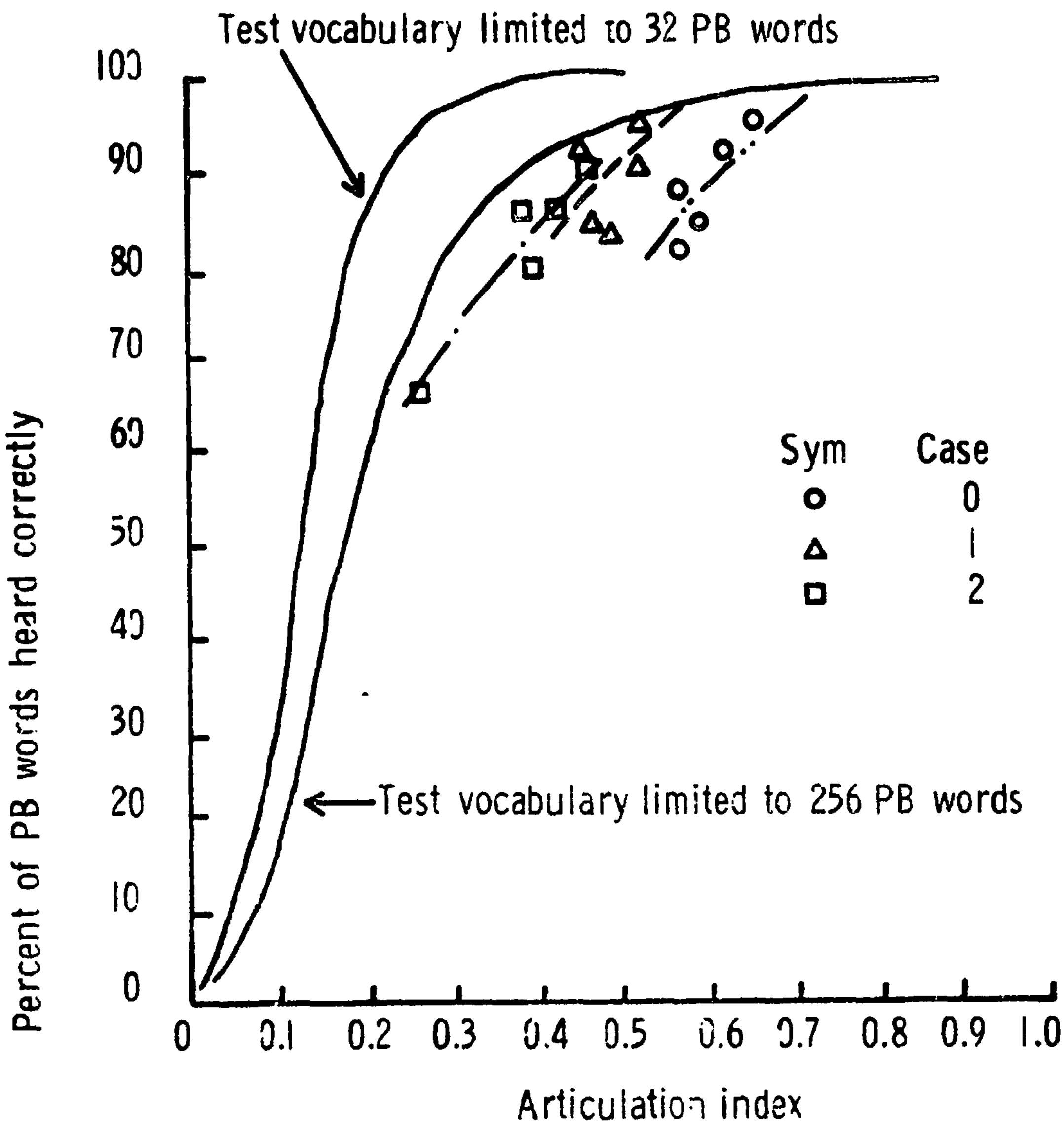


Figure 3.- Relation between AI and PB word speech intelligibility

REFERENCES

1. Stevens, S. S.; Miller, J. and Truscott, I.: The Masking of Speech by Sine Waves, Square Waves and Regular and Modulated Pulses. J. Acoust. Soc. America, vol. 18, 1946, pp 418-424

2. Rogers, E.: Sentence Intelligibility as a Function of Complexity and Amount of Time/Frequency Distortion. M. A. Thesis, Univ. Of Connecticut, 1970.

3. Ullrich, J. H.: An Experimental Study of the Acquisition of Information from Three Types of Recorded Television Presentations. Speech Monographs, vol. 24, 1957, pp. 39-45.

4. Williams, C., Stevens, K. N. and Klatt, M.: Judgement of the Acceptability of Aircraft Noise in the Presence of Speech. J. Sound and Vib., vol. 9, 1969, pp 263-275.

5. Bowman, N.: The Articulation Index and Its Application to Room Acoustic Design. J. Sound and Vib., vol. 32, 1974, pp. 109-129.

6. Williams, C., Pearsons, K., and Hecker, M.: Speech Intelligibility in the Presence of Time Varying Aircraft noise. J. Acoust. Soc. America, vol. 50, 1971, pp 426-434.

7. Methods for the Calculation of the Articulation Index. American National Standard S3.5. American National Standards Institute, New York, 1969.

8. Method for Measurement of Monosyllabic Word Intelligibility. USA Standard S3.2. American National Standards Institute, New York, 1960.

A PROPOSED METHOD FOR MEASURING THE ANNOYANCE DUE TO SPEECH INTERFERENCE BY NOISE

By

John A. Molino

National Bureau of Standards
Washington, D.C.

ABSTRACT

A method is proposed to measure both the interference of speech by noise and the annoyance caused by such interference. It is based upon a non-verbal preference procedure developed at the National Bureau of Standards called an "acoustic menu." Subjects listen to audible speech signals in a background of noise. At the same time the subjects are given a limited opportunity to select the particular type of background noise. By analyzing the preference structure for the various types of background noise, as well as the decrement in speech intelligibility suffered with each noise, information can be obtained on both relative annoyance and task interference.

A PROPOSED METHOD FOR MEASURING THE ANNOYANCE DUE TO SPEECH INTERFERENCE BY NOISE

John A. Molino

Institute for Basic Standards
National Bureau of Standards
Washington, D. C. 20234

INTRODUCTION

Certain noises may deliver intrinsically unpleasant acoustic sensation, and therefore be annoying. Other noises may interfere with ongoing human activity, and as a result generate annoyance. Most everyday noises, including aircraft noise, probably produce some proportion of both kinds of annoyance - hedonic unpleasantness and behavioral interference.

Noise is generally defined as "unwanted sound" (Harris, 1957). Central to an understanding of the "unwanted" properties of noise, i.e. the negatively reinforcing properties of noise, is some way to measure these two kinds of annoyance separately. (There may also be other kinds of annoyance, for example, that caused by a perception of misfeasance; but the present paper will treat only the two kinds mentioned at the outset - unpleasantness and interference.) Our research at the National Bureau of Standards (NBS) has led us to develop novel methods for measuring the negatively reinforcing properties of noise. At the same time these methods have built into them, for quite independent reasons, the ability to assess simultaneously the effects of noise on human performance.

They could easily be applied to the problem of measuring the annoyance due to speech interference by noise.

The research methods in use at NBS are directed at measuring human aversion for sound, i.e. the tendency for people to escape and avoid certain acoustic stimuli. As such they do not depend upon verbal reports of the annoyance experienced while listening to these soun . ut rather measure the behavioral effects that are likely to result fr posure. For example, our measurements of human aversion for sound might be expected to correlate with the tendency of people in noisy areas to alter their behavior patterns, to move away from those areas, or to complain about reduced market value of their homes. But more importantly, from a methodological point of view, these techniques offer a possible way to separate the hedonic and interference components of the human response to noise without requiring subjects to make subtle, difficult, and maybe even impossible verbal distinctions concerning the source of their annoyance. Imagine the difficulty subjects might encounter in complying with the following instructions from an experimenter: "You will hear several aircraft sounds while listening to messages from this loudspeaker. You should report how much of the annoyance you experience is due to the intrinsic unpleasantness of the aircraft noises and how much of it is due to interference with your listening task."

This difficulty is independent of the issue of how well such verbal reports of annoyance might correlate with actual behavioral responses to reduce, escape or avoid the noise. Preliminary evidence shows that, when forced to make judgements according to some verbally defined criterion, subjects may tend to exaggerate the differences along the abstract scale so defined as long as they can perceive any difference at all among the stimuli. Yet this judged difference may have little influence on the subjects' behavior with respect to the sound when given the opportunity to alter the sound (Zerdy and Molino, 1974). Non-verbal measures of human aversion to sound may be able to eliminate some of these difficulties.

BACKGROUND

Typical psychophysical experiments designed to assess the human response to noise require subjects to rate various sounds according to verbal descriptions that define a certain abstract quality of the sound. In some experiments subjects are asked to judge the "loudness" of the sounds (Stevens, 1961), and not to pay attention to other qualities, like "unpleasantness." In other experiments they are asked to judge the "annoyance" of the sounds (Spieth, 1956), supposedly independently of the "loudness" quality. Others use verbal descriptions defining qualities of "discomfort" (Hood and Poole, 1966), "dissatisfaction" (Keighley, 1970), or "unpleasantness" (Vitz, 1972), etc. Often these experiments suggest the establishment of a certain psychophysical scale that adjusts the physical components of the noise in a manner proportional to the human response to those components. If these procedures continue to proliferate, the number of possible scales might be limited only by the number of adjectives that can be used to describe sounds. Thus, in elaborating the concept of "perceived noisiness", a conglomerate of descriptions was employed in an attempt to avoid this problem. For example, in the verbal instructions given to the subjects in one experiment (Kryter and Pearsons, 1963), one may find the words "disturbing", "objectionable", and "acceptable", all appearing in a single paragraph. However, such a choice is by nature arbitrary and inexhaustive. Furthermore, the particular phrasing of the paragraph of instructions may give more emphasis to one word over another.

The hallmark of the methods being developed at NBS is that the human response is measured without any verbal descriptions of the sounds. Three procedures have been investigated thus far, all based upon a considerable body of research in experimental psychology (Honig, 1966). The first is an adjustment procedure, where subjects can earn decrements in sound intensity by tapping rapidly on a telegraph key. If the subjects do not tap, the sound intensity gradually increases 1 dB every 4 s. Thus the subjects are able to adjust the intensity to a tolerable level by working steadily on the key (Molino, 1974).

Let us investigate this adjustment procedure in more detail. In one such experiment, after two hours of training, each subject participated in 64 experimental sessions of 10 min duration (4 sessions per hr, 1 hr per day, for about 3 weeks). During each session, one of 16 acoustic stimuli (8 pure tones and 8 bands of noise) was present for the entire session. At the beginning of the session the intensity level was set at either a medium A-weighted sound level of 50 dB or a high A-weighted sound level of 90 dB. These initial levels were chosen so that all of the sounds at a given starting level would appear roughly equally loud to the subjects when the session began. Thereafter the intensity level was under the subject's control.

The results of the experiment are presented in Figs. 1 and 2. In Fig. 1, the average maintained sound pressure level (SPL) across stimuli, starting levels, and replications is shown as a function of time for each of the 14 subjects. The data points on each curve represent the mean of 64 measurements. The slopes of the intensity changes that would result from different average rates of responding are given in the arc near the top of the ordinate. These slopes indicate that the subject could maintain a constant SPL with a tapping rate of 3 responses/s. Most of the maintained SPL curves reached this constant intensity level after about 5 min of responding. However, different subjects maintained the average sound intensity at distinctly different levels.

In Fig. 2, the average maintained SPL across subjects, starting levels, and replications is shown for each of the eight 1/3-octave bands of noise. The data points on each curve represent the mean of 56 measurements. As is evident in the figure, a progressively lower maintained SPL was observed as the frequency was increased over the range from 63 to 500 Hz. For the higher frequencies, above 1000 Hz, there was little consistent difference in the maintained SPL for different frequencies.

These asymptotic maintained SPL values for the various frequencies may be regarded as equal aversion levels under the given experimental conditions. As such, they convey information about the relative human tolerance for the different frequency components of the stimuli. The asymptotic SPL results can then be compared with other determinations of constant human response as a function of frequency. Such a comparison is presented in Fig. 3. Here the curve connecting the solid circles represents the measurement of equal aversion levels (EAL) for the eight 1/3-octave bands of noise. Data are also shown for EAL levels for pure tones, as well as other data from other weighting contours: A-weighted sound level (SLA), "loudness" level (ISO), and "perceived noise" level (PNL). Thus the first procedure developed at NBS affords a determination of the relative aversiveness (annoyance) due to different frequency components of the sound.

The second procedure, a variable-interval escape schedule, can provide similar data by means of a quite different response contingency. The experimental session starts with an intense acoustic stimulus being presented to the subject. Instead of tapping rapidly on the telegraph key to earn decrements in sound intensity, in this instance a much slower rate of responding on the telegraph key will produce variable intervals of silence or soft background noise. If the subjects do not respond, they will remain exposed to the intense acoustic stimulus. Here the rate of responding on the key is taken as a measure of the aversiveness of the sound (Wakeford, 1974).

The third procedure determines the preference relations among various sounds by recording the proportion of time spent listening to them. We call this technique an "acoustic menu" (Zerdy and Molino, 1974). At any given time the subject can select either of a particular pair of sounds to be present. This pair is available to the subject during a 10 min experimental session. In addition, which sound of the pair is present alternates automatically on an intermittent schedule. Thus the subject must emit a number of responses in order to spend a larger proportion of the time in the preferred stimulus. By testing many such pairs, a preference structure may be ascertained for the collection of sounds.

In all of these procedures the subjects have some degree of control over the sound. However, no verbal descriptions are used to establish a criterion for what the subjects' response to the sound should be. We simply observe at what intensity level people begin to escape or avoid a given acoustic stimulus. Since such experimental sessions are rather unstructured and the subjects need not do anything with the sound if they do not want to, the subjects are typically simultaneously engaged with another task. Often, while the sounds are introduced, they will be learning to read and write Russian from a teaching machine. We have also employed programmed instruction in English and mathematics, as well as anagram and number games.

These additional tasks serve several functions. First, they eliminate boredom, which often results in the subjects manipulating the sound merely to avoid sensory deprivation. Second, these tasks provide a challenging activity that improves the motivation of the subjects toward overall participation in the experiment. Third, they make the laboratory situation a better simulation of the natural environment. When people are annoyed by noise, they are usually not concentrating on the noise alone in an otherwise impoverished sensory and intellectual surrounding. More realistically, they are probably engaged in some other activity that is holding most of their attention, and are attempting to ignore the sound as much as possible.

PROPOSED EXPERIMENT

The three procedures being developed at NBS could be easily adapted to provide information on the aversiveness (annoyance) due to speech interference by noise. Instead of the programmed instructional material presently used in the experiments, a speech recognition task could be substituted. At the same time the subjects could be permitted to alter the acoustic environment to a limited extent. Probably the most promising method in this regard would be a modified version of the "acoustic menu". With this technique, subjects could make pair-wise choices of which acoustic stimulus would be present for a majority of the time spent in the experimental session. Other procedures might be tried as well, such as adjustment techniques and interval schedules of reinforcement. The "acoustic menu" would be the most likely first candidate, however, because it is the least time-dependent of the procedures.

In any case, the main task of the subject would be the recognition of textual material or word lists presented either visually or aurally. During some experimental sessions the words would be presented visually over a closed-circuit television monitor. The words would appear in sequence, briefly, and one at a time. The subjects would be instructed to write down the words as they perceived them. During other experimental sessions, similar words would be aurally presented at the same rate over earphones or loudspeakers - the same transducers that would deliver the interfering noises. Again the subjects would write down the words perceived. If the "acoustic menu" is employed, during both types of sessions the subjects could select which of a pair of sounds would be present at any given time. These sounds could be continuous pure tones or one-third octave bands of noise in a theoretical study, simulated steady-state spectra of various types of aircraft noise, or recordings of actual aircraft fly-overs in a more applied investigation. The latter time-varying signals would present several additional, though not insurmountable, difficulties, however. The transient nature of the acoustic stimulus would make analysis of interference with the perception of verbal message more difficult. In addition, either a synchronization of the fly-over acoustic envelopes in both channels, or a refractory response period during a given fly-over envelope would have to be incorporated into the preference assessment portion of the "menu" procedure. By pairing a sample of noises at different intensity levels with each other and with some pleasant-sounding background sound, a preference structure could be generated for the sounds under investigation.

If the same preference structure is found for both visually presented and aurally presented work conditions, then the aversiveness of the sounds would be primarily due to the hedonic component. If the aurally presented word condition produces a significantly different preference structure, this difference would represent the unique contribution to the aversiveness of the sounds made by interference with perceived speech. In either case, speech interference measures, i.e. percentage of words perceived correctly, could be calculated for both verbal presentation conditions. The speech interference experienced with each of the sounds could then be compared with the relative preference for the sound to determine to what extent the least preferred sounds were also those that most interfered with speech intelligibility.

If the stimuli consist of pure tones or bands of noise at various intensity levels, more sophistication can be achieved. In this case, indifference contours can be determined in the frequency-intensity plane. That is, for each frequency an intensity level may be determined that is equally preferred or non-preferred to some intensity level at another frequency.

Thus, a psychophysical indifference function can be defined similar to an "equal loudness" or "equal noisiness" contour. Furthermore, two such indifference contours can be found, one for the aural condition and one for the visual condition. The difference between them would represent, at each frequency, the relative contribution of the aversiveness (annoyance) due to speech interference, as opposed to the aversiveness (annoyance) due to hedonic attributes alone. Likewise, two equal speech interference contours could be found, one for each condition. The difference between these interference contours would represent the relative contribution of interference with the aural perception of speech as opposed to interference with semantic processing in general (distraction). Thus the proposed experiment could assess the relative speech interference suffered at each frequency, and the relative aversiveness (annoyance) at each frequency due to that speech interference.

In this manner an algorithm might be generated to measure quantitatively both speech interference by noise and the resulting annoyance experienced by the listener. Such an algorithm might then be applied to the design of auditoria, classrooms, offices, or television viewing situations where noise interference is anticipated from aircraft, highways, railroads, or other noise sources.

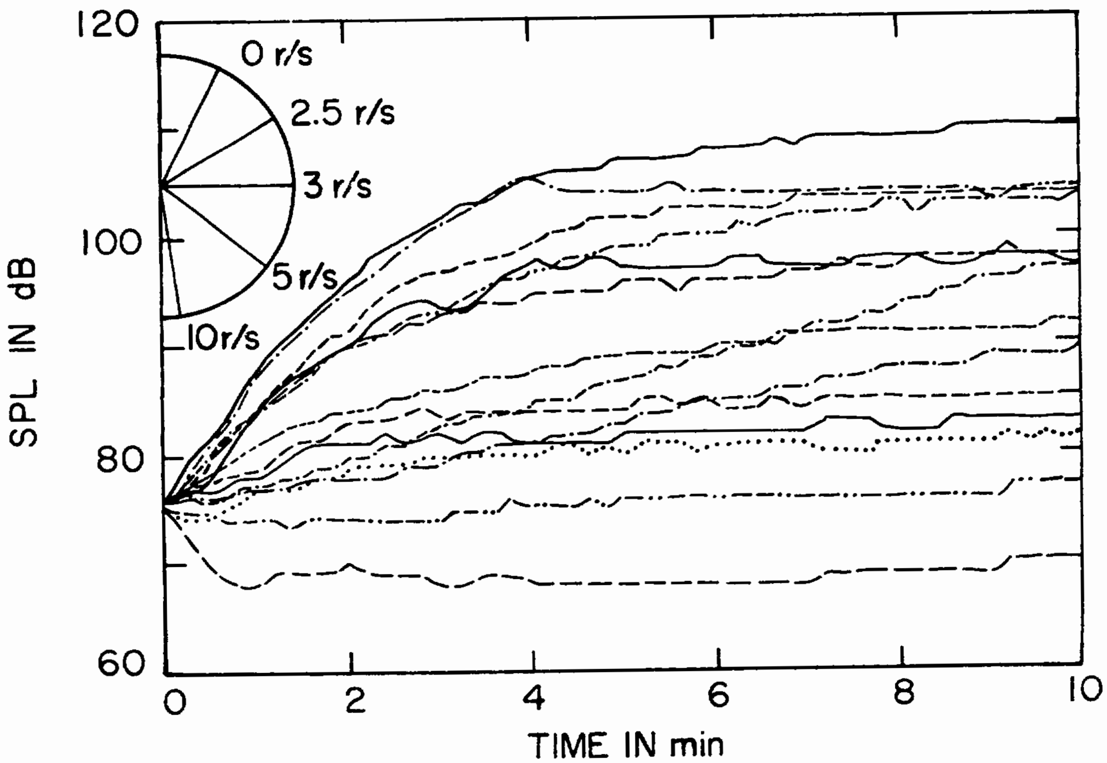


Figure 1 Average maintained sound pressure level (SPL) as a function of time for each of the 14 subjects in the adjustment procedure.

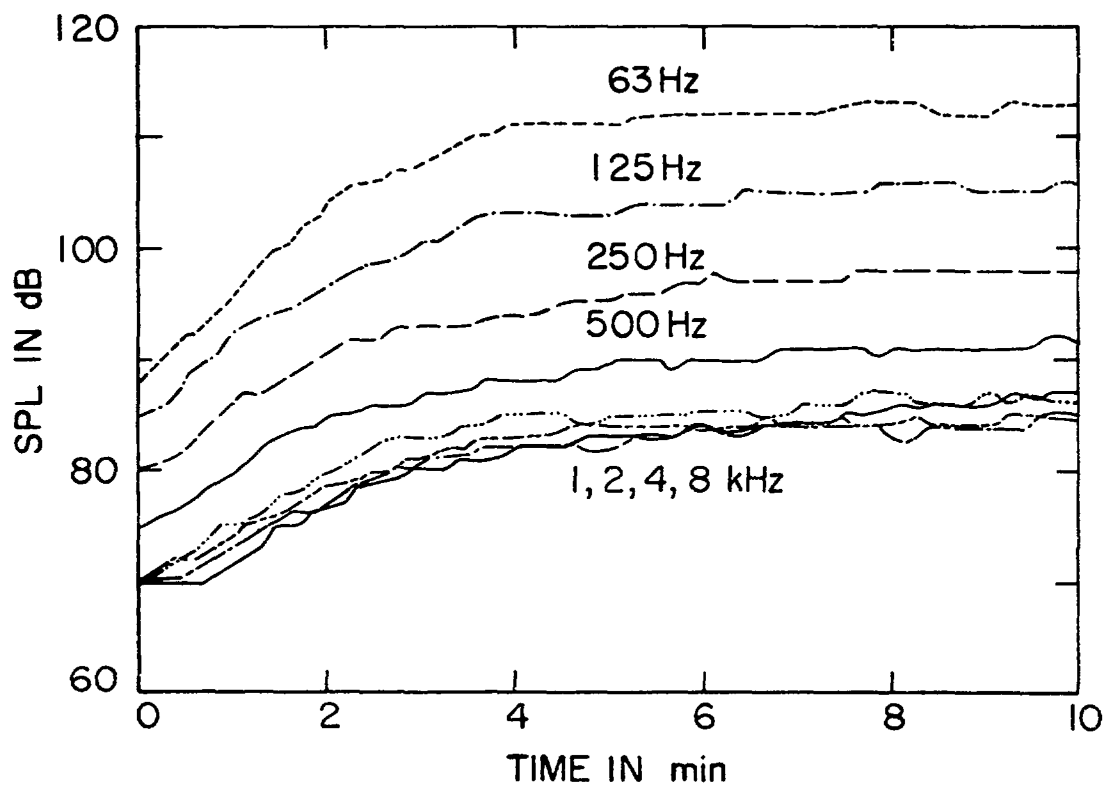


Figure 2 Average maintained sound pressure level (SPL) as a function of time for the group of subjects listening to each of eight 1/3-octave bands of noise.

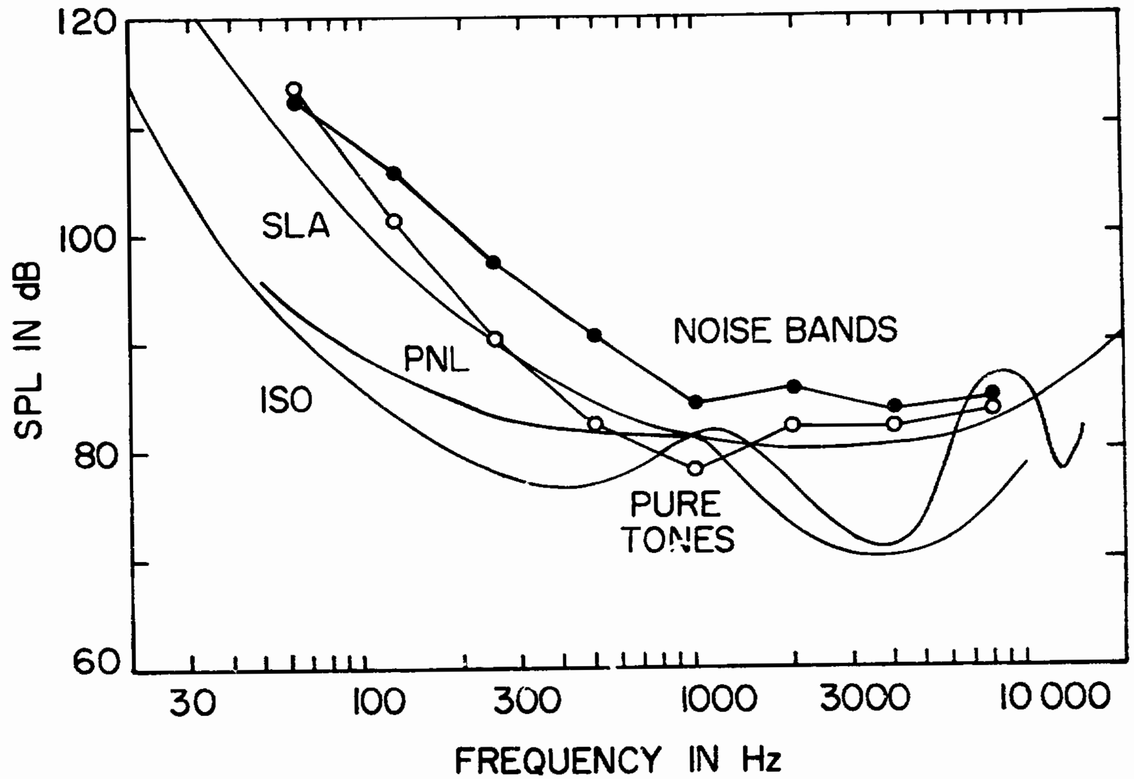


Figure 3 Sound pressure level (SPL) necessary for a constant human response.

REFERENCES

Harris, C. M. (Ed.) Handbook of noise control. New York:McGraw-Hill, 1957.

Honig, W. K. (Ed.) Operant behavior: areas of research and application. New York: Appleton-Century-Crofts, 1966.

Hood, J. D. and Poole, J. P., Tolerable limit of loudness: its clinical and physiological significance. J. Acoust. Soc. Amer., 1966, 40, 47-53.

Keighley, E.C., Acceptability criteria for noise in large offices. J. Sound & Vibr., 1970, 11, 83-93.

Kryter, K. D. and Pearsons, K. S., Some effects of spectral content and duration on perceived noise level. J. Acoust. Soc. Amer., 1963, 35, 866-883.

Molino, J. A., Equal aversion levels for pure tones and 1/3-octave bands of noise. J. Acoust. Soc. Amer., 1974, 55, 1285-1289.

Spieth, W., Annoyance thresholds of bands of noise. J. Acoust. Soc. Amer., 1956, 28, 872-877.

Stevens, S.S., Procedure for calculating loudness: Mark VI. J. Acoust. Soc. Amer., 1961, 33, 1577-1585.

Vitz, P. C., Preference for tones as a function of frequency (Hertz) and intensity (decibels). Perception and Psychophysics, 1972, 11, 84-88.

Wakeford, O. S., Measurement of human aversion to intense acoustic stimuli. _J. Acoust. Soc. Amer._, 1974, _56_, S64 (Abstract).

Zerdy, G. A. and Molino, J. A., Choosing among intense acoustic background stimuli - an acoustic "menu". _J. Acoust. Soc. Amer._, 1974, _56_, S64 (Abstract).

ANNOYANCE OF TIME VARYING NOISE WHILE LISTENING TO SPEECH

By

Karl S. Pearsons

Bolt Beranek and Newman Inc.
Canoga Park, CA.

SPEECH INTELLIGIBILITY

Most speech intelligibility testing has employed steady state noise as a masker. Unfortunately, most noise encountered in our home environments is of a time varying nature. To explore speech intelligibility in the more commonly encountered time varying noise, tests were conducted using recordings of traffic noise and shaped broadband noise as speech maskers.

Test Description

Six two-syllable (spondee) words were randomly presented to subjects during five minutes of recorded traffic noise. The words were presented in rapid succession and the subjects were asked to push one of six buttons corresponding to words they had heard. Ten different sets of six words were utilized for most of the traffic noise samples. However, average sound pressure levels of each of the ten groups varied by only $\pm$ 1 dB. Therefore, the small variation among the mean levels of the word sets permitted pooling of the data from the ten sets in determining the intelligibility of the words. A block diagram of the test setup is shown in Figure 1.

The traffic noise samples ranged in variability as shown by the samples of cumulative distribution in Figure 2. L_{10} - L_{50} values ranged from .4 dB(A) for the steady state shaped noise to 2 dB(A) for freely flowing relatively steady traffic noise. For the highly variable case, the L_{10} - L_{50} values were 8 dB(A).

Other tests were performed using the broadband shaped noise and 8 lists of 50 phonetically balanced (PB) words. To determine the effect of voice levels, the level of the word lists was varied to obtain various percentages of correct words for each of the word lists.

Results

In order to determine Articulation Index values, the speech spectrum was determined for the lists of PB words and spondee words. Figure 3 shows an example of the spectrum for 4 lists of spondee words. In addition the speech spectrum from the ANSI Standard S-3.5, 1969, is given for comparison. As might be expected, the speech spectrum used in the standard has a certain amount of smoothing since it is meant to represent an average of several different speakers. Figures 4 and 5 show the intelligibility functions for the time varying traffic noise utilizing spondee words. The lines on the figures indicate interpolated psychometric functions through the data points which are aggregate percent correct scores for a panel of 4 observers. The shapes of the functions are not unlike the more conventional functions utilizing steady noise. Figure 6 shows the results using PB words and shaped broadband noise. The function is not as steep, primarily because the number of words in the PB word list was greater than the closed set of 6 words employed in the spondee test. From Figures 4 and 5, the percent correct spondee words can be determined for various speech levels.

Discussion

To compare the intelligibility functions of the various samples of traffic noise, the results were all normalized to an L_{eq} of 60 dB(A) for the traffic noise samples. The shaded area on Figure 7 represents the range of all of the intelligibility functions for the various traffic noise samples.

The narrow width of the shaded area suggests that the variability of the traffic noise samples was not a factor in determining the intelligibility functions. The one exception was traffic noise sample No. 2 which, in general, required a higher speech level for obtaining a given percent correct of spondee words. However, there was no particular trend in the results which would indicate that a more or less variable noise was more or less interfering with speech communication. In fact, the standard deviation of noise levels required to produce a 90 percent intelligibility score was only 1.3 dB including the results of traffic noise sample No. 2. It should be remembered, however, that if the variability of the noise distribution is greater than for the samples of traffic noise utilized in this study, the effect of variability may become important. This might well be true for the aircraft noise situation.

Figure 8 shows the results of the study in terms of Articulation Index. Also shown in the graph are the results from other studies as depicted in the Articulation Index Calculation Standard (ANSI 3.5 - 1969). The results clearly show that for the spondee words which were tested 6 at a time, the percent correct versus Articulation Index is a very steep function, and people were able to score near 100 percent correct for a relatively low Articulation Index. The PB word intelligibility, however, was more nearly typical of other tests which have been conducted for speech intelligibility. As might be expected, the words which were chosen from a list of 400 appeared to have a greater intelligibility than the words taken from a list of 1000 according to Figure 8.

Recommendations

It appears that the major missing link in determining intelligibility of various time varying noise sources is an indication of the vocal levels or speech levels which are typically employed in every day situations. Most of the measurements of speech levels have been obtained utilizing recordings of word lists or continuous discourse in an anechoic chamber. This would suggest that recordings be made in a home situation using actual conversation rather than the reading of a word list or standard paragraph. In addition, some studies should be performed using aircraft noise source rather than traffic noise especially since the cumulative distribution functions would be significantly different from those employed for the traffic noise situations in this test. It would also be useful to obtain additional information on the intelligibility of word lists presented for the first time. Most of the work that has been done on intelligibility has utilized repeated presentations of a word list to overcome the learning effect. However, in every day conversation one would be interested in the intelligibility of the first utterance as opposed to establishment of a master list of words from which the word lists are derived.

ANNOYANCE

In addition to speech intelligibility per se, there is some annoyance associated with traffic noise either due to the speech interruption it causes or the annoyance of traffic noise itself. Additional tests were performed to investigate the annoyance of the time varying characteristics of traffic noise.

Test Description

The general setup is similar to that described under the speech intelligibility tests except that continuous discourse was used in addition to the spondee words for speech material. The continuous discourse consisted of articles taken from the Wall Street Journal and recordings of old radio shows. The traffic noise samples were similar to those employed in the speech intelligibility tests, but more extreme cases were utilized as indicated in Figure 9. For this test, annoyance ratings of the traffic noise samples which were 5 minutes in duration were obtained both with and without speech present. Three questions were asked about the speech material presented. This was mainly done to insure that the subjects would listen to the speech material. However, the answers to the questions were employed as a measure of the comprehension of the speech material presented.

Results

Figure 10 shows the results of annoyance ratings of a particular sample of traffic noise in which the speech level of spondee words was varied. As can be seen from the plot, the annoyance level decreases as the speech level increases. In other words, as the speech material becomes more and more intelligible, the annoyance of the traffic noise is lessened. This appears to be true at least for L_{eq} values of traffic noise 60 dB and lower. Figure 11 shows the annoyance ratings of the various traffic noise samples without speech present. Similarly, Figures 12 and 13 show the annoyance ratings of the same annoyance samples with speech present at varying degrees of comprehension. The plots indicate quite a bit of scatter in the test results.

However, in general, it appeared that for low and moderate comprehension, the annoyance values are higher than one finds for high comprehension or for no speech present at all. Because of the large scatter in the plots, the regression lines normally drawn through such a data were not employed. Rather the average sound levels for each of the annoyance category ratings were determined and are plotted in a summary graph as shown in Figure 14. Here it is clearly shown that for the low levels of traffic noise (less than L_{eq} = 60), the annoyance rating for cases of traffic noise where speech was present but at a low to moderate intelligibility, the annoyance rating was greater than for the ratings of traffic noise where no speech was present at all.

Figure 15 shows a plot of the number of questions correctly answered versus a measure of variability described by the difference of L_{10} and L_{50} measurements of the traffic noise. One can see from the figure that as the variability increases for a given level of L_{eq}, the comprehension of the speech material increases to a maximum value and then decreases slightly as the variability continued to increase. Actually, the decrease in comprehension as variability increases beyond 4 dB is probably a test artifact and that more realistically the comprehension might be expected to reach a plateau rather than decrease for the higher variability levels. As a further indication that annoyance is a function of speech comprehension in the presence of time varying traffic noise, Figure 16 provides a plot of the relation between annoyance rating and the number of questions correctly answered. As might be expected, as the number of questions correctly answered increases, the annoyance of the time varying traffic noise decreases.

Discussion

For levels of noise L_{eq} = 60 dB(A) or below, the level of speech can affect the annoyance rating of traffic noise. In other words if you find it difficult to hear the radio or TV or someone speaking, you would be more annoyed at a given level of background noise than if you were able to comprehend the speech material. Also, the variability of the traffic noise can affect its annoyance rating. Figure 17 shows a summary of the annoyance rating versus variability for conditions with and without speech present. For the case with speech present, it is clear that as the variability becomes higher the annoyance is reduced. This is in direct contradiction to the philosophy employed in the development of the Noise Pollution Level (NPL). Figure 17 also suggests that the increased variability also reduces the annoyance rating of traffic noise without speech present, however, the substantiating data is not as conclusive as for the case with speech present.

Recommendations

It is recommended that additional tests be conducted using aircraft noise as stimuli to check the annoyance ratings when speech is present, and also to determine the effect of variability utilizing aircraft noise samples instead of traffic noise samples.

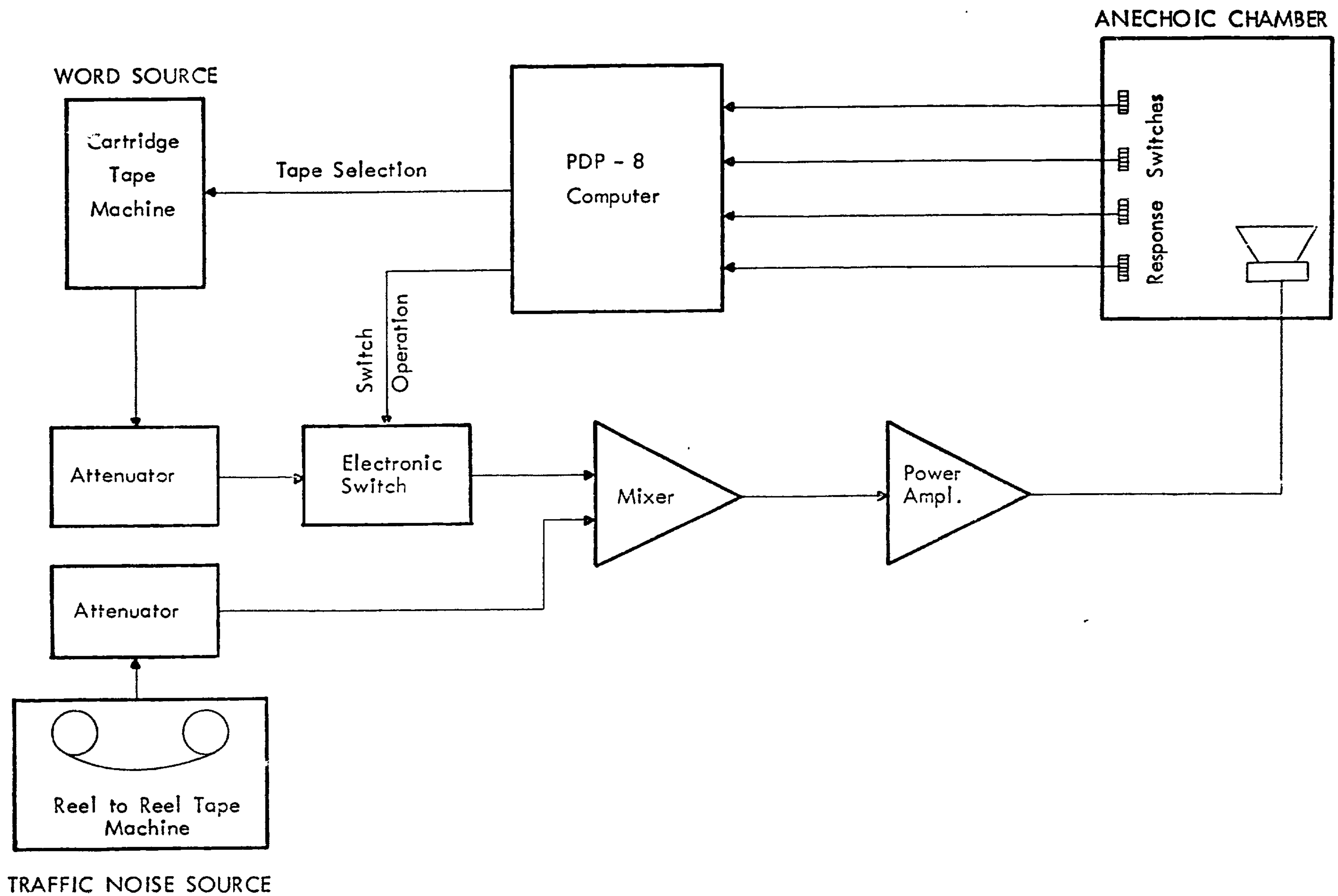


FIGURE 1. - EQUIPMENT USED FOR SPEECH INTELLIGIBILITY TESTING

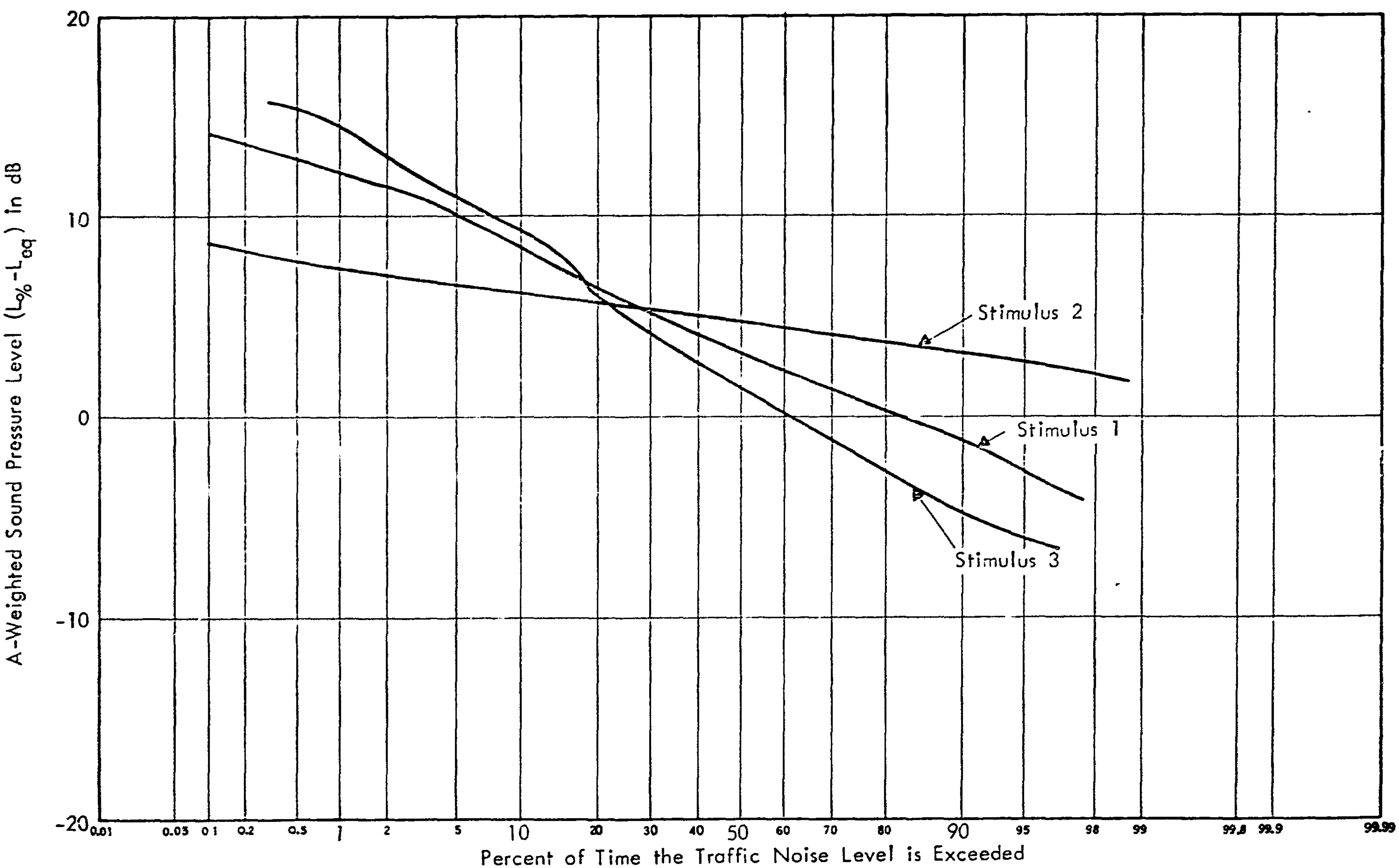


FIGURE 2.- CUMULATIVE DISTRIBUTION FOR SAMPLES OF TRAFFIC NOISE

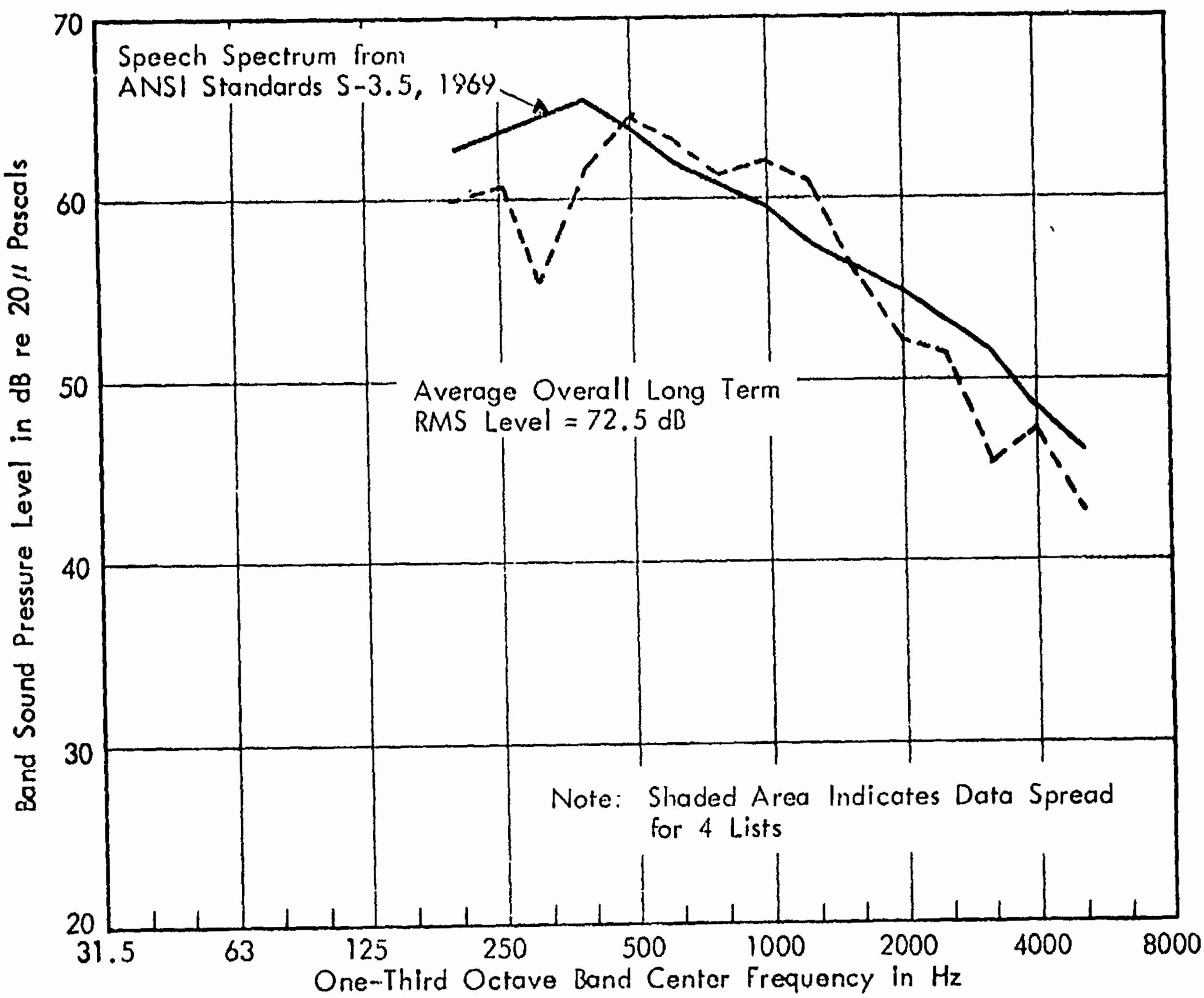


FIGURE 3.- SPEECH SPECTRA

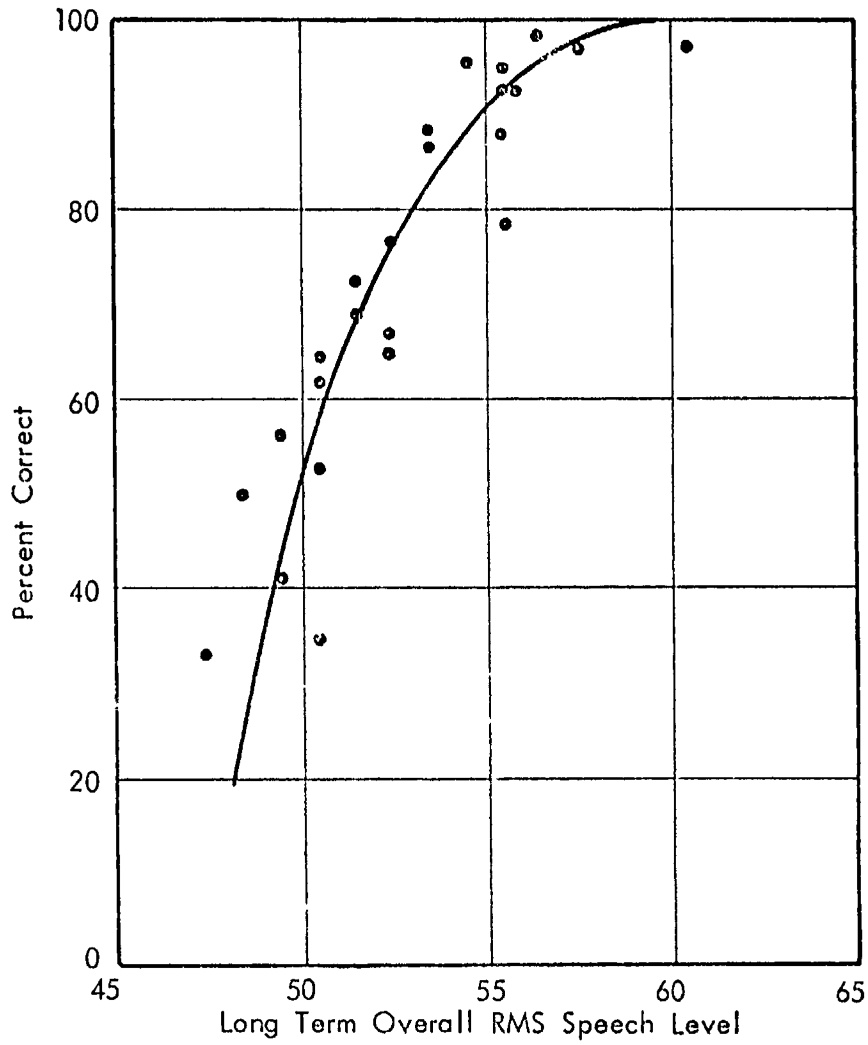


FIGURE 4.- INTELLIGIBILTY OF SPONDEE WORDS WITH TIME VARYING NOISE ($L_{eq} \stackrel{\sim}{=} 63.5$ dBA)

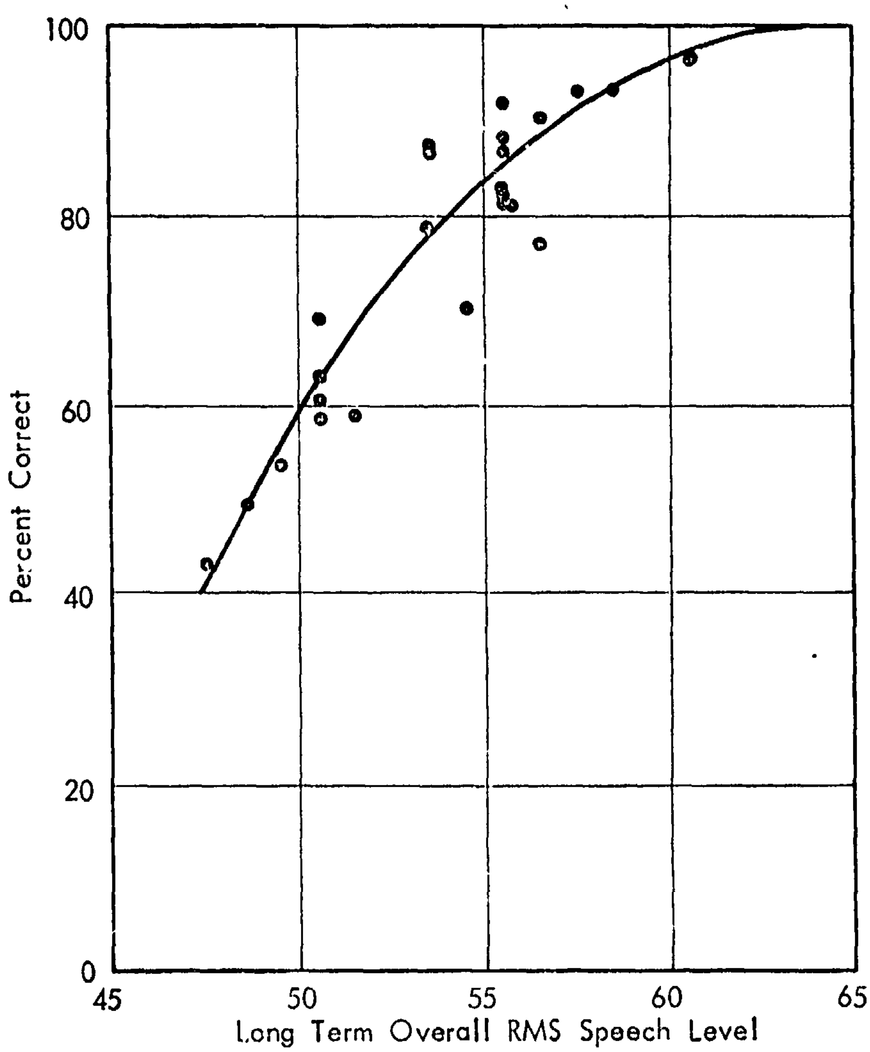


FIGURE 5.- INTELLIGIBILITY OF SPONDEE WORDS WITH TIME VARYING TRAFFIC NOISE (L_{eq} = 65 dBA)

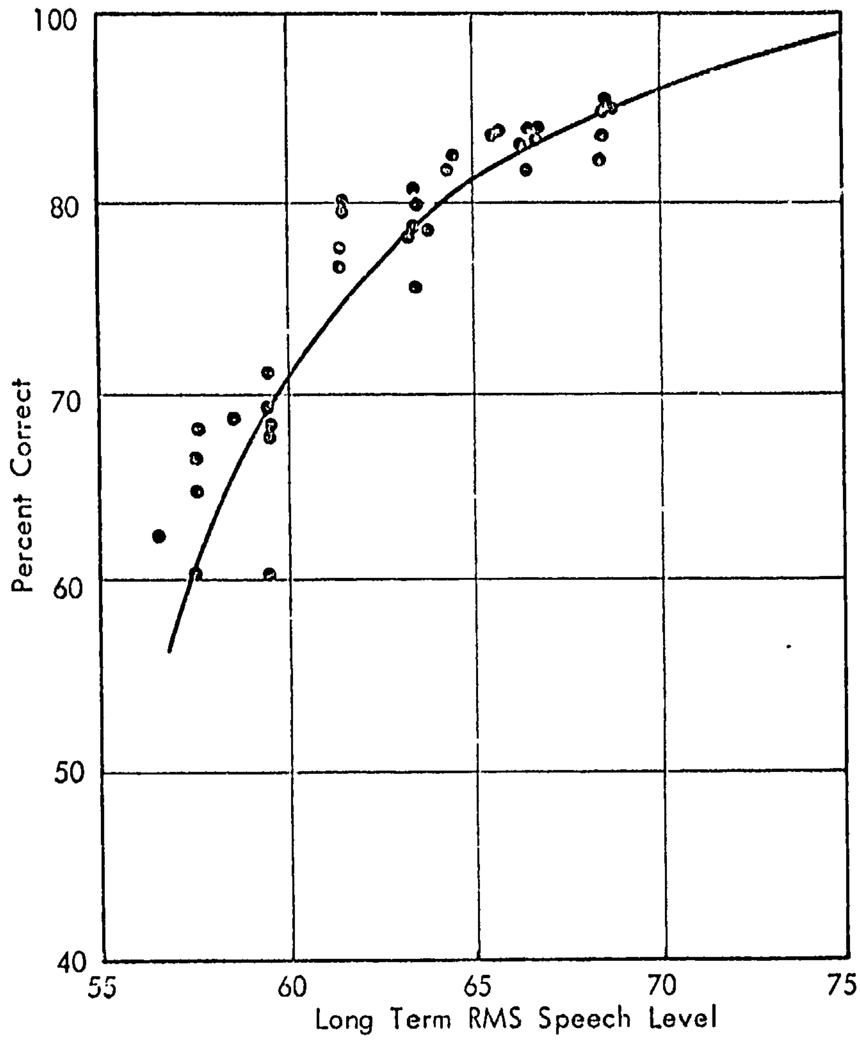


FIGURE 6.- INTELLIGIBILITY OF PB WORDS WITH STEADY NOISE (L_{eq} = 63.5 dBA)

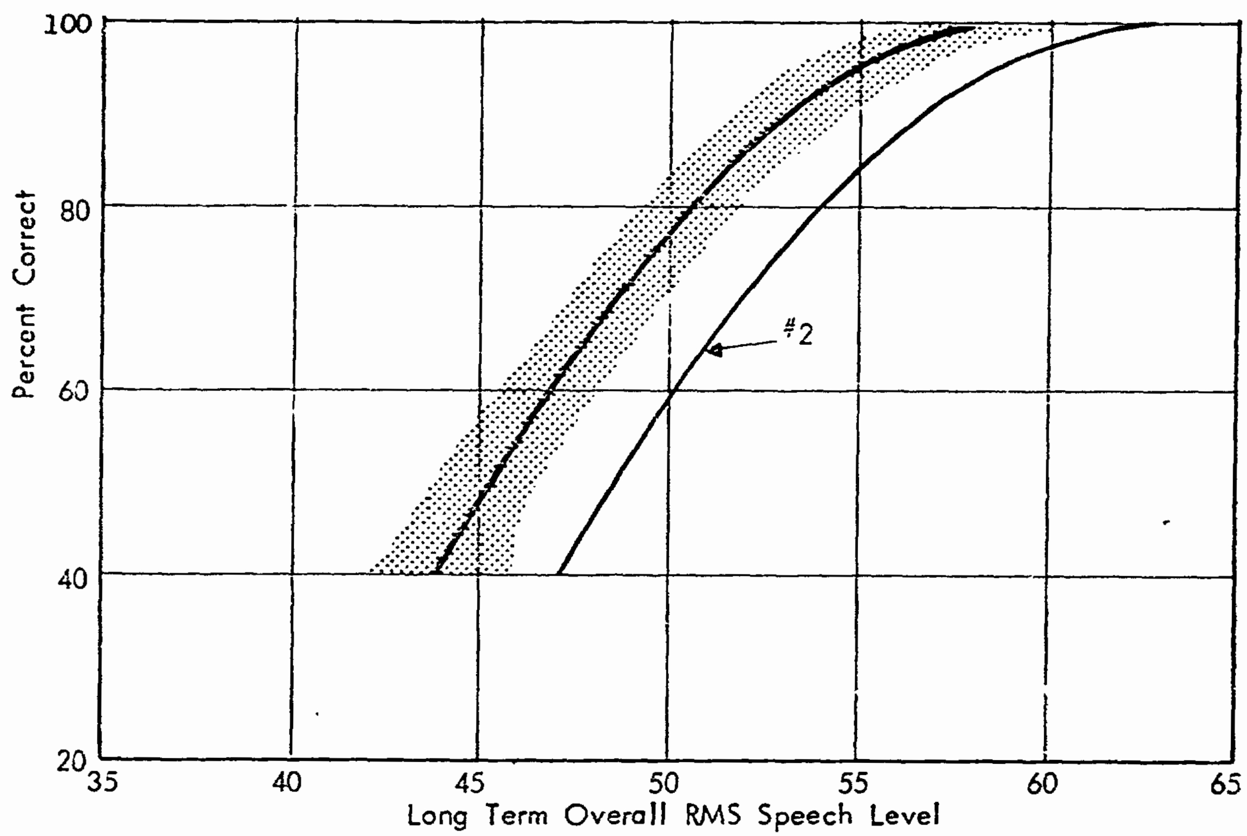


FIGURE 7.- INTELLIGIBILITY OF SPONDEE WORDS WITH VARIOUS TRAFFIC NOISE SAMPLES (L_{eq} = 60 dBA)

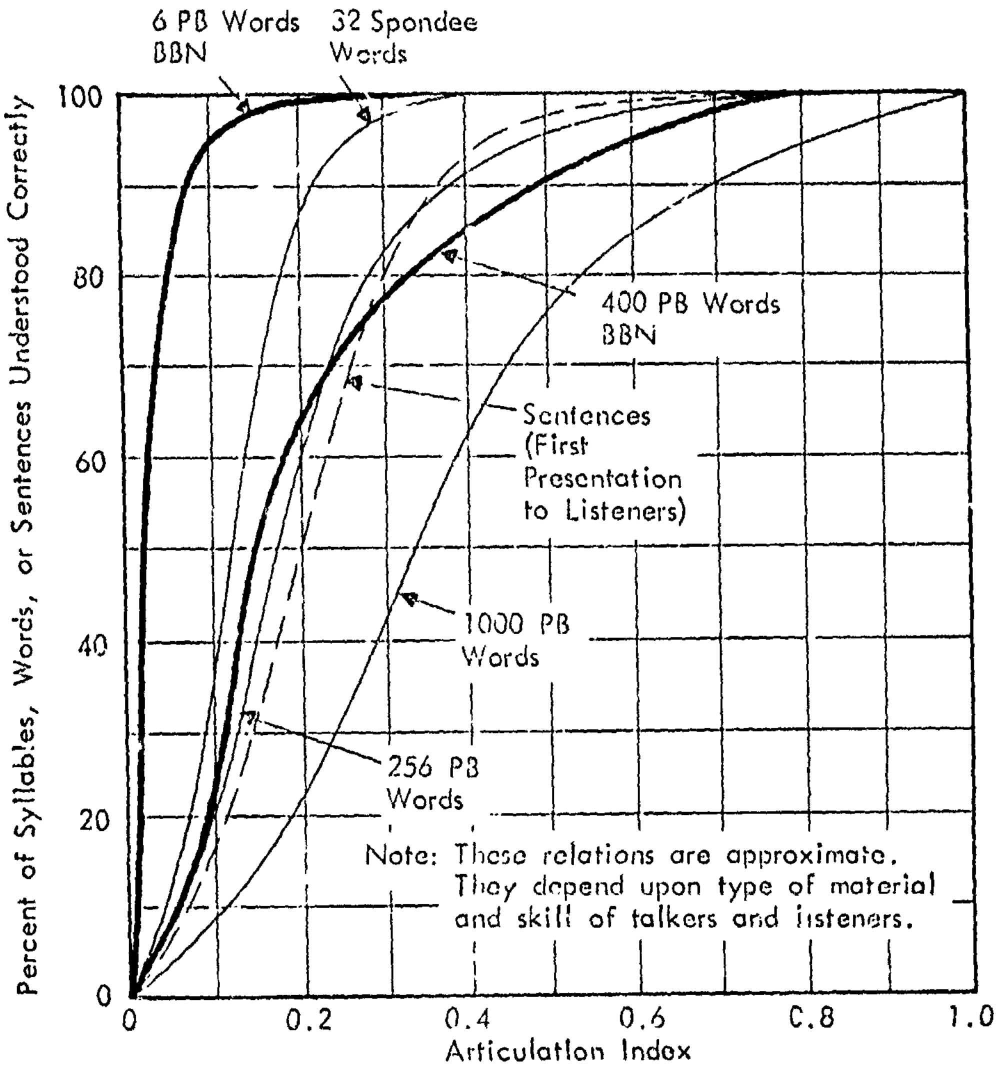
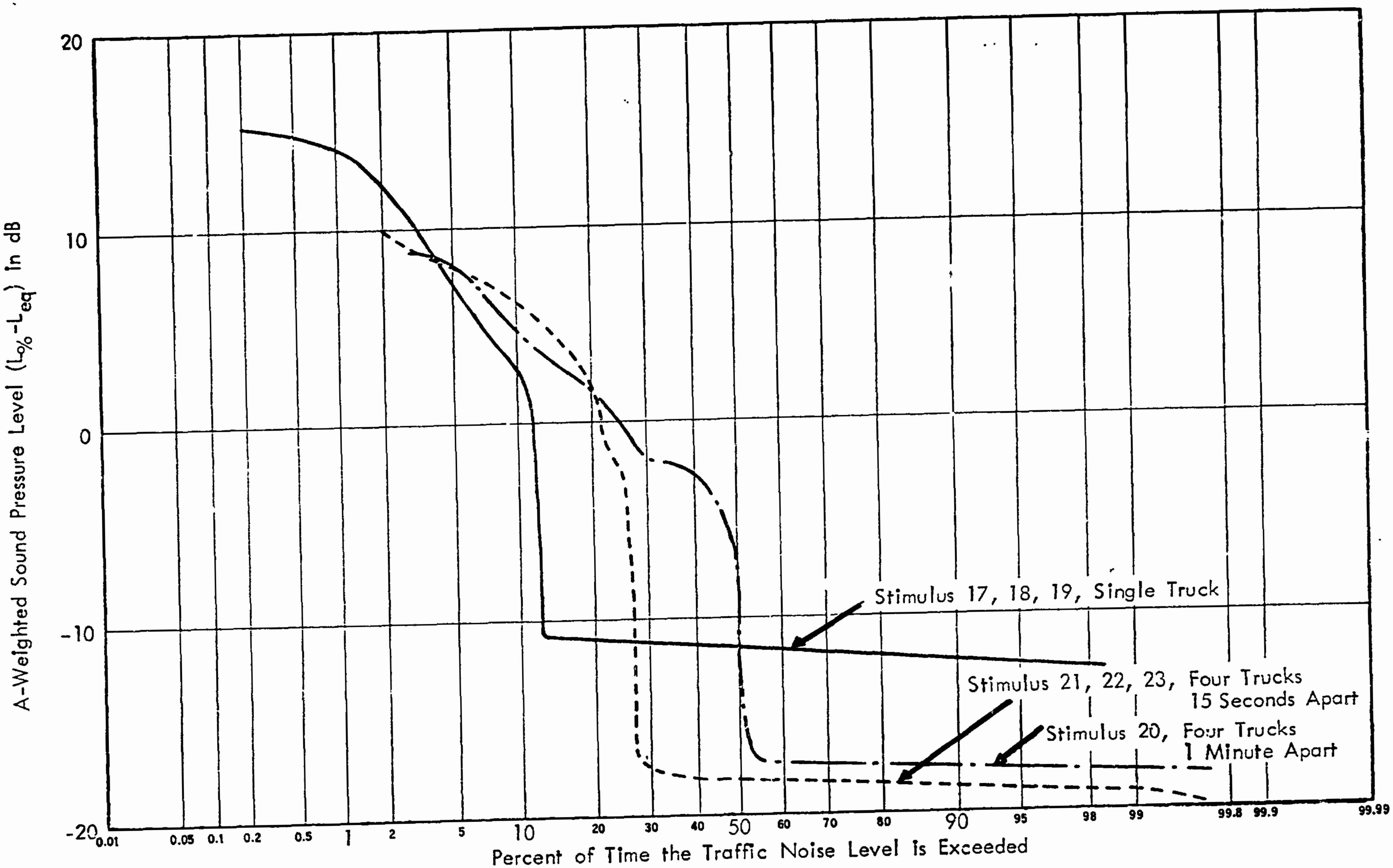


FIGURE 8. - COMPARISON OF VARIOUS MEASURES OF SPEECH INTELLIGIBILITY



9. - CUMULATIVE DISTRIBUTION FOR SIMULATED SAMPLES OF TRAFFIC NOISE

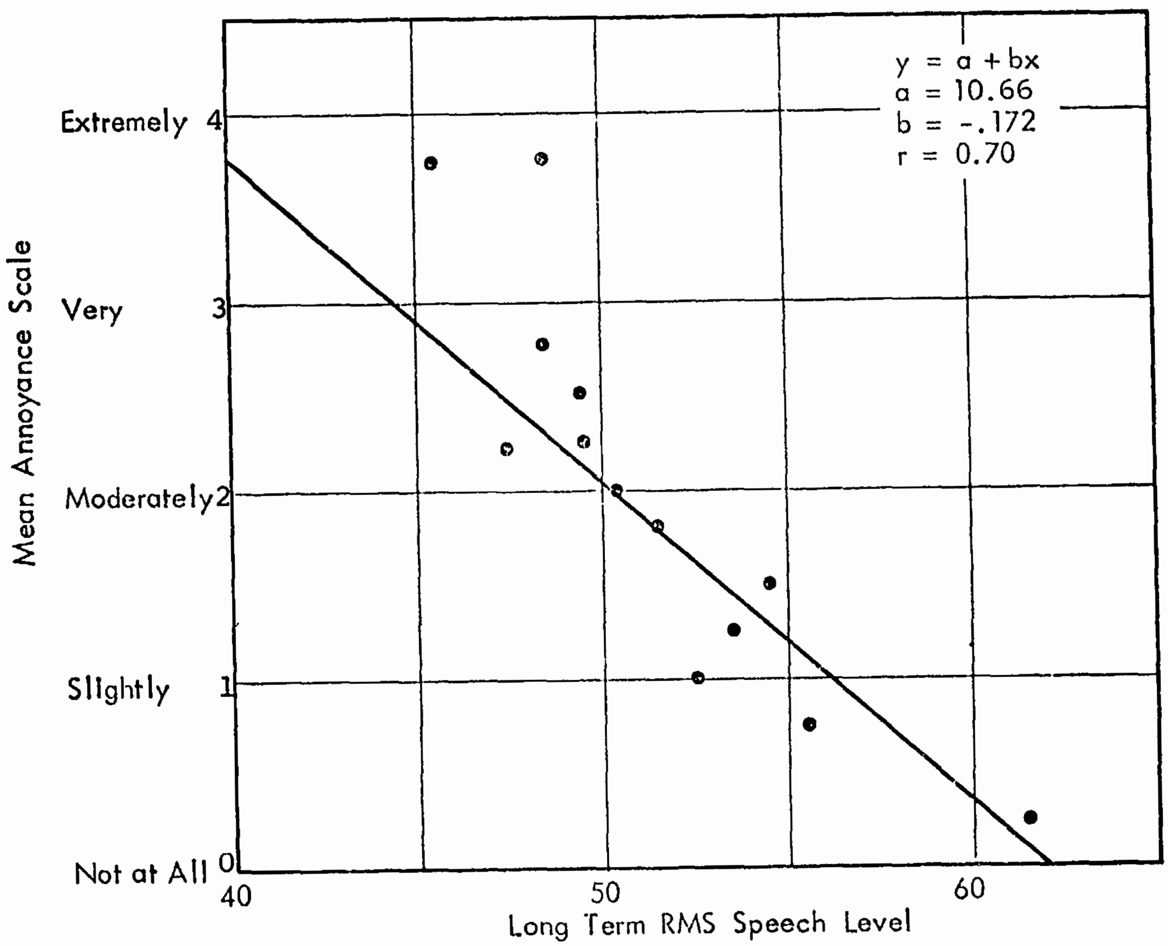


FIGURE 10.- ANNOYANCE RATINGS OF ONE TRAFFIC NOISE SAMPLE - L_{eq} = 59 dBA

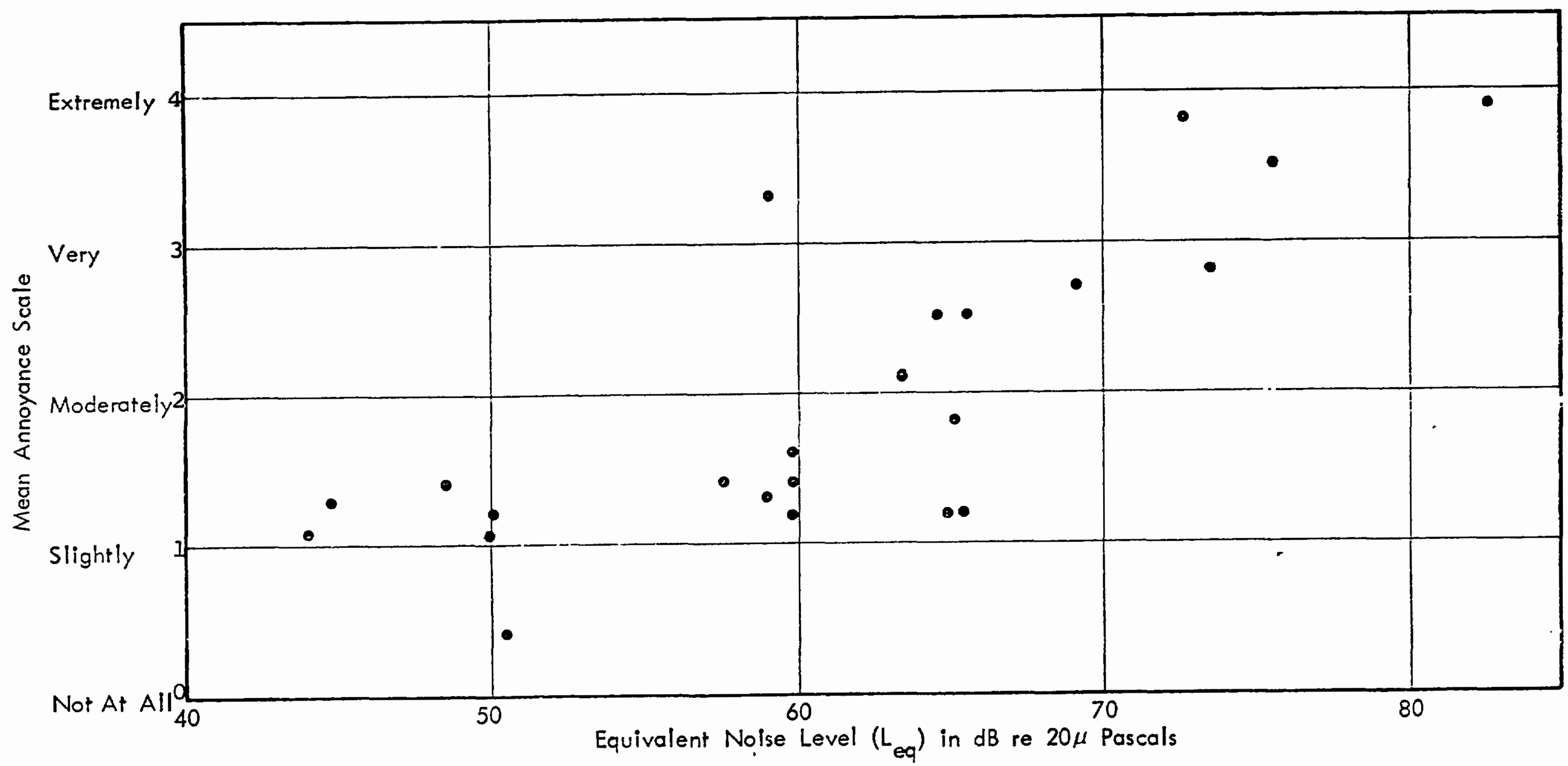


FIGURE 11.- ANNOYANCE RATING OF TRAFFIC NOISE - WITHOUT SPEECH

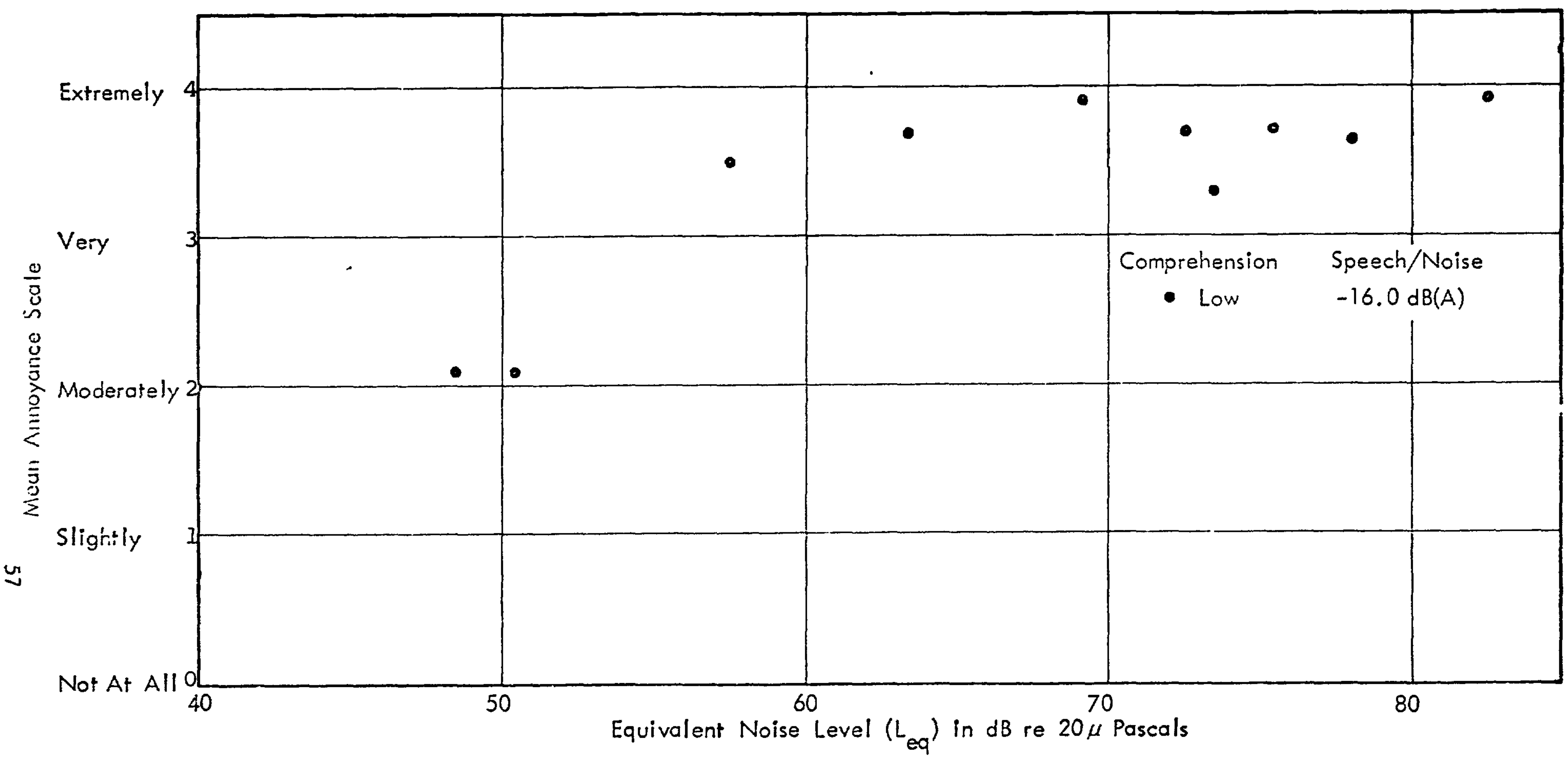


FIGURE 12.- ANNOYANCE RATINGS OF TRAFFIC NOISE - WITH SPEECH

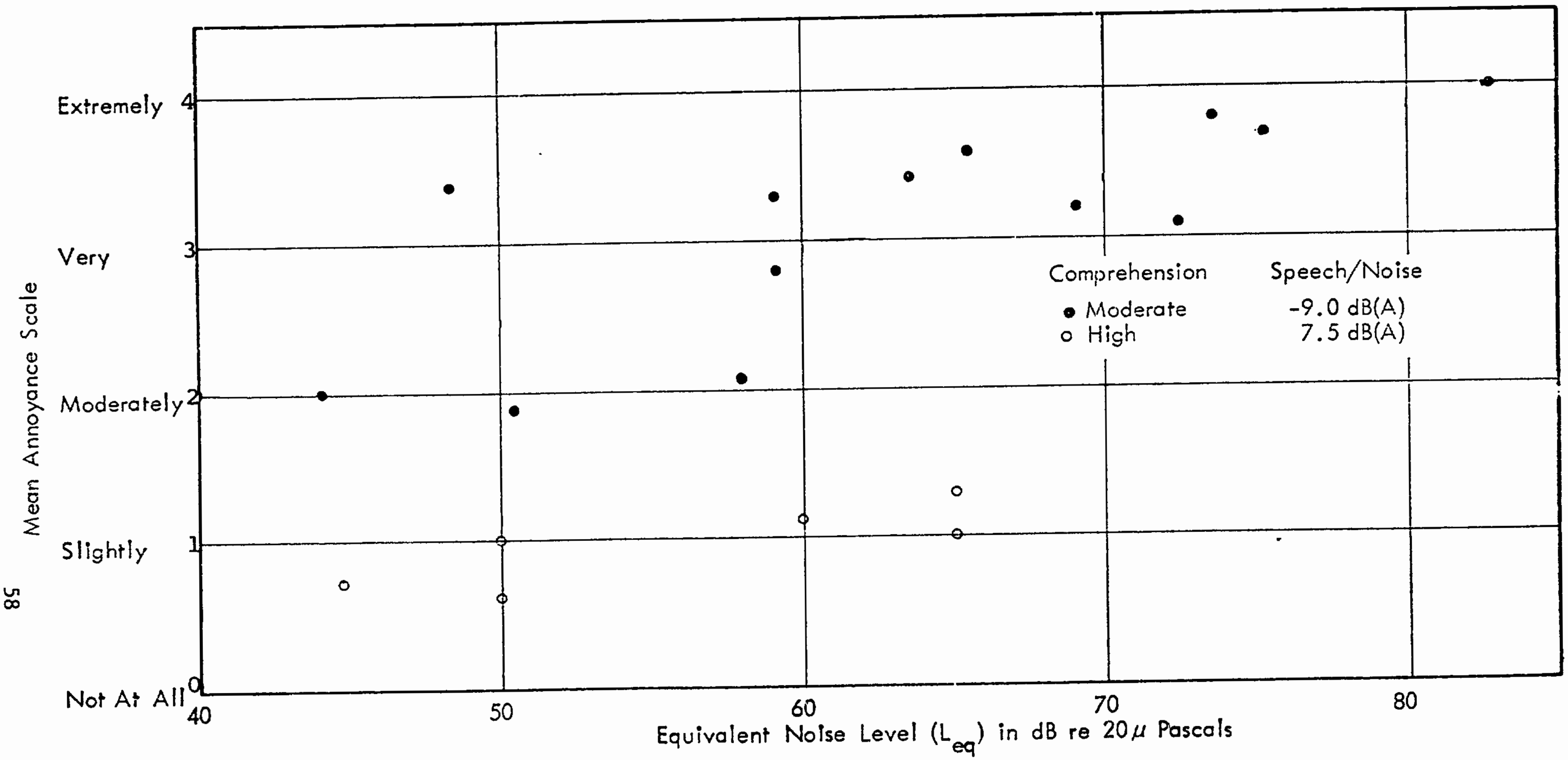


FIGURE 13.- ANNOYANCE RATINGS OF TRAFFIC NOISE - WITH SPEECH

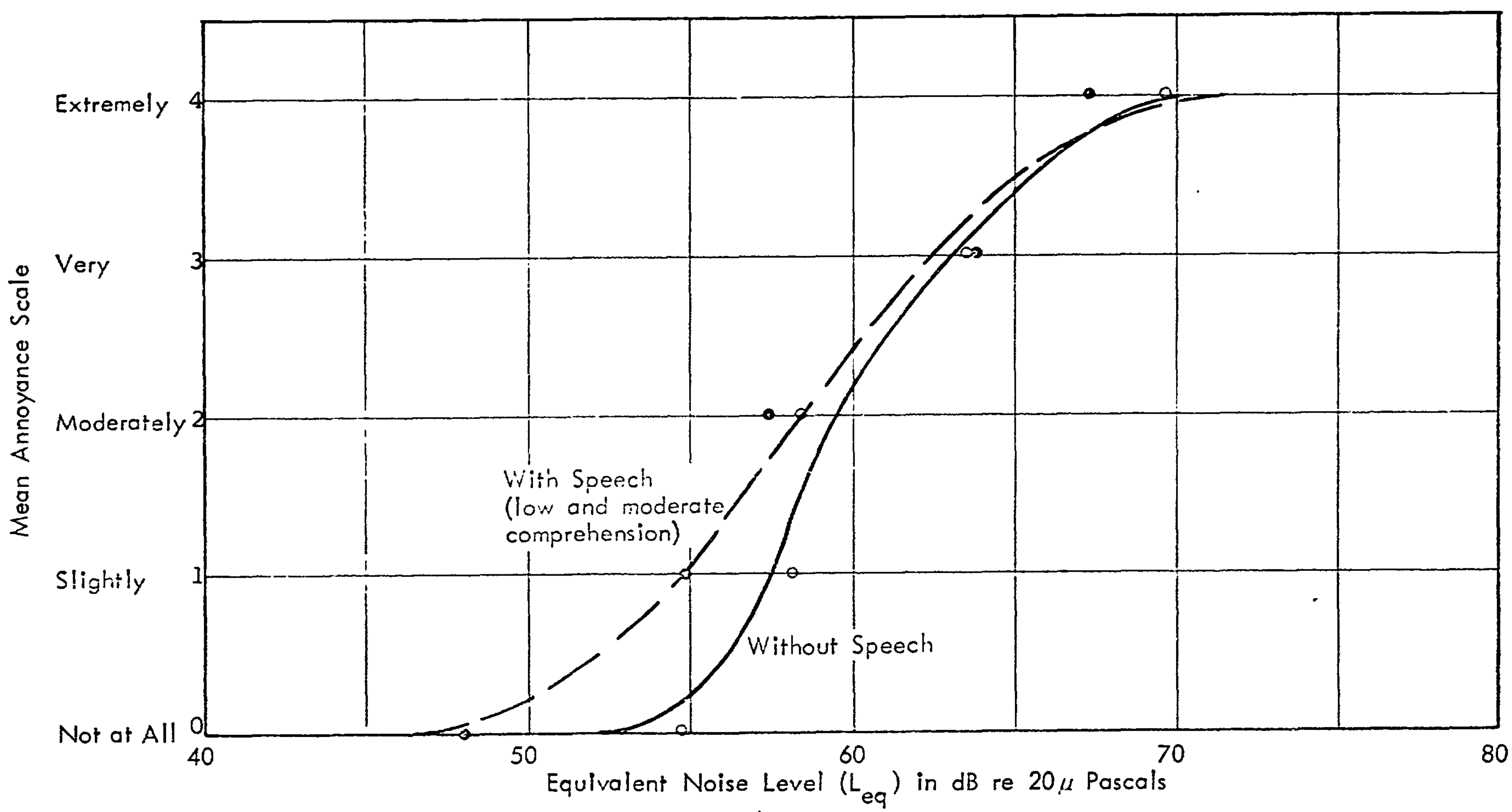


FIGURE 14.- ANNOYANCE RATING OF TRAFFIC NOISE - WITH AND WITHOUT SPEECH

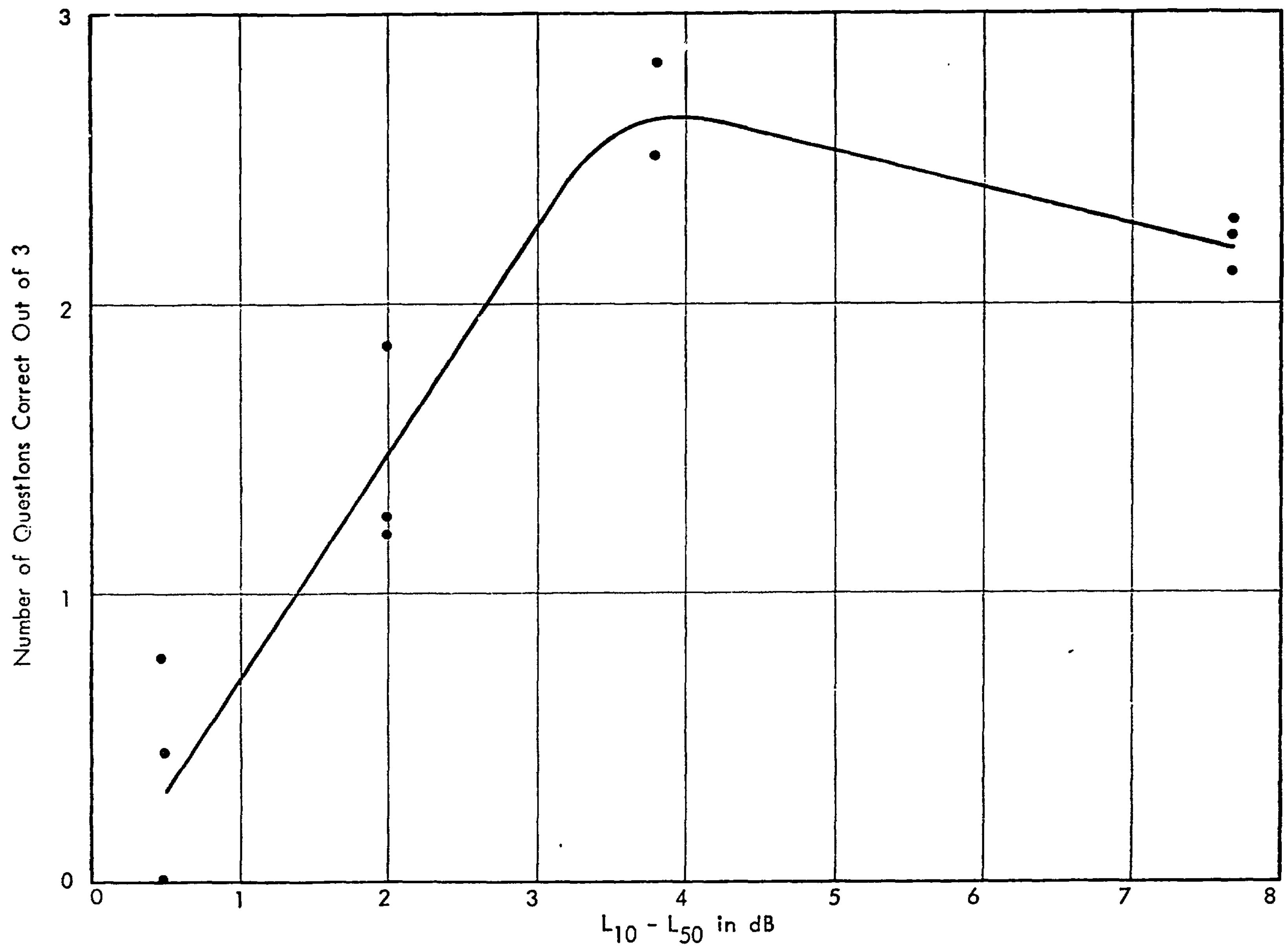


FIGURE 15. - SPEECH COMPREHENSION IN TRAFFIC NOISE FOR AN AVERAGE SPEECH/NOISE RATIO OF -9 dB

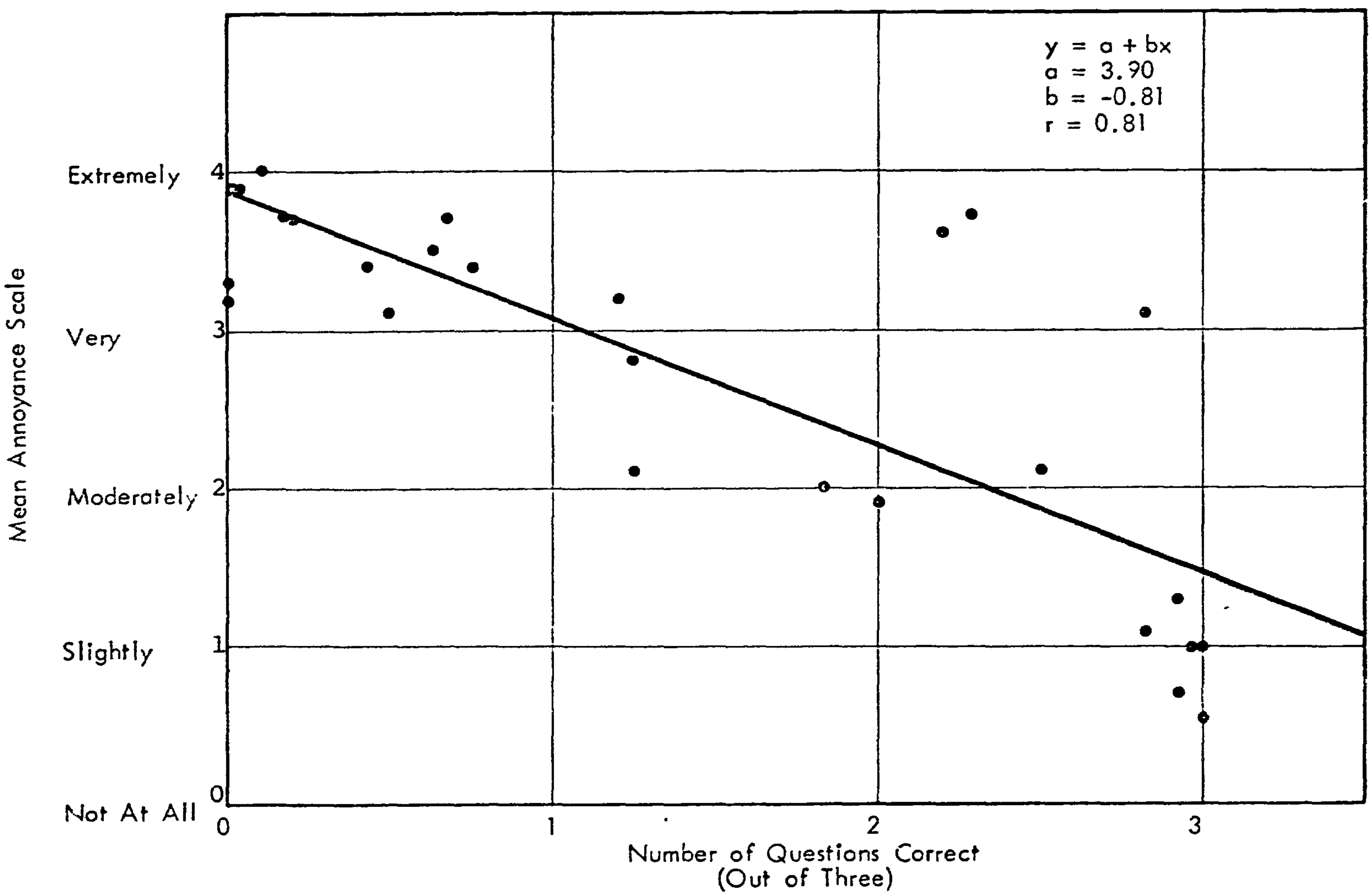


FIGURE 16.- ANNOYANCE RATING OF TRAFFIC NOISE FOR VARIOUS AMOUNTS OF SPEECH COMPREHENSION

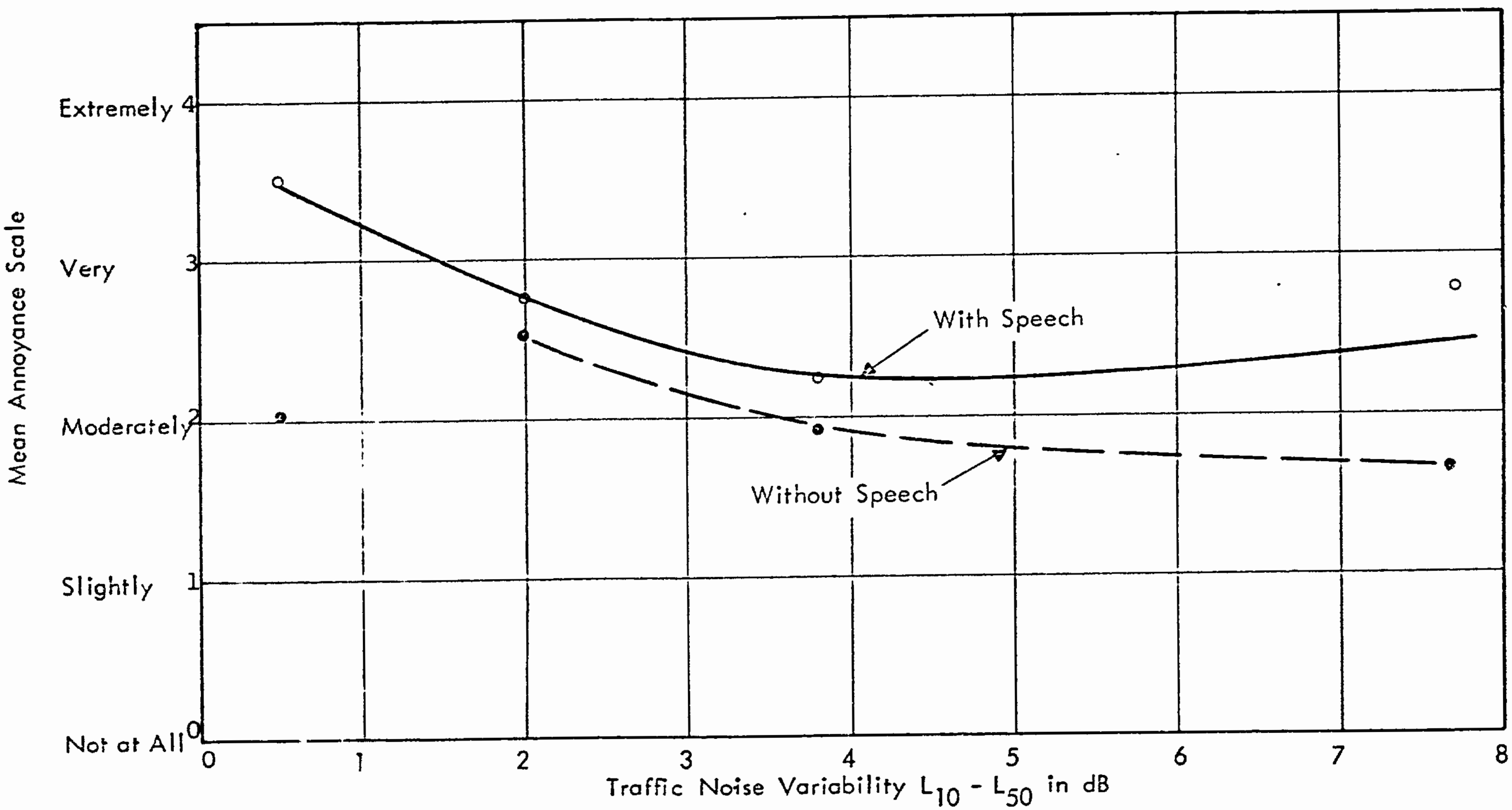


FIGURE 17.- ANNOYANCE RATINGS OF TRAFFIC NOISE - NORMALIZED TO L_{eq} = 60 dB(A)

EFFECTS OF THREE ACTIVITIES ON ANNOYANCE RESPONSES TO RECORDED FLYOVERS

By Walter J. Gunn and William T. Shepherd, NASA Langley Research Center, Hampton, Virginia, and John L. Fletcher, Memphis State University, Memphis, Tennessee

ABSTRACT

Subjects participated in an experiment in which they were engaged in TV viewing, telephone listening, or reverie (no activity) for a 1/2-hour session. During the session, they were exposed to a series of recorded aircraft sounds at the rate of one flight every 2 minutes. Within each session, four levels of flyover noise, separated by 5dB increments, were presented several times in a Latin Square balanced sequence. The peak level of the noisiest flyover in any session was fixed at 95, 90, 85, 75, or 70 dBA. At the end of the test session, subjects recorded their responses to the aircraft sounds, using a bipolar scale which covered the range from "very pleasant" to "extremely annoying." Responses to aircraft noises were found to be significantly affected by the particular activity in which the subjects were engaged. Furthermore, not all subjects found the aircraft sounds to be annoying.

INTRODUCTION

Interference with TV viewing is a major aircraft noise-related problem of airport community residents (ref. 1). Williams, Stevens, and Klatt (ref. 2) used a 10-point rating scale to obtain judgments of the acceptability of individual aircraft flyover noises while subjects either watched television or did not watch television. The ratings with or without TV viewing were almost identical. Langdon and Gabriel (ref. 3) conducted a series of experiments in which subjects watched videotaped television programs and, at the end of each period, rated the acceptability of the total noise exposure during that period. In these experiments, noise level was found to produce "significantly" less effect than predicted by the Williams, Stevens, and Klatt (ref. 2) data. The authors concluded further that "there is, however, almost certainly some positive effect, which contradicts a pure masking hypothesis." Given, however, the number of subjects per group and 95 percent confidence limits of about one unit, it is difficult to accept this conclusion without a test for significance. There is no obvious effect of level on acceptability which can be seen in their Experiments I and II data.

A model of human response to aircraft noise was recently developed by Gunn and Patterson (see Appendix A). This dynamic stress-reduction model predicts, among other things, that subjects engaged in different activities, when exposed to the same aircraft noise environment will respond with differing degrees of expressed annoyance. In order to test this hypothesis and learn the extent to which the specific activity engaged in effects one's annoyance reaction to aircraft noise, a laboratory experiment was performed as a part of a joint NASA/ Memphis State University research program and is described in this report.

PROCEDURE

Subjects

Subjects were 324 members of the university community at Memphis State University. All were screened for normal hearing and those with HL greater than 20 dB (ISO) were excluded from the study. Hearing of subjects was evaluated by a graduate student in audiology at the Memphis Speech and Hearing Center. Subjects were paid for their participation in this experiment.

Method

The 324 subjects were randomly divided into three groups of 108. Each of these groups were exposed (in subgroups of 6) to 1/2-hour of recorded aircraft landing noises. At the end of the 1/2-hour session, subjects were asked to indicate their general response to the aircraft sounds they had heard. The first group (reverie group), which was comprised of 18 subgroups of 6, simply sat and listened to the aircraft noises. The second group watched a preferred TV show during exposure to the aircraft noise and the third group listened to a recorded Modified Rhyme Test over a telephone during the aircraft noise exposure. In short, three groups of subjects were exposed to recorded aircraft noises and made judgments of annoyance at the end of the 1/2-hour session. The only difference in conditions between the three groups was the activity in which the subjects were engaged during the exposure to the aircraft noises. Table 1 shows the test sequence for each of the three groups.

Reverie

Subjects were ushered into the test room and seated. Seats were arranged before a loudspeaker so that the noise exposure would be equivalent for all subjects who were then left to themselves for a period of 15 minutes. This time was needed to provide a uniform experimental situation compared to the other two activities. Talking was permitted in this pretest period. Near the end of the 15-minute period, the experimenter reentered the room and read the instructions given in Appendix B. After this, the experimenter left the room and a tape recording of aircraft flyover sounds was activated. The same aircraft recording was used during all three activities. These flyover sounds and the method of presentation are described in the Apparatus and Stimuli sections of this report. At the end of the experimental session, the experimenter entered the room and distributed copies of the response sheet which is shown in figure 1. The scale used was bipolar and subject responses were not biased by the use of plus or minus signs at either end of the scale. Similarly, the flyover stimuli were never described as "aircraft noises" but rather as "aircraft sounds."

TV Viewing

Subjects were ushered into the test room and seated in an arc before a color television set. The TV set was situated in front of the loudspeaker mentioned previously, as it was in the no-task condition. These subjects had earlier indicated that the program they were about to watch was one of their favorite programs. The TV set was turned on and the subjects were read the instructions shown in Appendix C and the TV audio volume control was adjusted to a level acceptable to all subjects. Two minutes prior to the beginning of the program, the subjects were read the instructions shown in Appendix B. The TV set was then turned on to the selected program and the experimenter left the room. The aircraft flyover noise tape was immediately activated at the beginning of the TV program. After the last aircraft flyover in this session, the television set was left on so as not to cause changes in subjects' annoyance that would be unrelated to the flyover sounds. The experimenter quietly distributed copies of the response sheet shown in figure 1 and indicated that they were to complete this form according to the written instructions. After all subjects had completed this response form, the experimenter collected them and distributed copies of the response form shown in figure 2.

Telephone Listening

Prior to the beginning of this phase of the experiment, a pilot study was conducted with several listeners to determine the playback levels that would be required to achieve an average of about 90 percent correct on the speech interference tests, in quiet. This was done so that performance on the tests would be degraded even further during simulated aircraft flyovers. It must be remembered that the measure of primary concern here was annoyance related to the interference with telephone use, not speech intelligibility, per se. It was necessary to use an intelligibility test to provide a device that would hold subjects' attention to verbal stimuli.

Subjects in this phase of the study were ushered into the test room and seated. Beside each seat was a telephone handset. The subjects heard the instructions shown in Appendix D. The first instruction was read to the subjects by the experimenter. The second instruction was tape recorded and given to the subjects over the telephone handsets. Following these recorded instructions, the experimenter read to the subjects the instructions shown in Appendix B. (These latter instructions were read to all subjects in each phase of the experiment, thus providing maximum uniformity in instructions.) The experimenter then left the room and the recorded speech and aircraft noise stimuli were presented.

Six lists of the Modified Rhyme Test (MRT) as developed by House, et al., 1963 (ref. 4) were presented to subjects. The answer ensembles in these tests consist of six words each with a total of 50 ensembles per test. Prior to tape recording the tests, the correct word from each ensemble was selected by use of a table of random numbers. The tests used are shown in Appendix E. The recorded test word is underlined in each ensemble. Subjects' response forms were identical to the lists shown in Appendix E, except that no words were underlined, of course. Subjects were required to draw a line through the correct word in each ensemble per the instructions given in Appendix D. At the end of the experimental session, the experimenter collected the speech test response forms and distributed copies of the response form shown in figure 1. These forms were then completed by the subjects and collected by the experimenter.

Apparatus

The apparatus used in this experiment is shown in block diagram form in figure 3. During the TV viewing and reverie conditions, the speech track was disconnected at the tape recorder. The voltmeter was used to set noise and speech levels prior to each experimental session. The color TV set was positioned in front of the Klipschorn speaker in such a way that it did not significantly block the sound output from the speaker during presentation of aircraft flyover sounds. The test room was a 15 x 24 ft room furnished to resemble a living room. Ambient noise level in the room was 43 dBA as determined with a sound level meter set on slow reading position.

Stimuli

Aircraft noise.- Each subgroup of subjects was exposed to a 1/2-hour duration playback of recorded Boeing 747 landing sounds at the rate of one overflight every 2 minutes. In order to make the noise exposure a little more realistic, the peak levels of the individual flyover noise were varied from one overflight to the next. Within any session, there were four peak levels of aircraft noise, designated A, B, C, and D. There were 16 overflights during each 30-minute session and there were four overflights at each level A, B, C, and D, in a balanced Latin Square sequence. Table II shows the corresponding sound levels for each peak flyover level and figure 4 shows a plot of noise level, in dBA, versus time. For each activity, the aircraft noises, in general, were presented at six intensities, designated "Intensity 1, 2, 3, 4, 5, 6." As can be seen by inspection of Table II and figure 4, the most intense aircraft sound in intensity 1 is 70 dBA peak and the other peak levels within that session decrease to 55 dBA in 5 dB increments. Likewise, in intensity 2, the most intense aircraft sound is 75 dBA and the quietest is 60 dBA, and so on.

Speech stimuli.- The experiment involved the presentation of speech as well as aircraft flyover sound stimuli. The same flyover stimuli were presented during all three activities, i.e., reverie, TV viewing, and telephone listening. Controlled speech stimuli were presented only during the telephone listening phase of the experiment. The two sets of stimuli (aircraft and speech) were recorded on two tracks of a single tape. This provided synchrony between the speech and flyover stimuli. The speech stimuli were recorded in a commercially available sound treated room by a speaker of general American English. Speech stimuli were recorded at the rate of approximately one word every 6 seconds. The test word was appended to the phrase; "number _______ is ______," where the last blank corresponds to the position of the test word. The talker monitored his voice level with a VU meter during recording of speech stimuli. Speech stimuli were recorded on one tape track on a high quality audio tape recorder with a commercially available dynamic microphone. The recorded speech material is shown in Appendix E. Speech stimuli were played to listeners at constant level such that the speech peaks were approximately 50 dBA in the telephone handsets as measured in a 6cc coupler.

The aircraft flyover stimuli were recorded on the second track of the tape. The two tracks were juxtaposed so that the first word of the speech stimuli and the beginning of the first flyover occurred at about the same time. Flyover levels were calibrated in the test room using a sound level meter. A corresponding voltage for a calibration tone on the tape was observed and recorded. These voltages were used in subsequent sessions to set the correct flyover levels. These calibrations were checked periodically during the experiment to insure consistency of stimuli presentation. A diagram showing the level of stimuli presented to subjects and the activity they were performing is shown in Table III.

Stimuli analysis.- The aircraft flyover sounds were recorded as they occurred in the test room using commercially available acoustic analysis recording equipment. The sounds were recorded at the extreme levels of 95 and 70 dBA at several seat positions normally used by subjects. In addition, a recording of the speech signal was made with one of the handsets coupled to the microphone while the aircraft flyover sounds emanated simultaneously from the loudspeaker. These recorded stimuli will be analyzed at a computer facility and results will be available sometime in the near future for a more detailed analysis of the relationships between actual speech interference and the physical description of the noise.

RESULTS

Figure 5 shows the median annoyance scores versus session intensity level for each activity in which S's were engaged during the aircraft noise exposure. The three regression lines were significantly different from each other, i.e., the slope of the "telephone listening" line was significantly ($p < .05$ by t test) different than the slopes of the "TV Viewing" and "Reverie" regression lines and median values of the "TV Viewing" regression line differed significantly ($p < .05$ by median test) from those of the "Reverie" regression line. Median tests of the differences of annoyance at each session intensity show that annoyarce resulting from noise interruption of TV viewing at intensity 1 was significantly ($p< .05$) greater than that for either "Reverie" or "Telephone Listening," while at intensity level 5, the relation is reversed for "TV viewing" and "telephone listening." That is to say, in the session in which the loudest aircraft noise was 70 dBA peak, those subjects viewing TV expressed greater annoyance than those listening to speech stimuli on the telephone or those engaged in reverie (no task). As the aircraft noise intensity increased to the point where the loudest aircraft sound was 90 dBA peak, the annoyance of those engaged in the telephone listening task grew to the point where it was significantly greater than the annoyance of those engaged in the other two tasks.

Table IV shows the frequency distribution of annoyance scores for all intensity levels and activities. Note that 17 subjects (over 5 percent of the 324 who participated in this experiment) reported that the aircraft sounds were "pleasant" to hear.

DISCUSSION

The results suggest that the "telephone listening" task provides a much more sensitive indicator of peoples' overall annoyance response to aircraft noise than either "TV viewing" or "reverie" situations. While on the surface the results might at first seem to be at variance with past studies which show fairly high correlations between noise level and the resulting annoyance reaction in the no-task situation, careful consideration of the procedures and conditions of this experiment makes the results of this study more understandable. To begin with, it is widely known that laboratory subjects judging the loudness or noisiness of individual noises covering a given intensity range will quite neatly order the stimuli as an increasing monotonic function of the intensity level, clearly demonstrating that they can discriminate intensity levels, if nothing else. Note, however, that the subjects in these experiments made only one judgment of the effect of a 1/2-hour exposure to aircraft noises presented at various intensity levels at the rate of about one flight every 2 minutes. The experimental situation was contrived such that the subjects were not required to discriminate one intensity from another, but rather that they were to report their reactions to one specific exposure condition. This is not to say that the subjects did not use a standard against which to compare their reactions to the experimental stimuli. They could, conceivably, have an existing internal standard developed from real life experiences against which to compare the integrated effects of the laboratory noise exposure. The practice of obtaining only one response from each subject has much in common with the assessment of individual reactions of airport community residents to their own neighborhood noise environment. It is common practice in social surveys dealing with community response to aircraft noise to ask individuals to rate their own noise environment on various numerical category scales. In such studies, the respondents are not usually asked to rate more than one noise environment, their own. It is not surprising, therefore, that most such studies have found rather poor correlations between noise levels in the environment and reported annoyance reactions. It is clear from our data that the growth and absolute level of annoyance differ depending on which specific activity is interrupted by the intruding aircraft noise. With reference to the stress-reduction model of Appendix A, the data support the hypothesis that reaction to noise is modified by the nature of the activity engaged in at the time of the noise. A viable predictor of annoyance reaction to aircraft noise must then account for the "dominant" activity in a given community during each noise exposure period. It would not be surprising to find in future experiments still another (and totally different) psychophysical function relating annoyance and noise level which occurs during and possibly interrupts sleep. The same could be said for the reactions of people engaged in various other activities. While both our TV viewing task and telephone listening task involved aural communications, the telephone listening task differed in a number of important ways. Firstly, there was no redundancy built into the speech test presented over the telephone while there is a certain amount inherent in the usual TV show. Secondly, the importance of speech intelligibility was artifically increased in the telephone listening task by offering a bonus for superior speech reception scores. The differences in annoyance during TV viewing and reverie suggest a possible different basis for the annoyance reaction in each situation. One might speculate that the significantly greater annoyance reported by the TV viewers in intensity level 1 (where the loudest overflight was only 70 dBA peak) may have been due to distraction, rather than communication interference from masking, per se.

CONCLUDING REMARKS

It is concluded that the results of this experiment support the Gunn/Patterson Stress Reduction Model in that the degree of annoyance experienced by people exposed to aircraft noise depends upon the nature of the specific activity in which they are engaged at the time of the noise exposure. The finding that some laboratory subjects, over 5 percent, find the aircraft noises to be somewhat pleasant indicates the need for a closer look at the validity of laboratory studies, especially those in which subjects are required to respond on a unipolar scale of annoyance which does not allow for the possibility of some subjects who find the noises, at least in a laboratory setting, to be pleasant to hear. The speech communication task appears to be the most sensitive procedure for the laboratory assessment of the effects of different levels of aircraft noise exposure.

REFERENCES

1. Galloway, W. J.; and Bishop, D. E.: Noise Exposure Forecasts, Evolution, Evaluation, Extensions, and Land Use Interpretations. Bolt Beranek and Newman, Inc. Tech. Rep. FAA-NO-70-9, 1970.

2. Williams, C. E.; Stevens, K. N.; and Klatt, M.: Judgments of the Acceptability of Aircraft Noise in the Presence of Speech. J. Sound Vib., vol. 9, 1969, pp. 263-275.

3. Langdon, L. E.; and Gabriel, R. F.: Judged Acceptability of Noise Exposure During Television Viewing. J. Acoust. Soc. Am., vol. 56, 1974, pp. 510-515.

4. House, A. S.; Williams, C.; Hecker, M.; and Kryter, K.: Psychoacoustic Speech Tests: A Modified Rhyme Test. TDR No. ESD-TDR-63-403, Decision Sciences Laboratory, U. S. Air Force.

NAME________________________ SUBJECT NO.

PLEASE INDICATE YOUR GENERAL REACTION TO THE AIRCRAFT SOUNDS WHICH WERE PRESENTED DURING THE SESSION BY PLACING A CHECK MARK NEXT TO THE APPROPRIATE POINT ON THE SCALE SHOWN BELOW.

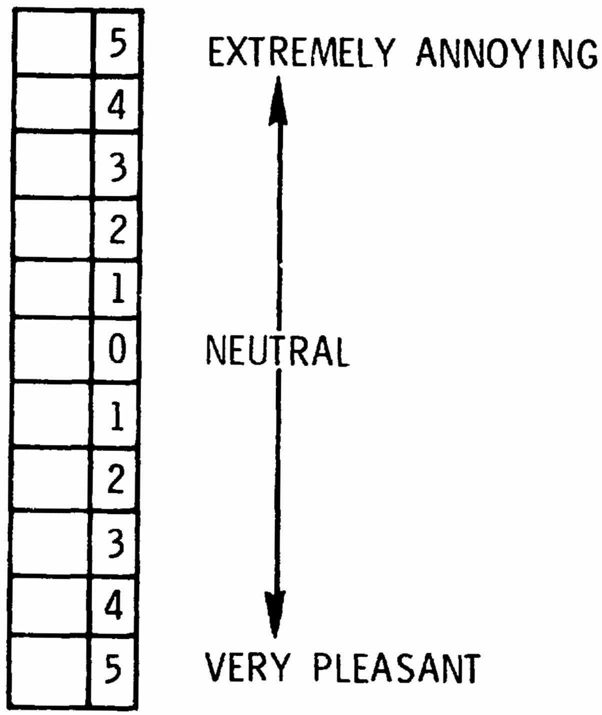


Figure 1.- Subject response sheet.

PLEASE ANSWER THE FOLLOWING QUESTIONS BY CHECKING THE APPROPRIATE BOX.

- HOW WOULD YOU RATE THE TV SHOW YOU WATCHED?

  ☐ EXCELLENT ☐ GOOD ☐ FAIR ☐ POOR

- HOW WOULD YOU RATE THE TV SOUND LEVEL?

  ☐ TOO QUIET ☐ JUST RIGHT ☐ TOO LOUD

- WHAT BOTHERED YOU THE MOST ABOUT THE AIRCRAFT SOUNDS? (WRITE A FEW WORDS TO DESCRIBE YOUR FEELINGS.)

Figure 2.- Subject response sheet 2.

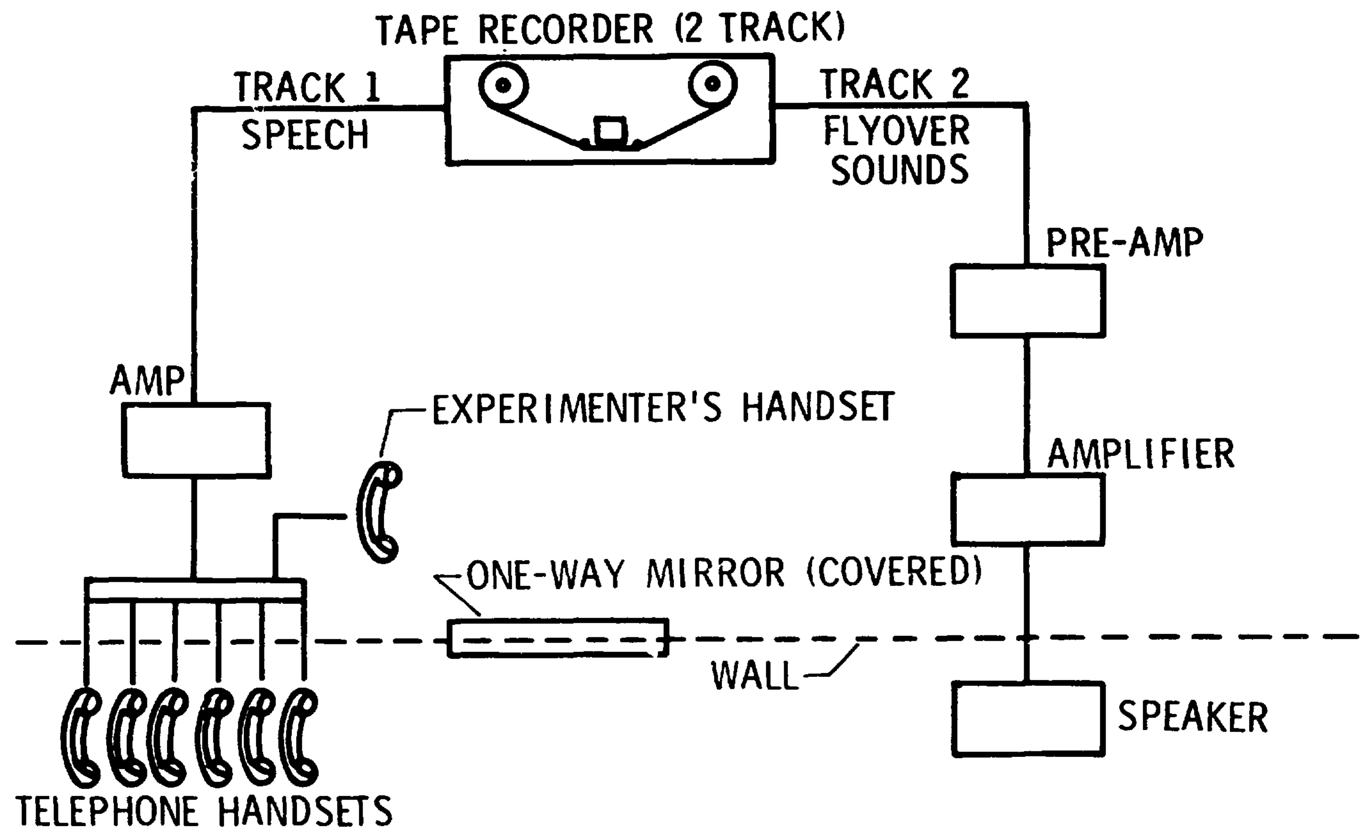


Figure 3.- Apparatus.

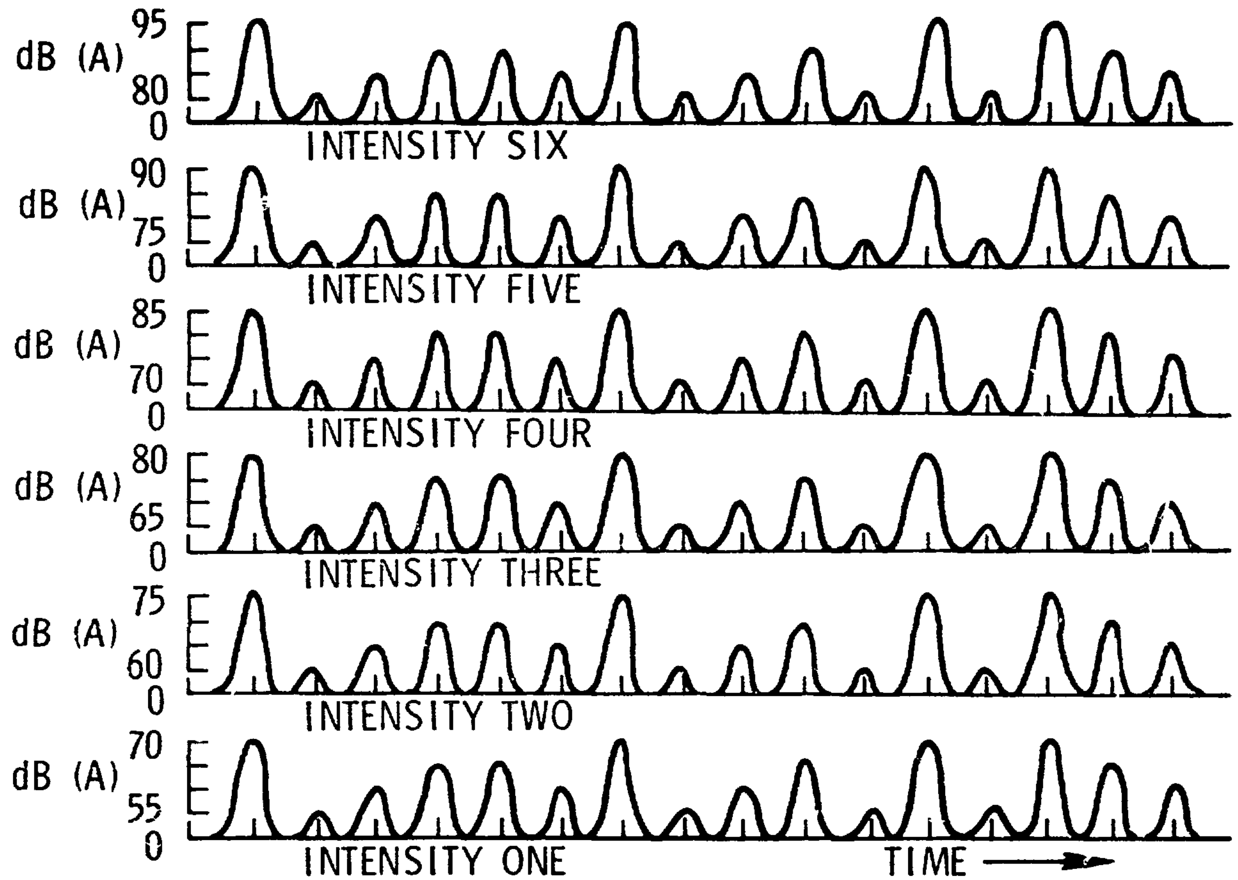


Figure 4.- Aircraft flyover noises.

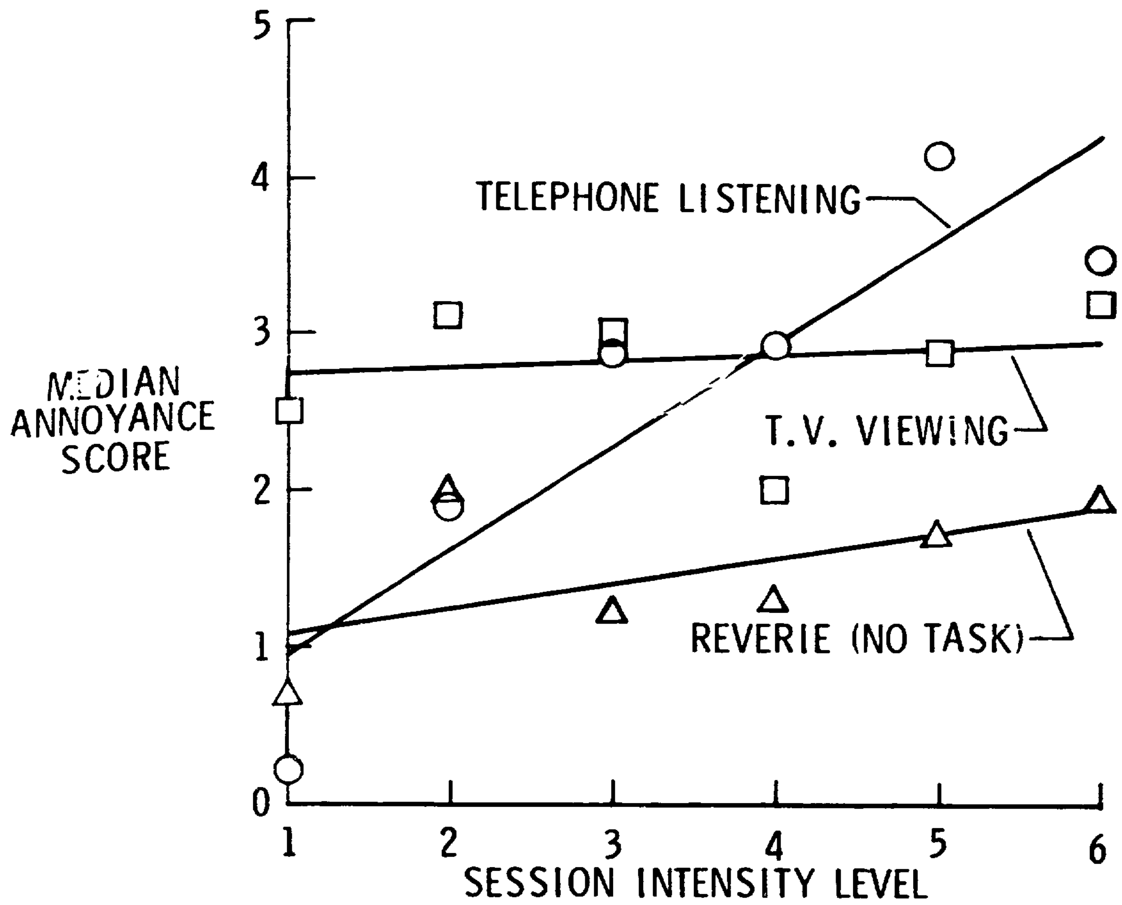


Figure 5.- Effects of activity interruption

TABLE I - TEST SEQUENCE

15 MINUTES	30 MINUTES	5 MINUTES	5 MINUTES
Reverie (no task)			
S's sit and talk freely, Instruction "A" read to S's	S sits; talking not permitted	S's complete Data Sheet 1	
TV Viewing			
TV audio adjusted and Instructions "B" and "A" read to S's	S views TV program previously selected	S's complete Data Sheet 1	S's complete Data Sheet 2
Telephone Listening			
Instruction "C" and practice given to S's; then instruction "A"	S listens to telephone for speech reception test	S's complete Data Sheet 1	

TABLE II - PEAK AIRCRAFT FLYOVER LEVEL IN dBA

Stimulus Designator	Session Intensity Level					
	1	2	3	4	5	6
A	70	75	80	85	90	95
B	65	70	75	80	85	90
C	60	65	70	75	80	85
D	55	60	65	70	75	80

TABLE III - SUBJECT ASSIGNMENTS

	Session Noise Intensity Level					
	1	2	3	4	5	6
Peak Level of Most Intense Aircraft Noise During Exposure, in dBA	70	75	80	85	90	95
Activity						
No Task	S1-S18	S19-S36	S37-S54	S55-S72	S73-S90	S91-S108
TV Viewing	S109-S126	S127-S144	S145-S162	S163-S180	S181-S198	S199-S216
Telephone Listening	S217-S234	S235-S252	S253-S270	S271-S288	S289-S306	S307-S324

TABLE IV - FREQUENCY DISTRIBUTION OF SCORES

Very Pleasant					Neutral					Extremely Annoying	Subject Response Scale	
-5	-4	-3	-2	-1	0	1	2	3	4	5	Median	Condition
				3	5	6		2	2		.67	70 Rev
					4	4	2	6	1	1	2.0	75 Rev
		1		1	2	1	2	2	2	1	1.2	80 Rev
1			2		1	6		2	5	1	1.3	85 Rev
	1	1			3	3	5	5			1.7	90 Rev
				1	4	1	1	4	1		1.93	95 Rev
					2	3	4	7	2		2.50	70 TV
				1			3	8	4	2	3.12	75 TV
					3	1	3	4	3	4	3.0	80 TV
				1	2	4	4	2	3	2	2.0	85 TV
						4	2	5	3	3	2.9	90 TV
						2	2	7	3	4	3.21	95 TV
		1	1		9	1	2	2	1		0.2	70 Tel
		1			1	5	5	2	1	3	1.9	75 Tel
					1	1	4	8	3	1	2.87	80 Tel
						2	4	7	1	4	2.93	85 Tel
						1	1	3	6	7	4.17	90 Tel
						1	4	4	4	5	3.5	95 Tel

APPENDIX A

THE GUNN/PATTERSON STRESS REDUCTION MODEL

Walter J. Gunn
NASA Langley Research Center
Hampton, Virginia

Harrold Patterson
Tracor, Inc.
Austin, Texas

In the development of a methodology for the assessment of community response to aircraft noise, an important concern is the identification of specific measurable changes exhibited by the exposed community. Following this, the psychophysical relationships between the cause (noise) and effect (community response) need to be determined. To increase the meaningfulness of the predicted response, relationships between response categories should also be determined. For example, if the mean annoyance of a given community is 4.8 (on a scale of 6) and this is designated as "very annoying," very little information regarding the actual state of mind of the average community resident is known. If, however, the relationship between annoyance, desire to move out of the neighborhood, health effects, sleep loss, hearing loss, activity interruption, and degradation of the perceived quality of life are predictable from knowledge of the degree of annoyance, for instance, then the information becomes considerably more meaningful to the various users, such as aircraft designers, airport operators, pilots, legislators, and public administrators.

Some of the specific measurable changes exhibited by airport community residents resulting from aircraft noise can be determined by answers to questions in social surveys, while certain behavioral changes can be directly observed or traced through official records, such as those of the telephone company, real estate offices, and hospitals. However, a specific model of individual reaction to aircraft noise is needed in order to determine better which specific changes may be anticipated and how they can be measured.

The initial attempt at formulation of a model* is shown in figure A1. This model is based upon the premise that individuals will attempt to reduce,

*The Stress Reduction Model was developed by W. J. Gunn of NASA, Langley Research Center and H. P. Patterson of Tracor, Inc.

avoid, or eliminate stress in their lives. Stress may be defined here as a general state of physical or psychological unrest. The model suggests that aircraft noise is perceived within two general contexts: situational and human factors. That is, qualities of the individual's physical, social, and psychological environments are important in his perception of the noise.

Only when the perception is "filtered" through the various meanings associated with the noise, through the interruption of activities and/or through evaluations of the aversive nature of the noise per se, is stress produced. The stress is manifested primarily in the development of negative feelings about the noise and in health problems. However, the individual will make every attempt to relieve this stress. Two methods are shown: overt behavior and internal adjustment. Overt behavior may be of various types, including complaint, retreating indoors or out of the neighborhood, and soundproofing the home. Internal adjustment is seen in adaptation, habituation, rationalization, and resignation to the noise. It is important to note that individuals who do not or cannot take overt action or who do not or will not make internal adjustments will develop more stress since the development of negative feelings and health problems themselves produce stress.

A. Stimulus Factors - The stimulus factors considered important in the model are divided into two general categories: noise and vibration.

(1) Noise

1. Level

2. Spectral characteristics

   a. General shape

   b. Discrete frequency content

3. Temporal characteristics

a. Time of occurrence

b. Duration

c. Impulsiveness

d. Dwell (temporal concentration)

4. Other characteristics

a. Rate of change of above

b. Directionality and movement

(2) Vibration

1. Level

2. Spectral content

3. Onset/offset characteristics

4. Correlation with the aircraft noise

5. Generation of secondary sounds (rattles, buzzes, etc.)

B. Situational Factors - The situational factors include the following: activity engaged in, setting, temporal factors, and other environmental conditions.

(1) Activity engaged in

The various activities which may be interrupted by aircraft noise are:

1. Relaxation (reverie)

2. Aural communications, whether active or passive, with or without visual cues

3. Sleep

4. Higher order cognitive functioning such as concentration, learning, problem solving, or reading

5. Physical activities

(2) Setting

The settings at times of noise exposure which may influence individual reaction are as follows:

1. At home or away

2. With others or alone

3. Indoors or out

(3) Temporal factors

The temporal factors which must be taken into consideration are:

1. Season

2. Day of week

3. Time of day

(4) Other environmental conditions

Other environmental factors which might effect stimulus conditions are as follows:

1. Presence and characteristics of nonaircraft sounds

2. Climatological conditions

   a. Temperature

   b. Relative humidity

   c. Atmospheric pressure

   d. Wind

   e. Precipitation

3. Illumination

4. Esthetics of surroundings, auditory, visual, tactile, and olfactory

C. Human factors - The human factors which may be influential in determining one's response to aircraft noise are divided into three general categories as follows: psychological factors, biological-physiological factors, and demographic factors.

(1) Psychological factors

There are at least seven psychological factors to be considered:

1. Attitudes
2. Intelligence
3. Traits
4. Needs
5. Self-concept
6. Values
7. State

(2) Biological-physiological factors

Important biological-physiological factors are:

1. Auditory sensitivity
2. Kinesthetic sensitivity
3. Condition: rested versus fatigued
4. General health
5/ State: relaxed versus tense

(3) Demographic factors

Possibly important demographic factors are:

1. Age
2. Sex
3. Occupation
4. Income
5. Education
6. Race
7. Class
8. Owner/Renter

9. Length of residence

10. Previous noise exposure

11. Dependence on aviation

D. Meaning associated with the noise - Kerrick, et al. (ref. A1) found that while noises from a variety of sources were rated equally on the basis of loudness or noisiness, they were not equally acceptable. Gunn, et al. (unpublished results of a study conducted by Langley Research Center personnel at NASA Wallops Station, Virginia) found that aircraft perceived as flying over an individual were rated as more annoying than aircraft perceived as flying off to the side, even at the same PNL. Connor and Patterson (ref. A2) found that "fear" of aircraft crashes was an important determinent of annoyance with aircraft noises. Wilson (ref. A3) found that aircraft noises were more acceptable and less noisy than motor vehicles at the same level. This suggests that the meaning associated with the source of the sound may have an important bearing on the degree of annoyance we feel about various sounds.

E. Activity interruption - In addition to the way we may feel about exposure to unpleasant sounds or the aversive meaning we attach to them, annoyance may result if the noise interferes with an ongoing activity, such as TV viewing, radio listening, sleeping, or activities requiring concentration. The extent of activity interruption could be assessed by questions on a social survey or through prediction based on controlled laboratory tests. There is good reason to think that interruption of these activities may contribute heavily to one's overall annoyance with aircraft noise.

F. Unpleasant characteristics of aircraft noise, per se - The range of possible feelings about the characteristics of a sound, per se, run the gamut from very pleasant, such as enjoyable music, to very unpleasant, such as a circular saw cutting sheetmetal. Similarly, certain aircraft sounds, at some levels, may actually be pleasant to hear, while other sounds may be perceived as neutral or unpleasant. Molino (ref. A4) developed what he calls "an equal aversiveness curve" for various bands of sound. The shape of the curve most closely resembled that of the inverse of the standard A-weighting characteristic. It is suggested that sounds above the threshold of aversiveness are "punishing" to the ear. Since the Molino data confounds aversiveness of the sound, per se, and interruption of concentration (the subjects were learning Russian during the experiment), the contour might be different under the condition of reverie. Clearly, there is a need to determine the psychophysical relationship between noise parameters and pleasantness or unpleasantness for various sounds. If a sound is perceived as being unpleasant to the ear, then continued exposure may lead to the development of stress in the unwilling listener.

G. Reported feelings - Airport community residents are often polled in order to determine how they feel about aircraft noise, airport operations, the people who are responsible, or the aircraft industry in general. The most commonly asked questions have to do with reported annoyance with aircraft noise. Sometimes people are asked for their overall annoyance, while in other cases they are asked about the annoyance they feel about the interruption of specific activities. In the latter case, the annoyance ratings for the interruption of various activities are usually combined in some way to form a single scale of annoyance. Although such a scale is typically well correlated with the single-question self-rating of annoyance (McKennell, ref. A5), it obviously represents only one particular dimension of annoyance and thus might best be termed "annoyance through disturbance of activities."

Questions are sometimes asked about feelings of "misfeasance" (feelings that those in authority are not doing all they could do to alleviate problems). Feelings of "fear of aircraft crashes" are also probed. The scales used to assess the various feelings are many and varied. Validity of the scales is, for the most part, assumed.

H. Health problems - While the evidence is scanty and sometimes in conflict, certain health-related problems resulting from aircraft noise may be:

1. Permanent hearing loss
2. Gastro-intestinal disorders
3. Increased nervousness
4. Cardio-vascular problems
5. Loss of sleep

Hospital and doctor's records might be helpful in assessing these aircraft noise related health effects.

I. Overt behavior - Few substantive studies have been conducted regarding the overt reaction of people to aircraft noise. Some important forms of overt behavior might be:

1. Moving family out of the noisy area
2. Complaints to authorities
3. Decrease in outdoor activities
4. Decrease in activities involving aural communications
5. Increased time spent out of neighborhood
6. Organizing to reduce the noise

J. Internal adjustment - The increased stress and the development of negative feelings and health problems represent an imbalance of the individual's normal or preferred state. In an effort to return to the normal state (homeostasis), the individual either takes overt action or makes internal adjustments, both of which serve to reduce the stress. Four types of internal adjustment are identified:

1. Adaptation
2. Habituation
3. Rationalization
4. Resignation

Thus, the individual may adapt to the noise or become habituated to it. Or, the individual may also rationalize his experience and convince himself that his situation is not so bad after all and that others are much worse off than himself.

K. <u>Feedback loops</u> - Every action or nonaction of the individual has a consequence. If the individual cannot or will not take overt action to reduce the stress, or if he does not make internal adjustments, then the development of negative feelings and health problems will themselves increase the stress. These relationships are shown in figure A1 by dashed lines from negative feelings and health problems back to stress. They represent positive feedback loops.

However, if the individual does take some overt action or makes an internal adjustment, then the stress will be relieved through an indirect process. Taking direct action has implications for both the stimulus and the situational factors. For example, through lobbying efforts, the individual may persuade the noise maker to reduce the noise or to change its characteristics so as to make it more tolerable. Or, the individual may change the situation by insulating his home, by spending less time outdoors (thereby decreasing his outdoor exposure time), or by moving out of the noise impacted area. If the individual makes an internal adjustment, this has implications for the human factors context. For example, the individual, in response to stress, may develop qualities of an "imperturbable" person. Such a person would deny that the noise <u>ever</u> bothered him and, in fact, might report difficulty in even perceiving the noise. These consequences of overt behavior and internal adjustment are represented by dashed lines back to the stimulus and situational factors for the former and back to human factors for the latter. Both are negative feedback loops.

L. <u>The nature of the "filter" variables</u> - As shown in the model diagram, there are no feedback loops to the boxes representing "meaning," "activity interruption," and "unpleasant characteristics." This means only that later elements within the model are not thought to affect these elements. Certainly, events outside the model have an effect. For example, if an aircraft crashes in the near vicinity, the individual may very well associate the next flyover event with a feeling of fear of crash. In a like manner, outside events are thought to produce a certain condition within the individual which tends to "color" his perception of aircraft noise. At any one point in time, these conditions work to predispose individuals to react in certain ways. Over time, however, the conditions can change and the individual's predispositions take on a dynamic character.

M. <u>Hypotheses</u> - A number of specific hypotheses are suggested by the stress reduction model. These are as follows:

1. Increased stimulus from aircraft operations will result in:
   a. increased development of negative feelings about the noise and/or
   b. increased development of health problems.

These results will be obtained provided the following elements are held constant:

(1) Situational factors

(2) Human factors

(3) Meaning associated with the noise

(4) Activity interruption

(5) Unpleasant characteristics of the noise, per se

2. The greater the development of negative feelings about the noise

a. the greater the amount of overt behavior directed toward reducing or eliminating the noise, and/or

b. the greater the internal adjustment of the individual.

The model thus suggests that once the situational and human factors are "controlled," and once the individual's perceptions are "filtered," then the following typical outcomes would be expected:

(1) A reduction in outdoor activities

(2) An exodus of noise sensitive individuals from the noise impacted area (provided there is an opportunity to move)

(3) An increase in overt behavior to reduce the noise exposure, e.g., soundproofing

(4) An increase in health problems

(5) A rise in atypical living habits, e.g., less conversation

(6) An increase in positive attitudes toward the noise source for those who make an internal adjustment

(7) An increase in indicators of other types of stress, e.g., family arguments

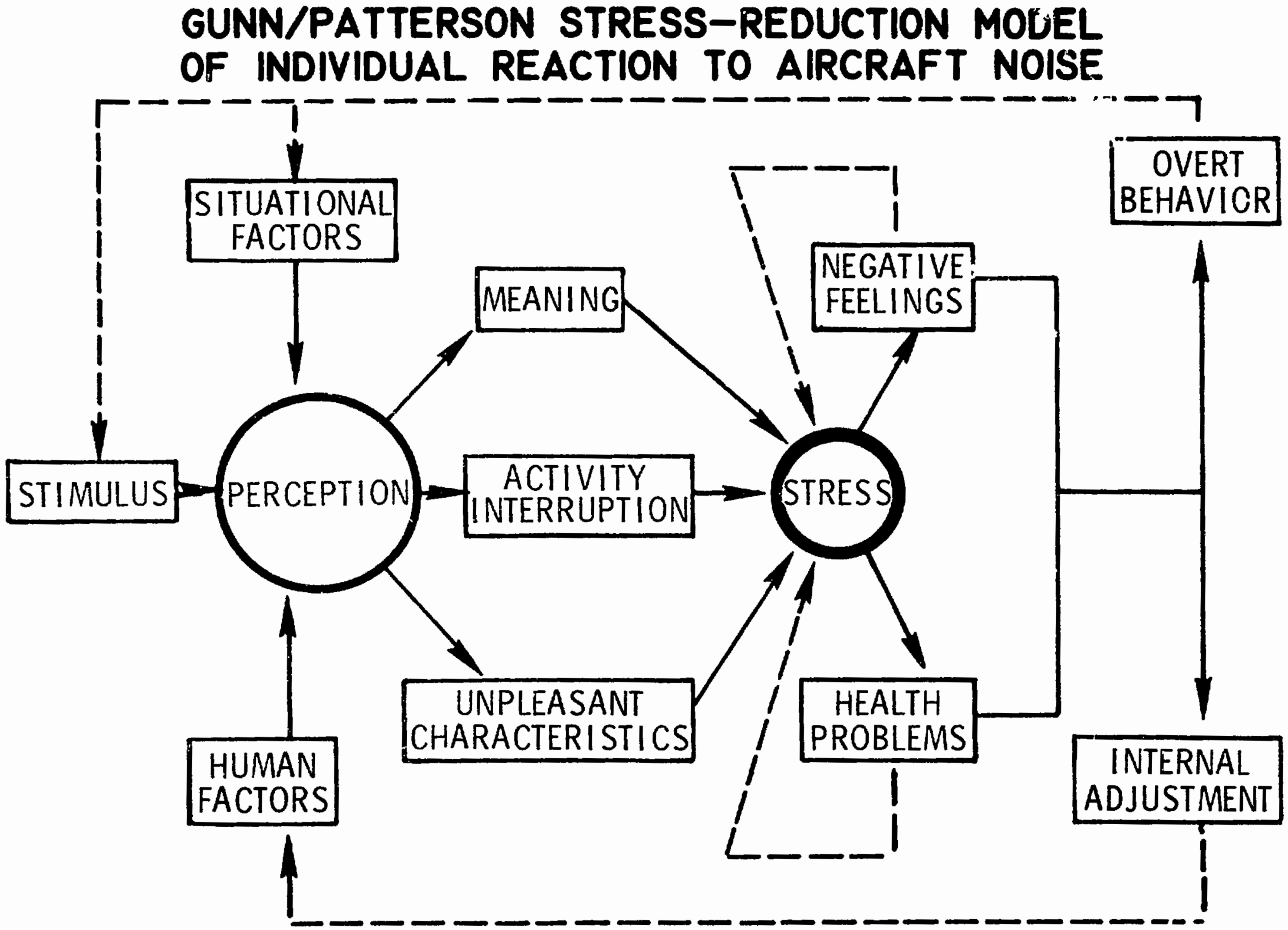


Figure A1.- Gunn/Patterson stress reduction model of individual reaction to aircraft noise.

C-2

REFERENCES

A1. Kerrick, J. S.; Nagel, D. C.; and Bennett, R. L.: Multiple Ratings of Sound Stimuli. J. Acoust. Soc. Am., Vol. 45, 1969, pp. 1014-1017.

A2. Connor, William K.; and Patterson, Harrold P.: Community Reaction to Aircraft Noise Around Smaller City Airports. NASA CR-2104, 1972.

A3. Wilson, A. H.: Noise. Her Majesty's Stationery Office, London, 1963.

A4. Molino, John A.: Equal Aversion Levels for Pure Tones and 1/3-Octave Bands of Noise. J. Acoust. Soc. Am., Vol. 55, 1974, pp. 1285-1289.

A5. McKennell, A. C.: Methodological Problems in a Survey of Aircraft Noise Annoyance. The Statistician 19:(1), 1968.

APPENDIX B

INSTRUCTION A

"We would like you to help us in this experiment which has to do with how you feel about the airplane sounds you will hear during the next 30 minutes. During the experiment, you are not to talk to each other. You will be asked for your reaction to the airplance sounds at the end of the session, which, as I said, will last about 1/2-hour."

APPENDIX C

INSTRUCTION B

"We will need to set the listening level of the TV so that it is acceptable to your group. Let's try to find a level which is a good compromise and generally comfortable for all of you."

EXPERIMENTER - FIND ACCEPTABLE LEVEL BY CONSENSUS (IN QUIET).

THEN TURN OFF TV

"Do not readjust the level during the program, please. It is imperative for the purpose of the study that the sound level stay where it is presently set."

APPENDIX D

INSTRUCTIONS TO SUBJECTS IN LISTENING PHASE OF THE EXPERIMENT

Instructions to Subjects in Telephone Listening Phase of the Experiment

"You are about to take a listening test in which you will be identifying words spoken over the telephone. The two best scoring subjects on the test will receive $7 each. The four lower scoring subjects will receive $4 each. If you will pick up your telephone, you will receive more detailed instructions. Remember, during the test, do not cover your open ear and do not switch the phone to the other ear. Listen for the item number that accompanies each word. Some words may be completely masked out in the background noise. Make sure you are checking off a word in the correct box."

Recorded Instructions

"Your attention, please.
You are going to hear some one syllable words presented along with different loudness levels of background noise, each word will be presented in a carrier phase giving its particular item number. For example, you will hear phrases like the following:

NUMBER ONE IS TREE
NUMBER 46 IS MILE

The word presented will be one of the six words printed in a block on your answer sheet for that particular item number. Your task is to identify the word by drawing a line through it on your answer sheet. Look now at the answer sheet marked practice.

Here are some practice words:

NUMBER THREE IS TOW

Within block no. 3 is the correct word tow.

If this is the word you thought you heard, you will have drawn a line through "tow" on the practice answer sheet.
Here is another word.

NUMBER 14 IS BAT

In this case, the correct word was "bat." If this is the word you thought you heard, you will have drawn a line through "bat" within block 14 on the practice answer sheet. In the following exercise, some words will be easier to hear than others.

If you are not sure what the word is--guess. Always draw a line through one of the six words for each item number. If there are any questions, please ask the person in charge now. (Pause)

Please turn now to the answer sheet marked number one and prepare to begin. Remember, always draw a line through a word even if you must guess. After drawing a line through a word, move down to the next numbered block and prepare for the next word. After completing each of the 50 items, turn to the next answer sheet and continue, starting again with item no. 1.

A total of 300 words will be given at the rate of approximately one word every 6 second. The exercise will begin in about 30 seconds."

APPENDIX E

WORD LISTS

1	lick	pick	tick
	wick	sick	kick
2	seat	meat	beat
	heat	neat	feat
3	pus	pup	pun
	puff	puck	pub
4	look	hook	cook
	book	took	shook
5	tip	lip	rip
	dip	sip	hip
6	rate	rave	raze
	race	ray	rake
7	bang	rang	sang
	gang	hang	fang
8	hill	till	bill
	fill	kill	will
9	mat	man	mad
	mass	math	map
10	tale	pale	male
	bale	gale	sale
11	sake	sale	save
	same	safe	sane
12	peat	peak	peace
	peas	peal	peach
13	king	kit	kill
	kin	kid	kick
14	sad	sass	sag
	sat	sap	sack
15	sip	sing	sick
	sin	sill	sit
16	sold	told	hold
	cold	gold	fold
17	buck	but	bun
	bus	buff	bug
18	lake	lace	lame
	lane	lay	late
19	gun	run	nun
	fun	sun	bun
20	rust	dust	just
	must	bust	gust
21	pan	path	pad
	pass	pat	pack
22	dim	dig	dill
	did	din	dip
23	wit	fit	kit
	bit	sit	hit
24	din	tin	pin
	sin	win	fin
25	teal	teach	team
	tease	teak	tear
26	tent	bent	went
	sent	rent	dent
27	sung	sup	sun
	sud	sum	sub
28	red	wed	shed
	bed	led	fed
29	hot	got	not
	tot	lot	pot
30	dud	dub	dun
	dug	dung	duck
31	pip	pit	pick
	pig	pill	pin
32	seem	seethe	seep
	seen	seed	seek
33	day	say	way
	may	gay	pay
34	rest	best	test
	nest	vest	west
35	pane	pay	pave
	pale	pace	page
36	bat	bad	back
	bath	ban	bass
37	cop	top	mop
	pop	shop	hop
38	fig	pig	rig
	dig	wig	big
39	tap	tack	tang
	tab	tan	tam
40	cave	cane	came
	cape	cake	case
41	game	tame	name
	fame	same	came
42	oil	foil	toil
	boil	soil	coil
43	fin	fit	fig
	fizz	fill	fib
44	cut	cub	cuff
	cuss	cud	cup
45	feel	eel	reel
	heel	peel	keel
46	dark	lark	bark
	park	mark	hark
47	heap	heat	heave
	hear	heath	heal
48	men	then	hen
	ten	pen	den
49	raw	paw	law
	saw	thaw	jaw
50	bead	beat	bean
	beach	beam	beak

1	went	sent	bent	dent	tent	rent
2	hold	cold	told	fold	sold	gold
3	pat	pad	pan	path	pack	pass
4	lane	lay	late	lake	lace	lame
5	kit	bit	fit	hit	wit	sit
6	must	bust	gust	rust	dust	just
7	teak	team	teal	teach	tear	tease
8	din	dill	dim	dig	dip	did
9	bed	led	fed	red	wed	shed
10	pin	sin	tin	fin	din	win
11	dug	dung	duck	dud	dub	dun
12	sum	sun	sung	sup	sub	sud
13	seep	seen	seethe	seek	seem	seed
14	not	tot	got	pot	hot	lot
15	vest	test	rest	best	west	nest
16	pig	pill	pin	pip	pit	pick
17	back	bath	bad	bass	bat	ban
18	way	may	say	pay	day	gay
19	pig	big	dig	wig	rig	fig
20	pale	pace	page	pane	pay	pave
21	cane	case	cape	cake	came	cave
22	shop	mop	cop	top	hop	pop
23	coil	oil	soil	toil	boil	foil
24	tan	tang	tap	tack	tam	tab
25	fit	fib	fizz	fill	fig	fin
26	same	name	game	tame	came	fame
27	peel	reel	feel	eel	keel	heel
28	hark	dark	mark	bark	park	lark
29	heave	hear	heat	heal	heap	heath
30	cup	cut	cud	cuff	cuss	cub
31	thaw	law	raw	paw	jaw	saw
32	pen	hen	men	then	den	ten
33	puff	puck	pub	pus	pup	pun
34	bean	beach	beat	beak	bead	beam
35	heat	neat	feat	seat	meat	beat
36	dip	sip	hip	tip	lip	rip
37	kill	kin	kit	kick	king	kid
38	hang	sang	bang	rang	fang	gang
39	took	cook	look	hook	shook	book
40	mass	math	map	mat	man	mad
41	ray	raze	rate	rave	rake	race
42	save	same	sale	sane	sake	safe
43	fill	kill	will	hill	till	bill
44	sill	sick	sip	sing	sit	sin
45	bale	gale	sale	tale	pale	male
46	wick	sick	kick	lick	pick	tick
47	peace	peas	peak	peach	peat	peal
48	bun	bus	but	bug	buck	buff
49	sag	sat	sass	sack	sad	sap
50	fun	sun	bun	gun	run	nun

1	gold	hold	sold
	told	fold	cold
2	lame	lane	lace
	late	lake	lay
3	bust	just	rust
	dust	gust	must
4	did	din	dip
	dim	dig	dill
5	sin	win	fin
	din	tin	pin
6	sun	sud	sup
	sub	sung	sum
7	lot	not	hot
	got	pot	tot
8	pill	pick	pip
	pit	pin	pig
9	may	gay	pay
	day	say	way
10	pave	pale	pay
	page	pane	pace
11	pop	shop	hop
	cop	top	mop
12	tang	tab	tack
	tam	tap	tan
13	keel	feel	peel
	reel	heel	eel
14	heal	heap	heath
	heave	hear	heat
15	paw	jaw	saw
	thaw	law	raw
16	pub	pus	puck
	pun	puff	pup
17	meat	feat	heat
	neat	beat	seat
18	kit	kick	kin
	kid	kill	king
19	cook	book	hook
	shook	look	took
20	race	ray	rake
	rate	rave	raze
21	bill	fill	till
	will	hill	kill
22	sap	sag	sad
	sass	sack	sat
23	gale	male	tale
	pale	sale	bale
24	peas	peal	peach
	peat	peak	peace
25	rent	went	tent
	bent	dent	sent
26	sun	nun	gun
	run	bun	fun
27	bus	buff	bug
	buck	but	bun
28	tick	wick	pick
	kick	lick	sick
29	sin	sill	sit
	sip	sing	sick
30	name	fame	tame
	came	game	same
31	safe	save	sake
	sale	sane	same
32	map	mat	math
	mad	mass	man
33	gang	hang	fang
	bang	rang	sang
34	sip	rip	tip
	lip	hip	dip
35	beach	beam	beak
	bead	beat	bean
36	hen	ten	then
	den	men	pen
37	cuff	cuss	cub
	cup	cut	cud
38	park	mark	hark
	dark	lark	bark
39	fizz	fill	fib
	fin	fit	fig
40	soil	toil	oil
	foil	coil	boil
41	came	cape	cane
	case	cave	cake
42	wig	rig	fig
	pig	big	dig
43	ban	back	bat
	bad	bass	bath
44	test	nest	best
	west	rest	vest
45	seen	seed	seek
	seem	seethe	seep
46	dun	dug	dub
	duck	dud	dung
47	led	shed	red
	wed	fed	bed
48	tease	teak	tear
	teal	teach	team
49	bit	sit	hit
	wit	fit	kit
50	pad	pass	path
	pack	pan	pat

No.						
1	kick	lick	sick	tick	wick	pick
2	neat	beat	seat	meat	feat	heat
3	pun	puff	pup	pub	pus	puck
4	hook	shook	book	took	cook	look
5	lip	hip	dip	sip	rip	tip
6	rake	rate	ray	raze	race	rave
7	fang	bang	hang	sang	gang	rang
8	will	hill	kill	bill	fill	till
9	map	mat	math	mad	mass	man
10	pale	sale	bale	gale	male	tale
11	sane	sake	safe	save	same	sale
12	peak	peach	peas	peal	peace	peat
13	kin	kid	kick	king	kit	kill
14	sack	sad	sap	sag	sat	sass
15	sit	sip	sill	sick	sin	sing
16	fold	sold	gold	hold	cold	told
17	but	bug	bus	buff	bun	buck
18	late	lake	lay	lame	lane	lace
19	run	bun	fun	sun	nun	gun
20	dust	gust	must	bust	just	rust
21	path	pack	pass	pat	pad	pan
22	dip	dim	din	dill	did	dig
23	fit	hit	bit	sit	kit	wit
24	tin	fin	sin	win	pin	din
25	tear	teal	teak	team	tease	teach
26	dent	tent	rent	went	sent	bent
27	sup	sub	sud	sum	sun	sung
28	wed	fed	bed	led	shed	red
29	pot	hot	lot	not	tot	got
30	duck	dud	dung	dun	dug	dub
31	pit	pin	pig	pill	pick	pip
32	seethe	seek	seen	seed	seep	seem
33	say	pay	may	gay	way	day
34	best	rest	nest	vest	test	rest
35	page	pane	pace	pave	pale	pay
36	bass	bat	ban	back	bath	bad
37	hop	cop	shop	mop	pop	top
38	dig	wig	big	fig	pig	rig
39	tack	tam	tab	tan	tang	tap
40	cake	came	cave	cane	case	cape
41	tame	came	fame	same	name	game
42	toil	boil	foil	coil	oil	soil
43	fig	fizz	fit	fib	fin	fill
44	cuss	cud	cup	cut	cub	cuff
45	heel	peel	keel	feel	eel	reel
46	mark	bark	dark	lark	hark	park
47	heath	heave	heap	heat	heal	hear
48	then	den	ten	pen	hen	men
49	law	saw	paw	jaw	raw	thaw
50	beat	beak	beach	beam	bean	bead

1	sent rent dent	tent bent went
2	told fold cold	gold hold sold
3	pass pat pack	pan path pad
4	lay lame lake	lace late lane
5	sit kit wit	fit hit bit
6	just must dust	gust rust bust
7	team tease teach	tear teal teak
8	dill did dig	dip dim din
9	shed bed wed	fed red led
10	win pin din	tin fin sin
11	dung dun dud	dub duck dug
12	sud sum sub	sung sup sun
13	seed seep seem	seethe seek seen
14	tot lot pot	hot got not
15	nest vest west	rest best test
16	pick pig pit	pin pip pill
17	bath ban bass	bat bad back
18	gay way day	say pay may
19	rig dig pig	big fig wig
20	pace pave pane	pay page pale
21	cape cake case	cave cane came
22	mop pop top	hop cop shop
23	boil soil coil	oil foil toil
24	tab tan tam	tap tack tang
25	fill fig fin	fit fib fizz
26	fame same came	game tame name
27	reel heel eel	keel feel peel
28	bark park lark	hark dark mark
29	hear heath heal	heap heat heave
30	cud cuff cut	cub cup cuss
31	saw thaw jaw	raw paw law
32	den men pen	hen ten then
33	puck pun pus	pup pub puff
34	beak bead beam	bean beach beat
35	beat heat meat	feat seat neat
36	hip tip sip	rip dip lip
37	kid kill king	kit kick kin
38	rang fang gang	hang sang bang
39	shook look took	cook book hook
40	man map mass	math mad mat
41	rave rake race	ray raze rate
42	sale sane same	safe save sake
43	till will fill	kill bill hill
44	sick sin sing	sit sip sill
45	sale tale gale	male bale pale
46	sick tick li	pick kick w
47	peach peat pea	peace peas peak
48	buff bun buck	but bug bus
49	sass s k sat	sap sag sad
50	nun fun run	bun gun sun

1	cold	gold	fold
	sold	told	hold
2	lace	late	lane
	lay	lame	lake
3	gust	rust	bust
	just	must	dust
4	dig	dip	did
	din	dill	dim
5	fin	din	win
	pin	sin	tin
6	sub	sung	sum
	sun	sud	sup
7	got	pot	tot
	lot	not	hot
8	pin	pip	pill
	pick	pig	pit
9	pay	day	gay
	way	may	say
10	pay	page	pale
	pace	pave	pane
11	top	hop	pop
	shop	mop	cop
12	tam	tap	tan
	tang	tab	tack
13	eel	keel	heel
	peel	reel	feel
14	heat	heal	hear
	heath	heave	heap
15	jaw	raw	thaw
	law	saw	paw
16	pup	pub	puff
	puck	pun	pus
17	feat	seat	neat
	beat	heat	meat
18	kick	king	kid
	kill	kin	kit
19	book	took	shook
	look	hook	cook
20	raze	race	rave
	rake	rate	ray
21	kill	bill	hill
	till	will	fill
22	sat	sap	sack
	sad	sass	sag
23	male	bale	pale
	sale	tale	gale
24	peal	peace	peat
	peak	peach	peas
25	bent	dent	sent
	rent	went	tent
26	bun	gun	sun
	nun	fun	run
27	bug	buck	buff
	bun	bus	but
28	pick	kick	wick
	sick	tick	lick
29	sing	sit	sin
	sill	sick	sip
30	came	game	same
	name	fame	tame
31	sake	sale	save
	same	safe	sane
32	math	mad	mat
	man	map	mass
33	sang	gang	rang
	fang	bang	hang
34	rip	dip	lip
	hip	tip	sip
35	beam	bean	bead
	beat	beak	beach
36	ten	pen	den
	men	then	hen
37	cub	cup	cuss
	cud	cuff	cut
38	lark	hark	park
	mark	bark	dark
39	fib	fin	fill
	fig	fizz	fit
40	foil	coil	boil
	soil	toil	oil
41	case	cave	cake
	came	cape	cane
42	big	fig	wig
	rig	dig	pig
43	bad	bass	bath
	ban	back	bat
44	west	rest	vest
	test	nest	best
45	seek	seem	seed
	seep	seen	seethe
46	dub	duck	dug
	dung	dun	dud
47	fed	red	led
	shed	bed	wed
48	teach	tear	tease
	teak	team	teal
49	hit	wit	sit
	kit	bit	fit
50	pack	pan	pat
	pad	pass	path

SOME ASPECTS OF INTERACTIONS BETWEEN SPEECH AND NOISE

By

J. C. Webster

Naval Electronics Laboratory Center, San Diego, CA.

INTRODUCTION

I would like to talk about three different but related topics today; (1) optimum simple measurements of the speech interfering aspects of steady state noises (2) the rationale for selecting among various kinds of voice communication systems for shipboard use, and (3) the effects of communication masking on annoyance judgment.

I have probably talked the speech-interference aspects of noise nearly to death and if I didn't keep changing my mind we could probably bury the subject. However, before the burial let me dress the subject up in its latest tailormade suit. In doing this I will quote quite liberally from two recent papers presented at the Eighth International Congress of Acoustics and the Inter-Noise 74 Conference, namely Webster and Cluff, (1974a, 1974b). The question being addressed is what octaves should be used in calculating the Speech Interference Level (SIL) and/or what frequency weighting network could be used or added to a sound level meter to measure the speech interfering properties of noise?

SPEECH INTERFERING ASPECTS OF NOISE

As stated in Webster and Cluff, 1974a, "Webster, 1973, showed that the best sets of octaves for calculating the SIL were centered at 500, 1000, 2000, and 4000 Hz. The lower three (3L) Webster, (1969) and the upper three (3H) Kryter; (1972) have also been proposed. Webster (1973) showed that the (3L) SIL is best when predicting marginal performance (AI = 0.2), the four-octave SIL (4) is best for good systems (AI = 0.5) and the 3H SIL is best for exceptional systems (AI = 0.8). At an AI of 0.2 a 50% Modified Rhyme Word score would be expected, at 0.5 a 75% PB score, and at 0.8 a 90% nonsense syllable score. Criticisms of Webster's (1973), generalization centered on his choice of 16 (Navy) noises for his tests. Cluff (1969), collected 112 industrial noises and adjusted the levels of each to give AI values of 0.1, 0.2 . . . 0.9. He reconfirmed that lower frequency bands predicted low AI values better, while higher frequency bands predicted high AI values better.

Webster and Cluff, 1974b reevaluated Cluff's 112 AI-equated noises in terms of the 3L, 4 and 3H SIL's and the A-weighting and three proposed speech inter-(SI) sound level meter weighting contours, shown in Figure 1. The development of these contours are discussed in Webster (1964a, 1964b, and 1973). It was hypothesized that the 3L SIL and SI-70 weighting would best predict AI's of 0.2; that the 4-octave SIL and SI-60 would be optimal at an AI of 0.5 and the best compromise for all AI's; and the 3H and SI-50 contour would best predict an AI of 0.8.

The basic procedure consisted of (1) adjusting the levels of the 112 noises via electronic computer to arrive at AIs (determined by ANSI-1969 procedures for 1/3 octave bands) of 0.1, 0.2, . . . 0.9, assuming a generalized conversational speech spectrum, (2) measuring the resulting levels by a variety of single-number measurement techniques and (3) analyzing the central tendency and dispersion characteristics of the 112 "equally-speech-interfering" levels for each single-number measurement technique at each AI. In addition to looking at the 112 noises as a whole, subgroupings based on differences between C-weighted and A-weighted (C-A) levels were analyzed.

The five sets of data in figure 2 show the noise spectra of Cluff's (1969) noises when categorized by C-A groupings. Shown are means, standard deviations, ranges, and comparisons to Botsford's (1969) categorizations of Karplus and Bonvallet's (1953) noises. Note that with few exceptions, and none that don't average out for the four crucial octaves, Cluff's noise spectra agree with the Karplus and Bonvallet's (1953) spectra when categorized by Botsford's (1969) C-A categorizations. The only obvious non-agreement is for "up-sloped" noises (when C-A is negative). For these groupings the sample is small for both sets of data. Figure 3 shows explicitly how SILs based on some combinations of the octaves centered at 500, 2000 and 4000 Hz, and various actual or potential frequency weighting networks for sound level meters, measure the levels of the various C-A noise groups when adjusted to give three values of AI. The three lower octaves, (3L), show the least variation with spectral change at an AI of 0.2; all four octaves, SIL(4), are best at an AI of 0.5; and the highest three octaves, SIL(3H), are best at an AI of 0.8. This is shown in two ways: in average (mean) level (the line most closely approaching the horizontal in figure 3) and in dispersion (the smallest standard deviations around the general mean of the noises) see Table 1. The standard deviations around the specific mean for each C-A category shown in Table 1 follow the same general rules except that (1) SIL (3L) is always best for "low frequency" noises (C-A = 15dB) and (2) at an AI of 0.2, SIL (4) is less variable than SIL (3L) for all positive C-A values except 15 dB.

Concerning weighting networks in sound level meters, the results in Figure 3 show the SI-60 network to be superior for AIs of 0.2 and 0.5 and the SI-50 best at an AI of 0.8. The A-weighting is the second best frequency weighting at AIs of 0.5 and 0.8. In general the SI-60 is a better predictor of speech interference than A-weighting, although neither is ever as good as the SIL(4).

The calculations so far shown and discussed tacitly assume that the AI is a valid measure of speech intelligibility and Kryter (1962) has summarized the data showing this to be generally true. Remember however, that no intelligibility testing was performed. It is therefore necessary to compare these analyses to at least one set of data where AI calculations and word intelligibility testing were both performed, such as Klumpp and Webster's (1963) data. To make these comparisons two steps need to be taken: (1) four noises (#6, 11, 12, and 16) are eliminated because they are either extremely time dependent (non-steady state) or contain narrow-band or tonal components (spectra lines) and therefore are not good candidates for AI predictions, and (2) C-weighting minus A-weighting (C-A) categories are established. The resultant comparisons are shown in Figure 4. The top data in Figure 4 are the C-A values of the 12 Klumpp and Webster (1963) noises (top abscissa) and the corresponding values for Cluff's (1969) environmental noises. Both the C-A sorting rules and the mean values of the Cluff (1969) noises are shown on the bottom abscissa (as well as solid circle in the top plot). The middle data in Figure 4 show how well the AI predicts the 50% Fairbanks (1958) Rhyme Test (FRT) score for the Klumpp and Webster (1963) data. The AI represented by the hollow circles is based on a 20-band analysis using the actual speech spectrum as measured; whereas the diamond-symbol-analysis are based on a generalized speech spectrum and octave bands. The lower data in

Figure 4 show the A-weighted and SIL(4) measures of levels adjusted to give 50% FRT scores on the Klumpp and Webster (1963) noises and an AI of 0.2 for the Cluff (1969) noises.

Note from Figure 4 that the AI fails to predict 50% FRT scores for the Klumpp and Webster (1963) noises in a direction and manner very similar to the difference in SIL(4) between the Klumpp and Webster and the Cluff data. This shows that the SIL(4) predicts AI quite well, but AI errs somewhat in predicting word scores, particularly for low frequency noises. The A-weighting over-estimates both the AI and the 50% FRT scores for both high and low-frequency noises.

CONCLUSION

The four-octave (500, 1000, 2000, and 4000 Hz) SIL is the best predictor of speech interference for all levels of intelligibility followed by the SI-60 and A-weighting networks in that order for Cluff's 112 noise as it was for Klumpp and Webster's (1963) 16 noises.

SELECTING THE PROPER COMMUNICATION MODE

Finding the optimum measure of the speech interfering properties of noise is only the first step in selecting the best method for conveying voice information. The next questions to be asked are: how are face-to-face communications limited by noise; and how can electrical or electronic communications systems be optimized to function in noise. This last question can be further broken down into two sub parts that concern (1) the selection of transducers (microphones, loudspeakers, and earphones) and (2) speech or language processing. This last question will not be considered in detail in this task.

Concerning the limiting effects of ambient noise on face-to-face communications there are two major factors to be considered, the decrease in sound pressure level of (spoken) sound with distance, and (2) the effects of ambient noise level on the talkers own voice level (vocal effort), Webster (1969, 1973, 1974b) using Beranek's (1954) voice-level, noise-level and communicating distance table and using two criteria of vocal effort versus noise level constructed a chart, Figure 5, summarizing the major limiting factors in noise-limited face-to-face communications. The more subtle effects of room acoustics (reverberation) talker (articulatory) effectiveness, lip-reading, language redundancy, etc. need to be considered if such factors are known for any specific application. The contents of this chart have been used as a guide in specifying that noise levels should not exceed 70 dBA in ship or aircraft spaces where peoples jobs require them to converse face-to-face at distances no greater than three feet.

Concerning the choice of transducers in various levels of noise, I will not give any specifics because (1) I have no recent evaluations to report, and (2) those I have summarized in the past are available in Webster and Gales (1970) and Webster (1971). However a summary chart, Figure 6, shows the general limitations. Note for example that until the noise level exceeds 90 dBA no real transducer limitations are serious but that a wide-band (300-6000 Hz) should be used, which implies that telephone usage becomes difficult (also see Figure 5). If telephones are used in noise levels between 90 and

110 dBA an acoustic booth and push-to-talk switch should be provided (see bottom box). If noise levels exceed 130 dBA the best method to communicate is visual.

Factors other than noise limitations must also be considered in selecting a communication mode and Figures 7 and 8 are suggested ways of aiding in this process. In Figure 7 note that the first consideration is the type of information to be passed. If it is a pictorial or alphanumeric, it should be communicated by some visual method. (Visual communication needs will not be discussed here). If it is language-related auditory communication methods are indicated. In the auditory path the next factor to be considered is whether or not the message originates and/or terminates at a fixed location. If language-related information is to be transmitted to and from a fixed location face-to-face, telephone, or intercom is indicated. If face-to-face is preferred refer to Figure 5 for limitations. In choosing between telephone and intercom, the first consideration is the number of potential subscribers. Telephones are conventionally routed via switchboards to have random access among hundreds of subscribers, whereas intercoms are conventionally hard-wired into fixed networks of up to 20 subscribers. In the conventional mode the number of subscribers dictates the choice point between the need for a telephone or an intercom. However, if multiplexing techniques or switching techniques are used for intercoms, the number of subscribers is not a key choice point.

The next factor that helps decide whether telephone or intercoms should be used involves message density. If there are more than ten messages per hour, intercom is the preferred method, unless the average message duration is greater than 15 seconds. If there are fewer than two messages per hour* or if messages are routinely longer than 2 minutes, face-to-face communication is indicated.

The next factor determining choice of auditory communication method is whether the spaces/functions to be interconnected can be expected to stay in the same location from deployment to deployment. At present permanency of space location cannot be assumed, and if modern equipment practices prevail, dial telephones might become the logical choice even if many short-duration messages were to be passed among 20 or fewer subscribers. With telephones the only communications change between deployments is the listings in the telephone directory.

A further factor is the requirement for message privacy. If the message to be passed is not for the ears of everyone, the handset on an intercom or a telephone is indicated. The remaining factor is ambient acoustic noise level, and reference should be made to Figure 6.

*This is an arbitrary figure used to give some reason for not requiring a telephone in every single manned space. It is open to argument and must be viewed in a total picture as to where the closest telephone or other means of communication is located, how remote and isolated the space may be, etc.

The reasons behind some of these points will now be discussed. Concerning message duration - if the message were routinely 15 seconds in duration or less, it would take about as long to place the call as it would to transmit the message since it takes up to 10 seconds to establish a communications circuit on a telephone (lift handset, receive dial tone, dial three digits, ring, wait for answer). On an intercom it takes from 1 to 5 seconds to place a call (select station, press to talk, speak).

Measurements of message duration on attack aircraft carriers (CVA's) have demonstrated that if intercoms are available they are indeed used for very short question/answer communications. The median duration of an intercom message was found to be five seconds and very few were longer than 15 seconds. The use of the intercom for a short message keeps the blocking ratio (occurrences of stations busy) acceptably small. Telephone communications tend to and should be used for private and/or detailed instructions, etc. that usually last from 15 seconds up to 1 or 2 minutes. Longer usage of telephones can result in unacceptable blockage ratios.

Messages longer than 2 minutes should be face-to-face. In some instances a note or memo should be written and either mailed or hand-carried.

Figure 8 shows some factors to be considered if the communicating person's) is(are) not in a fixed location. The final question concerns time critically. If neither time nor noise are prime considerations face-to-face is recommended on board ships. If distances are in miles, not feet, then face-to-face is not practical and as far as the logic chart is concerned it is a time critical situation.

If the information to be passed is time critical then questions of intercept and message security and hands-free operation become the limiting factors in choice of design. And as always the final consideration is ambient noise.

These logic charts are included to show that even in my case I realize that noise is not the only factor that determines the choice of a proper communication mode. Noise is still very important however in the design of a system once the proper mode is chosen.

THE MASKING OF COMMUNICATIONS AND ANNOYANCE

I would like to conclude this talk with a short discussion of the role played by the masking properties of noise on speaking and listening in determining noise annoyance. Bomsky (1973) says for example that "The most disruptive and widespread effect of noise is masking or the interference with the reception of speech. This interference is a major contributory factor to problems of aircraft noise annoyance. Social surveys in airport neighborhoods, for example, have found more people to be annoyed from aircraft sounds due to speech interference, either in face-to-face conversations, telephone use, or radio and TV listening, than any other form of noise disturbance. In schools, office buildings and churches, where speech and listening activities are a vital ongoing function, the intrusion of aircraft noise has been decisive in forcing either the closure of the facilities or expensive acoustic treatment for noise control."

Hazard (1971) in an airport noise study found that daily activities bothered most by noise were listening to TV/radio/records-tapes (30%), telephone and face-to-face conversations (29%) relaxing (23%), sleeping (8%), reading (6%) and eating (4%).

Everyone who studies the general problem of community annoyance with noise finds that moderator (intervening or social) variables are about as important as physical measures of noise in determining annoyance. A very recent study by Finke, Guski, Martin, Rohrmann, Schumer, and Schumer-Kohrs (1974) around Munich airport show the relationships among moderator, independent (physical) and dependent (response) variables very meaningfully. A response, or reactor, variable of interest to our discussion is the interference with speaking and/or listening. The relationships between dependent variable - responses or reactions; independent variable - physical measures of aircraft flyover noise(s); and intervening variables or moderators (M) are shown in Figure 9 in the form of a vector diagram showing a two-factor vari/max rotated solution. The stimulus (S) factor is shown along the abscissa and the moderator (M) factor on the ordinate. Datum points that lie toward the top of the diagram are strongly influenced by moderators, those toward the right are strongly influenced by the physical stimuli or aircraft flyover noises.

Finke, et al (1944) found, as did Hazard (1971), Robinson (1971) and others that the particular physical measure of noise used to define the stimulus aircraft flyover noise was relatively unimportant as long as account was taken of the number of noise incidents, flyovers.

The moderator variables show very little correlation with the physical stimulus, and this is particularly true of those labeled a, b, c, and d. Moderators e, f, and g bear some relationship to the stimulus (noise) but only the "fear" moderators share much relationship to noise intensity.

The eleven reaction datum plotted form a Global Reaction, (R) vector that lies midway between the moderator and stimulus vectors. The reactor variables are numbered in the rank order that they correlate with the "global reaction". Note that three of the "best" four correlate higher with moderator (intervening) variables than with the stimulus (independent variable or flyover intensity). They could quite properly be classified as annoyance reactions. Finke, et al, (1974) lump their a, b, and c moderators into a single "sensitivity to noise" moderator which correlates (r = -.56) with global reaction, R. They then show that "noise sensitive" individuals as opposed to non-noise sensitive individuals show stronger relationships between noise and emotional reactions (vice cognitive reactions for the non-noise sensitive); complained more about noise; and score higher on indicators of social class.

There are some reactions that are minimally influenced by moderators. Note (1) that the reactor that correlates highest with physical intensity is "disturbance of communication (#3)" - speaking and/or listening, and (2) that loudness (#10) although not highly correlated with intensity is the least influenced by moderators. It should be pointed out that this "loudness" is not a classically defined psychophysically determined loudness.

The Bolt, Beranek, and Newman staff in a series of reports on vehicle noise, see Jones (1971), Bishop and Simpson (1971), Horenjeff and Findley

(1971) and Galloway (1971) sampled 1200 individuals living near roadways in Boston, Detroit and Los Angeles. Even this population, chosen to reflect the effects of vehicular noise, answered that the noisiness in their neighborhoods was caused by motor vehicles (55%), aircraft (15%), and TV, radio and conversation (14%). Their lists of activities annoyed were in order: sleep; listening to TV, radio, or recordings; mental activity; driving; conversing; and walking.

Langdon and Gabriel (1974) actually used interference with TV viewing as an activity against which the aversiveness of noise could be evaluated in a laboratory situation. They found that within a group of viewers (listeners) who (1) heard one duration of flyover noise but at rates of 7.5, 15, 30, or 120 per hour, or (2) heard one rate but at durations of 2, 4, 8 or 16 seconds the "acceptability" decreased by two units (roughly equivalent to doubled loudness) as the maximum level on the integrated duration or rate increased by 10 dB. This equal energy rule agrees well with results obtained using more conventional psychophysical tests.

This section is not intended to be complete or conclusive and is included to show that the interference of noise with speech is not only real and measurable but also highly annoying in the totality of daily living.

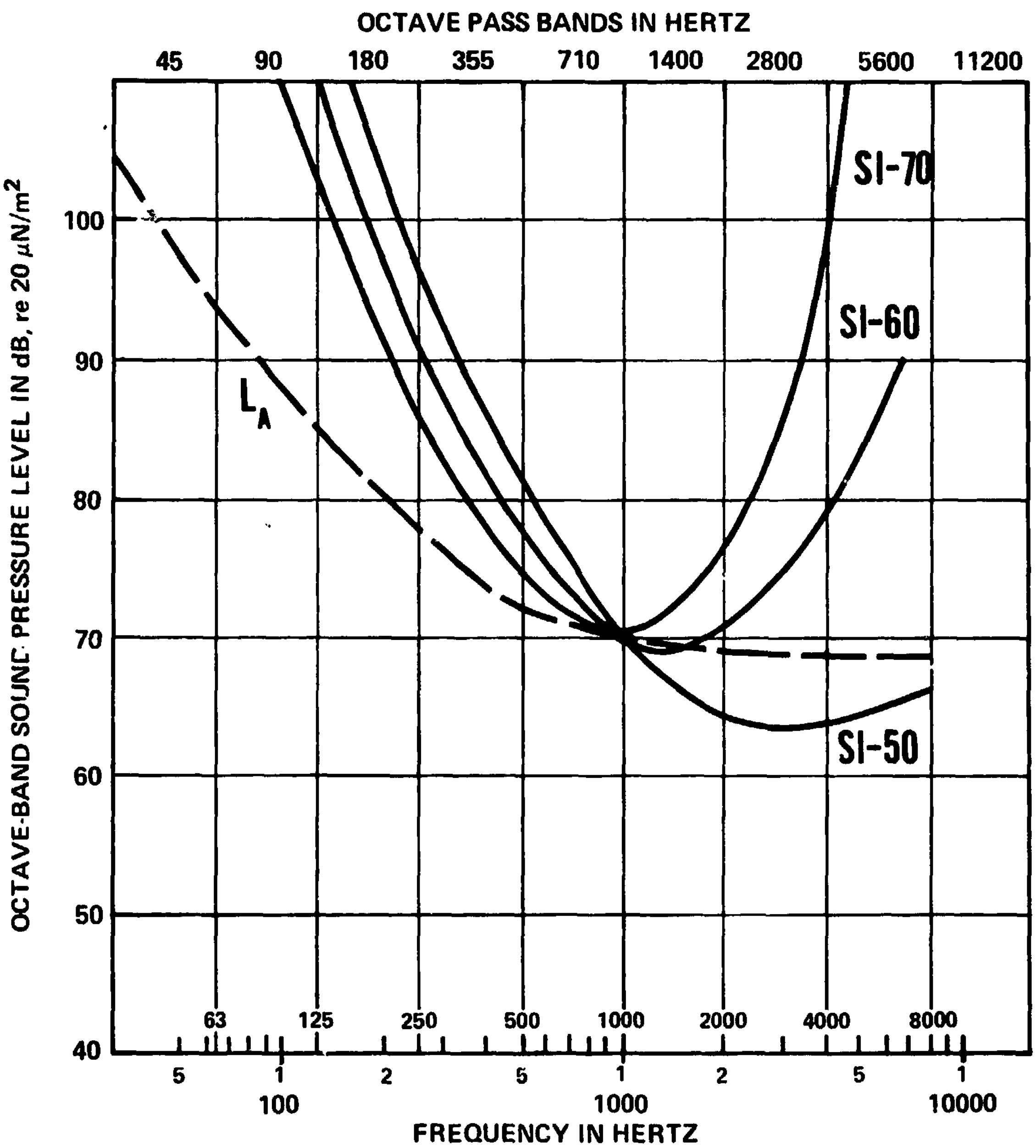


Figure 1.— Actual (A weighting L_A) and proposed frequency weighting networks for use in sound level meters for measuring the speech interfering aspects of noise.

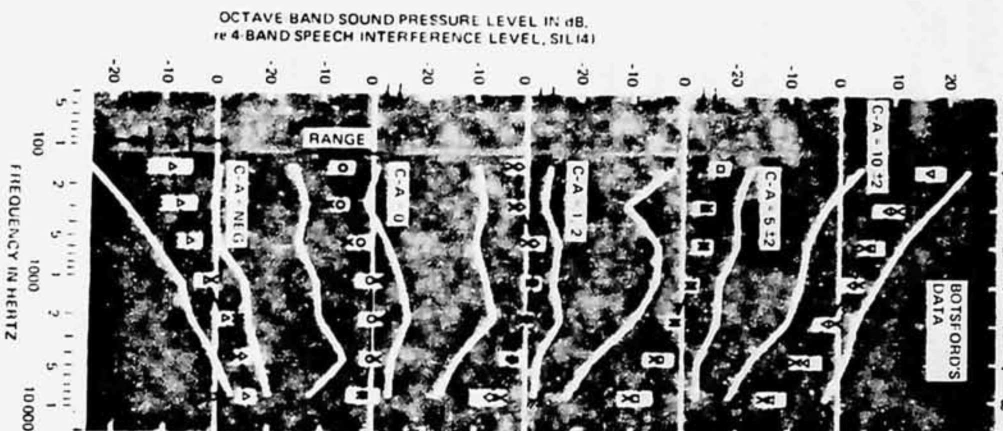


Figure. 2- Means, standard deviations, and ranges of 112 environmental noises collected by Cluff (1969) sorted by C-weighting minus A-weighting (C-A) categories. Also shown, as x's are spectra sorted by Botsford (1969) on 953 manufacturing, neighborhood, and earth-moving vehicle noises collected by Karplus and Bonvallet (1953).

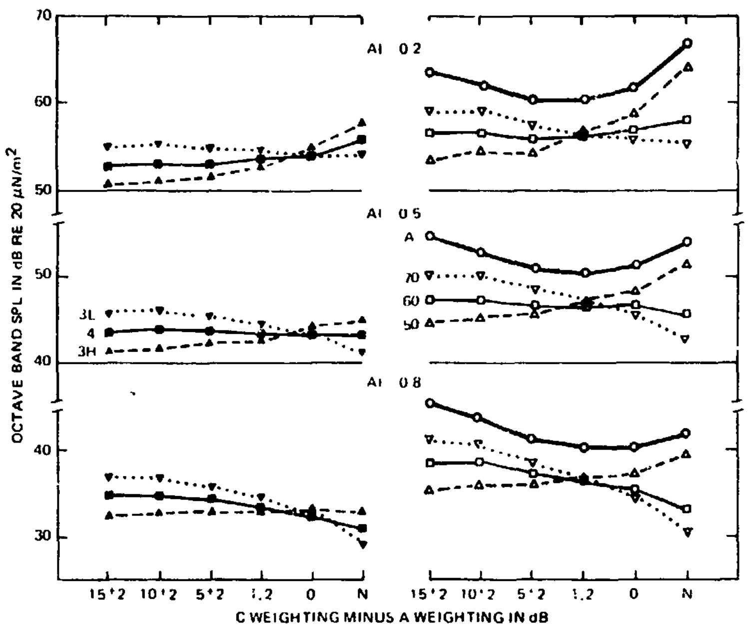


Figure 3.- Mean values of seven single-number measurement methods for predicting speech interference. Measurement methods include three ways of calculating SIL from octave bands centered at 500, 1000, 2000, and 4000 Hz namely using the lower three, 3L; all four (4); or the higher three (3H); and four actual or potential frequency weighting networks for sound level meters namely A-weighting and speech interference contours SI-70, SI-60, and SI-50.

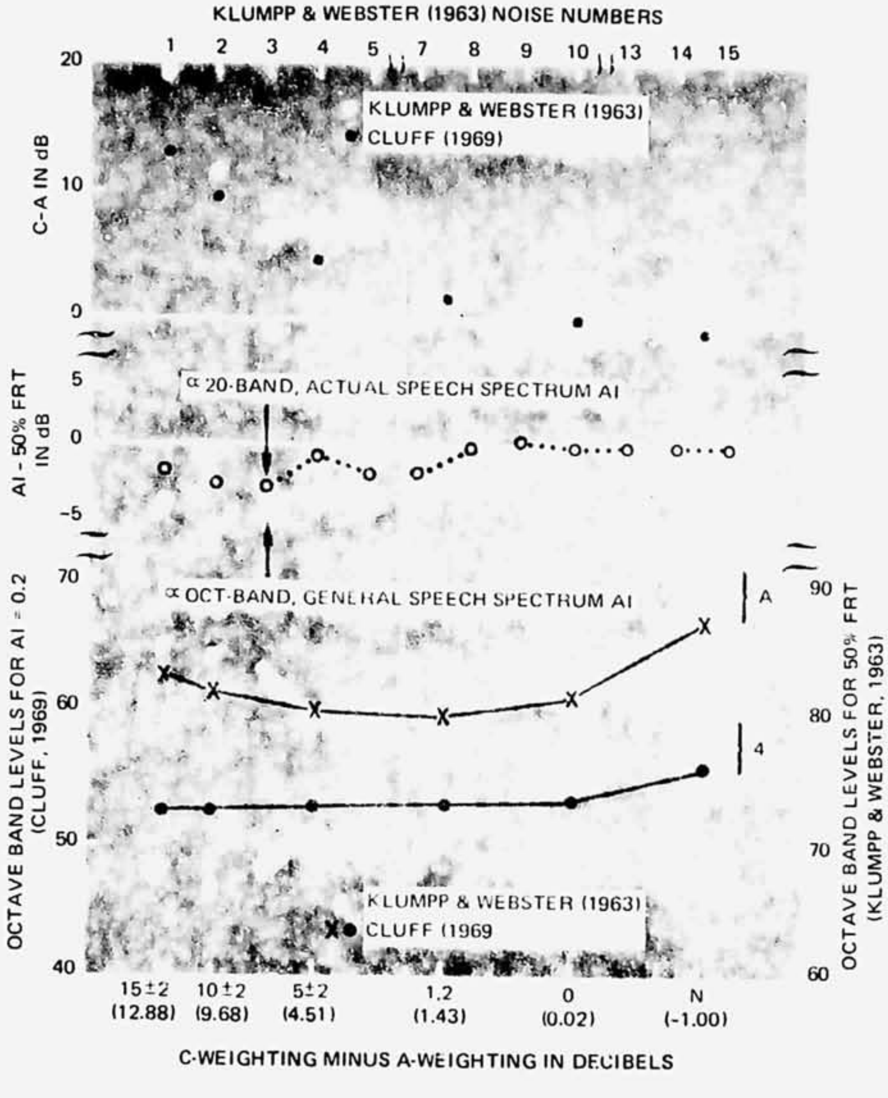


Figure 4.- Comparisons among various parameters from the Klumpp and Webster, 1963 data (K & W) and selected data from Figure 3. At the top, C-A on 12 of the 16 K & W noises vice mean C-A values on Cluff's 112 noises. In the middle, difference between AI calculations (of two degrees of complexity) and experimentally determined 50% word scores on the Fairbanks Rhyme Test (FRT) on the K & W data. The reference or zero line is the AI score (in dB where AI=1.0 = 30 dB; AI = 0.5 = 15 dB; etc) for noise #10 (thermal, flat). At the bottom A-weighted and 4-band SIL calculations on K & W noises adjusted for 50% FRT scores and on Cluff's noises adjusted for AI of 0.2.

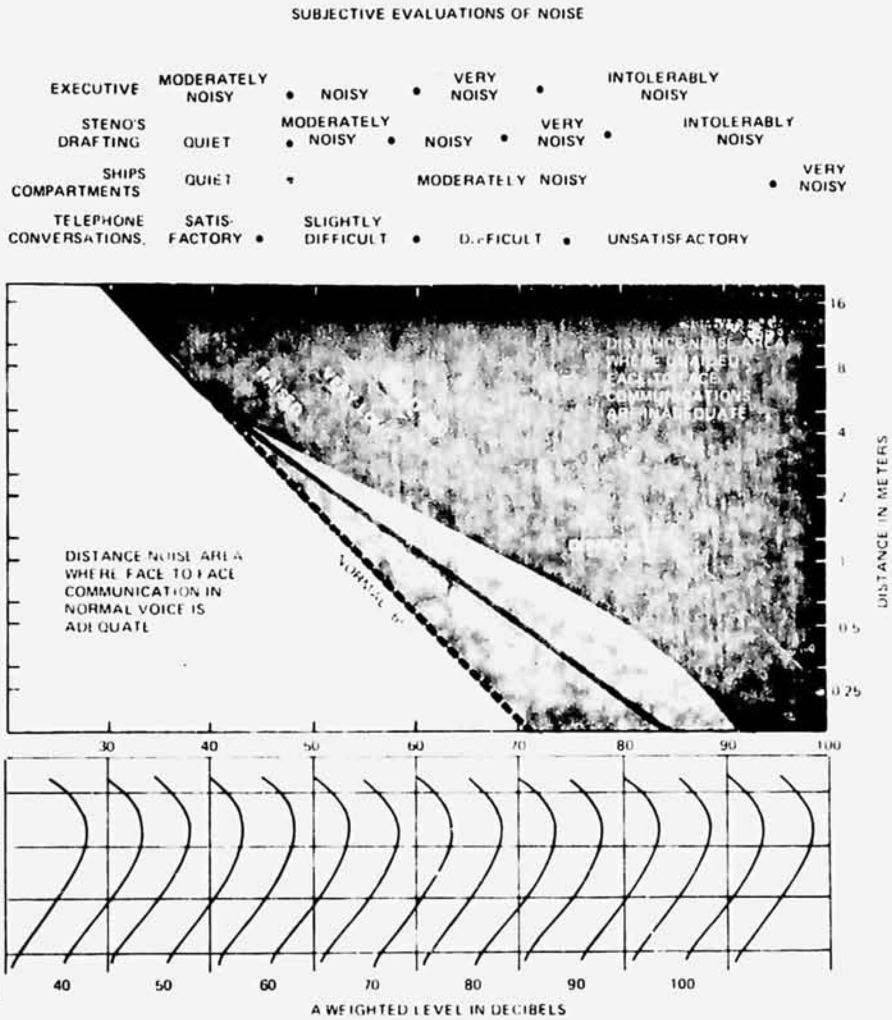


Figure 5.- Necessary voice levels as limited by ambient noise for selected distances between talker and listener for satisfactory face-to-face communication. Along the abscissa are various measures of noise, along the ordinate distance, and the parameters are voice level. At levels above 50 dB(A) people raise their voice level as shown by the "expected" line if communications are not vital or by the "communicating" line if communications are vital. Below and to the left of the "normal" voice line communications are at an AI level of 0.5, 98% sentence intelligibility. At a shout, communications are possible except above and to the right of the "impossible" area line. To use the A-weighted (L_A) noise measure corrected for the difference between the C-weighting and L_A (C-A): Find the measured C-A along the lower ordinate, say 5, follow that horizontal line across to the L_A measure, say 77, and enter the chart directly above this intersection which in this example corresponds to an SIL(4) of 70.

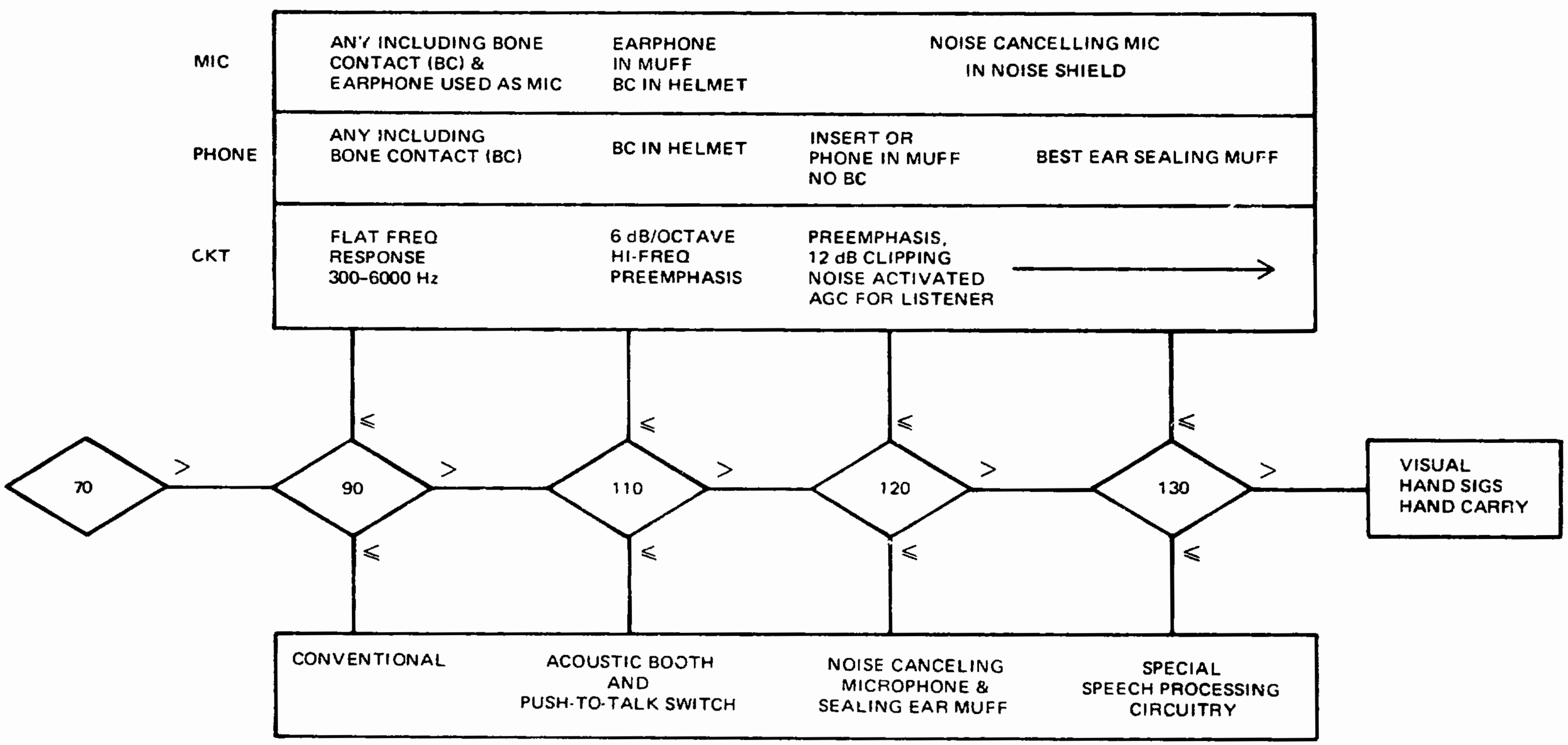


Figure 6.- Voice-communication-equipment in noise chart. Transducer and circuit design parameters for radio or special intercoms are shown above the noise level categories, and telephone parameters below.

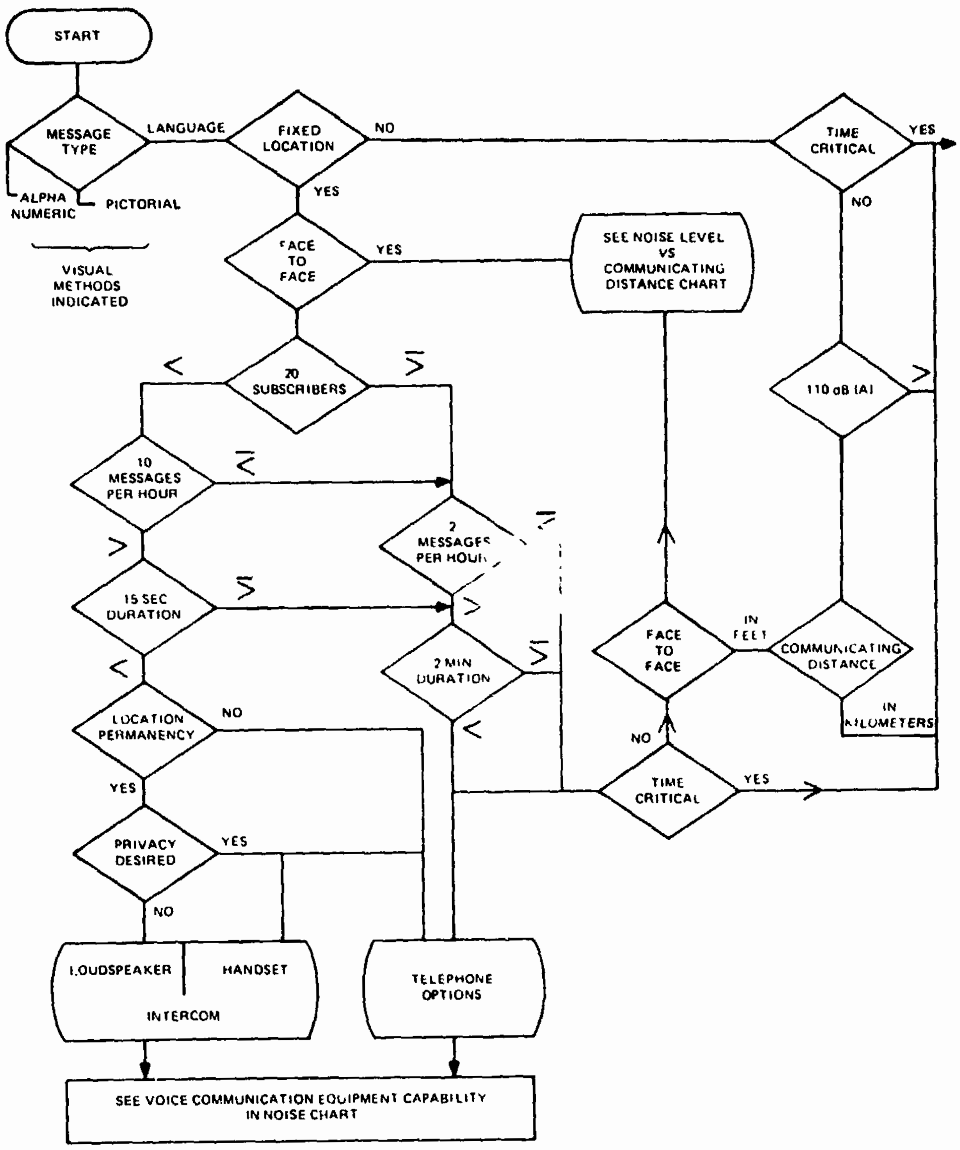


Figure 7.- Logic flow chart for selecting voice communication modes (face-to-face, intercom, telephone) for people in fixed locations.

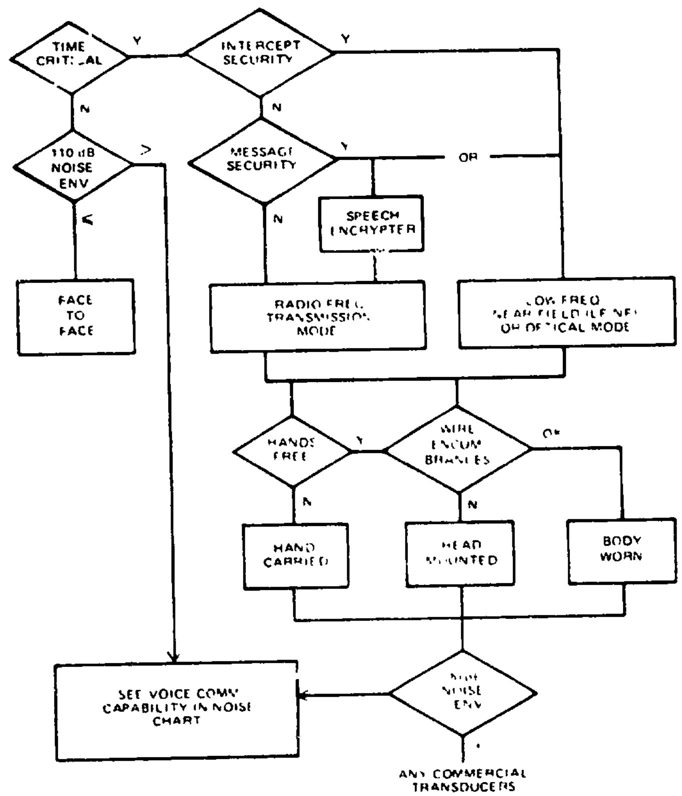


Figure 8.- Logic flow chart for selecting voice communication modes (face-to-face, radio) for people in non-fixed locations.

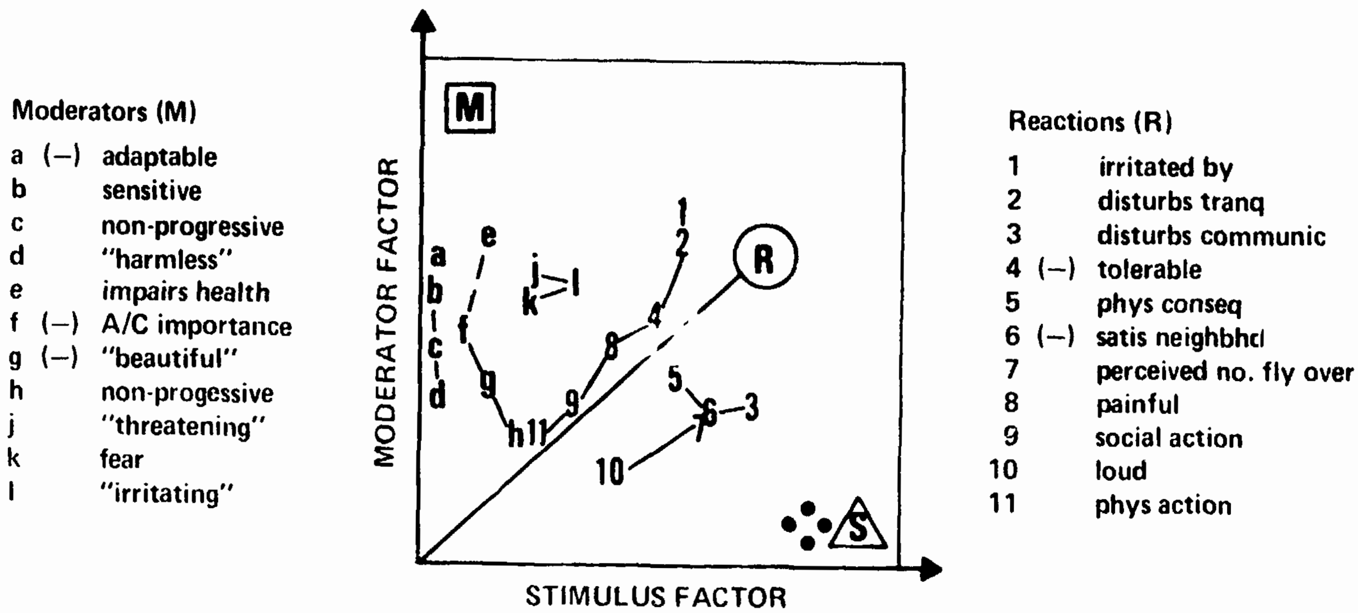


Figure 9.- Rotated varimax factor analysis of Finke, Guski, Martin, Rohrman, Schümer, and Schümer-Kohrs (1974) Munich airport study. All variables have been transposed into the positive quadrant. The stimulus factor vector (independent variable) increases from left to right (all physical noise measure load about equally and highly positive on this factor). The moderator factor vector (intervening variable) increases upward from the origin and shows three groups as concerns correlation with the stimulus factor; hardly any (a, b, c, d); very little (e, f, and g); and some (h, j, k, l). The moderators that best determine the total moderator factor are those furtherest from the origin (e, a, b, j, k, l).The reactor factor vector (dependent variable) lies midway between the others and increases on the diagonal away from the origin. The strength of the relationships between the individiual reactor factor and the total or global reaction (R) can be determined by drawing perpendicular lines from the datum to the diagonal and are purposely (re)labeled to show the strength of this relationship, #1 being highest and #11 the lowest. The relationship of the reactor variables to the other two factors can also be seen by drawing perpendiculars. Perpendiculars dropped on the stimulus vector show #3 to correlate the highest and #11 the least. Perpendiculars across to the moderator factor show that #1 correlates highest with the moderator factor and #10 the least.

REFERENCES

1. Beranek, L. L. (1954) Acoustics, New York: McGraw Hill.
2. Bishop, D. & Simpson, M. (1971), "Community Noise Measurements in Los Angeles, Detroit and Boston." BBN Report No. 2078 to Automobile Manufacturers Association.
3. Borsky, P. N. (1973) "A new field laboratory methodology for assessing human response to noise" NASA CR-2221, March.
4. Botsford, J. H. (1969), "Using sound levels to gauge human response to noise", Snd. and Vib. 3(10) 16-28.
5. Cluff, G. L. (1969), "A comparison of selected methods of determining speech interference calculated by the Articulation Index", J. Auditory Res., 9, 81-88.
6. Fairbanks, G. (1958), "Test of phonemic differentation: The rhyme test," J. Acoust. Soc. Amer. 30, 596-600.
7. Finke, H. O., Guski, R., Martin, R., Rohrmann, B., Schumer, R., and Schumer-Kohrs, A. (1974). "Effects of aircraft noise on man", Preprint of paper presented at Symp. on Noise in Transportation, Southampton, 22-23 July.
8. Galloway, W. (1971), "Motor Vehicle Noise: Identification and Analysis of Situations Contributing to Annoyance." BBN Report No. 2082 to Automobile Manufacturers Association.
9. Hazard, W. R. (1971), "Predictions of Noise Disturbance near large airports," J. Sound Vib. 15(4), 425-445.
10. Horonjeff, R., & Findley, D., (1971) "Noise Measurements of Motorcycles and Trucks." BBN Report No. 2079 to Automobile Manufacturers Assn.

11. Jones, G., (1971) "A Study of Annoyance from Motor Vehicle Noise." BBN Report No. 2112 to Automobile Manufacturers Association.

12. Karplus, H. B. and Bonvallet, G. L. (1953), "A noise survey of manufacturing industries". Am. Ind. Hyg. Assoc. Quart 14, 235-263 (1953)

13. Klumpp, R. G. and J. C. Webster (1963), "Physical Measurements of equally speech interfering Navy noise," J. Acous. Soc. Amer. 35, 1328-1338.

14. Kryter, K. D., (1962) "Validation of the articulation index". J. Acoust, Soc. Amer., 34 1698-1702.

15. Kryter, K. D. (1972), "Speech Communication" Chap 5 in Human Engineering Guide to Equipment Design, Revised Edition, Eds H. P. Van Cott & R. G. Kinkade, Gov. Printing Office, Lib. Congr Catalog Card #72-600054.

16. Langdon, L. E., & Gabriel, R. F. (1974), "Judged acceptability of noise exposure during television viewing". J. Acoust. Soc Am 56, 510-515.

17. Robinson, D. W. (1971). "Towards a unified system of noise assessment". Journal of Sound and Vibration. 14, 279-298.

18. Webster, J. C. (1964a), "Generalized speech interference noise contours", J. Speech and Hearing Research, 7, 133-140.

19. Webster, J. C. (1964b). "Relations between speech-interference contours and idealized articulation index contours". J. Acoust. Soc. Amer. 36, 1662-1669.

20. Webster, J. C. (1969). "Effect of noise on speech intelligibility". P 49-73 in National Conference on Noise as a Public Hazard. Proceedings. 13-14 June 1968, The American Speech and Hearing Association (ASHA) Reports 4.

21. Webster, J. C. (1971). "Flight deck noise levels and their effects". NELC TR-1762 of 30 April 1971.

22. Webster, J. C. (1973). "The effects of noise on the hearing of speech". Proc. International Congress on Noise as a Public Health Prob. 13-18 May 1973, US Envir. Prot. Agency, US Gov. Prntng. Off. 550/9-73-008, p. 25-42.

23. Webster, J. C. and Cluff, G. L. (1974a), "Validation of the four octave preferred frequency speech interference level (SIL)". Proceeding of the 8th International Conference on Acoustics, 23-30 July, London.

24. Webster, J. C. and Cluff, G. L. (1974b). "Speech interference by noise". 1974 International Conference on Noise Control Engineering. p. 553-558. Washington, D. C. 30 Sept. - 2 Oct.

25. Webster, J. C., and R. S. Gales, (1970) "Noise rating methods for predicting speech communication effectiveness". Transportation Noises A Symposium on Acceptability Criteria. University of Washington Press. p 85-103.

UNITS FOR THE ASSESSMENT OF NUISANCE DUE TO TRAFFIC NOISE IN A SPEECH ENVIRONMENT

By

C. G. Rice

ISVR, The University,
Southampton, U. K.

INTRODUCTION

A laboratory study of nuisance due to traffic noises in a speech environment has recently been carried out (1) (2), in which it was suggested that L_{10}dB(A) might be the most suitable unit for relating the indoor intrusion caused by the traffic noise to its physical characteristics.

Further analyses of these results enabled other physical parameters of the noises to be taken into account, and these in turn led to the formulation of a 'goodness factor' which enabled the efficiency of the different rating scale units to be reassessed.

The model used is particularly important in assessing the merits of such units as L_{10}, L_{eq} and L_{NP} in the formulation of the optimum unit for use in the general assessment of urban noise.

LABORATORY STUDY

The study was designed to investigate the effects which a variety of traffic noise situations had on the appreciation of speech in a controlled environment. Subjects were asked to adjust the intensity level of an intruding time-varying traffic noise signal until they considered it to be just "unacceptable" for relaxed listening to speech. A criterion of speech interference was not used, rather subjects were asked to select the level at which the traffic noise just began to be noticeably unacceptable.

The traffic signals were representative of sounds produced indoors near roads with varying percentages of heavy vehicles superimposed upon a high flow of light vehicles. Three conditions were chosen (12%, 4% and 1.3% heavy vehicles in a 6000 v/hr light traffic flow) at each of two peak-steady noise levels (5 dB and 20 dB) and two durations (20 dB down points of 5 and 15 seconds). The thirteenth condition was the steady light traffic flow of 6000 v/hr. The speech signals were thirteen separate male voice recordings of short stories of topical interest.

Each of the 13 traffic noises were presented to each subject. In order to balance out the possible effects due to different speech recordings or to changes in the subject's tolerance during a test session a 3-way balanced design was needed. This ensured that each noise situation was paired an equal number of times with each and every speech recording, and was presented an equal number of times in each and every presentation order position.

These requirements were achieved by using a design based on two 13 x 13 balanced Graeco-Latin squares, which required 13 speech signals and 26 subjects. The Graeco-Latin square design is shown in Table 1.

Subject No.	Presentation Order 1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th
I	1m	21	13a	3k	12b	4j	11c	5i	10d	6h	9e	7g	8f
II	2a	3m	1b	41	13c	5k	12d	6j	11e	7i	10f	8h	9g
III	3b	4a	2c	5m	1d	61	13e	7k	12f	8j	11g	9i	10h
IV	4c	5b	3d	6a	2e	7m	1f	81	13g	9k	12h	10j	11i
V	5d	6c	4e	7b	3f	8a	2g	9m	1h	101	13i	11k	12j
VI	6e	7d	5f	8c	4g	9b	3h	10a	2i	11m	1j	121	13k
VII	7f	8e	6g	9d	5h	10c	4i	11b	3j	12a	2k	13m	11
VIII	8g	9f	7h	10e	6i	11d	5j	12c	4k	13b	31	1a	2m
IX	9h	10g	8i	11f	7j	12e	6k	13d	51	1c	4m	2b	3a
X	10i	11h	9j	12g	8k	13f	71	1e	6m	2d	5a	3c	4b
XI	11j	12i	10k	13h	91	1g	8m	2f	7a	3e	6b	4d	5c
XII	12k	13j	111	1i	10m	2h	9a	3g	8b	4f	7c	5e	6d
XIII	131	1k	12m	2j	11a	3i	10b	4h	9c	5g	8d	6f	7e

1-13-13 test signals

a-m-13 speech recordings

I-XIII-13 subjects

TABLE I Graeco-Latin square design

The settings of the attenuator controlling the traffic noise level chosen by each subject as his "just acceptable" level for each test situation were noted. These were related to physical means of the test signals made both as heard in the listening chamber (in the absence of a subject) and in the equivalent outside facade position. Using real time analysis and computational facilities, over eighty rating scale units were evaluated to see which 'best' related the physical characteristics of the noises to the judged subjective responses. The criterion of 'best' is not easy to define, but in the context of the study it was considered that it was not unreasonable to expect the 'ideal unit' to be one which would give the same numerical value for all thirteen noise signals when subjectively lined up at the average levels chosen by subjects. The results obtained for a selection of units in terms of both F-ratio and standard deviations (in parentheses) are shown in Table II.

Although the L_{10}dB(A) measure at the facade of the building appears to be the most appropriate unit and supports the Noise Advisory Council's recommendation based on Building Research Station researches (3), it is clear that none of the units examined comes close to being 'ideal'; in particular all 'F' ratios from the analysis of variance are significant which indicates the inability of any of the units to satisfactorily account for the physical characteristics in the noises when judged to be subjectively equal.

DISCUSSION

Of the other favoured units which are often reported in the literature L_{eq} was well rated provided it was calculated using the energy mean or by using the B & K Noise Dose Meter. L_{NP} was not as successful, nor were NNI or TNI. Of particular interest however is the approximated formula (based on the assumption that noise levels from road traffic are normally distributed) which was used in the calculation of L_{eq} (see Table II). Not all the traffic noises were normally distributed and that by using such an approximation a large F value and standard deviation were obtained. Further detailed investigation of the properties of such non-normally distributed noises is currently being carried out, and preliminary results reveal that the skewness of the distribution may be an important factor worthy of inclusion. For example, the standard deviation of the L_{10}dB(A) result in Table II can be reduced from 1.8 depending upon the form of the skewness correction. Extrapolation below the L_{10} level also indicates that levels between L_5 and L_{10} further reduce the standard deviation to below 1 dB. These significant changes will be reported elsewhere in more detail in the future.

The analysis of variance tables also showed that the temporal

TABLE II

F-ratios for selected units

	dB(A)	dB(B)	dB(D)	PLdB
Measured as heard inside				
L_{10} Statistical distribution analyser	5.54(1.8)	7.53(2.1)	7.26(2.0)	
Peaklevel recorder r.m.s. maximum value	9.24(2.3)	7.67(2.1)	7.25(2.0)	
Maximum integrated ½ second by computer	9.60(2.3)	8.12(2.2)	7.61(2.1)	7.84
L_{50}dB(A)	69.70(6.3)			
L_{eq1}-Energy mean dB(A) by computer	6.55(1.9)		9.00	
L_{eq2} - Dosemeter	7.91(2.1)			
$L_{eq3} - L_{50} + (L_{10} - L_{90})^{2/57}$	36.50(4.5)			
$L_{NP1} - L_{eq3} + (L_{10} - L_{90})$	30.0(4.1)			
$L_{NP2} - L_{eq3} + 2.56\sigma$	21.75(3.5)			
$L_{NP3} - L_{eq2} + 2.56\sigma$	34.94(4.5)			
NNI - PNL_{max} + 15logN - 20 where $N = \frac{720}{(I+1)}$	58.17(5.7)			
TNI - $L_{90} + 4(L_{10} - L_{90}) - 30$	590.55(18.3)			
Measured outside				
$L_{10\%}$ Statistical distribution analyser	4.54(1.6)	5.19(1.7)	5.30(1.7)	

TABLE II (Cont'd)

Peak-level recorder r.m.s. maximum value	9.54(2.3)	8.95(2.3)	9.13(2.3)

Levels of significance: 5% $F(12,276) = 1.8$

1% $F(12,276) = 2.3$

Results indicate that no unit satisfactorily rates the subjective judgements.

distributions of the traffic noises are not well accounted for by the existing units. The somewhat regular occurrence of the noises enabled an interval correction to be added to the peak values. This empirical correction takes the form $n \log_{10}(I/m)$ where n and m are integers and I is the time interval in seconds between the pass-by peaks. The final unit becomes

$$dBI = dB_p - 5 \log_{10} \left(\frac{I'}{5}\right)$$

where dB_p is the peak rating scale unit value, and $I' = I$ for $I' > 5$ secs and $I' = 5$ for $I' < 5$ secs.

Table III shows that this condition lined up the test signals with a non-significant scatter that could be attributed to random error, suggesting that a peak or maximum measure coupled with a rate of occurrence correction might be the best unit solution. However, how much the regularity of the signals affected subjects' judgements is not known, and in practice freely flowing traffic with varying concentrations of heavies is not regular. Bunching occurs causing a randomness which may be very hard to physically define, although under certain circumstances, such as 'worst mode', these conditions might be quantifiable.

GOODNESS FACTOR MODEL

The 'ideal unit' concept previously defined may not necessarily be the correct way of identifying the physical rating scale unit which best describes the subjective reactions to the noises concerned.

Consideration should also be given to the way in which the unit is sensitive to changes in the physical characteristics of the noises. If the noises in this study were lined up on their background levels ($L_{90\%}$) the approximate ranges covered when measured by different units were: L_{eq} - 12 dB, $L_{10\%}$ - 17 dB, Peak and NNI - 20 dB, L_{NP} - 25 dB, TNI - 55 dB.

TABLE III

Summary analysis of variance table for a selection of weighted values measured inside

Source of variation	Degrees of freedom	F-ratios L_{10} (dB(A))	Peak dB(A)	Leak dB(A)I*	L_{eq1}	L_{NP2}	Max* PNLI	TNI
Subjects	25	78.8	78.8	78.8	78.8	78.8	78.8	78.8
Order	12	4.1	4.1	4.1	4.1	4.1	4.1	4.1
Speech	12	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Noise	12	5.54	9.24	1.6	6.55	21.8	1.0	590.6
Interval (I)	2	5.5	43.8	0.2	0.8	15.3	0.1	461.8
Peak (P)	1	9.6	1.7	5.0	27.5	71.4	0.4	2957.3
Duration (D)	1	25.7	4.1	4.8	24.1	9.7	0.1	98.0
Residual	276							
TOTAL	337							

Levels of significance: 5% $F(25,276) = 1.6$, $F(12,276) = 1.8$

1% $F(25,276) = 1.9$, $F(12,276) = 2.3$, $F(2,276) = 4.7$, $F(1,176) = 6.7$

*Interval corrected

This infers that units such as TNI and L_{NP} can much more sensitively measure changes in noise characteristics than do L_{eq} or L_{10}. Because this is a desirable quality in a noise unit, more account should perhaps be taken of this fact. It is therefore proposed that the best unit may be the one whose 'Goodness Factor' (GF) is the smallest where

$$GF = \frac{\sigma \text{ of unit values at subjective equality levels}}{\sigma \text{ of unit values of the noise set}} = \sigma_s / \sigma_p$$

The best unit measure is therefore the one which allows maximum flexibility and sensitivity of physical measurement (i.e. large σp) with minimum subjective scatter (i.e. small σs). Application of the goodness factor to a selection of the results of the traffic noise study yields the values shown in Table IV.

TABLE IV

GOODNESS FACTOR RESULTS

L_5-L_{10} $^{(x)}$ dBA	0.15 - 0.3
L_{NP}	0.4
L_{eq}	0.4 - 0.8
TNI	0.8

(x) Depending upon form of skewness correction.

These results change the rank ordering suggested in Table II, most noticeable being the relegation of L_{eq}. L_{NP} now ranks slightly superior to L_{eq} and this result needs further consideration in the light of recent trends towards the adoption of L_{eq} as national units in other European countries and in the USA.

CONCLUDING REMARKS

This study has indicated that the 'A' weighted units such as L_{5-10} and L_{eq} may be adequate measures for expressing the physical characteristics of traffic noises causing nuisance in a speech environment. However in seeking a unified index for community noise annoyance L_{eq} does not appear to be as effective as $L_{NP}(4)$ where combined aircraft and traffic noise environments are concerned.

It also seems that other factors based on the skewness and statistical time distribution properties of the noises may be necessary. Evidence of the importance of this in the speech environment is also provided by Gordon in 1971, who recommended that at least two points on the time domain curve might be needed such that

(1) the articulation index should not deteriorate below 0.4 for more than 10% of the time, and

(2) the articulation index should not fall below 0.6 for more than 50% of the time.

These two criteria are therefore separated by about 6 dB(A) (a change of 3 dB(A) corresponds to a change of articulation index of .1).

REFERENCES

1. C. G. Rice; Brenda M. Sullivan; J. G. Charles; J. A. John. 1974 "A Laboratory Study of Nuisance Due to Traffic Noise in a Speech Environment" J. Sound Vib. 37(1), 87-96.
2. C. G. Rice. 1973 "Units for the Assessment of Nuisance Due to Traffic Noise in a Speech Environment" Paper presented to British Institute of Acoustics Meeting on "Urban Noise Measurement and Evaluation" 13/14 November, Southampton, England.
3. W. E. Scholes, J. W. Sargent. 1971 "Designing Against Noise from Road Traffic" Applied Acoustics 4, 203-234.
4. C. A. Powell, C. G. Rice. 1975 "Judgements of Aircraft Noise in a Traffic Background" J. Sound Vib. 38(1), 39-50.
5. C. G. Gordon. 1971 Personal Communication - Wolfson Unit for Noise and Vibration Control. ISVR, The University, Southampton, England.

A NEW LOOK AT MULTIPLE WORD TEST ITEMS FOR

EVALUATING TALKERS, LISTENERS, AND COMMUNICATION SYSTEMS.*

By

Carl E. Williams

James D. Mosko

and

James W. Greene **

Naval Aerospace Medical Research Laboratory

*A portion of this paper was presented at an AGARD Aerospace Medical Panel Specialists Meeting in Pozzuoli, Italy, September 16-20, 1974.

**Paper presented by James D. Mosko

SUMMARY

Word recognition performance for double-word and triple-word Modified Rhyme Test (MRT) items is not appreciably different from that for single-word MRT items. Having individuals give confidence ratings of their response choices does not influence their overall performance. Because of their more representative message length and decreased testing time (less than one-half the time required for the regular single-word format of the MRT), the triple-word test items (TMRT) appear to hold promise as suitable speech materials for use in the development of an efficient reliable test for assessing the hearing capabilities of aircrew personnel. The multiple-word closed-response test format may also be appropriate for evaluating talkers and listeners in general and communication systems.

INTRODUCTION

It has been recognized for some time that hearing tests used in the selection and retention of aircrew personnel do not measure the type of hearing ability required for the efficient performance of flying duties. An individual's ability to hear pure tones in quiet or to hear whispered speech at some standard distance from a talker has little, if any, relation to how well he can perceive loud speech in the presence of high levels of ambient noise.

The Acoustical Sciences Laboratory, NAMRL is currently conducting a series of studies directed toward the development of an efficient reliable test that will adequately assess an aviator's ability to hear speech in his operational environment. The investigations center around the utilization of multiple-word Modified Rhyme Test (ref. 1) items. This paper discusses two studies undertaken to determine whether the use of multiple-word Modified Rhyme Test items influences the intelligibility function of test words relative to their presentation as single-word test items, to obtain general information concerning the ability of individuals to perform the multiple-word recognition task, to explore possible word position effects, and to examine the possibility of having subjects rate the confidence with which they make their response choices.

The six basic lists of the Modified Rhyme Test, hereafter called the MRT, were randomized and reconfigured in a manner to provide two words per test item and three words per test item. The double-word MRT (DMRT) lists contain 25 two-word items and the triple-word MRT (TMRT) lists contain 17 three-word items. Since the latter required 51 words in order to balance the number of words per item, one word was chosen at random to be repeated as the third word in the last item of the test. The repeated word was not scored during subsequent data analysis.

MATERIALS AND METHODS

High quality recordings were made of an adult male talker reading the six word lists of the MRT, DMRT, and TMRT. The talker was experienced in the recording of materials for use in listening tests. The words were spoken without instrumental monitoring with the talker attempting to maintain a constant level of vocal effort throughout each list. The test words were spoken in the context of a carrier phrase which can be seen in figure 1, along with examples of MRT, DMRT, AND TMRT items. The talker attempted to read the test items in a manner and rhythm analogous to aircraft radio messages. While there was no attempt to establish a specific time interval between test words within an item, the speaker attempted to give discrete productions for each word. The interstimulus time between test items was approximately 3 seconds. On the average, the total elapsed time for the different tests was: 5 minutes for the MRT, 3 minutes for the DMRT, and 2.3 minutes for the TMRT.

Two response forms were constructed for each test so that the same form would not have to be used each time a particular word list was presented. Examples of response formats for each of the three tests may also be seen in figure 1.

Graphic level tracings were generated from each of the 18 master tape lists in order to equate the relative levels of the lists for experimental presentation and to establish the speech-to-noise ratios selected for study: +4 dB, 0 dB, and -4dB. A 1 kHz discrete frequency tone recorded at a constant voltage level prior to each test list was used to derive the relative levels of each of the test words in the different lists. For a given list, the level was derived by averaging the peak rms values for the 50 words in the list. Measurements from graphic level tracings of sub-master recordings of the level-equated lists indicated an average level deviation between lists of no more than ± 1dB. To provide the experimental tapes, the level-equated lists were played back on a high-quality tape recorder and mixed with white Gaussian noise shaped to simulate the spectrum of aircraft noise. The spectrum of the noise is shown in figure 2. The desired speech-to-noise ratios were obtained by keeping the level of the speech constant and varying the level of the noise relative to the level of the 1 kHz reference signal.

A preliminary study, Study I, was conducted to provide the investigators with general information concerning the ability of individuals to perform the multiple-word recognition task, to determine if there were any word position effects, and to examine the possibility of having subjects rate the confidence with which they made their response choices. Pollack and Decker (ref. 2) and Clarke(ref. 3) have indicated the efficacy of such rating procedures to determine the performance criteria of listeners in intelligibility testing, particularly since additional data are obtained with no apparent increase in experimental testing time. Since this type of analysis was being considered for future experiments, the inclusion of the rating procedure in Study I permitted us to determine whether the additional task would degrade the overall word recognition performance of the listeners. A four point scale was used to obtain the ratings: 1) "I know I heard the word correctly;" 2) "I think I heard the word correctly;" 3) "I don't think I heard the word correctly;" and 4) "I know I did not hear the word correctly."

Following Study I, a larger scale study, Study II, was conducted to provide a direct comparison of the double-word MRT and triple-word MRT with its parent test, the MRT. If listener scores for the multiple-word item tests are not significantly below those for the regular MRT (one word per item), it would appear that such modifications could be incorporated into the test without reducing its overall effectiveness. Moreover, the time required for administering the test would be considerably shortened. Conversely, if scores on the multiple-word item tests are significantly below those for the MRT, perhaps the increased degradation could be utilized to provide a more sensitive test instrument. The reasons for the increased degradation would, of course, have to be explored.

RESULTS AND DISCUSSION

Table I shows the test formats, test conditions, and number of test subjects utilized in Study I and Study II. The order of presentation of the test lists and different formats (MRT, DMRT, and TMRT) was randomized. The test lists were presented via earphones (diotically) at a sound pressure level of 80 dB. Group testing was employed with the subjects seated in a ten-man sound-treated booth. For each test item, the subjects responded by drawing a line through the word of their choice in the appropriate word ensemble boxes. In those instances where the subjects were asked to rate their responses, they wrote their rating scale numbers to the right of each ensemble box.

Table I. Test formats, test conditions and number of test subjects utilized in Study I and Study II

	Study I	Study II
Test Formats	DMRT, TMRT	MRT, DMRT, TMRT
Test Conditions	MRT Lists:	MRT Lists:
Quiet	A, B*	
+4 dB	C, D*	A, B, C, D, E, F
0 dB	E, F*	A, B, C, D, E, F
-4 dB	A, B*	A, B, C, D, E, F
Test Subjects+	5	10

*Subjects were asked to give a confidence rating following each of their responses.

+Subjects were male volunteers from the laboratory staff and young Naval officers in flight training. With the exception of one subject who had a moderate high frequency hearing loss, all subjects exhibited hearing within normal limits.

The mean percent correct listener scores obtained in Study I for the DMRT and TMRT formats at the different test conditions are shown in Table II. There were no significant differences between scores obtained with the two multiple-word test item formats for either the different speech-to-noise ratios or the rating and non-rating conditions.

Table II. Mean percent correct scores for the five subjects in Study I.

MRL LIST	TEST CONDITION	DMRT	TMRT
A	Quiet	100	98
B	Quiet (Rating)	100	98
C	0 dB	78	82
D	0 dB (Rating)	78	80
E	+4 dB	88	92
F	+4 dB (Rating)	88	92
A	-4 dB	64	64
B	-4 dB (Rating)	60	56

Table III displays the mean percent correct listener responses obtained in Study II with the MRT, DMRT, and TMRT formats for the six MRT lists at the three speech-to-noise ratios. While listener scores are comparable across lists for a given speech-to-noise ratio, there were some significant differences, both between lists within a given format and between formats within a given list. In general, a difference of about eight percentage points between any two mean scores is statistically significant at the .05 level of confidence. Possible list differences and subject learning during testing may account for some of the differences. While it has been shown that repeated exposure to the MRT does not change the level of average response in any appreciable way, this may not hold true for such modifications to the test as the DMRT and TMRT.

Table III. Mean scores and standard deviations (in parentheses) for the 10 subjects in Study II averaged according to test list, format, and speech-to-noise ratio. Grand means (GM) for each format are shown at the bottom.

	+4 dB			0 dB			-4 dB		
List	MRT	DMRT	TMRT	MRT	DMRT	TMRT	MRT	DMRT	TMRT
A	92 (4)	90 (4)	86 (6)	80 (6)	82 (8)	80 (6)	66 (4)	56 (10)	54 (10)
B	92 (4)	92 (4)	90 (6)	86 (6)	78 (8)	74 (12)	70 (6)	60 (10)	60 (8)
C	92 (2)	78 (10)	90 (4)	82 (6)	78 (4)	80 (8)	64 (6)	58 (8)	64 (10)
D	88 (4)	84 (4)	80 (6)	70 (6)	72 (8)	72 (8)	56 (4)	60 (10)	54 (8)
E	90 (6)	92 (6)	88 (4)	82 (6)	78 (6)	78 (10)	68 (6)	58 (8)	54 (10)
F	84 (6)	82 (6)	90 (4)	78 (4)	76 (6)	82 (8)	56 (6)	54 (6)	60 (12)
GM	90	86	88	80	78	78	64	58	58

With only one exception, for each list there were the typical changes in percent correct response as a function of speech-to-noise ratio. The one exception - List C, +4 dB, DMRT format - was always the first test to be administered. The significantly lower score obtained for List C at this condition is probably attributable to the subjects' initial learning and adjusting to their listening task.

Tabulations of the number of incorrect responses as a function of word position (totalled across speech-to-noise ratios and test lists) are displayed in Table IV for both Study I and Study II. As can be seen, whereas the non-rating condition exhibits word position effects, the rating condition does not. An examination of the number of incorrect responses with respect to whether a test word occurred during the first half of a test list or the last half of a test list revealed no large differences. For the DMRT format in both Study I and Study II, there were more incorrect responses for the second word. For the TMRT format, the position bias appears to be evenly distributed between the first and second words in Study I, and between the second and third words in Study II. The percentages of the total number of incorrect responses (non-rating) at the two DMRT word positions were 44 and 56 percent, respectively, in Study I and 46 and 54 percent in Study II. For the three TMRT word positions, comparable percentages were 45, 31, and 24 percent, respectively, in Study I and 27, 36 and 37 percent in Study II. The total number of incorrect responses for the DMRT and TMRT formats were not widely divergent in either study. They were, however, considerably larger than the total number of incorrect responses for the MRT, also shown in Table IV.

Table IV. Number of incorrect responses at the different word positions, totalled across word lists and speech-to-noise ratios for Study I and Study II.

	MRT	DMRT			TMRT			
Study I		Word 1	Word 2	Total	Word 1	Word 2	Word 3	Total
Without Rating		79	100	(179)	89	61	48	(198)
With Rating		98	97	(195)	64	63	66	(193)
Study II								
Without Rating	(2034)	1096	1253	(2359)	640	851	895	(2396)

The comparability of listener responses for the three test formats can be seen most clearly when the data are collapsed across test lists and plotted as a function of speech-to-noise ratio. Such a plot is presented in figure 3.

The largest divergence in scores among the three formats, about six percent, occurs at the poorest speech-to-noise ratio (-4 dB) where the mean score for the MRT is seen to be slightly better than the mean scores for the two multiple-word tests. The rate of change in percent correct response as a function of speech-to-noise ratio appears comparable across formats. Also shown in figure 3 are mean scores obtained for the two multiple-word test formats in Study I.

CONCLUSIONS

In conclusion, the data obtained in these two studies indicate that for the speech-to-noise ratios employed word recognition performance on multiple-word Modified Rhyme Test items is not appreciably different from that for the regular single-word format of the MRT. Having individuals given confidence ratings of their response choices in multiple-word item closed-response tests does not influence subject performance. Because of their more representative message length and decreased testing time (less than one-half the time required for the regular format of the MRT), the triple-word MRT (TMRT) test items appear to hold promise as suitable speech materials for use in the development of an efficient reliable test for evaluating the hearing capabilities of aircrew personnel.

FUTURE RESEARCH

Further data to be obtained utilizing the multiple-word item format with the Modified Rhyme Test materials and other closed-response test materials which test vowel as well as consonant intelligibility, should indicate the feasibility of using such a format in the evaluation of not only aircrew personnel but also talkers and listeners in general and communication systems. Also to be obtained are data relating to what role, if any, short-term memory plays in such a test procedure.

Single Word Test Item (MRT)

"Nine, do you read <u>look</u>? Over."

9

cook	book	hook
shook	look	took

Double Word Test Item (DMRT)

"Ten, do you read <u>fit</u>, <u>cut</u>? Over."

10

fizz	fill	fib		cuff	cuss	cub
fin	fit	fig		cup	cut	cud

Triple Word Test Item (TMRT)

"Four, do you read <u>saw</u>, <u>safe</u>, <u>hold</u>? Over."

4

saw	thaw	jaw		sale	sane	same		told	fold	cold
raw	paw	law		safe	save	sake		gold	hold	sold

Figure 1. Examples of single word (MRT), double word (DMRT), and triple word (TMRT) test items and response formats.

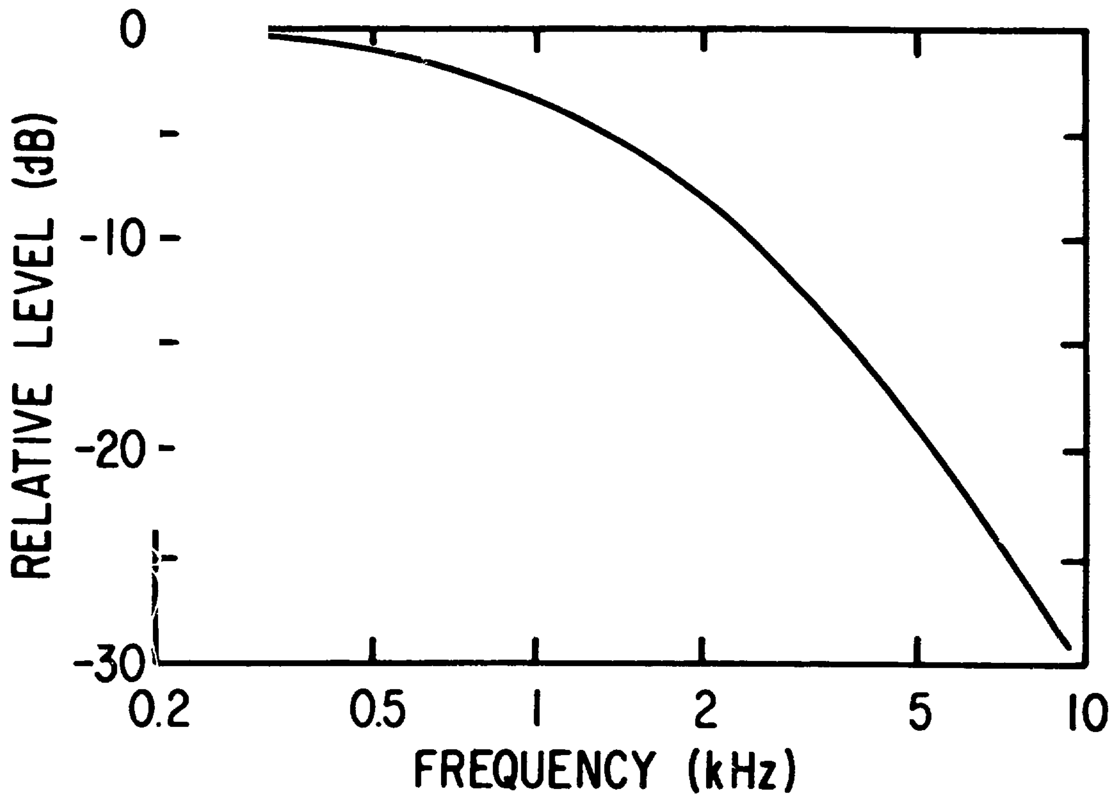


Figure 2. Spectrum of noise used in the experiments discussed in the text.

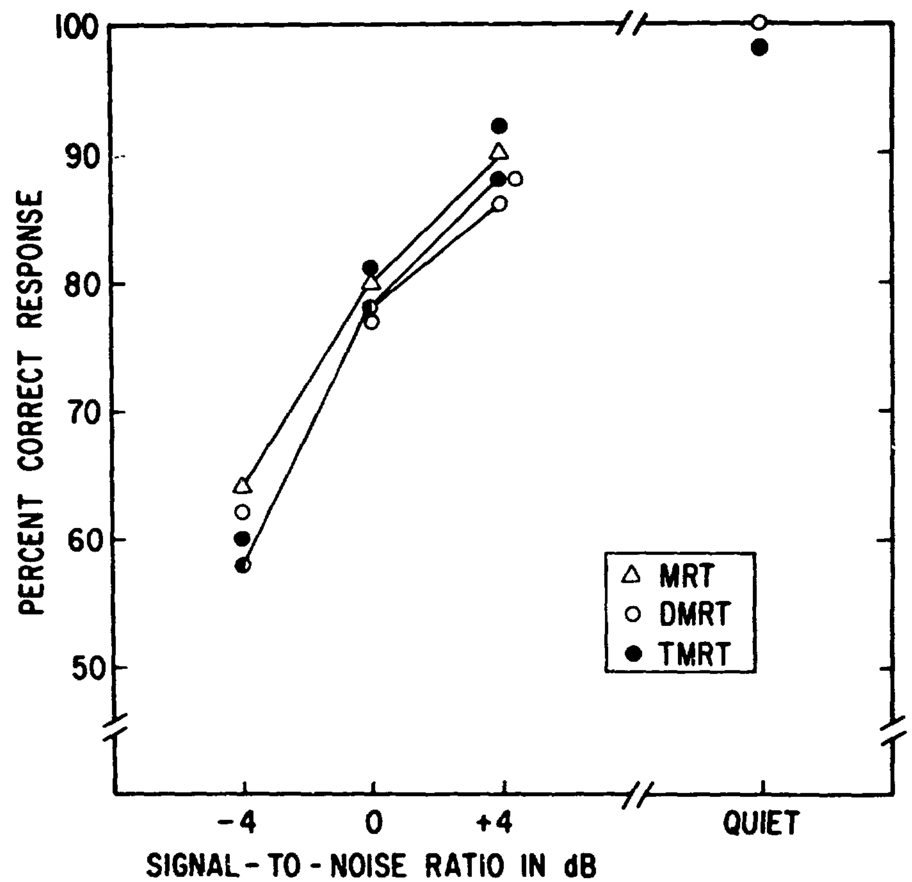


Figure 3. Mean percent of correct responses, averaged over test lists, as a function of speech-to-noise ratios for Study I (unconnected data points) and Study II (connected data points). Only the DMRT and TMRT formats were employed in Study I.

REFERENCES

1. House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D.: Articulation-Testing Methods: Consonantal Differentiation with a Closed Response Set. J. Acoust. Soc. Amer., vol. 37, 1965, pp. 158-166.
2. Pollack, I. and Decker, L. R.: Confidence Ratings, Message Repetation, and the Receiver Operating Characteristic. J. Acoust. Soc. Amer., vol. 30, 1958, pp. 286-292.
3. Clarke, F. R.: Confidence Ratings, Second-Choice Responses, and Confusion Matrices in Intelligibility Testing. J. Acoust. Soc. Amer., vol. 32, 1960, pp. 35-46.

FIGURE CAPTIONS

Figure 1. Examples of single word (MRT), double word (DMRT), and triple word (TMRT) test items and response forms.

Figure 2. Spectrum of noise used in the experiments discussed in the text.

Figure 3. Mean percent correct responses, averaged over test lists, as a function of speech-to-noise ratios for Study I (unconnected data points) and Study II (connected data points). Only the DMRT and TMRT formats were employed in Study I.

A TRI-WORD TEST OF THE INTELLIGIBILITY OF SPEECH

By

Russell L. Sergeant

Director, Communication Sciences Laboratory
Hunter College of CUNY
New York, NY

A. INTRODUCTION

Communication by speech involves the transfer of ideas or thoughts from the talker's to the listener's brain. Many things can interfere with that process. Some are linguistically oriented, some physiological, others acoustical, still others are oriented to electronic disruptions. Loud noise masks the intelligibility of speech. A pilot's helmet can either restrict the talker's ability to correctly articulate sounds or it can distort the acoustic signal that reaches the listener's ear. Vibrations in certain transportation vehicles can be a problem, or different hardware components in a voice communication system can be faulty. In the excitement of an emergency, the rapid speech of someone from East Dover, Vermont may not be understood by someone from the deep South. One type of distortion to speech is caused by high amounts of reverberation. Although the specific sources of speech distortion are nearly endless, they can be classified for simplicity into different categories, such as those oriented to talker, hardware, medium and listener.

Communications obviously is a vital part of any situation where people work together, and the most natural as well as efficient means of communication is speech. Therefore it is important in military operations to properly assess the existing communicability as well as its importance to the success of the operation at hand. There also is a need for critical and detailed evaluations during the development of the hardware to be used for communicating. The "Intelligibility Test" has been the principle bench-mark metric for evaluating the effect of different types of distortions caused by passing speech through various components of communication systems. The test material can be sentences, words, or nonsense syllables. Typically recordings are made of talkers reading materials from specially constructed speech tests, and then the recordings are passed through communications equipment to panels of listeners. The percent correct responses by the listeners is the intelligibility score, and it describes the efficiency for various combinations of the talker, the listener, and the effects of any distortions occuring between, i.e., from the hardware components or the medium through which the speech signals are transmitted.

Prior to the development in 1958 by Fairbanks of the Rhyme Test (RT), it was a tedious and time consuming task to obtain intelligibility scores. When used to evaluate hardware, test results depended on the talker's ability to speak clearly and the listener's training and experience in taking intelligibility tests. Williams, et al (1964) noted "Practical testing procedures that are convenient to administer and score, and at the same time are short and reliable, are not in general use."

House, et al (1963) revised Fairbanks' RT and called their version the Modified Rhyme Test (MRT). Using six rhyming lists of words, they introduced the closed response set. They found that the MRT was less affected by naive talkers and listeners than previous tests of intelligibility. In a restricted sense their modification also permitted the assessment of phonemic confusions.

In 1967 Griffiths (1967) modified the MRT into a simple diagnostic articulation test (DAT). His major addition was to improve the quality of phonemic comparisons by including all the minimal feature contrasts in English so that the efficiency of performance by a particular speech system could be estimated for conditions of natural speech. The DAT's capability for phonemic analysis can be applied to the construction of special vocabularies for use in specific situations where communication requirements are high but distortions are extreme, a situation which precludes unrestricted use of language. Like the MRT, the DAT is easy to administer and score, it produces stable responses with minimal learning effects from talker and listener, and it yields a useful index of the efficiency of communication components.

When measuring the performance of communication systems, intelligibility testing requires listeners to respond to speech stimuli. As an alternative method to evaluate hardware efficiency, communication engineers have developed a measure (French & Steinberg, 1947) based on levels of speech and noise in 20 equally contributing frequency bands. Called the Articulation Index (AI), reliable estimates can be made of intelligibility scores that would be obtained with the more cumbersome use of panels of listeners. There are corrections to the basic AI formula for different kinds of distortion, such as reverberation. However, Sachs, et. al (1969) found that for one reverberation-like distortion the AI fails to predict adequately results that are obtained with traditional articulation testing. A brief description of that distortion follows.

When an acoustic signal is transmitted through the ocean, a type of distortion in the time domain exists which is similar to reverberation. However, it differs from the traditional descriptions of reverberation which are familiar to room acousticians. Figure 1 summarizes several multipaths of a transmission as it travels from Point A to Point B. One path goes in a straight line from A to B. Another path includes reflections from the surface and/or bottom bounces, e.g., A to C to D to E to B. These two paths might be heard as the initial signal and its echo. A third type of path can travel from

A along several lines of sight and reflect off "area" F to B. Since area F is not a point source, the signal arriving at B may be comprised of an infinite number of reflections. The distorted signal which reaches B by this path has been smeared in the time domain. Speech distorted in this way is called "smeared speech".

The distortion of smeared speech, as well as a number of other types of reverberant speech, reveals an inherent difficulty in the traditional single word intelligibility test. Such tests do not take into account the influence of adjacent speech signals upon the speech signal under test. Consider a stimulus word which stands alone, i.e., without a lead-in or follow-up phrase. Time smearing distortions to the initial phoneme could occur from a backward smearing of the remainder of the word, but not from the silence preceding the phoneme. A similar analogy exists for the final phoneme. This type of distortion could also affect whole words. If the speech stimulus were a sentence, the initial word can be distorted by the rest of the sentence, the final word by the preceeding speech, and the middle words by both preceeding and following speech. In other words, there are pre-, per- and post- word distortions caused by time-smearing which can reduce the intelligibility of spe . Existing tests of intelligibility have not been designed to evaluate properly this type of distortion.

B. PURPOSE

The purpose of this study was to develop an intelligibility test which would account for unusual distortions caused by reverberant-like conditions. The test should have the desireable features of speed and ease of administering and scoring as well as a capability for diagnostically evaluating contrasts in distinctive features among phonemes typically used in natural speech.

C. DESCRIPTION OF TRI-WORD TEST OF INTELLIGIBILITY (TTI)

The TTI is composed of three lists. Each list contains 50 tri-word items. Different DAT lists are utilized for each of the three word positions. Table I shows which of the five DAT lists were used to produce each of the three TTI lists, and Appendix A presents the three complete TTI lists. Appendix B

Table I. Lists of the Griffiths' (1967) Diagnostic Articulation Test used to produce the Tri-Word Test of Intelligibility (TTI).

	DAT LIST USED		
TTI LIST	Initial Words	Middle Words	Final Words
A-1	A	B	C
A-2	D	E	B
A-3	E	C	D

is the listener's 50-item response form for all three TTI lists. Every item contains three 5-word response sets, one for each word position in a tri-word item. The five words comprising a particular response set are the rhyming words which make up the equivalent items across the five DAT lists. The order of words within each 5-word set have been randomized.

Tape recordings of the TTI lists were made in an anechoic chamber using a high q ality microphone and an Ampex PR-10 Tape Recorder. The talker, experienced in intelligibility testing, was raised in the San Francisco Bay area and spoke with a General American dialect typical of that region. Ten tri-word items were recorded immediately following, and with an attempt to maintain the same vocal effort as, a carrier phrase which was spoken with attempts to maintain peak VU readings of -3. There were intervals of approximately 2 sec between the carrier phrase and the first item, and between each of the other nine tri-word items. Each item was spoken as a monotonic three word phrase. This procedure was repeated for additional sets of ten tri-word items until all three TTI lists were recorded.

Preliminary presentations of the TTI lists to several panels of listeners with varied intervals between items indicated that a rate of presentation of one tri-word item every 9 sec was the most comfortable rate for groups of naive listeners to respond. Therefore, the final TTI stimulus tape followed that rate of presentation. In order to eliminate any effects of preceeding or following speech on the initial and final stimulus words, there was no carrier phrase surrounding the tri-word items.

D. EVALUATION OF TTI: PROCEDURES

Stimulus tapes were made of three lists from the Modified Rhyme Test and three lists from the CHABA Sentence Intelligibility Test (Silverman and Hirsh, 1955). The same talker recorded for the TTI lists was used for these recordings. All of the stimulus tapes were presented both in quiet and combined with different levels of noise. Measurements were made of each item with a Graphic Level Recorder, and the mean item level for each list was calculated for use in determining speech-to-noise (S/N) ratios. Noise was shaped by passing the output of a General Radio Random Noise Generator through a General Radio Multifilter set to pass frequencies from 300 to 3500 Hertz (Hz) with a down-slope of -6dB per octave. Listening Panels 1-3 heard the nine lists in quiet according to a semi-random Latin square design. The panel size and order of presentation of lists is shown in Table II. Note that each intelligibility list was heard by two listening panels, or approximately 40 listeners. Listening Panels 4-6 heard the same nine intelligibility lists combined with various levels of noise according to a semi-random Latin square. Table III shows the order of presentation of S/N and list, and the panel size for Panels 4-6 and 7. A different set of six S/N's determined from preliminary testing was used for each of the three types of tests in order to equate the range of difficulty of response among the tests and also to eliminate ceiling and/or cellar effects. A 7th panel heard an additional S/N condition with the TTI to more fully cover the range of correct responses to that test. S/N's varied in 5 dB steps from +5 to -20 dB for the CHABA lists, +10 to -15 dB for the MRT lists, and +20 to -10 dB for the TTI lists. Mean level of speech was set at a 70 dB Sound Pressure Level in the phones for all testing.

The seven listening panels were 136 Naval enlisted men who had passed a screening test for hearing at 15 dB ISO from 250 to 6000 Hz at the Naval Submarine Medical Center in New London. All intelligibility testing was done there also. The listeners received no special training in intelligibility testing procedures. Panels were presented the different test materials monaurally in a group testing room which contained 20 matched PDR-8 phones in MX/41-AR cushions. Listeners marked their responses to the TTI on the Response Form

Table II. Order of presentation of different intelligibility lists in quiet, showing panel size and the obtained mean percent correct responses.

LISTENING PANEL	PANEL SIZE	PRESENTATION ORDER	LIST	MEAN PERCENT CORRECT RESPONSES		
1	20	1	CHABA F		99.9	
		2	CHABA H		99.6	
		3	MRT B		95.0	
		4	MRT A		98.6	
		5	TTI A-1	90.4	89.7	91.8*
		6	TTI A-2	91.7	87.8	92.3
2	20	1	CHABA A		99.3	
		2	CHABA F		99.8	
		3	MRT B		96.2	
		4	MRT C		96.9	
		5	TTI A-2	92.0	86.1	94.0
		6	TTI A-3	90.5	87.7	94.2
3	20	1	CHABA A		99.2	
		2	CHABA H		99.0	
		3	MRT A		98.9	
		4	MRT C		98.4	
		5	TTI A-1	95.8	92.1	92.7
		6	TTI A-3	93.6	88.9	94.4

*The three mean percent correct responses for a TTI list are for the first, middle and last words of the 50 item tri-word list.

Table III. Order of presentation and condition of S/N for different intelligibility lists, showing panel size and the obtained mean percent correct responses.

LISTENING PANEL	PANEL SIZE	PRESENTATION ORDER	S/N RATIO	LIST	MEAN PERCENT CORRECT RESPONSES		
4	17	1	-5 dB	CHABA A		90.2	
		2	-10	CHABA F		64.8	
		3	-15	CHABA H		52.8	
		4	+10	TTI A-1	81.1	78.8	85.6*
		5	+5	TTI A-2	74.6	68.6	75.6
		6	0	TTI A-3	55.4	49.6	59.6
5	20	1	+5	MRT A		84.1	
		2	-5	MRT B		54.7	
		3	-15	MRT C		18.6	
		4	-5	TTI A-1	44.4	39.6	42.7
		5	+20	TTI A-2	83.4	79.0	87.9
		6	+15	TTI A-3	81.2	79.9	85.8
6	19	1	+10	MRT A		86.8	
		2	0	MRT B		70.2	
		3	-10	MRT C		47.4	
		4	0	CHABA A		97.3	
		5	+5	CHABA F		97.3	
		6	-20	CHABA H		95.3	
7	20	-	-10	TTI A-1	34.6	36.7	36.9

*The three mean percent correct responses for a TTI list are for the first, middle and last words of the 50 item tri-word list.

in Appendix B, a standard response form was used for the MRT, and responses were written on a blank sheet of paper for the CHABA sentences.

E. EVALUATION OF TTI: RESULTS

The mean percent correct responses to all tests are presented in the final columns of Tables II and III. Overall means in quiet were 99.5% for the CHABA lists and 97.3% for the MRT. Overall means for the TTI in quiet were 92.3%, 88.7% and 93.2% for the first, middle and final words respectively. The results by lists for different S/N's presented in Table III are shown graphically in Figure 2. The abscissa is S/N, the ordinate is mean percent correct responses. The random chance response differs among the three tests because of the small closed response sets used on the forms for the MRT and TTI. Therefore the following correction factors, Q, were applied to the obtained means (M):

$$\text{TTI:}\quad Q = .125\ (M - 20)$$

$$\text{MRT:}\quad Q = .120\ (M - 16.7)$$

$$\text{CHABA:}\quad Q = .100\ (M - 0)$$

Figure 3 shows the same data in Figure 2 replotted after Q-corrections. A Q-score of 5 represents a 50% mean correct response after correction for chance. The S/N ratios obtained for that point were -14.5 dB for the CHABA lists, and -5.2 dB for the MRT lists. For the TTI lists, the corrected 50% point was obtained for S/N's of -1.8,-3.0 and -0.3 dB for the first, middle and last words respectively. Analysis of variance indicated that significant (.05 level) trends exist among the three tests for changes in S/N, but these trends are not parallel from test to test. In addition, the mean responses to different S/N ratios among tests were quite different. As expected, the CHABA sentences were least affected by the level of background noise, and the TTI most affected.

Trend analysis for the three positions of test words in the TTI indicated parallel trends for changes in S/N. The mean correct responses for the word positions according to S/N also differed significantly. In the presence of noise the final word was easiest to identify, the middle most difficult, and the initial word was between the two. Based on this result, if one wanted to select the most intelligible words in 3-word phrases, he would choose the final words.

In the initial words of each TTI list, 25 items have response sets which differ only with regard to the initial phoneme. Consequently, for these words only the initial phoneme can be evaluated. Likewise, the response sets of 25 of the third words in the tri-word items differ only on the final phoneme. Comparisons can be made between the 25 initial and 25 final phonemes in a TTI list. Results of such comparison are presented as a function of S/N in Figure 4. Statistical analysis revealed significant trends with increased level of noise for both phoneme positions, but these trends were not parallel. The obtained F-ratio for testing the mean differences did not meet requirements for significance at the .05 level of confidence. It appears that the aberrant shape of the S/N function for the initial phoneme (see Figure 4) disrupts the parallelism between trends of the initial and final phonemes. Otherwise the responses for the two phoneme positions appeared similar.

F. SUMMARY AND CONCLUSIONS

The most usual means of assessing the efficiency of communication systems makes use of speech intelligibility tests. However, there are certain conditions of distortion for which traditionally used tests are not suited. Reverberation is one such condition. This report describes the Tri-word Test of Intelligibility (TTI) which was developed specifically to evaluate distortions to speech which are caused by reverberant-like interferences. There are three equated lists in the TTI, each consisting of 50 tri-word items. A list produces three intelligibility scores based upon the percent correct responses to the initial, medial and final words in the 50 items. Furthermore, in each list scores determined from 25 of the initial phonemes in the items can be compared with 25 final phonemes.

Taped recordings of the TTI, the Modified Rhyme Test, and CHABA Sentence Intelligibility Lists were played to 136 listeners divided into 7 listening panels. Results are presented for and comparisons made among responses to different equated lists of the tests for conditions of quiet and different levels of background noise. These results provide comparative data for future users of the TTI.

It was concluded from this study that the TTI is quick and easy to administer and score, it permits evaluations within a framework of phonemic distinctive features, and it provides different intelligibility scores for word position and phoneme position within tri-word items. Although a major feature incorporated into the design of the TTI is the capability for precise evaluation of distortions of speech caused by reverberation, the test should be equally efficient for assessing communicability under many other types of distortion as well.

G. ACKNOWLEDGEMENT

The major part of the study reported here was conducted in the Auditory Research Branch of the Naval Submarine Medical Research Laboratory prior to August, 1974, when the author was a staff member of the laboratory. The opinions or assertions contained in the report are the private ones of the author and are not to be construed as official or reflecting the views of the Navy Department or the Naval service at large.

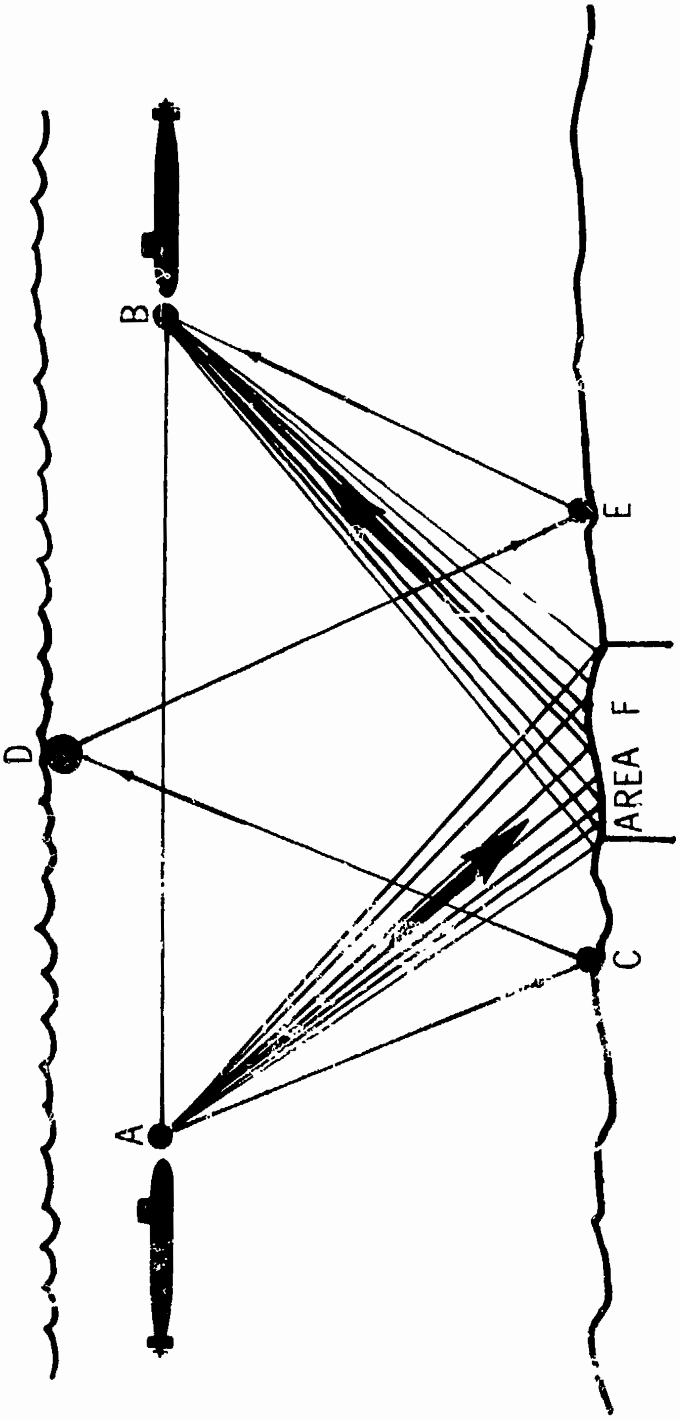


FIGURE 4.

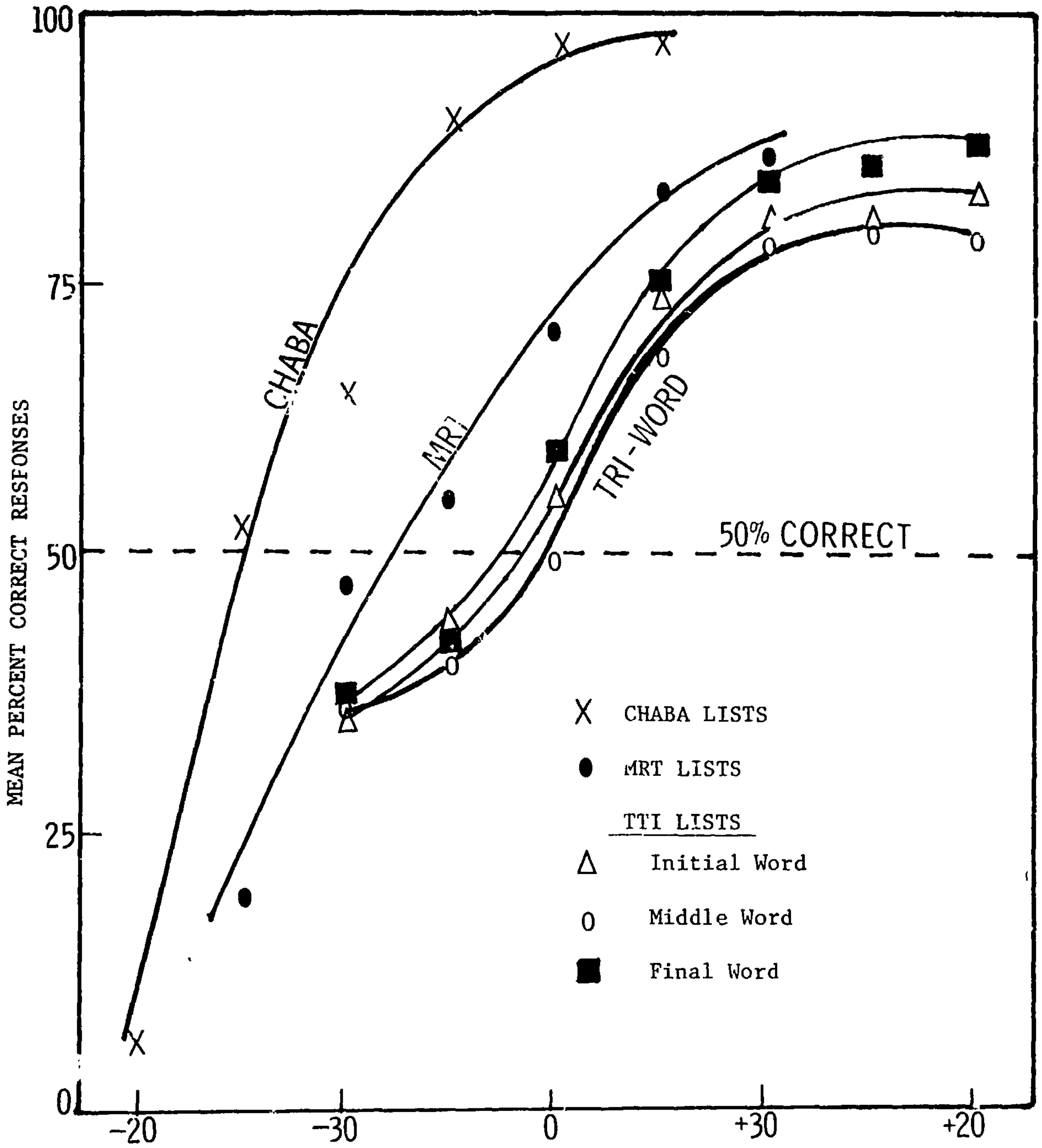


FIGURE 2.- SPEECH-TO-NOISE RATIO IN dB

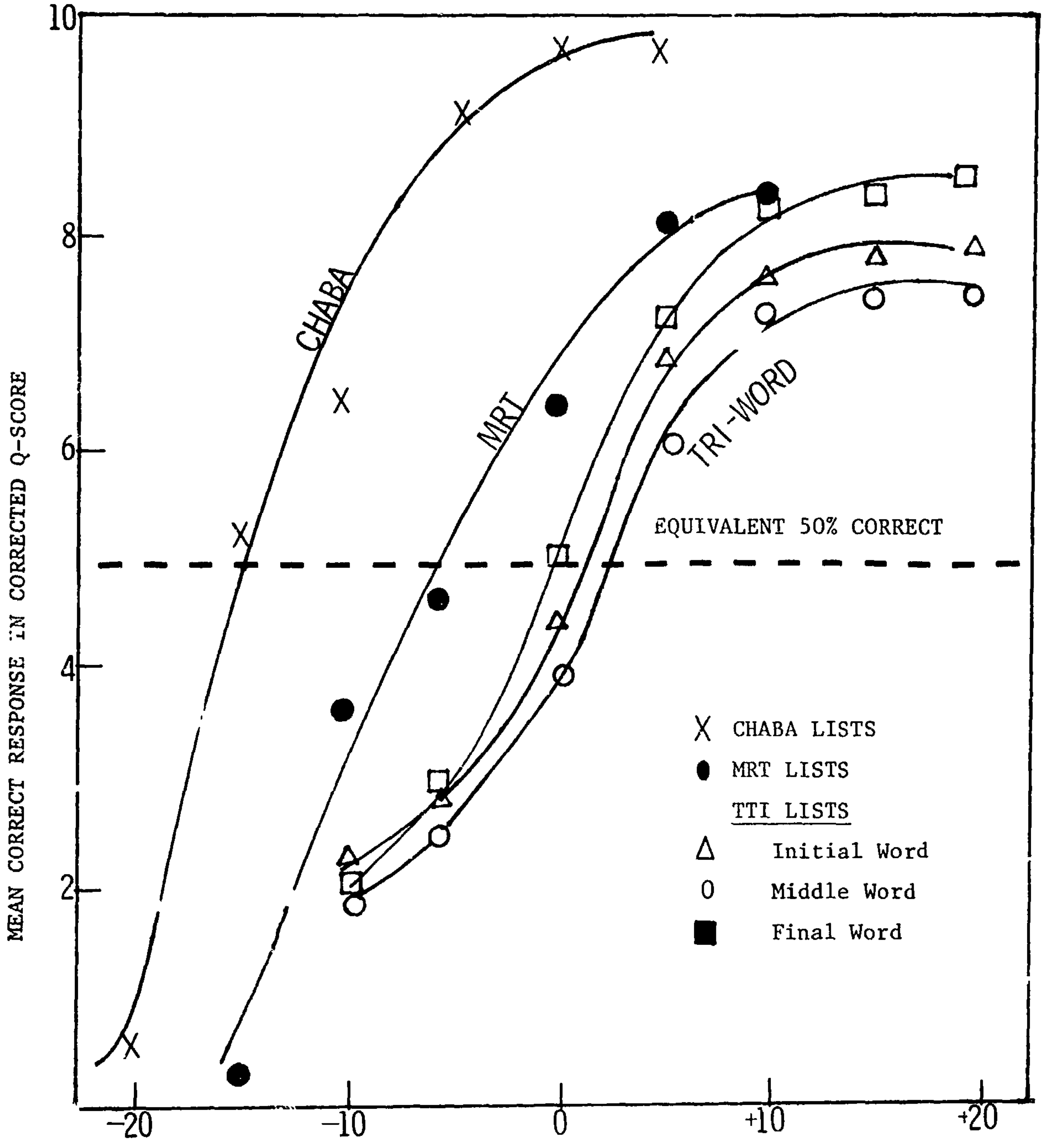


FIGURE 3.-SPEECH-TO-NOISE RATIO IN dB

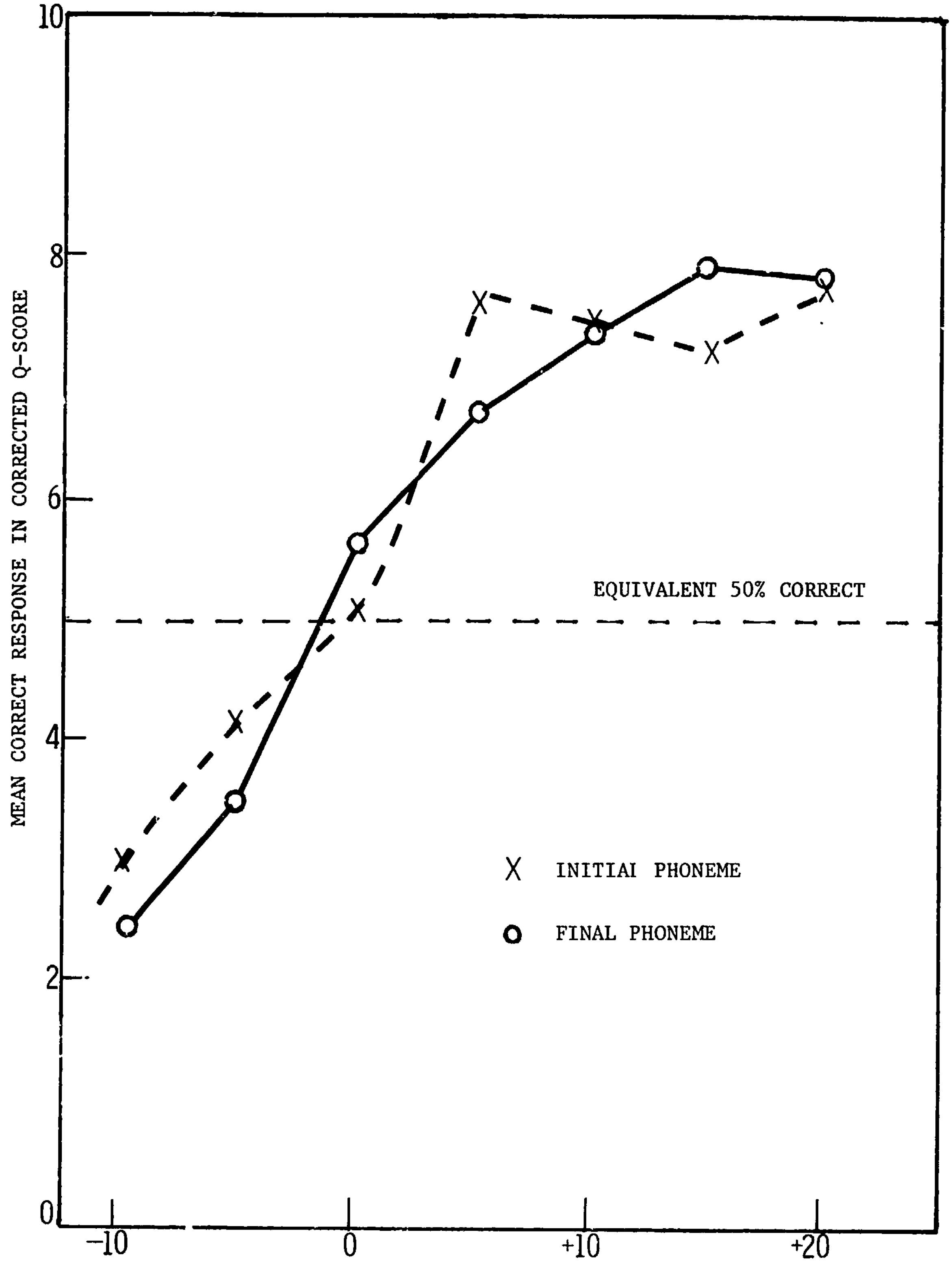


FIGURE 4.- SPEECH-TO-NOISE RATIO IN dB

H. REFERENCES

1. FAIRBANKS, G. 1958. "Test of phonemic differentiation: The rhyme test", J. Acoust. Soc. Amer., 30: 596-600.
2. FRENCH, N. R. AND STEINBERG, J. C. 1947. "Factors governing the intelligibility of speech sounds", J. Acoust. Soc. Amer., 19: 90-119.
3. GRIFFITHS, J.D. 1967. "Rhyming minimal contrasts: A simplified diagnostic articulation test", J. Acoust. Soc. Amer., 42: 236-241.
4. HOUSE, A. A., WILLIAMS, C., HECKER, M.H.L. and KRYTER, K.D. 1963. "Psychoacoustic speech tests: A modified rhyme test", Tech. Documentary Rpt. No. ESD-TDR-63-403, Electron. Systems Div., USAF Sys. Com., Hanscom Field, Bedford, Mass.
5. SACHS, M.B., O'BRIEN, G.J., SERGEANT, R. L. and RUSSOTTI, J.S. 1969. "Speech intelligibility in a stationary multipath channel", Naval Submarine Medical Research Laboratory and Naval Underwater Sound Laboratory Jnt. Rpt. No. 574, Naval Submarine Medical Center, Groton, Conn.
6. SILVERMAN, S. R. and HIRSH, I. J. 1955. "Problems related to the use of speech in clinical audiometry", Ann. Otol. Rhinol. and Laryngol., 64: 1234.
7. WILLIAMS, C. E., HECKER, M.H.L and KRYTER, K.D. 1964. "Methods for psychoacoustic evaluation of speech communication systems", Tech. Documentary Rpt. No. ESD-TDR-65-153, Electron. Sys. Com., Hanscom Field, Bedford, Mass.

APPENDIX A

TRI-WORD TEST OF INTELLIGIBILITY,

LISTS A-1, A-2, and A-3.

	A	B	C
1.	bat	base	that
2.	laws	cub	sin
3.	wig	batch	tam
4.	dumb	sin	seal
5.	cuff	just	came
6.	dig	lack	sub
7.	dun	peas	mark
8.	fill	dud	half
9.	leave	bent	pub
10.	toss	puff	hold
11.	lash	liege	vest
12.	mat	rip	tip
13.	beige	long	red
14.	pass	din	sag
15.	peak	mad	wit
16.	pick	sum	pip
17.	pup	best	went
18.	hath	pen	lee
19.	we're	weal	pop
20.	sad	cold	den
21.	sheen	path	big
22.	sing	sheave	dung
23.	sud	tear	cut
24.	tab	sip	kill
25.	teeth	dee	weave

TRI-LIST A-1

	A	B	C
26.	led	tan	pack
27.	sold	may	bayed
28.	dig	sat	log
29.	kick	chick	tale
30.	fin	dark	tong
31.	bark	game	lass
32.	gale	feel	sheathe
33.	peel	tin	tease
34.	will	fig	leach
35.	feel	with	chin
36.	hame	hop	fin
37.	ten	pit	tin
38.	pin	tin	shin
39.	thin	wig	bash
40.	thee	hill	eel
41.	rent	lip	doth
42.	hip	pale	did
43.	top	shed	peal
44.	yore	reel	fie
45.	vie	hash	wore
46.	zip	thy	gay
47.	next	vat	thick
48.	bust	dub	math
49.	mat	taj	rust
50.	way	gore	nip

TRI-LIST A-1

	D	E	B
1.	bass	bays	vat
2.	lodge	cud	sip
3.	witch	badge	tan
4.	duff	fin	reel
5.	cup	dust	game
6.	dim	lath	sum
7.	dub	peat	dark
8.	fizz	dug	hash
9.	leash	tent	puff
10.	talks	pus	cold
11.	laugh	lead	best
12.	man	lip	rip
13.	bathe	lob	shed
14.	pad	dill	sat
15.	peach	mass	with
16.	pig	sung	pit
17.	puck	west	bent
18.	have	then	dee
19.	weed	wean	hop
20.	sack	gold	pen
21.	sheath	pat	wig
22.	sit	sheaf	dub
23.	sun	teeth	cub
24.	tang	sick	hill
25.	teel	zee	weal

TRI-LIST A-2

	D	E	B
26.	wed	tap	path
27.	told	nay	base
28.	rig	sap	long
29.	pick	sick	pale
30.	kin	park	taj
31.	lark	tame	lack
32.	bale	keel	sheave
33.	heel	thin	tear
34.	till	fib	liege
35.	zeal	wick	tin
36.	same	shop	fig
37.	hen	pitch	sin
38.	win	gin	tin
39.	shin	pig	batch
40.	knee	bill	feel
41.	dent	ship	dud
42.	dip	male	din
43.	cop	fed	peas
44.	lore	veal	thy
45.	thigh	has	gore
46.	gyp	high	may
47.	rest	rat	chick
48.	gust	dove	mad
49.	fat	tog	just
50.	they	roar	lip

TRI-LIST A-2

	E	C	D
1.	badge	bayed	fat
2.	lob	cut	sit
3.	wick	bash	tang
4.	dove	tin	zeal
5.	cud	rust	same
6.	dill	lass	sun
7.	dug	peal	lark
8.	fib	dung	have
9.	lead	went	puck
10.	tog	pub	told
11.	lath	leach	rest
12.	mass	tip	dip
13.	bays	log	wed
14.	pat	did	sack
15.	peat	math	witch
16.	pitch	sub	pig
17.	pus	vest	dent
18.	has	den	knee
19.	wean	weave	cop
20.	sap	hold	hen
21.	sheaf	pack	rig
22.	sick	sheathe	dub
23.	sung	tease	cup
24.	tap	sin	till
25.	teeth	lee	weed

TRI-LIST A-3

	E	C	D
26.	fed	tam	pad
27.	gold	gay	bathe
28.	pig	sag	lodge
29.	sick	thick	bale
30.	thin	mark	talks
31.	park	came	laugh
32.	male	eel	sheath
33.	keel	shin	teel
34.	bill	fin	leash
35.	veal	wit	shin
36.	tame	pop	fizz
37.	then	pip	win
38.	fin	chin	kin
39.	gin	big	bass
40.	zee	kill	heel
41.	tent	nip	duff
42.	lip	tale	dim
43.	shop	red	peace
44.	roar	seal	thigh
45.	high	half	lore
46.	ship	fie	they
47.	west	that	pick
48.	dust	doth	man
49.	rat	tong	gust
50.	nay	wore	gyp

TRI-LIST A-3

APPENDIX B

FORM A RESPONSE SHEET

FOR TRI-WORD TEST OF INTELLIGIBILITY

SPEECH INTELLIGIBILITY TEST
TRI-WORD LIST

NSMRL/NUSC

FORM A RESPONSE SHEET

Name______________________________ Score____________

Date____________

1.	BADGE	BATHE	MAT	8.	FIN	DUB	HALF
	BATCH	BASE	FAT		FIB	DUNG	HAS
	BASS	BAYED	THAT		FIG	DUG	HASH
	BAT	BAYS	RAT		FILL	DUN	HATH
	BASH	BEIGE	VAT		FIZZ	DUD	HAVE
2.	LAWS	CUT	SIP	9.	LEAD	DENT	PUP
	LOG	CUB	SICK		LEAVE	RENT	PUCK
	LOB	CUFF	SIN		LEIGE	WENT	PUB
	LODGE	CUP	SING		LEASH	TENT	PUFF
	LONG	CUD	SICK		LEACH	BENT	PUS
3.	WIT	BADGE	TAP	10.	TONG	PUP	TOLD
	WICK	BAT	TAN		TAJ	PUB	SOLD
	WITH	BASS	TAB		TOSS	PUCK	COLD
	WITCH	BATCH	TAM		TALKS	PUS	GOLD
	WIG	BASH	TANG		TOG	PUFF	HOLD
4.	DUMB	WIN	VEAL	11.	LATH	LEAD	NEST
	DUFF	TIN	ZEAL		LAUGH	LEAVE	REST
	DOTH	PIN	REEL		LASH	LIEGE	BEST
	DOVE	SIN	FEEL		LACK	LEACH	WEST
	DUB	FIN	SEAL		LASS	LEASH	VEST
5.	CUP	JUST	CAME	12.	MAT	LIP	DIP
	CUB	BUST	SAME		MAN	DIP	TIP
	CUT	GUST	GAME		MATH	HIP	RIP
	CUD	RUST	SHAME		MAD	RIP	HIP
	CUFF	DUST	TAME		MASS	TIP	LIP
6.	DILL	LAST	SUD	13.	BEIGE	LONG	LED
	DIG	LACK	SUB		BATHE	LOG	FED
	DIN	LAUGH	SUN		BAYED	LAWS	SHED
	DID	LATH	SUM		BASE	LOB	WED
	DIM	LASH	SUNG		BAYS	LODGE	RED
7.	DUN	PEAT	LARK	14.	PAT	DILL	SACK
	DUG	PEAS	DARK		PAD	DIM	SAG
	DUD	PEAL	BARK		PASS	DIG	SAD
	DUNG	PEAK	PARK		PATH	DIN	SAT
	DUB	PEACE	MARK		PACK	DID	SAP

Name: ______________________________ **Date:** __________

15.	PEAT PEAK PEACE PEAS PEAL	MAD MASS MAT MAN MATH	WITH WIT WIG WICK WITCH
16.	PIT PIP PICK PITCH PIG	SUN SUD SUM SUB SUNG	PIP PIT PICK PIG PITCH
17.	PUB PUFF PUP PUS PUCK	BEST VEST NEST REST WEST	TENT RENT BENT WENT DENT
18.	HAVE HATH HASH HAS HALF	TEN THEN DEN HEN PEN	DEE ZEE KNEE LEE THEE
19.	WEED WEAL WE'RE WEAN WEAVE	WE'RE WEAN WEAL WEAVE WEED	COP TOP POP SHOP HOP
20.	SACK SAG SAD SAT SAP	HOLD SOLD TOLD COLD GOLD	PEN TEN THEN DEN HEN
21.	SHEAF SHEATH SHEEN SHEAVE SHEATHE	PASS PAT PATH PACK PAD	PIG WIG BIG DIG RIG
22.	SIP SING SIN SIT SICK	SHEEN SHEATH SHEATHE SHEAVE SHEAF	DUN DUG DUB DUD DUNG
23.	SUN SUM SUD SUNG SUB	TEASE TEAR TEETHE TEETH TEEL	CUB CUFF CUP CUT CUD
24.	TAM TAN TANG TAB TAP	SIP SING SIN SIT SICK	BILL HILL WILL KILL TILL
25.	TEAR TEETHE TEEL TEASE TEETH	KNEE DEE ZEE THEE LEE	WEAN WE'RE WEED WEAL WEAVE
26.	RED WED LED FED SHED	TAN TANG TAB TAM TAP	PACK PAD PATH PASS PAT
27.	HOLD SOLD GOLD COLD TOLD	NAY MAY WAY GAY THEY	BAYED BASE BEIGE BATHE BAYS
28.	PIG WIG RIG BIG DIG	SAD SAP SAT SACK SAG	LODGE LONG LAWS LOB LOG
29.	KICK THICK CHICK SICK PICK	SICK CHICK KICK THICK PICK	MALE BALE GALE PALE TALE
30.	THIN FIN KIN TIN SHIN	BARK LARK DARK PARK MARK	TALKS TOG TOSS TAJ TONG

Name:______________________ Date:______________

31.

PARK	CAME	LACK
MARK	TAME	LASS
DARK	GAME	LASH
BARK	SHAME	LAUGH
LARK	SAME	LATH

32.

MALE	PEEL	SHEAF
TALE	HEEL	SHEATH
GALE	EEL	SHEAVE
BALE	FEEL	SHEATHE
PALE	KEEL	SHEEN

33.

EEL	KIN	TEAR
HEEL	FIN	TEASE
FEEL	TIN	TEETH
PEEL	THIN	TEETHE
KEEL	SHIN	TEEL

34.

KILL	FIZZ	LEAVE
HILL	FIG	LEASH
WILL	FILL	LEAD
TILL	FIB	LEACH
BILL	FIN	LIEGE

35.

ZEAL	WIG	SHIN
FEEL	WIT	TIN
SEAL	WITCH	CHIN
REEL	WITH	GIN
VEAL	WICK	THIN

36.

TAME	SHOP	FIG
SAME	TOP	FIZZ
SHAME	HOP	FILL
CAME	COP	FIN
GAME	POP	FIB

37.

HEN	PITCH	TIN
DEN	PIG	SIN
TEN	PIT	FIN
PEN	PICK	PIN
THEN	PIP	WIN

38.

TIN	SHIN	TIN
PIN	CHIN	FIN
SIN	THIN	SHIN
WIN	TIN	KIN
FIN	GIN	THIN

39.

THIN	WIG	BASS
CHIN	PIG	BADGE
TIN	BIG	BAT
SHIN	DIG	BATCH
GIN	RIG	BASH

40.

DEE	HILL	PEEL
THEE	BILL	HEEL
ZEE	KILL	EEL
LEE	WILL	FEEL
KNEE	TILL	KEEL

41.

BENT	SHIP	DUFF
DENT	ZIP	DOTH
WENT	LIP	DUD
TENT	GYP	DOVE
RENT	NIP	DUMB

42.

TIP	GALE	DILL
DIP	BALE	DIN
RIP	TALE	DIM
LIP	PALE	DIG
HIP	MALE	DID

43.

SHOP	WED	PEAK
HOP	SHED	PEAT
TOP	LED	PEAS
POP	FED	PEAL
COP	RED	PEACE

44.

GORE	SEAL	THY
WORE	REEL	THIGH
ROAR	FEEL	FIE
YORE	ZEAL	HIGH
LORE	VEAL	VIE

45.

THY	HAVE	GORE
HIGH	HAS	WORE
FIE	HASH	YORE
VIE	HALF	LORE
THIGH	HATH	ROAR

46.

GYP	THIGH	NAY
NIP	HIGH	GAY
LIP	VIE	MAY
SHIP	THY	WAY
ZIP	FIE	THEY

C-3

Name: ______________________________ Date: ________________

47.	REST	RAT	CHICK
	VEST	FAT	SICK
	NEST	VAT	KICK
	BEST	THAT	PICK
	WEST	MAT	THICK
48.	GUST	DUFF	MATH
	BUST	DUMB	MAD
	RUST	DOVE	MAT
	DUST	DOTH	MASS
	JUST	DUB	MAN
49.	VAT	TOSS	GUST
	THAT	TALKS	JUST
	RAT	TOG	RUST
	FAT	TONG	DUST
	MAT	TAJ	BUST
50.	MAY	YORE	GYP
	NAY	ROAR	SHIP
	THEY	WORE	ZIP
	GAY	GORE	LIP
	WAY	LORE	NIP

IS INTELLIGIBILITY ENOUGH?

By

David C. Nagel

NASA - Ames Research Center

Moffett Field, CA.

In this conference we are concerned with how noise, specifically aircraft noise, affects the communication process among people and how this disruption in turn is related to noise-induced annoyance. The main point that I hope to make here is that if we wish to predict the amount of annoyance that will result from undue noise exposure, it may not be sufficient to only consider measures of speech <u>intelligibility</u> as indicators of communication effectiveness. Further, I hope to show that the conceptual framework known as information processing can be a productive vehicle for beginning to understand the complete effects that noise and perhaps other stresses produce in human cognition.

It has widely been suggested that disruption of communication is a major component leading to noise dissatisfaction. This evidence has come from at least two sources: social survey work (e.g. Borsky (1961); McKennell, 1963; Hazard, 1971), and laboratory experiments (e.g. Williams, Stevens and Klatt, 1969). These studies have clearly pointed to communication disruption as a strong determinant of annoyance. Indeed, the study by Williams, and coworkers has established some relatively reliable relationships between noise level and rated annoyance with a given noise environment.

The question that I wish to entertain here, however, is somewhat different. Specifically, what is the proper way to measure the amount of disruption of the speech communication process caused by any particular noise environment? This question has previously been approached from a number of viewpoints but most often in the context of the assessment of the quality of electronic communication systems. A number of categories of communication system tests have been identified, including articulation tests, intelligibility tests, speech interference tests, and speech comprehension tests (Chambers, 1973). The major emphasis of these tests has been on the accuracy of immediate identification of speech sounds at the phonetic, phonemic or syntactic levels. However, little attention has been focused on the _efficiency_ with which information is communicated, although speech comprehension tests partially address this question. To be sure, intelligibility is the most obvious thing to examine initially, if we cannot hear a spoken message or understand individual words, further processing is difficult or impossible. However, recent advances in the modeling of human information processing (e.g. Norman and Lindsay, 1973) suggest that reduction of intelligibility may be only the most obvious manifestation of the disruption of the speech understanding process. Even in situations where intelligibility is perfect, interference with the total communication process may be taking place.

Actually, telephone engineers and others associated with the design of advanced electronic communication systems, have been aware for some time of the need for assessment tools that address more subtle issues than intelligibility. The need to quantify communication system quality, for example, has led to a number of test paradigms. Pollack and Decker (1958) asked subjects to rate how confident they were that the message they reported in a sentence comprehension experiment was in fact the one that was transmitted. Confidence ratings were found to be reliably related to average percent correct message reception even when signal-to-noise ratio was varied. Even more importantly, however, this study illustrates an attempt to assess how satisfactory the process of communication is, from the point of view of the receiver. Such a measure might well depend on factors other than simple intelligibility provided by the system. Munson and Karlin (1962) suggested that equal preference contours could be constructued on a two-dimensional grid of speech level and noise spectrum level so that different speech/noise combinations could be effectively ranked in terms of quality of the transmission system. Richards and Swaffield (1953), (cited in Broadbent, 1958) among others, have suggested that the level of effort that must be expended by individual speakers and listeners is a good subjective measure of speech communication system quality. Nakatani (1971) proposes the intelligibility of speech in the presence of interfering speech as a good index of effectiveness of a telephone system of high quality. It is my intention to suggest that these kinds of measures, although they might be considered secondary measures of speech communication effectiveness, nevertheless be integrated into any assessment of annoyance due to aircraft noise, and that their inclusion is especially important where intelligibility is essentially perfect.

The Information Processing Model

I have asserted that information processing is the conceptual framework that will best explain (and predict) some of the more subtle effects that noise and other stresses may produce in cognition.[1] To make it clear why this should be so I would like to very briefly review some of the major elements of this metatheory.

Figure 1 is a schematic version of a model proposed by Norman and Rumelhart in 1970 to explain how people process very simple visual stimuli (e.g. letters of the alphabet in a recognition task). Although somewhat

[1]This view has been proposed before. See particularly Broadbent (1958, 1971).

removed from the kinds of speech processing we are discussing here, the model is exemplary of the class of information processing models. The major points that this model illustrates are the following:

(1) sensation, perception, memory and thought are mutually interdependent,

(2) perceptual response is assumed not to be an immediate and direct consequence of a stimulus but rather is assumed to have gone through a number of stages of processing, each of which takes time to organize or traverse,

(3) increased time to perform a task reflects either an increase in complexity of processing or a decrease in processing efficiency,

(4) processing is limited by capacities of the information processing channels or the central processor, the contents of the stimulus, and/or the prior experience and condition of the observer, and,

(5) the role of memory and memory processes is emphasized because information is recoded and preserved, with varying degrees of fidelity, at each of the stages in the overall process.

To be sure, the processing of speech is somewhat more complicated in a number of respects, analysis of the meaning as well as the surface structure of the stimulus being necessary. More complicated models have been constructed for processing tasks of greater complexity. Nevertheless, the essential features of this class of models as noted above is assumed to hold.

Given this metatheory, the variety of ways that noise (or any other stressor) might interfere with the processing of speech information may be made clearer.

As noted above, this model is for a relatively simple perceptual task, e.g. tachistoscopic recognition of visually presented material. The processing of speech is clearly a more complex business involving processing of the meaning as well as the surface structure of the verbal stimulus. Yet, the general features of the model (the limited capacity, recoding and temporal emphasis notions, for example) suggest: (1) a variety of ways in which a stressor such as noise might affect the processing of information and (2) a variety of ways to measure these effects. In fact, there exists a body of literature that illustrates some of these more subtle noise effects, quite distinct from the more traditional changes in intelligibility, and that are nicely consistent with the general information processing model. I will review some of these below.

Noise and Information Processing

An important notion is that increased time to perform a task represents either an increase in processing complexity or a decrease in processing efficiency. In either case, the expression, "an increase in processing load", is often used. Pollack and Rubenstein (1963) administered a standard articulation task to observers with broadband noise of various levels mixed into the communication circuit. In no case was the noise of sufficient level to cause decrements in measured intelligibility. The response time, nevertheless, was a monotonic increasing function of the noise level. It thus appears that noise which has little effect on overall recognition performance might produce an increase in processing load.

Holloway (1970) reasoned that if accuracy can be traded off against response speed when processing complexity or load is increased, then restricting the time allowed for responding should lead to a reduction of intelligibility performance. Observers were given an immediate recognition task for monosyllabic words. The syllables were presented in five levels of noise and at six presentation rates, from 24 to 112 words per minute. Results are shown in Figure 2. Although in this case, noise did markedly affect intelligibility at the lower speech-to-noise ratios, the important result is that there is an interaction between speech-to-noise ratio and presentation rate. Specifically, decreases in intelligibility are more pronounced for fast presentation rates than for slow. The result is therefore consistent with the idea that, to a degree, greater accuracy may be achieved if more time is allowed for processing. Noise adds to processing complexity in addition to acting as a masker for these subjects.

Other examples of these more subtle effects of noise on the communication process are provided by Rabbitt (1966; 1968). In a first experiment, (1966) subjects were presented lists of four letter nouns over a loudspeaker. The words were either presented in quiet or mixed with pulse modulated noise. Subjects had no trouble shadowing (e.g. repeating aloud) each word as it was presented, whether in quiet or in the noise. When given a delayed recognition task, however, in which both target and distractor words were presented subsequent to an initial presentation of a target list, subjects misidentified more of the distractor words presented in noise than they did in quiet. The correct identification rate for target words remained about the same in either case. Table I shows the results. The two indices labeled, respectively, d' and β, are theoretical parameters which correspond to observer accuracy and observer criterion. It should be noted that both accuracy and criterion parameters are reduced when the test words are presented in noise. Experiment 2 suggests that the locus of the noise effect is the time at which the words are first presented for memory. This is shown in the lower half of Table I where "quiet/noise" denotes that the original list was memorized in quiet but tested in noise. The accuracy index is higher for this condition than when the words had initially to be memorized in noise but recognized in quiet.

Table I

RECOGNITION: MEMORY FOR WORDS CORRECTLY HEARD IN NOISE (Rabbitt, 1966)

Mean number correct and false positive scores, with calculated d' and β for four noise/quiet conditions

	Mean correct	mean false alarm	mean d'	mean β
Experiment 1				
quiet (N=17)	12.71	2.20	2.24	5.69
noise (N=29)	12.17	3.92	1.93	3.77
Experiment 2				
quiet/ (N=12) noise	11.59	2.50	2.05	3.77
noise/ (N=29) quiet	12.08	4.33	1.87	3.52

Recall of words initially learned in noise is similarly affected. Rabbitt (1968) asked subjects to recall lists of eight digits which were initially presented for memorization in either quiet or mixed with "0 dB S/N" noise. Immediate recognition was virtually unaffected as shown in the upper half of Table II. Delayed recall (in which observers must reproduce or "recall" the digit sequence at some point following the initial presentation) is differentially affected by quiet and noise, however. Sequences were found to be more difficult to recall if lists were initially heard in noise.

Table II

MEAN NUMBER OF LISTS OF EIGHT DIGITS CORRECTLY REPRODUCED
RECOGNITION AND RECALL (Rabbitt, 1968)

	Digits presented in quiet	digits presented in noise with 0 dB S/N
Recognition (and transcription)	10.00 (S=0.0)	9.64 (S=0.48)
Recall	4.02 (S=3.9)	2.84 (S=4.20)

An additional experiment by Rabbitt (1968) appears to suggest that the increased difficulty of recall can be attributed to a reduction in observer's capacity to <u>rehearse</u> the digit sequences when they are heard initially in noise. In this respect the results for recall are the same as those for delayed recognition; the decreased performance appears to be due to a decrease in cognitive capacity (specifically a decrease in ability to commit the information to storage) produced by the noise.

Thus, rote memory tasks appear to be performed less efficiently by subjects when they are forced to listen to the memory items in noise, even though intelligibility may remain essentially perfect. Are other more complex aspects of the communication process affected as well? Rabbitt (1968) performed a further experiment in which subjects were read one of two prose passages and then asked questions about the content of these passages. Ten questions in all were asked, five from the first half of each passage and five from the second half. In the first condition of the experiment both halves of each passage were recorded through a simulated telephone link of relatively high fidelity and low noise. In the second condition the first half of each passage was recorded as previously while the second half was mixed with noise that was maintained at an instantaneous noise level 5 dB below that of the speech signal. The results of the experiment are shown in Table III. Interestingly, the no-noise subjects performed significantly better than the quiet/noise even on the first half of each passage. Apparently attention to a continuous stream of new verbal data must be shared with rehearsal and other cognitive processes associated with the assimilation of what has already been heard. If more attention must be allocated to processing of later material, less capacity is available for continued processing or development of understanding of earlier material and recall may be impaired.

Table III

MEAN NUMBERS OF QUESTIONS ANSWERED CORRECTLY RE SCIENTIFIC AMERICAN EXTRACTS (Rabbitt, 1968)

	N	First Half of Passage	Second Half of Passage
Passage A			
No Noise	36	2.1 (S=1.7)	3.2 (S=1.9)
Quiet/Noise	36	1.7 (S=1.4)	2.5 (S=1.6)
Passage B			
No Noise	26	1.8 (S=1.6)	2.6 (S=1.5)
Quiet/Noise	26	1.2 (S=1.1)	2.4 (S=1.8)

One final line of evidence points toward a pre-emptive effect that noise may have on cognitive processing. In an experiment reported by Broadbent (1958) subjects were required to share their attention between a visual tracking task and a standard articulation test. Two forms of signal distortion were chosen (simple filtering and frequency translation) which produced the same level of performance as the articulation test, in the absence of the tracking task. The distortions were either applied singly or in combination and performance on both the articulation test and the tracking task monitored. Results are shown in Table IV. The articulation task scores are shown in the top half of this table, (Table IVa) tracking task scores shown at the bottom (Table IVb). The important result is that the tracking and articulation scores are essentially independent. I have circled the interesting comparisons. Note that for the two conditions circled (with dashed lines) where visual tracking scores are identical, articulation scores vary from 67 to 81%. Similarly, the solid circles show conditions which produce essentially equivalent tracking performance but greatly varying articulation scores.

NOISE LOAD AND SUBSIDIARY TASK PERFORMANCE (Broadbent, 1958)

Table IVa. The percentage of words correctly heard with a simultaneous visual tracking task

High Pass Filtering (cutoff-Hz)	Frequency Transposition -300 Hz	-200 Hz	0 Hz
0	63	86	97
660	67	58	81

Table IVb. The mean score on the visual tracking task while listening to various distortions of communication channel

High Pass Filtering (cutoff-Hz)	Frequency Transposition -300 Hz	- 200 Hz	0 Hz
0	336	333	365
660	363	347	363

This last experiment illustrates, in a most graphic fashion I believe, the concept of processing capacity and processing strategy. One can maintain performance on a particular task at the expense of performance on a subsidiary task. Maintenance of high performance on the primary task in most cases can only be achieved at the expense of extra effort. Is it unreasonable to suppose that people are aware of this kind of cognitive cost and that this awareness may lead to annoyance?

The available evidence is suggestive on this point but hardly conclusive. Rabbitt (1966) reports that subjects who were able to maintain high articulation scores in a noisy environment nevertheless spontaneously exhibited a high degree of annoyance because of the increased difficulty they experienced in attempting to remember the material.

What conclusions may be drawn from these studies? First, intelligibility and other measures of communication efficiency, as may be reflected in increased processing time, reduced capacity for performing other tasks or reduced memory retention abilities, may be relatively independent. Secondary measures of communication efficiency may exhibit greater sensitivity to noise disruption than intelligibility. If subjective ratings of annoyance are in any way tied to these, annoyance may be underestimated by intelligibility scores. Second, the kinds of disruption of the communication process we have been discussing may well be representative of the action of noise as a stress rather than noise as a masker. If this is the case, it may well be helpful to consider such effects within the general context of an information processing model such as the one discussed. Finally, the studies I have reviewed have failed to deal in any quantitative manner with noise parameters and sizes of the various effects for the categories of disruption discussed. If these kinds of effects are deemed important enough to warrant further study in the context of aircraft noise, carefully selected information processing paradigms should be used to establish relationships between noise parameters, information processing abilities, and subjective ratings of annoyance.

SUMMARY

Intelligibility may be only the most obvious measure of the disruption effect that aircraft noise produces within the context of speech communication. The literature outlining some of the secondary effects of noise on human information processing and a conceptual model for interpreting these effects are reviewed. It is concluded that secondary measures of communication efficiency (i.e. information processing performance) may prove to be more sensitive indicators of noise disruption and noise induced annoyance than primary measures such as intelligibility.

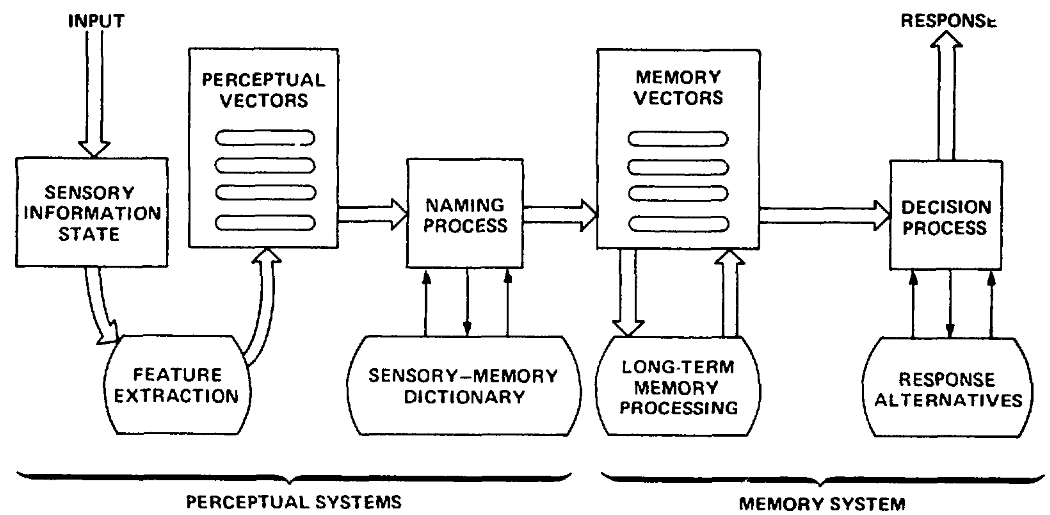


Figure 1. Information processing model for tachistoscopic recognition of letters of the alphabet (Norman and Rumelhart, 1970).

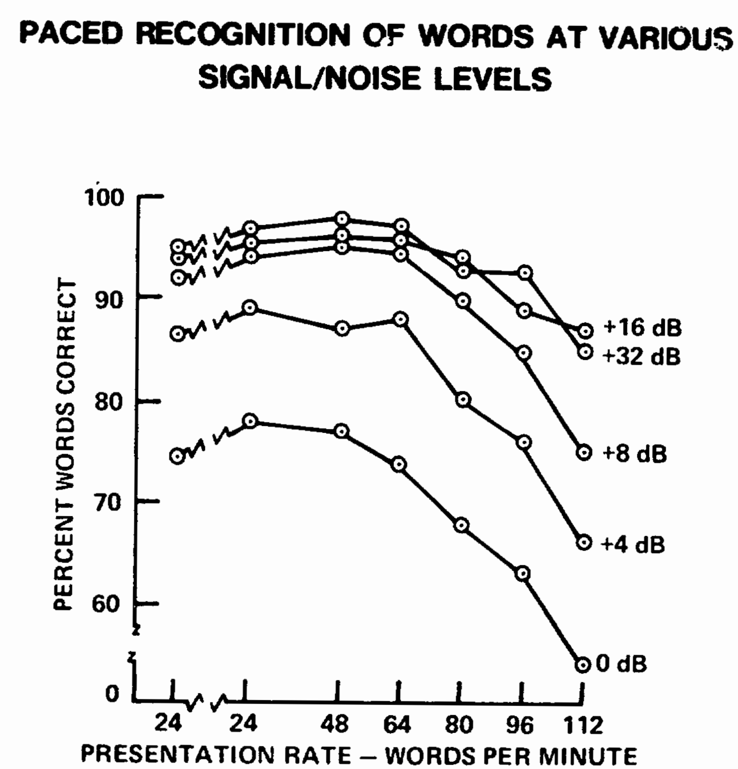


Figure 2. Recognition performance for words presented at increasingly quickened rates of presentation and at varying signal/noise levels. The difference in performance that can be attributed to the effects of noise is greater at the presentation rate of 112 wpm than at 24 wpm.

REFERENCES

Borsky, P. N. Community reactions to Air Force noise, I: basic concepts and preliminary methodology; II: data on community studies and their interpretation. NADD Report 60-689, 1961.

Broadbent, D.E. Perception and Communication, Pergammon Press, Oxford, 1958.

Broadbent, D.E. Decision and Stress, Academic Press, London, 1971.

Chambers, A. A review of tests for the evaluation of speech communication with particular reference to high speed low level strike aircraft. Royal Aircraft Establishment, Technical Memorandum EP543, 10 May 1973.

Hazard, W.R. (1971) Predictions of noise disturbance near large airports. J. Sound and Vibration, 15 (4): 425-445.

Holloway, C.M. (1970). Paced recognition of words masked in white noise. J. Acoust. Soc. America, 47 (6): 1617-1618.

Lindsay, P.H. and Norman, D.A. Human Information Processing, Academic Press, New York, 1972.

McKennell, A.C. Aircraft noise annoyance around London (Heathrow) Airport. S. S. 337, Central Office of Information, 1963.

Munson, W. A. and Karlin, J.E. (1962). Isopreference method for evaluating speech transmission circuits. J. Acoust. Soc. America, 34: 762-774.

Nakatani, L. Measuring the ease of comprehending Speech, Proceedings, VIIth International Congress on Acoustics, Budapest, 1971.

Norman, D.A. and Rumelhart, D.E. A system for perception and memory, in Models of Human Memory, D.A. Norman (Ed.) Academic Press, New York, 1970.

Pollack, I. and Decker, L.R. (1958). Confidence ratings, message reception, and the receiver operating characteristics. J. Acoust. Soc. America, 30: 286-292.

Pollack, I. and Rubenstein, H. (1963). Response times to known message sets in noise, Language and Speech, 6: 57-62.

Rabbitt, P.M.A. (1966). Recognition: memory for words correctly heard in noise. Psychonomic Science, 6: 383-384.

Rabbitt, P.M.A. (1968). Channel capacity, intelligibility and immediate memory. Q. Journal Experimental Psychology, 20: 241-248.

Williams, C.E., Stevens, K. N. and M. Klatt (1969). Judgments of the acceptability of aircraft noise in the presence of speech. J. Sound and Vibration, 9(2): 263-275.

OBJECTIVITY-SUBJECTIVITY CONTINUUM IN INTELLIGIBILITY TESTING

By

G. C. Tolhurst, Professor

University of Massachusetts

ABSTRACT

At the present time there are no speech testing methods that truly predict speech communication efficiency. There does exist a considerable body of data concerning speech reception. This data should be collated and abstracted into meaningful transfer functions. In the most experimentally rigid studies, there remain plaguing subjective factors contributing to prediction variability. Hence, it is suggested that a frankly subjective scaling method of speech testing may offer some advantages over present techniques.

OBJECTIVITY-SUBJECTIVITY CONTINUUM IN INTELLIGIBILITY TESTING

The long history of speech testing and the continued application of a variety of methods used to evaluate communication systems, either whole or in part, indicate the non-universality of a single, acceptable procedure. The following discussion is an honest evasion of a "true" answer to the question, "can speech interference, or speech intelligibility 'really' be predicted." At the present time, obviously, there is no single unequivocal answer. And before any meaningful discussion can be initiated, any possible answers will hinge upon the interpretation of the word "really" in the question above, as well as for several other terms.

One of these other terms needing more precise specification is "speech intelligibility". This phrase and its synonym "articulation testing" afford very little information as to the focus of experimental attention under investigation, i.e., the ends of the talker-listener continuum. To reduce ambiguity in reporting of experimental procedures and data, it is suggested that the investigator use the term speech reception (scores or values) if message reception is the dominant factor being explored, or the output of a system being assessed. If the experimental variable is some characteristic of the talker, i.e., dialectal differences, education, modification of auditory feedback, environmental or "internal" stressors, etc., then the appropriate descripter term would be speaker intelligibility.

If the word "really", in the first paragraph above, means validly and reliably predicting message transference under all permutations of talkers, listeners, noise environments and communication equipments, the answer must be a blunt, "No". Even in well controlled laboratory situations, with only one element of the "communication chain" allowed to vary systematically, the variance is often unacceptable. When all of the elements can be affected simultaneously, as in most operational environments, it is fortunate that spoken language is so highly redundant.

However, if the word "really" can mean adequately predicting listener reception efficiency (either operationally or pragmatically), then the answer is a reasonably firm "Yes". If intelligibility means speaker intelligibility, the answer must be a reasonably firm "No". There exists a rather large body of experimental data concerning listener reception. While there is a considerable number of studies exploring speaker intelligibility, these usually lack the statistical wealth of subjects representing populations as found in listener reception studies. This speaker-subject condition is due, in part, to the numbers of listeners whose responses must be used to validate the output of a single talker; costly in terms of manpower and time.

Since there exists such a substantial corpus of data relating listener reception efficiency to a wide variety of speech samples (testing materials), environmental conditions, psychological and physiological factors, it should be possible to abstract and collate the findings up to the present with a goal of constructing transfer functions which would allow listener reception productions across several of the reception influencing factors before other more objective testing methods are sought. For example:

There is a well-known family of monotonic curves relating percent correct speech reception versus Articulation Index (AI) abstracted from various investigations that used speech test materials varying in difficulty level from Spondee words, sentences, rhyme tests, multiple-choice, PB, to CVC nonsense syllables. The results on the same sort of speech tests have been plotted for percent correct reception versus signal-to-noise ratios (S/N) yielding similar monotonic functions. In neither case are the functions linear, but have the usual sigmoidal shape. Now, if one were to carefully evaluate all of the contributing data abstracted in the two [illegible]ies of curves and found the data points reliable, it should be possible to [illegible]ne the data to derive compound relationships. By taking a particular percent correct score, i.e., 25, 50 or 75, and plotting these points for each test along S/N and AI axes, a preliminary hypothetical set of functions probably would look like those shown in Figure 1. Each percent line has a different slope, but the speech test type relationships are roughly linear. The 50 percent line does seem to approach a slope of one.

To examine further the linearity of the speech test type relationship, it is possible to construct an AI versus AI function using the same percent correct points as above for each type of speech test. This hypothetical comparison might be similar to the plot shown as Figure 2. Ideally, the three lines should have a slope of one but be separated at two regions along the diagonal. The percent correct line slopes do not deviate far from one, but there is considerable curve overlap, i.e., no separation.

Another graphic summary which should prove interesting, if the data were carefully evaluated, equated, and properly plotted, would be to use the 50 percent value found for each of the various types of speech tests determined under a variety of noises, varying in complexity and band width, and hold the S/N constant.

In restructuring the presently available data one of the restrictions that critical examination should reveal is that many investigators have modified the output speech signal, usually deliberately. In otner words, the data-base would have to consist of studies in which the signals were presented over a relatively broad-band system (0.2 - 8 KHz) and be unprocessed, that is, not peak clipped (Licklider, 1945), time delayed (Thompson, et. al., 1972), pseudo-dichotic modifications (Tolhurst, 1971) or by other types of release from masking techniques.

Nor is it possible at the present time to construct various transfer functions concerned with generalizing predictions of S/N ratios between the conditions in which the signal level varies as the independent variable under several levels of noise (one level at a time) and the conditions in which the independent variable is the masking noise level during which the speech signal remains constant. There has not been enough of the later type studies to make the comparison valid. Signal-to-noise "should be" signal-to-noise regardless of which components of the ratio is varied, but there are indications of differences from linearity at the extremes of the intensity range.

The veiled optimism regarding listener reception prediction expressed above and the nearly complete pessimism of predicting speaker intelligibility efficiency remain. This may be because of the vast number of variables to be controlled at any one instant of experimental time (Berman, et. al., 1970). Webster (1972) has indicated one factor contributing to listener reception variation: "There are, in fact, at least 10 standardized tests that can, when properly chosen, give reliable (repeatable) scores varying from 50% to 90% on the same 'test system'. This apparent anomaly exists --- because of the extreme redundancy of spoken language." In addition to this and other language factors inherent in speech testing, and somewhat regardless of the type of test(s) to be used and the sophistication of the experimenter or of the experimental procedures he may employ, there is always a certain residual (amount, degree) of subjective factor(s). Test results can be affected by the language sample utilized (Schultz, 1972), introducing factors of subject variability whether they be classed as psychological or physiological (Boothroyd, 1968), or by the selection and use of experimental instrumentation.

Several investigators have commented on the variability due to language sampling. Speech communication is a series of message units spoken in a sequence (Egan, 1957). These signals are probablistic in nature in that a wide variety of inputs may give rise to the same phonemic perception and identical inputs can give rise to different phonemic perceptions (Schultz, 1972). Even when the tests are composed of meaningful monosyllabic words, in which there are few contextual cues, subjects do not eliminate all such cues (Boothroyd, 1968).

Some of the psychological factors that keep speech reception testing from being as objectively predictable as investigators and the consumers of their studies would like are the intelligence ranges of the subjects (Speaks, 1972; Broadbent, 1967) and a corollary of intelligence, educational level and maturation (Boothroyd, 1968). In addition, there is the factor of both the immediate and long-term psychological "set" of the receiver who listens and makes his best guess as to the message sent. His accuracy is influenced by the confusability among the message subsets, either open or closed

(Egan, 1957), word probability within a language, item difficulty (Speaks, et. al., 1972), and the acoustic coarticulation effects of phonemic probabilities between diagram and trigram combinations (Boothroyd, 1968). There is also the potent subjective factor of the criterion level the listener adapts under any particular experimental situation (Egan, 1957; Berman, et. al., 1970). The subjective criterion level can be shifted in either direction by the varient sorts of behavior of the investigator in structuring the experimental design and/or during the running of subjects. The results are often given the blanket category of "experimentor error". And as in any list, there are always the etceteras.

Physiological factors which may intrude, in addition to the characteristics of the masking noise(s) and their effects upon hearing, comprise a long list and they are more conjectural in nature than the acoustic and psychological modifiers outlined above. Definitive experiments are more difficult to do even in the laboratory. Adequate assessments under operational conditions are generally not feasible. However, it is impossible to overlook the "case-history" indications of the effects upon speech production and reception of fatigue. The state may be defined as the result of stresses of long duration task performance, short or long-term high task loading or complete physical-emotional exhaustion and/or excessive sleep deprivation. There are, almost undoubtedly, short-term and accumulative effects upon the function or malfunction of the organism due to diet and the extension of that continuum, drugs, either prescriptive or social. Other environmental changes affecting the human physiology can be reflected to psychological modifications and affect communications efficiency both as to perception and production of speech.

The contributions to subject variability attributable to inadequate and/or intimidating instrumentation are more or less obvious to most investigators. These factors can be reduced to having minimal effects by using reasonable scientific accumen and expenditure of time and funds. Hence, no further listings will be attempted here.

The preceeding sections of this paper have been an effort to gather evidence, opinion and assumptions that no speech testing procedure can be objective, truly. Since there is a wide range of variable subjectivity in any presently employed testing methodology, it may be expedient to obtain estimates of communication transmission efficiency by a technique that more or less "exploits" subjectivity. This procedure has been experimentally tested and reported by Speaks, et. al. (1972). This study was an extension and refinement of earlier research of Hawkins and Stevens (1950) in which they had the subjects vary the amplitude of a running sample of continuous speech until the subjects reported they heard something versus not hearing anything. This level they labeled as the Threshold of Detectability (TD) for speech. They then increased the continuous speech presentation level in small increments until the listeners reported they could "just understand" the meaning of most words and phrases in the speech sample. This average level was termed, Threshold of Intelligibility (TI). The differences in presentation level between the two thresholds was not large, only 9 dB. Other examples of the use of "scaling techniques" to find thresholds of running speech are found in the reports of Falconer and Davis (1959), O'Neil (1954), and others including Dahle, Hume and Haspiel (1968).

Speaks, et. al. (1972) employed a limited number of trained subjects to adjust the level of running speech, mixed with noise, using a "Bekesy" technique, until they could report they were understanding the speech at some fixed percentage of understanding, i.e., 25, 50, 75 and 100 percent. These investigators had their listeners adjust the level of speech during which the background white noise was kept at a constant level and then a separate series of judgments in which the noise levels were adjusted while the speech presentation level was kept constant. Their results are reported in percentage correct values which differ from the previous studies using scaling methods. Reliability estimates of the subjects' judgments were high with the standard deviations ranging from 0.8 to 1.3 dB for the 25, 50 and 75 percent scaled values, which means that their subjects did not vary significantly when they had a similar level of training.

Two series of tests were run to assess the comparability between intelligibility estimation judgments and sentence repetition (shadowing) percent correct scores, the latter a common method of determining speech reception accuracy. From data obtained under various signal-to-noise ratios, the correlation coefficient between the two sets of data was .93, later corrected to .84 showing that scaling judgments and sentence repetition are highly related. Additional comparisons, using other types of speech reception tests, should be made.

While Speaks, et. al. (1972) did not explicitly define the time expenditure of the subject training period, it cannot exceed the time needed to train a listening panel to truly adhere to the ANSI standard for articulation testing using PB monosyllabic words. It is very probable that subject training need be no more extensive or rigorous than with other articulation testing methods.

Once trained, the listening panel could rapidly determine reception functions under a wide variety of noise spectra, each noise at several S/N ratios. It should be possible to explore a number of conditions of language usage or operational vocabularies at any specified level of face validity. Additionally, this speech reception scaling would allow an investigator to survey rapidly various "release-from-masking" techniques.

Unless psycho-acousticians wish to extend and/or utilize an instrumental analysis of speech combined with noise weighting factors as developed by Licklider, et. al. (1959) which yielded an index proportional to AI, it may seem unrealistic to continue to seek objective, universal predictors based on human response data. Since obtaining on-line electrophysiological or biochemical indexes of human speech perception is unlikely in the near future, it may be worthwhile to exploit reliable subjective methods, blatantly and frankly.

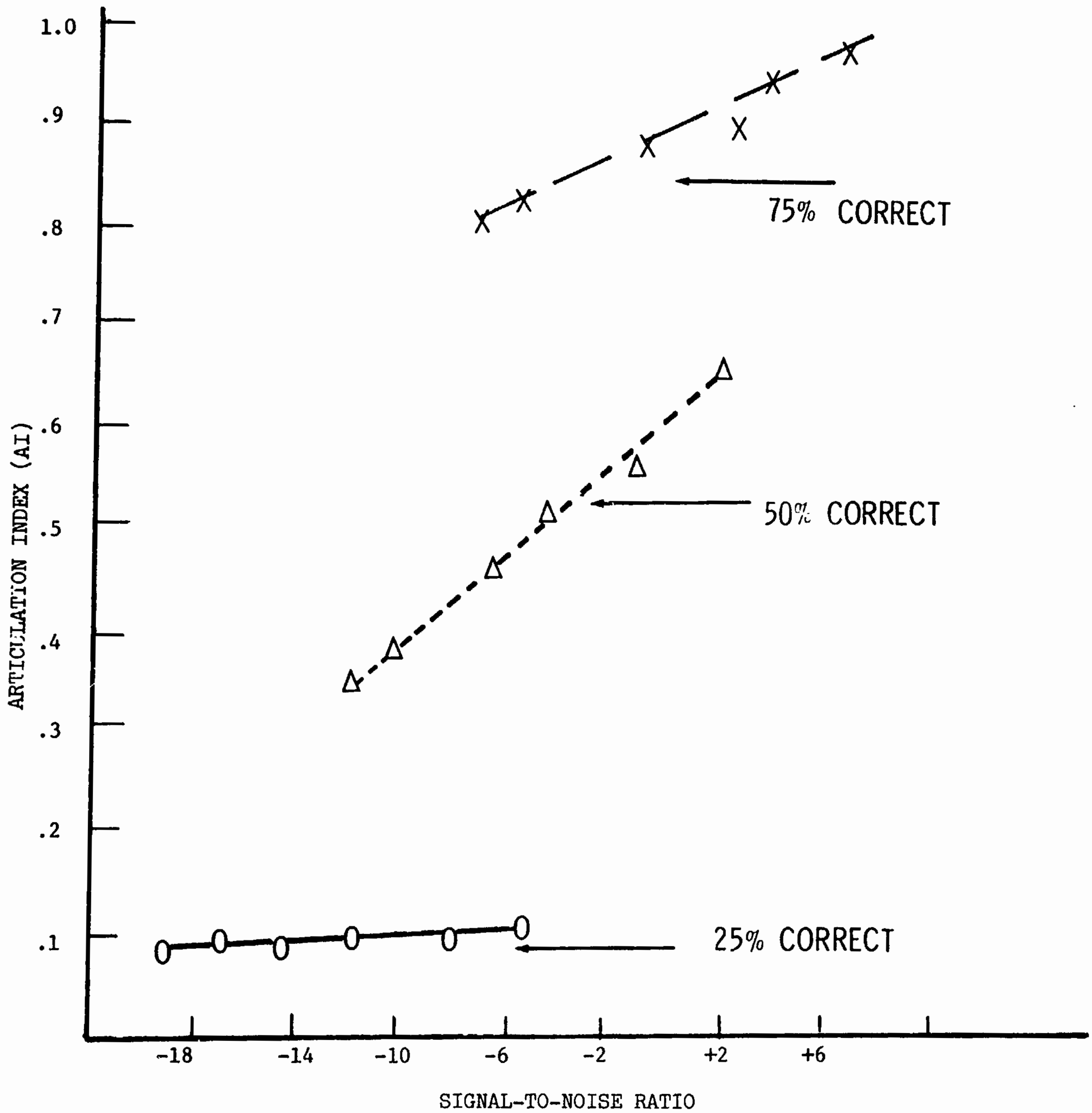


Figure 1. 25, 50, and 75% correct reception score values obtained from six different speech tests as a function of AI and S/N. The speech tests: spondee, sentence, rhyme, multiple-choice,PB, and nonsence syllables are always plotted in that order.

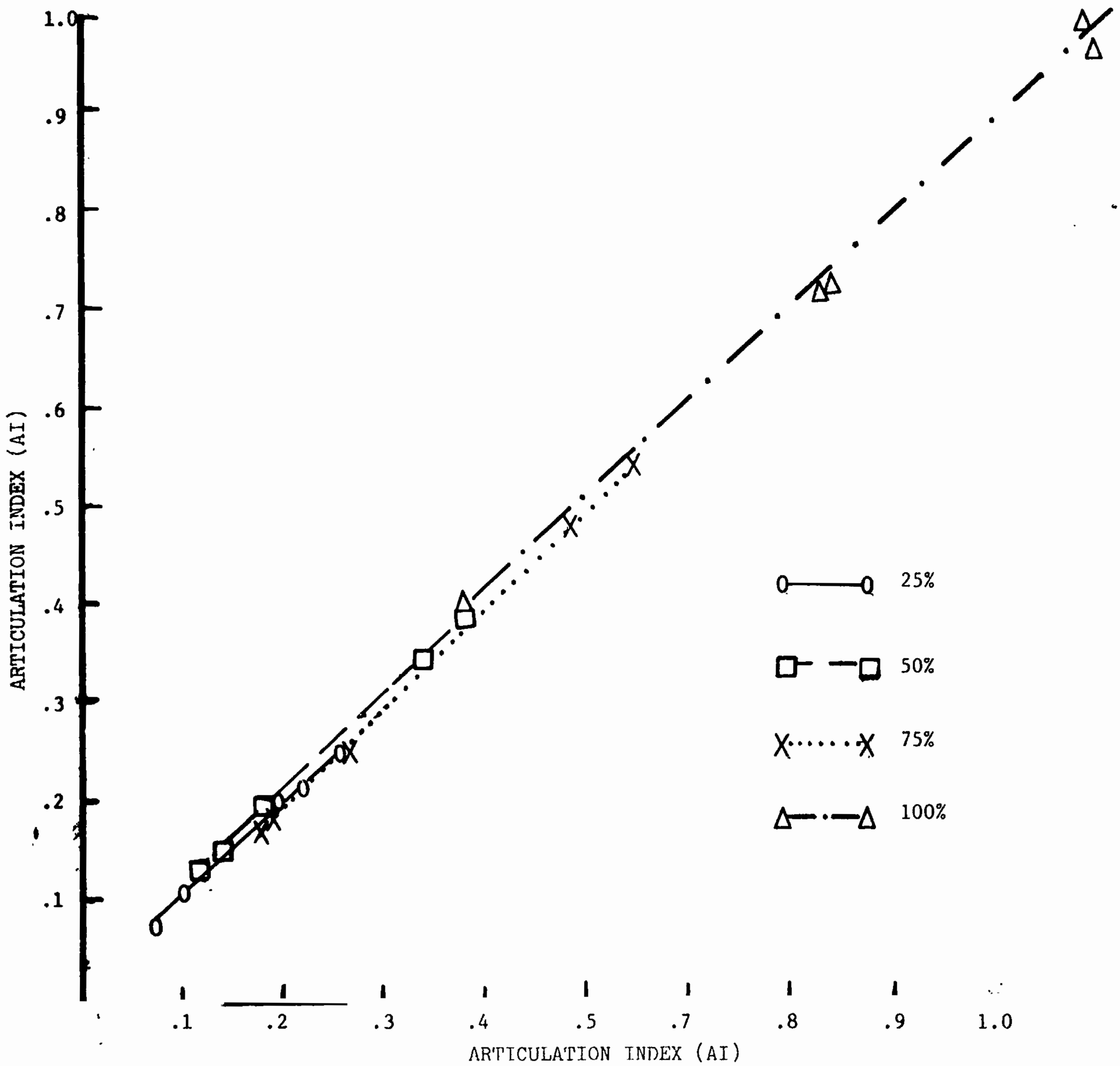


Figure 2. 25, 50, 75 and 100% correct reception score values obtained from five different speech tests as a function of AI versus AI. The speech test: spondee, sentence, rhyme, PB, and nonsense syllables are always plotted in that order.

BIBLIOGRAPHY

1. Berman, M. S., Schultz, M. C., and Tanner, W. P., Jr. (1970). Speech intelligibility: A model and an experimental study. J. acoust, Soc., Amer., 47, 74(A).

2. Boothroyd, A. (1968). A statistical theory of the speech discrimination score. J. acoust. Soc. Amer., 43, 362-367.

3. Broadbent, D. E. (1967). Word-frequency effect and response bias. Psych. Rev., 74, 1-15.

4. Chaiklin, J. B. (1959). The relation among three selected auditory speech thresholds. J. Speech Hear. Res., 2, 237-243.

5. Dahle, A. J., Hume, W. G., and Haspiel, G. S. (1968). Comparison of speech - Bekesy tracings with selected clinical auditory measures. J. Aud. Res., 8, 225-228.

6. Egan, J. P. (1957). Monitoring task in speech communication. J. acoust. Soc. Amer., 29, 482-489.

7. Falconer, G. and Davis, H. (1947). The intelligibility of connected discourse as a test for the threshold of speech. Laryngoscope, 57, 581-595.

8. Hawkins, J. E., Jr. and Stevens, S. S. (1950). The masking of pure tones and of speech by white noise. J. acoust. Soc. Amer., 22, 6-13.

9. Licklider, J. C. R., Bisberg, A., and Schwartzlander, H. (1959). An electronic device to measure the intelligibility of speech. Proc. Nat'l Electronic Conf., 15, 1-6.

10. Licklider, J. C. R. and Miller, G. A. (1951). The perception of speech. Chapter 26 in S. S. Stevens (ed.) HANDBOOK OF EXPERIMENTAL PSYCHOLOGY, N. Y., John Wiley & Sons, Inc., 1040-1074.

11. O'Neil, J. J. (1954). LISTENER JUDGMENTS OF SPEAKER INTELLIGIBILITY. Joint Report NM 001 064.01.28, The Ohio State Univ. Res. Found. and U.S. Nav. Sch. Aviat. Med., Pensacola, Florida.

12. Schultz, M. C. (1972). A critique of speech recognition testing preliminary to hearing therapy. J. Speech Hear. Dis., 37, 195-202.

13. Speaks, C., Parker, B., Harris, C., and Kuhl, P. (1972). Intelligibility of connected discourse. J. Speech Hear. Res., 15, 590-602.

14. Thompson, C. L., Stafford, M. R., Cullen, J. K., Hughes, L. F., Lowe-Bell, S. S., and Berlin, C. I. (1972). Interaural intensity differences in dichotic speech perception. J. acoust. Soc. Amer., 52, 174(A).

15. Tolhurst, G. C. (1971). Factors for more efficient communications. Naval Res. Rev., xxiv, 11-19.

16. Webster, J. C. (1972). COMPENDIUM OF SPEECH TESTING MATERIAL AND TYPICAL NOISE SPECTRA FOR USE IN EVALUATING COMMUNICATIONS EQUIPMENT. Tech. Doc. 191, U.S. Nav Electr. Lab. Ctr., San Diego, CA.