

UNCLASSIFIED

AD NUMBER

ADA071930

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

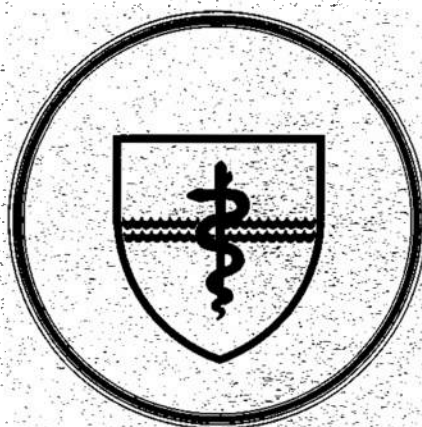
Distribution authorized to U.S. Gov't. agencies only; Test and Evaluation; 03 AUG 1977. Other requests shall be referred to Naval Submarine Medical Research Laboratory, Submarine Base, Groton, CT.

AUTHORITY

NMRDC per DTIC form 55

THIS PAGE IS UNCLASSIFIED

**NAVAL SUBMARINE MEDICAL
RESEARCH LABORATORY
SUBMARINE BASE, GROTON, CONN.**



REPORT NUMBER 873

The NSMRL TRI-WORD TEST OF INTELLIGIBILITY (TTI)

by

L. Sergeant and P. G. Lacroix

with

James E. Atkinson

New London Laboratory, Naval Underwater Systems Center

and

Naval Sea Systems Command

Naval Medical Research and Development Command
Research Work Unit M4305.08-3003DAC9

Released by:

R. A. Margulies, CDR MC USN
Commanding Officer
Naval Submarine Medical Research Laboratory

1 June 1979

Distribution limited to U. S. Gov't. agencies only. Other requests for this document must be referred to the Commanding Officer, Naval Submarine Medical Research Laboratory.

The NSMRL tri-word test of intelligibility (TTI)^{a)b)}

L. Sergeant^{c)}

Naval Submarine Medical Research Laboratory, U. S. Naval Submarine Base, Groton, Connecticut 06340

James E. Atkinson

New London Laboratory, Naval Underwater Systems Center, New London, Connecticut 06320

Paul G. Lacroix

Naval Submarine Medical Research Laboratory, U. S. Naval Submarine Base, Groton, Connecticut 06340

(Received 3 August 1977; revised 3 July 1978)

A version of Griffiths's Diagnostic Articulation Test (DAT) using three-word items is described. The test is applicable where monosyllabic words or sentence lists are undesirable or inappropriate. Each word of an item is drawn from a separate set of five monosyllabic real words differing only in the initial or final element. For each item, subjects underline one word in each of the three sets of words for that item on the answer sheet. The test examines reception of 150 words in 7 min as compared with 50 words in 5 min by the usual single-word format and preserves, moreover, the effects of gross temporal distortions and masking that occur within and between words in consecutive discourse. Discrimination scores for the tri-word test of intelligibility (TTI) as compared with the Modified Rhyme Test (MRT) and the C. I. D. sentence lists all taped with the same talker were lower when tested in quiet and even relatively lower in noise. The multiple-choice closed-set response permits easy administration, scoring, and analysis of confusions; enunciation of three semantically unrelated words in coarticulatory succession preserves interword transitions while limiting the effects of memory and linguistic redundancy. The use of the DAT lists permits a somewhat more detailed analysis of errors than use of the three-word format with the MRT as proposed by Williams *et al.* (*Aviat. Space and Environm. Med.* 47, 154-158 (1976)). The NSMRL TTI is proposed as a relatively ease of administration and scoring are desirable.

PACS numbers: 43.70.Ep

INTRODUCTION

The ideal speech reception and discrimination test for evaluating components (talker, channel, listener) of a communication chain must meet two basic requirements: it must be a valid sample of the speech the chain carries or is about to carry, and it must make possible an analysis of errors. Not many speech tests approach the ideal in both these regards. Dozens of sentence intelligibility tests have been constructed, but these are always cumbersome to administer and score, and furthermore, even with key word emphasis within the sentences they do not lend themselves readily to error analysis. Dozens of single-word intelligibility lists have been constructed which are quick and easy to administer and score, and make for very precise error analysis, but represent only poorly the speech material of direct interest.

^{a)}The opinions in this paper are solely the author's and do not necessarily represent the views of the U.S. Navy.

^{b)}This paper is taken from portions of a previous report by J. E. Atkinson, "Measuring Speech Intelligibility in a Multipath Channel," Tech. Rep. No. 5661, U. S. Naval Underwater Systems Center, New London, CT 06320 (1 September 1977). A preliminary draft of the construction of this test was presented to NASA (R. L. Sergeant, "A Tri-Word Test of Intelligibility of Speech," *Proc. Symp. in Speech Interference*, edited by W. Shepherd NASA-TM-X-72696, NASA Langley Research Center, Hampton, VA (1975), and the Acoustical Society of America (J. E. Atkinson, R. L. Sergeant, and P. G. Lacroix, "Speech intelligibility in a stationary multipath channel," *J. Acoust. Soc. Am.* 58, S129 (A) (1975).

^{c)}Present address: Department of Speech and Theater, Hunter College, New York, NY 10021.

A major difficulty with the use of sentence intelligibility tests is that the variance among listeners' responses depends very heavily on the match between each listener's condition (experience, intelligence, etc.), and the vocabulary and content of the message. Single-word closed-set response tests, in which all allowable choices are given on the answer sheet, reduce this variance so far as possible. On the other hand, tests with single word lists do not at all sample the acoustic and prosodic transitions between and within words which are so much a part of colloquial speech.

Carhart and Porter (1969) first proposed a compromise combining the virtues of a single-word test with the flow of sentential approximations. They tape-recorded items of three words each, taken from the CNC lists of Peterson and Lehiste (1962); this tape was used successfully by Vargo (1977) in studying amplitude compression in hearing aids. Speaks and Jerger (1965) constructed ten-word sentential approximations which progressively approached colloquial linguistic complexity. Subjects responded by identifying in a multiple-choice format the sentences which they heard. Although this approach controls linguistic variables and eliminated word-for-word writedown responses, it is, as the authors point out, a test of sentence identification rather than of intelligibility.

Beasley and Shriner (1973) followed Speaks and Jerger in constructing word strings grouped into first-, second-, and third-order sentences, in which each order more closely approximated the linguistic constraints of complete sentences. In their tests the words were drawn from the 250 words of Griffiths's (1967) Diagnos-

tic Articulation Test (DAT). Subjects wrote down all words heard and remembered. This format is amenable to rather complete error analysis, but there may be appreciable serial order effects. At least for clinical purposes, a shorter word-string would be preferable.

A real advance was made by Williams *et al.* (1974), who adapted the multiple-word format to the closed-set response feature of the Multiple Rhyme Test (MRT) of House *et al.* (1963). Strings of two and three words were compared with the usual single-word format. For each of the words in an item, subjects had an answer sheet on which they could indicate which of six possibilities they heard. The lists were recorded in the context of an item number and carrier phrase: "One, do you read *sake*? Over." "Ten, do you read *fit*, *cut*? Over." "Fourteen, do you read *saw*, *safe*, *hold*? Over." In enunciating multiple-word lists the talker adopted a manner and rhythm appropriate for a message, but he tried to give "discrete productions" for each word. The differences in scores for the one-, and three-word formats were shown to be negligible across speech-to-noise ratios; savings were demonstrated in that, in the three-word format, 150 words could be presented in 7 min, as against 50 words in the usual test in 5 min.

For many purposes, an advantage exists of the Griffiths DAT over the MRT, in that the DAT incorporates more difficult discriminations involving numbers of instances where the foil words differ from the target word in only one dimension; foils in the MRT could differ in perhaps all three of the dimensions of *manner of articulation*, and *voicing*. Care was taken in the construction of the DAT to include all possible contrasts, and as a result the DAT provides a more precise notion of the underlying nature of a listener's errors or of the deficiencies of a channel in a communication chain.

In our work a need appeared to assess a communications chain involving time smears and other temporal distortions, and internal masking between the phonemes of a sentence both within and across words. Complete and accurate analysis of errors was necessary, and yet the material had to be approximately of sentence length. We combined the efficiency of the three-word format with the precision of error analysis which the DAT makes possible, and created the NSMRL tri-word test of intelligibility (TTI).

Because experience is limited with the three-word format, it was desirable to validate the TTI by comparing it on the same subjects and under the same general conditions with more widely used intelligibility tests before offering it as an acceptable test of speech intelligibility. Since the TTI partakes of the advantages both of a single-word list and of a sentence list, it should be validated against a sample of each type. We chose the MRT of House *et al.* primarily because it had been recommended by a National Research Council committee (Kreul *et al.*, 1968) as an archetype of tests using monosyllables in a closed-set response format. We chose the C.I.D. sentence lists (Silverman and Davis, 1970) because they had been written following the rules laid down by a National Research Council

committee (Silverman and Hirsh, 1955) for constructing samples of colloquial speech.

This paper describes in detail the construction of the NSMRL TTI and compares listener performance using a selection of speech-to-noise (S/N) ratios with performance both with the MRT and with the C.I.D. sentence lists.

I. GENERAL METHOD

A. Tri-word test of intelligibility

Each TTI list consisted of 50 three-word items, for a total of 150 words. The three individual words in any list were drawn at random from one of Griffiths's five DAT lists, with the restriction that each of the three words come from a different DAT list. Griffiths's lists each contain 50 response sets as follows:

	List				
	A	B	C	D	E
1.	bat	batch	bask	bass	badge
2.	laws	long	log	lodge	lob
13.	beige	base	bayed	bathe	bays
49.	mat	vat	that	fat	rat
50.	way	may	gay	they	nay

where the words within each response set (Rows 1-50) differ only in the initial or in the final consonant.

To construct a TTI item, three words are randomly selected (e.g., "Badge-bayed-mat") from lists E, C, and A, respectively, (TTI Lists 1, 2, and 3 are available).¹

For every tri-word item, five-word response sets were printed in three columns, one for every word position in the tri-word item. The listener's task was to cross out the word heard from each of the three sets. For example, if the stimulus was "badge-bayed-mat," the three response sets for that item were as follows:

<i>badge</i>	<i>bathe</i>	<i>mat</i>
batch	base	fat
base	<i>bayed</i>	that
bat	bays	rat
bash	beige	vat

Correct answers are indicated for the convenience of the reader. An 8-s subject response interval was allowed between items.

B. Subjects

Listeners were 136 young enlisted men drawn at random from candidates for the USN Submarine School. All had hearing threshold levels ≤ 15 dB from 0.5-8 kHz.

TABLE I. Mean discrimination score by listening panel and list.

List	Listening panel No.			Difference in mean score	Average
	20	20	20		
	1	2	3		
CID No. 1	X	99.3	99.2	0.1	99.25
CID No. 2	99.9	99.8	X	0.1	99.85
CID No. 3	99.6	X	99.0	0.6	99.3
MRT A	98.6	X	98.9	0.3	98.75
MRT B	95.0	96.2	X	1.2	95.6
MRT C	X	96.9	98.4	1.5	97.65
TTI No. 1	90.6	X	93.5	2.9	92.05
TTI No. 2	90.6	90.7	X	0.1	90.65
TTI No. 3	X	90.8	92.3	1.5	91.55

C. Stimuli

Master tapes were made using a high quality microphone in an anechoic chamber and an Ampex PR-10 tape recorder. The talker (RLS) was a man experienced in intelligibility testing who spoke with a general American dialect. Three TTI tests were recorded, three of the MRT, and three of the C.I.D. sentences. Tri-word items were spoken as monotonic three-word phases, not as "discrete productions." A calibration tone at 1 kHz was recorded on each tape.

From one experiment in which speech material was mixed with noise at a selection of speech-to-noise ratios, the ratios were determined as follows: prior to presentation, measurements using a General Radio graphic level recorder were made of each item at the input to the earphones. The mean item level for each list was computed and used in determining S/N. Speech-spectrum noise from a General Radio 901B noise generator was mixed with the speech in an appropriate circuit and adjusted to give the desired S/N for any condition (for S/N's used, see Table II).

Seven listening panels of approximately 20 men each were composed. All were presented with materials monaurally in a room fitted with 20 matched Permoflux PDR-8 earphones in MX-41/AR cushions. Presentation level for all tests was set by adjusting calibration tones to 70 dB SPL, as determined for the median phone of the room in an NBS 9-A coupler. Listeners marked multiple-choice answer sheets for the MRT and TTI, and wrote down responses to the sentences on a blank sheet of paper.

II. EXPERIMENT I

Three listening panels were each presented with two lists of each test in quiet (see experimental design in Table I). This provided both baseline data and an indication of intergroup reliability for every list.

A. Results and discussion

Table I shows mean discrimination score (DS) and mean differences between panels for all conditions (see also Fig. 1). Overall mean DS's in quiet (last column)

show that, in general, the TTI is a more difficult test (mean DS=91.4) than either the sentences (Mean DS=99.5) or the MRT (mean DS=97.3).

The mean DS difference between panels (column 4) indicates that interpanel variability was relatively small; only 1 of 9 comparisons between panels exceeded 1.5 percentage points difference.

III. EXPERIMENT II

Order of presentation of lists, S/N ratios, and listening panels were randomized with four listening panels. Based on preliminary testing, different levels of noise were mixed with the various speech materials to equate degree of difficulty (see Fig. 1). Each panel heard six test S/N conditions.

A. Results and discussion

Table II lists the mean DS and standard deviations for all conditions. Figure 1 presents mean DS corrected for chance.

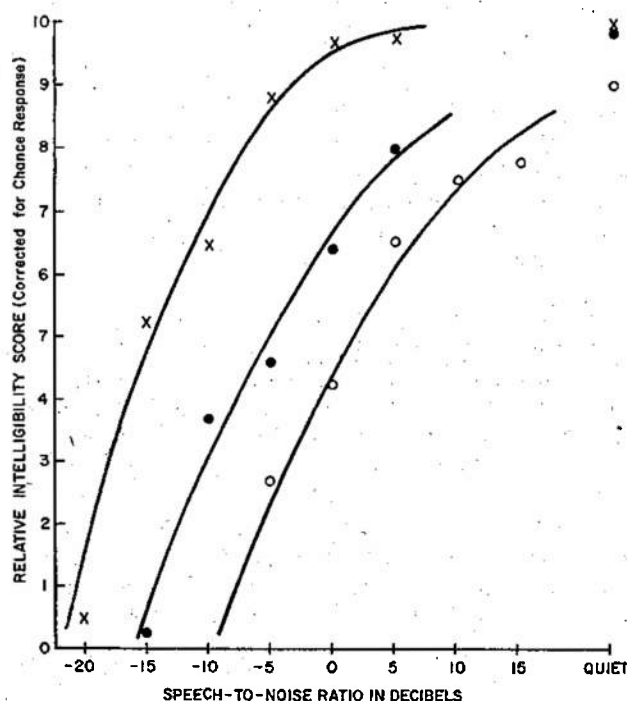


FIG. 1. Mean per cent correct intelligibility scores (corrected for chance) as a function of Speech-to-Noise ratio for three speech materials. Crosses: C. I. D. sentences. Filled circles: MRT. Open circles: TTI. Scores in per cent correct are not directly comparable for the three types of test. On the TTI, a random (guessing) score would be 20% correct, while on the MRT it would be 16.7% correct and 0% for the sentences. In order to render the data comparable across tests, raw scores were adjusted by the following formulas:

Sentences: Score corrected for chance = $0.1 (X - 0)$

MRT: Score corrected for chance = $0.12 (X - 16.7)$, and

TTI: Score corrected for chance = $0.125 (X - 20)$

where X is percent correct response. These formulas transform the raw scores to a common scale ranging from 0 for completely random responses to 10 for a perfect score. Scores corrected for chance permit a more meaningful comparison of the types of test.

TABLE II. Mean discrimination scores and standard deviations as a function of speech-to-noise ratio for experiment II.

Test	S/N in dB			Mean DS			S. D.	
CID sentences	-20				5.3			5.5
	-15				52.8			14.7
	-10				64.8			12.7
	-5				90.2			10.4
	0				97.3			3.1
	+5				98.0			4.2
MRT	-15				18.6			11.4
	-10				47.4			9.5
	-5				54.7			7.9
	0				70.2			7.5
	+5				84.1			5.2
	+10				86.8			4.8

Word				Word			
	1st	2nd	3rd	1st	2nd	3rd	
TTI	-10	34.6	36.7	36.9	6.8	6.7	7.8
	-5	44.4	39.6	42.7	7.7	6.5	6.0
	0	55.4	49.6	59.6	7.4	8.3	10.2
	+5	74.6	68.6	75.6	11.2	8.8	7.9
	+10	81.1	78.8	85.6	8.1	7.5	4.9

The sentences are of course more intelligible in noise, due to their greater inherent redundancy, than either the MRT or TTI. The TTI is the least intelligible as a function of S/N, presumably as a result of greater complexity both of task and perceptual processing.

Since subject-by-subject variability is an index of test reliability, it is useful to compare variability about the mean at a point of equal difficulty for each test. S/N in dB for the mean 50%-correct intelligibility points were therefore determined: these were -14.5 for the sentences, -5.2 for the MRT, and +1.0 for the TTI (see Fig. 1). The experimental conditions corresponding most closely to these points were -15 for sentences, -5 for the MRT, and 0 for the TTI. The associated standard deviations (see Table II) were 14.7 for sentences, 7.9 for the MRT, and 8.6 for the TTI.

For all practical purposes, the variability about the mean is no greater for the TTI than for the MRT, but it is considerably less than for the sentences. The fact that sentences provide a less stable mean DS is simply a reflection of the greater variance associated with the linguistic complexity of the material, a feature reduced intentionally in the other tests.

The variability inherent within a speech test derives from two sources, (1) response complexity and (2) task complexity; the former refers to the influence of open versus closed-set response biasing, and the latter to the perceptual processing demand which a task places on the listener. The CID sentences in this scheme are higher in response complexity by virtue of the open-set response format, but lower in task complexity due to inherent linguistic redundancy. The converse is true for the MRT and TTI; however, for the TTI there is an additional variable which contributes to task complexity,

that of serial order effects. Since the listener must perceive and store three words in succession and then recall them in order, the task variable is more complicated than in the single-word MRT Test. Overall, the TTI is a more suitable test of intelligibility because it presents the listener with a longer and more difficult task while limiting response complexity. This provides less variability and lower scores resulting in a stable test without "ceiling effects".

ACKNOWLEDGMENTS

Sincere gratitude is expressed to J. D. Harris for invaluable assistance in the preparation of this manuscript. Financial support for this research was provided by the Naval Underwater Systems Center under their Independent Research and Development Program and by the Naval Sea Systems Command Exploratory Development Program in underwater acoustic communications.

¹NSMRL TTI Lists 1, 2, and 3 are available from the American Society for Information Science, National Auxiliary Publications Services, by addressing: ASIS/NAPS, c/o Microfiche Publications, 305 East 46th St., New York, NY 10017. Hard copy of microfiche prints may be ordered by the public at modest cost. A tape recording of the lists as taped and used in this study may be obtained by writing to the Chief, Auditory Division, NavSubMedResLab, Submarine Base, Groton, CT 06340.

Beasley, D. S., and Shriner, T. H. (1973). "Auditory analysis of temporally distorted sentential approximations," *Audiol.* 12, 262-271.

Carhart, R., and Porter, L. (1969). "Development of test materials for CAMP studies," Auditory Research Laboratories, Northwestern U., Evanston, IL (unpublished) (cited by Vargo, 1977).

Griffiths, J. D. (1967). "Rhyming minimal contrasts; a simplified diagnostic articulation test," *J. Acoust. Soc. Am.* 42, 236-241.

House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (1963). "Psychoacoustic speech tests: a modified rhyme test," *Electronic Systems Div., USAF Systems Command, Hanscom Field, Bedford, MA, Tech. Doc. Rep. No. ESD-TDR-63-403*. See also: House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (1965). "Articulation-testing methods: consonantal differentiation with a closed-response set," *J. Acoust. Soc. Am.* 37, 158-166.

Kreul, E. J., Nixon, J. C., Kryter, K. D., Bell, D. W., Land, J. S., and Schubert, E. D. (1968). "A proposed clinical test of speech discrimination," *J. Speech Hear. Res.* 11, 536-552. See also Kreul, E. J., Bell, D. W., and Nixon, J. C. (1969). "Factors affecting speech discrimination test difficulty," *J. Speech Hear. Res.* 12, 281-287.

Peterson, G. E., and Lehiste, I. (1962). "Revised CNC lists for auditory test," *J. Speech Hear. Dis.* 27, 62-70.

Silverman, S. R., and Davis, H. (1970). *Hearing and Deafness*, 3rd ed. (Holt, Rinehart and Winston, New York).

Silverman, S. R., and Hirsh, I. J. (1955). "Problems related to the use of speech in clinical audiometry," *Ann. Otol. Rhinol. Laryngol.* 64, 1234-1244.

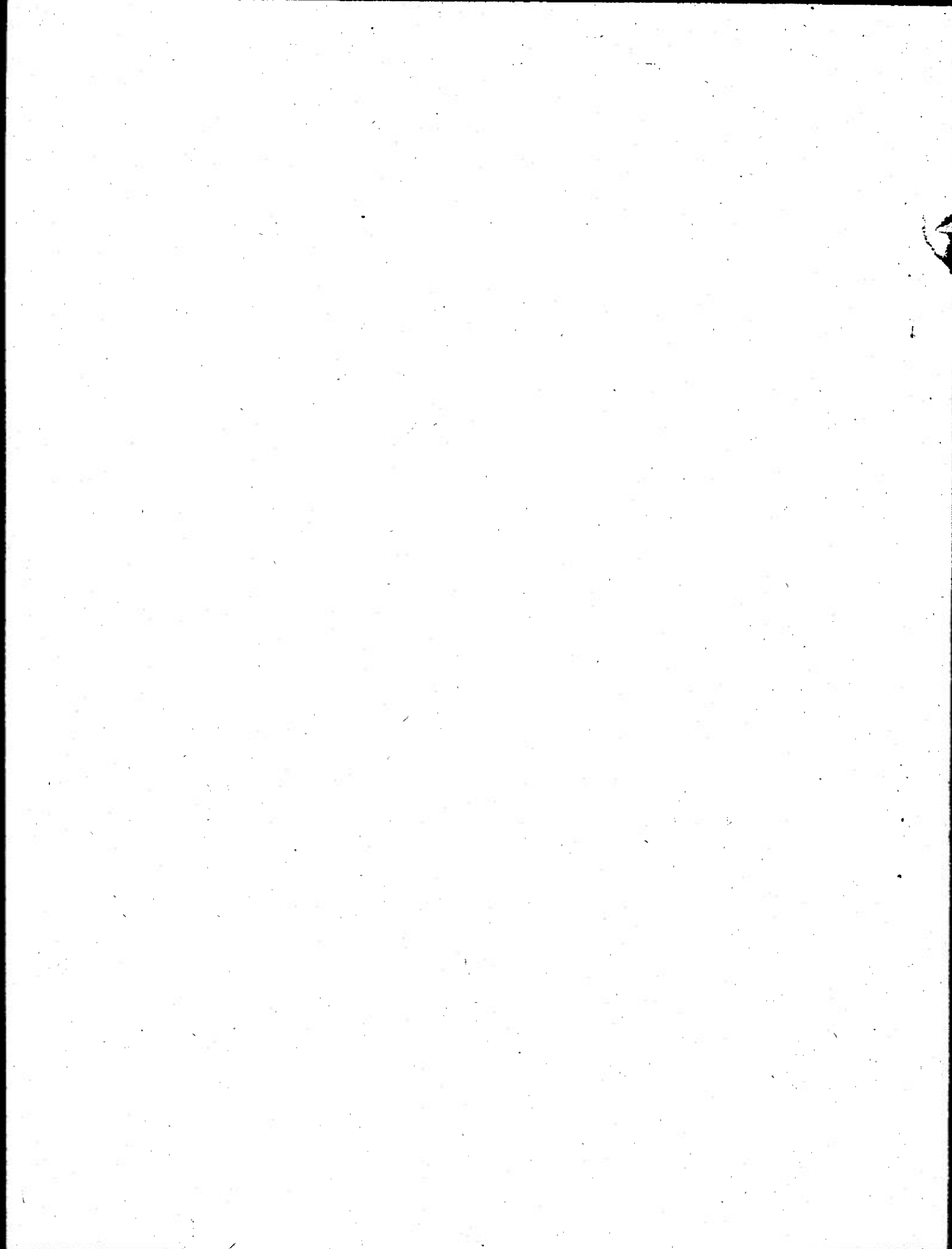
Speaks, C., and Jerger, J. (1965). "Methods for the measurement of speech identification," *J. Speech Hear. Res.* 8, 185-194.

Vargo, S. W. (1977). "Intelligibility of amplitude compressed speech in quiet: normal and pathological hearing groups," *Audiol. Hear. Educ.* 3(2), 16-17.

Williams, C. E., and Mosko, J. D. (1974). "Evaluating the

ability of aircrew personnel to hear speech in their operational environments," Program AGARD Aerospace Medical Panel Specialists meeting, Pozzuoli, Italy (September). See also Williams, C. E., and Mosko, J. D., (1975). "A new look at multiple-word test items for evaluating talkers, listeners, and communications systems," *Proc. Symp. in*

Speech Interference, edited by W. Shepherd, NASA-TM-X-72696, NASA Langley Research Center, Hampton, VA. See also Williams, C. E., Mosko, J. D., and Greene, J. W. (1976). "Evaluating the ability of aircrew personnel to hear speech in their operational environments," *Aviat. Space Environ. Med.* 47, 154-158.



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NSMRL Report Number 873	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The NSMRL Tri-Word Test of Intelligibility (TTI)		5. TYPE OF REPORT & PERIOD COVERED Interim report
7. AUTHOR(s) L. Sergeant, J. E. Atkinson and P. G. Lacroix		6. PERFORMING ORG. REPORT NUMBER NSMRL Report 873
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Submarine Medical Research Laboratory Box 900 Naval Submarine Base Groton, Connecticut 06340		8. CONTRACT OR GRANT NUMBER(s)
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command National Naval Medical Center Bethesda, Maryland 20014		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS M4305.08-3003DAC9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE 14 June 1979
		13. NUMBER OF PAGES 5
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution limited to U. S. Government agencies only; other requests for this document must be referred to the Commanding Officer, Naval Submarine Medical Research Laboratory.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) tri-word test of intelligibility; Diagnostic Articulation Test (DAT); Modified Rhyme Test (MRT);		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A version of Griffith's Diagnostic Articulation Test (DAT) using three-word items is described. The test is applicable where monosyllabic words or sentence lists are undesirable or inappropriate. Each word of an item is drawn from a separate set of five monosyllabic real words differing only in the initial or final element. For each item, subjects underline one word in each of the three sets of words for that item on the answer sheet. The test examines reception of 150 words in 7 minutes as compared with 50 words in 5 minutes by the usual single-word format and		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract: continued

preserves, moreover, the effects of gross temporal distortions and masking that occur within and between words in consecutive discourse. Discrimination scores for the tri-word test of intelligibility (TTI) as compared with the Modified Rhyme Test (MRT) and the C.I.D. sentence lists all taped with the same talker were lower when tested in quiet and even relatively lower in noise. The multiple-choice closed-set response permits easy administration, scoring, and analysis of confusions; enunciation of three semantically unrelated words in coarticulatory succession preserves interword transitions while limiting the effects of memory and linguistic redundancy. The use of the DAT lists permits a somewhat more detailed analysis of errors than use of the three-word formant with the MRT as proposed by Williams et al (Aviat. Space & Environ. Med. 47, 154-158 (1976). The NSMRL TTI is proposed when a relatively ease of administration and scoring are desirable.