

AD 684608

TRACOR Project
079 001 01
Document Number
TRACOR 68-1152-U

DIAGNOSTIC EVALUATION OF SPEECH
PROCESSING SYSTEMS

by

William D. Voiers

TRACOR, Inc., 6500 Tracor Lane, Austin, Texas 78721

Contract No. F19628-68-C-0068
Project No. 4610
Task No. 461002
Work Unit No. 46100201

FINAL REPORT

Period Covered: 15 September 1967 - 14 September 1968

December 1968

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

Contract Monitor: Caldwell P. Smith
Data Sciences Laboratory

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

DLG
MAR 28 1969

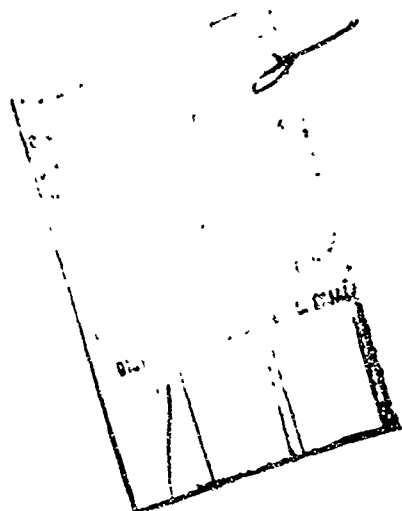


6500 Tracor Lane, Austin, Texas 78721, AC 512/926-2800

Prepared for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS 01730

Qualified requestors may obtain additional copies from the Defense Documentation Center. All others should apply to the Clearinghouse for Federal Scientific and Technical Information.



AFCRL-68-0568



6500 TRACOR LANE, AUSTIN, TEXAS 78721

TRACOR Project
079 001 01
Document Number
TRACOR 68-1152-U

DIAGNOSTIC EVALUATION OF SPEECH
PROCESSING SYSTEMS

by

William D. Voiers

TRACOR, Inc., 6500 Tracor Lane, Austin, Texas 78721

Contract No. F19628-68-C-0068
Project No. 4610
Task No. 461002
Work Unit No. 46100201

FINAL REPORT

Period Covered: 15 September 1967 - 14 September 1968

December 1968

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

Contract Monitor: Caldwell P. Smith
Data Sciences Laboratory

Prepared for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS 01730



6500 TRACOR LANE AUSTIN, TEXAS 78721

TABLE OF CONTENTS

Introduction -----	1
Publications, Meetings, and Personnel -----	1
Technical Accomplishments -----	2
The Present State of Digital Vocoding Technique -	Appendix I
Minor Experimental Studies -----	Appendix II



6500 TRACOR LANE, AUSTIN, TEXAS 78721

ABSTRACT

This report presents the results of a survey in which the Diagnostic Rhyme Test was used to evaluate the present state of digital technique for speech processing and communication. Also presented are the results of a series of minor studies concerned with the methodology of intelligibility evaluation.



6500 TRACOR LANE AUSTIN, TEXAS 78721

INTRODUCTION

Described here are the major accomplishments of the Human Communications Section of the Life Sciences Department, Sciences and Systems Division of TRACOR, Inc. under Contract No. F19628-68-C-0068, 15 September 1967 to 14 September 1968.

PUBLICATION, MEETINGS AND PERSONNEL

Publications

Voiers, William D., The Present State of Digital Vocoding Technique, IEEE Transactions on Audio and Electroacoustics, Vol. AU-16, No. 2, 275-279 (1968).

Presentations

Voiers, William D., On the Problems of Reliability, Sensitivity and Diagnostic Value in Tests of Speech Intelligibility 1967 Conference on Speech Communications and Processing, Cambridge, Mass.

Voiers, William D., Effects of Masking Speech on the Discriminability of Elementary Consonant Attributes 75th Meeting, Acoustical Society of America, Ottawa, Ontario, Canada, May 1968.



6500 TRACOR LANE AUSTIN TEXAS 78721

Meetings Attended

1967 Conference on Speech Communication and Processing,
Cambridge, Mass. Dr. Voiers, Mr, Mickumas, Miss Miethe.

74th Meeting of the Acoustical Society of America,
Miami Beach, Florida, November 1967, Dr. Voiers.

75th Meeting of the Acoustical Society of America,
Ottawa, Ontario, Canada. May 1968, Dr. Voiers.

Technical Personnel

Dr. William D. Voiers (17September 1967 - 14 September
1968).

Mr. Joseph Mickumas, Jr. (Sept. 1967 - March 1968).

Miss Virginia Miethe (Sept. 1967 - March 1968).

Mr. Lonnie Wilson (March 1968 - Sept. 1968).

Mr. Alan Sharpley (March 1968 - Sept. 1968).

TECHNICAL ACCOMPLISHMENTS

For convenience in discussion the work accomplished
under this contract is divided into three categories:

1. The preparation and conduct of a survey of present-day speech processing and communication systems as



6500 TRACOR LANE, AUSTIN TEXAS 78721

a feature of the 1967 Conference on Speech Communication and Processing.

2. Diagnostic evaluation of a series (32) of experimental vocoder systems from output tape recordings provided by the Data Sciences Laboratory, Digital Speech Compression Branch, of the Air Force Cambridge Research Laboratories.
3. Basic Research on selected speech perception phenomena and their implications for the methodology of communications equipment evaluation from the stand point of speech intelligibility.

The Speech Analysis/Synthesis Survey

The survey was conducted by the TRACOR Human Communications Section under direction by a sub-committee of the Technical Program Committee of the 1967 Speech Conference. The members of the Committee were:

Caldwell P. Smith, Chairman

Air Force Cambridge Research
Labs. (CRBS), L. G. Hanscom
Field, Bedford, Mass.

Ben Gold

Lincoln Laboratory
Lexington, Mass. 02173



6500 TRACOR LANE AUSTIN TEXAS 78721

Michael Hecker

Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Mass. 021138

Joseph Mickumas, Jr.

TRACOR, Inc.
6500 Tracor Lane
Austin, Texas 78721

Joseph Tierney

Lincoln Laboratory
Lexington, Mass. 02173

William D. Voiers

TRACOR, Inc.
6500 Tracor Lane
Austin, Texas 78721

Carl Williams

Bolt, Beranek & Newman, Inc.
50 Boulton Street
Cambridge, Mass. 02138

B. Woods

Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Mass. 02138

The objectives of the survey were twofold:

- (1) To provide a demonstration of the present state of speech processing techniques for conference attendees and



6500 TRACOR LANE AUSTIN, TEXAS 78721

(2) To provide an objective, diagnostic evaluation of present-day speech processing techniques from the stand point of speech intelligibility.

For purposes of the survey all potentially interested individuals and organizations were invited to participate in the demonstration and/or the objective evaluation. Respondents were supplied high quality tape recordings of selected continuous speech materials, recorded by three different male speakers, for purposes of the demonstration. Recordings of two randomizations of the Diagnostic Rhyme Test materials, recorded by a single trained speaker, were supplied to those who wished to participate in the objective phase of the survey. Twenty eight samples of processed speech (more than one from some organization) were received for inclusion in the demonstration. (See Appendix I.) Eighteen samples were submitted for purposes of objective evaluation.

All participants were subsequently provided with copies of the demonstration tape. In addition, participants in the objective phase of the program received the results of the intelligibility tests conducted at TRACOR with the tapes which they submitted for evaluation. Proprietary considerations preclude the public disclosure of results for individual systems. However, for one class of systems, 2400 bps digital vocoder, there was sufficient representation to permit an evaluation of the state-of-the art for this class without violating the proprietary



6500 TRACOR LANE, AUSTIN, TEXAS 78721

rights of the participant individuals and organizations. The results of this evaluation have been published (Voiers, 1968) and are also presented in Appendix I.

Evaluation of AFCRL Experimental Systems

These evaluations were all "blind" evaluations performed on unidentified systems. The results have been transmitted to the contract sponsor under separate cover.

Methodological Studies of Speech Evaluation Techniques

A number of minor studies were conducted in attempts to resolve various issues in the methodology of speech evaluation. Among these was "The effects of vowel context on the discriminability of elementary consonant attributes." (Appendix II). The results of this investigation indicate that vowel context significantly influences the discriminability of some of the six consonant attributes treated in the Diagnostic Rhyme Test, but has no influence on others.

Another study involved an item analysis of the Diagnostic Rhyme Test which culminated in the development of a set of formally equivalent, equally difficult, abbreviated versions of the DRT, primarily for use in circumstances where transmission time is at a premium. Validation studies of these abbreviated versions of the DRT are continuing under TRACOR sponsorship. The results obtained thus far (Appendix II) attest to the feasibility of abbreviating the DRT.



6500 TRACOR LANE. AUSTIN, TEXAS 78721

Further validation of the standard Diagnostic Rhyme Test was accomplished through a study in which speech level was the independent variable (See Appendix II). The effects of reduced speech level were, as hypothesized, similar to the effects of Gaussian noise masking of the speech signal.

A study of the effects of speech interruption on DRT scores (See Appendix II) was conducted to test hypotheses generated by previous work with vocoded speech and speech masked speech. The results obtained were consistent with the hypothesis that discriminability of the attributes, graveness and sustention is heavily dependent upon the temporal continuity of the speech signal.

Another study (Appendix II) was conducted in an attempt to isolate the mechanism of the performance changes that listeners exhibit with increased exposure to the Fairbanks Rhyme Test materials. It was hypothesized that, as listeners accumulate experience with the corpus of Fairbanks test words, they tend increasingly to select their responses, correct or incorrect, from that corpus of test words. The results of this experiment, involving the successive administration of 32 Fairbanks Tests (250 items each) to a naive crew of male listeners, failed to verify this hypothesis, but further research on the subject is indicated.



6500 TRACOR LANE, AUSTIN, TEXAS 78721

APPENDIX I

THE PRESENT STATE OF DIGITAL VOCODING TECHNIQUE:
A DIAGNOSTIC EVALUATION



6500 TRACOR LANE, AUSTIN, TEXAS 78721

INTRODUCTION

Conventional intelligibility scores of .90 and higher are now quite common in digital vocoder systems operating at data rates in the neighborhood of 2400 bps. Understandably, therefore, there is some tendency among vocoder engineers and scientists to regard the intelligibility problem as effectively solved and, in turn, to concentrate their efforts primarily on such problems as speaker recognizability and aesthetic acceptability. It is possible, moreover, that the difficulty of demonstrating significant differences in gross intelligibility among competitive vocoder designs, or between experimental variations of the same design, serves to impart additional momentum to the trend away from primary concern with intelligibility.

It is not altogether clear, however, that the state of the art does in fact justify such an attitude of complacency with respect to the problem of intelligibility. For, implicit in this attitude is the assumption that the deficiencies of present-day vocoders tend to be distributed more or less uniformly with respect to the various crucial features of the speech signal. On this assumption it may in turn be argued that attempts to remedy any specific deficiency will not, in general, result in significantly improved over-all intelligibility nor, in any event, warrant the necessary investigative effort and perhaps the increased cost, size and complexity of



6500 TRACOR LANE, AUSTIN, TEXAS 78721

essential hardware. However, the question can be raised as to whether such an assumption can in fact be justified. A review of the relevant literature (e.g., Voiers, et al,¹ and Dolansky et al²) reveals a number of indications to the effect that it cannot. Rather there is substantial evidence to suggest that the shortcomings of present-day vocoders are not of a general character, but, rather, take the form of deficiencies in the transmission of a limited number of relatively specific speech features. Given this to be the case, the issue of the intelligibility of digital vocoders would seem to merit further examination. The present investigation was undertaken in an attempt to resolve this issue.

Under the auspices of the 1967 Conference on Speech Communication and Processing, sponsored jointly by the Air Force Cambridge Research Laboratories and the IEEE Audio and Electroacoustics Group, all agencies and individuals known to be active in the speech processing field were invited to submit samples of processed speech for demonstration at the conference and/or evaluation by means of the Diagnostic Rhyme Test (DRT).

Nine agencies submitted samples of 2400 bps, digitally vocoded speech for evaluation with the DRT. The input tape recordings of the DRT materials utilized by the various participants were all made by the same speaker, an experienced radio announcer who had a medium pitched voice and a general American dialectal background. All materials were recorded in a sound treated room



6500 TRACOR LANE, AUSTIN, TEXAS 78721

with a General Radio ceramic microphone (type 1560-P5) positioned at a distance of 20 cm. from the speaker's lips. Because of some indication of system malfunction and/or defective recording, one sample was excluded while test data for the remaining eight systems were compiled in an attempt to provide a characterization of the present state of digital vocoding technique from the standpoint of speech intelligibility.

METHOD AND MATERIALS

The Diagnostic Rhyme Test

The Diagnostic Rhyme Test³ is a two-choice test of consonant discriminability which yields a gross measure of speech intelligibility and additional scores relating to specific aspects of the performance of the speaker, listener or system under test. It utilizes a corpus of 192 words (96 rhyming pairs selected such that the initial consonants of each pair differ in terms of a single phonemic attribute) each of which is normally presented twice in the course of a testing session. In a given instance, one member of a rhyming word pair serves as the stimulus. The listener's task is then to indicate which member of the pair has been spoken. A correct choice indicates that the listener has, in effect, discriminated the state of one of six perceptual attributes of English consonant phonemes. An incorrect choice indicates that the listener has failed to discriminate the source state of one of the six attributes. When, for example, the stimulus word is "zeal" and the choices open to the listener



6500 TRACOR LANE, AUSTIN, TEXAS 78721

are "zeal" and "seal," a correct response indicates that the listener has discriminated (and/or the system under test has transmitted) the attribute voicing. A correct choice between "feet" and "peat" may likewise indicate that a system under test has faithfully transmitted the acoustic speech features which distinguish the sustained consonants (e.g., /fv²/) from their interrupted counterparts (/pbd/). Depending on the word pair involved, each item thus serves to test for the discriminability of one of the following perceptual phonemic attributes:

Voicing
Nasality
Sustention
Sibilation
Graveness
Compactness

In addition to a gross score for the discriminability of each attribute, separate scores for the discriminability of each state of each attribute may be obtained by appropriate analyses of listener response data. For example, one score representing the detectability of the presence of sustention (as opposed to interruption) may be obtained. A second score, representing the detectability of interruption, may then also be obtained. The average of these two scores thus provides a gross indicant of the discriminability of the attribute, sustention, while the difference between them serves as an indicant of bias



6500 TRACOR LANE, AUSTIN, TEXAS 78721

in favor of one state or the other.

All scores are adjusted for the effects of chance by means of the formula:

$$D = \frac{R-W}{T}$$

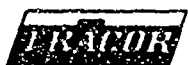
where D is percent correct discriminations, R is the number of correct responses, W is the number of incorrect responses and T is the total number of responses.

Listening Crew

The listening crew for the present series of test consisted of eight normal hearing male high school students, selected on the basis of pure tone audiometric tests and consistency of discrimination performance as measured by the DRT. All had in excess of ten hours previous experience with the DRT in general and had been given one exposure to each of the nine scramblings of the DRT materials used in the present series of tests.

Testing Facilities

The testing facilities included two double-walled IAC rooms, each partitioned into four listening booths. Test materials were presented diotically via an Ampex 602-2 tape recorder, a high quality custom built amplifier and Koss PRO-600 head phones at a comfortable listening level (roughly estimated at 75-80 dB SPL for vowel peaks).



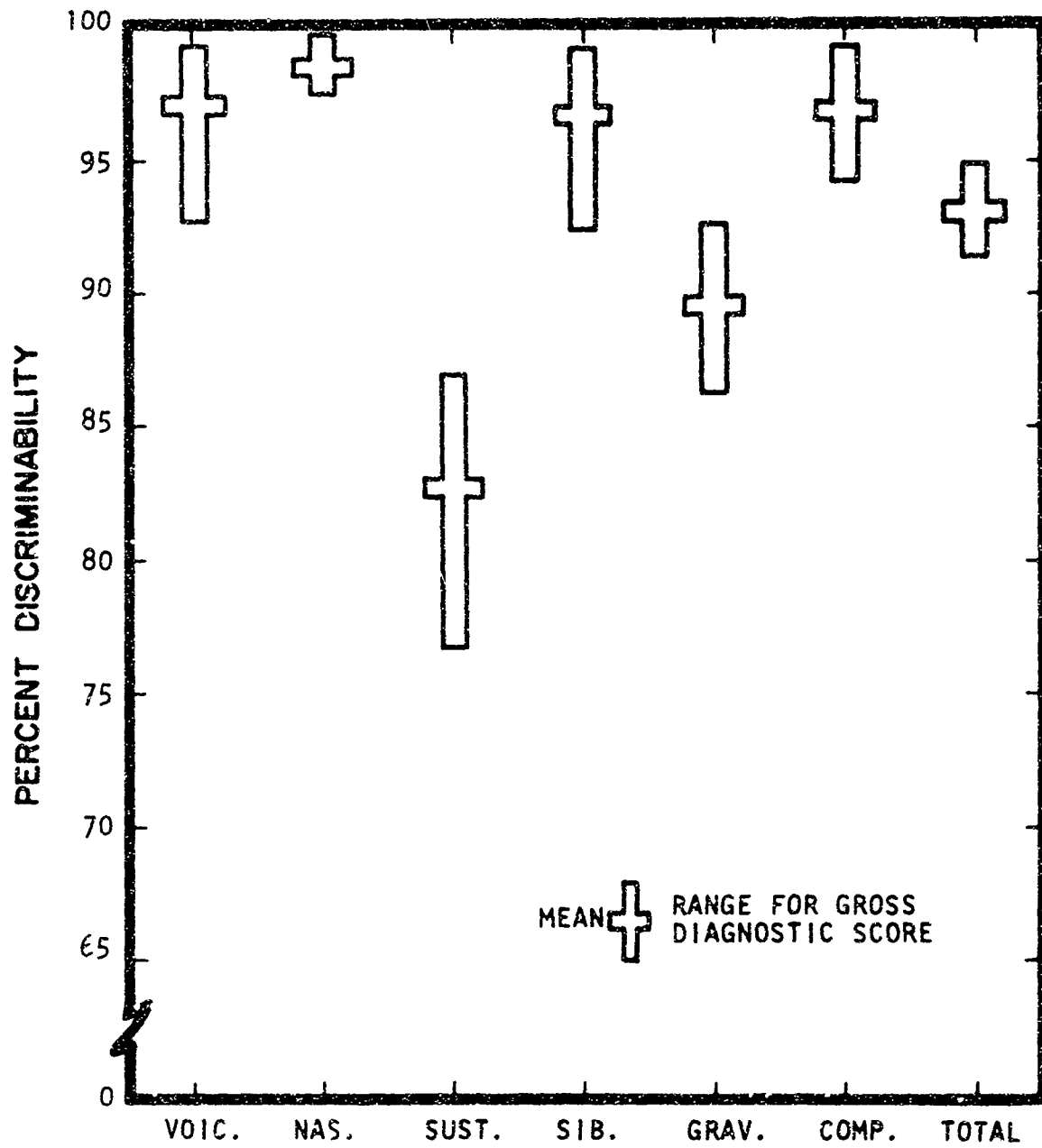
6500 TRACOR LANE, AUSTIN, TEXAS 78721

Testing Procedure

Two scramblings of the DRT materials were obtained for each system under consideration. These were used in a counter-balanced series of listening tests which extended over several weeks, eight tests being conducted in each of two to three sessions per week. The 2400 bps systems were randomly ordered, along with several other systems, and tested using one of the two scramblings available for each system. The testing order was then reversed and each system retested using the remaining scrambling of the DRT materials.

RESULTS

Figure 1 summarizes the Diagnostic Rhyme Test results obtained for eight 2400 bps digital vocoders. Shown in the figure are the mean and range for the total DRT score and for each of the six gross diagnostic scores. It may be seen that present-day digital vocoders can cope quite adequately with the physical correlates of some consonant attributes. Their performance with respect to others leaves something to be desired, however. For all of the systems tested, listeners are able to discriminate the speaker's intent with respect to the attribute voicing at levels comparable to those typically observed with unprocessed, high-fidelity speech recordings. The case of nasality is similar to that of voicing. All systems tested



CONSONANT ATTRIBUTES

Fig. 1. Gross discriminability of six consonant attributes for digitally vocoded speech.



transmit the stimulus correlates of this attribute with effectively perfect fidelity.

It is in the case of sustention (i.e., the fricative-plosive opposition) that present-day digital vocoder technique is most significantly deficient. No system tested permits gross discrimination of this attribute at better than the .88 level of accuracy, the average being .83, and even these levels are not achieved without certain sacrifices, as will be discussed below.

The attribute, sibilant (which distinguishes the /s-θ/, /z-·/, /ʃ-k/ and /ʒ-g/ opposition) proves relatively impervious to the vocoding process, the average system permitting .97 correct discrimination and no system yielding poorer than .95 gross discrimination of this attribute.

In the case of graveness (as exemplified by the /p-t/ and /f-θ/ oppositions) the average system permits correct discrimination of the speaker's intent at .90 better than chance level, while the best system in this respect yields a discrimination score of no higher than .93. However, in the case of compactness which, like graveness, is generally considered to depend, albeit a rather complicated manner, upon the nature of the second formant transition, listeners are able to discriminate the speaker's intent with a level of accuracy which approaches that for the case of unprocessed speech.



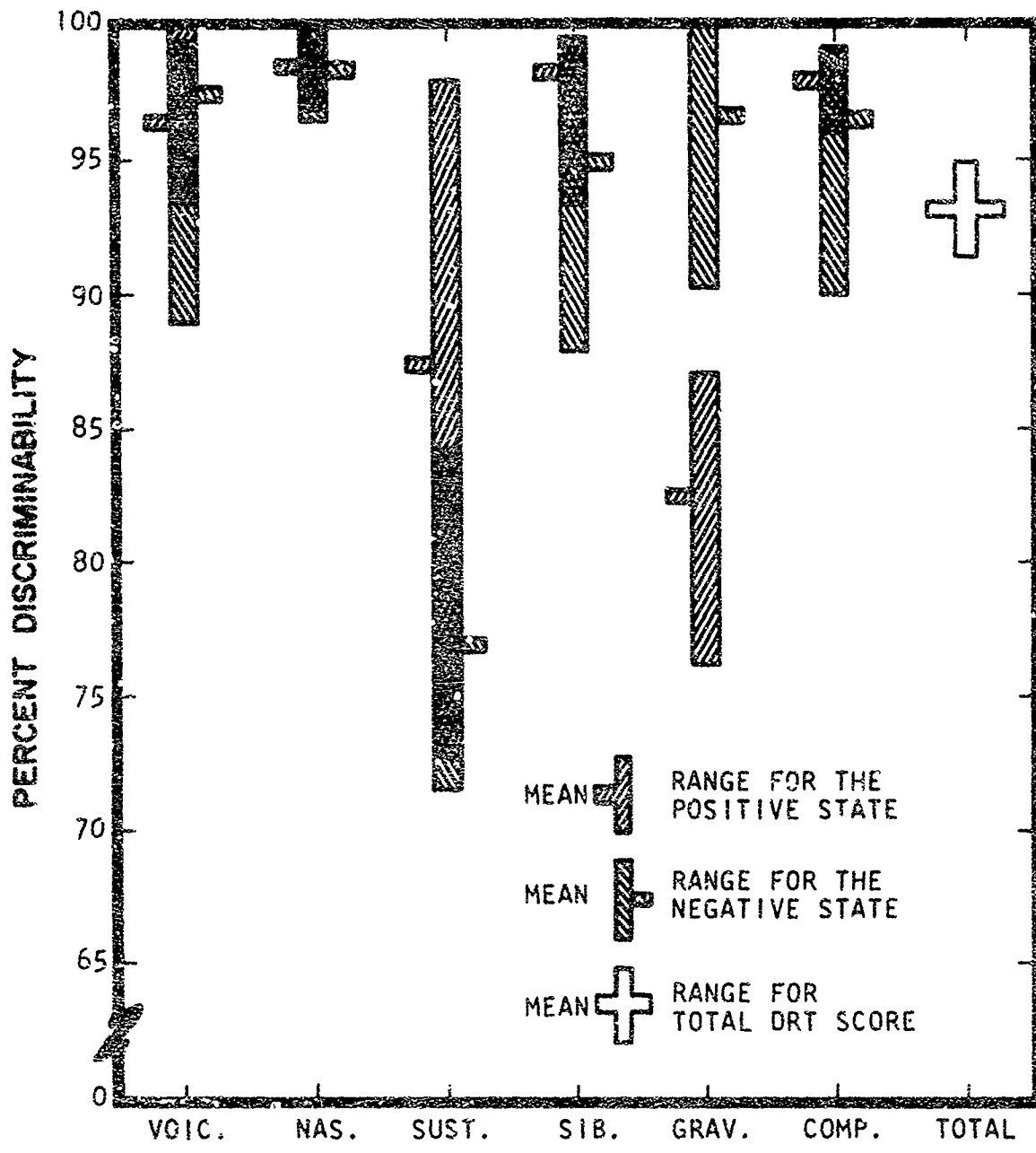
6500 TRACOR LANE, AUSTIN, TEXAS 78721

It is not feasible, here, to attempt extensive generalizations as to the specific system features which account for the deficiencies noted above. However, a more detailed examination of the nature of these deficiencies may yield insights of some value.

Figure 2 permits separate examination of listener performance in discriminating each of the two extreme states, of the six attributes recognized in the DRT. It thus reveals the extent to which the vocoding process may bias listener perception of the various attributes.

In the cases of voicing and nasality no significant bias is evident, at least for the typical vocoder. Listeners discriminate the positive states (e.g., voiced) of these attributes as well as their negative states (e.g., unvoiced).

In the case of sustention, however, a drastic bias is evident. It appears, here, that digital vocoding not only acts generally to obscure important cues used by listeners in discriminating the state of this attribute, it tends, typically, to bias listener perception significantly in favor of the positive (i.e., sustained state of this attribute). It should be noted, however, that among the eight systems involved virtually none conform closely to the statistical norms indicated in Figure 2. Several approached the extreme in positive bias within the ranges indicated in the figure for the case of



CONSONANT ATTRIBUTES

Fig. 2. Discriminability of each state of six consonant attributes for digitally vocoded speech.

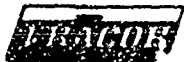


6500 TRACOR LANE, AUSTIN, TEXAS 78721

sustention. Thus with some systems listeners were able to discriminate correctly the positive (sustained) state of this attribute with better than .95 accuracy, but evidently achieved this level of performance partially through significant tendency to perceive interrupted consonants as their sustained cognates (to perceive /b/ as /v/, /t/ as /θ/, etc.). With several other systems, however, a pronounced, if not so extreme, bias in the opposite direction was observed. This variation in degree and direction of bias in discriminability of the sustained-interrupted opposition serves in fact to define the one dimension of performance in which present-day vocoders differ most drastically among themselves. It should not be forgotten, however, that in no case was performance with respect to the attribute sustention altogether acceptable by any absolute standard.

Further perusal of Figure 2 reveals that the typical digital vocoder of today is conducive to a slight if statistically significant, bias in favor of the positive state of the consonant attribute, sibilation (i.e., a tendency to perceive /θ/ as /s/, /g/ as /^hg/ and so on). With regard to direction, these were no exceptions to this trend among the eight vocoders tested. However, the magnitude of this bias varied from negligible to the extremes possible within the ranges shown in Figure 2.

In the case of graveness a pronounced negative bias is evident. On the average, the negative (acute) state of this attribute is correctly discriminated at the .97 level of accuracy



6500 TRACOR LANE, AUSTIN, TEXAS 78721

and no system tested yielded a score of lower than .89 in this dimension of performance. However, no system in the survey permitted discrimination of the positive state of this attribute at better than the .87 level, the average being only .82 for the eight systems.

Specific reasons for this pronounced bias are rather difficult to find in current theory concerning the perceptually relevant acoustical correlates of phonemic features or attributes. However, an accumulation of results from diagnostic tests of speech processing devices serve to provide an insight of practical significance, at least. No speech processing device utilizing spectrum sampling or frame rates of forty frames per second has yet yielded a discrimination score significantly above .90 for the positive or grave state of the attribute, graveness. Frame rates of this order tend, ceteris paribus, to cause grave consonants (/pbfvwm/) to be perceived as their acute cognates (i.e., as /tdθ:rn/ respectively). Although it is conceivable that techniques for the enrichment of the speech signal by means of other, nontime dependent cues to the state of the attribute, graveness, could serve to compensate for this deficiency of present-day digital vocoders, the most direct remedy is evidently an increase in spectrum sampling rate.

In the case of compactness the typical system exhibits only negligible bias, and even the exceptional cases permit discrimination scores of .90 or greater for both states of this attribute.



6500 TRACOR LANE. AUSTIN, TEXAS 78721

In summary, the typical 2400 bps digital vocoder of 1967 achieves adequate or better fidelity in transmitting the physical correlates of the perceptual phonemic attributes, voicing, nasality, sibilant and compactness. In the case of two attributes, however, its performance is significantly deficient. It tends not only to reduce the gross discriminability of the attribute, sustentation and graveness, but also to bias seriously listener perception of the states of these attributes. These deficiencies are not characteristic of analog vocoder systems and must, therefore, be attributed to the inadequacies of current digital vocoding practices. The remedy is not altogether clear, but the best available evidence suggests that increased sampling (i.e., frame) rates, even if achieved at some sacrifice in short term spectral fidelity, could substantially improve the situation.

Finally, it should be recalled that the results described here were obtained under highly controlled conditions, such that the variations observed among the different systems can confidently be attributed to intrinsic differences among them. Such ideal conditions are likely to prevail only rarely; if ever, in the operational communications environment. It remains to be demonstrated, therefore, that vocoders which can cope adequately with high fidelity recordings by trained speakers will also perform satisfactorily over a range of speakers, microphones ambient sound conditions and transmission techniques.



6500 TRACOR LANE, AUSTIN, TEXAS 78721

REFERENCES

1. Voiers, W. D., Cohen, M. F., and Mickunas, J., Performance Evaluation of Speech Processing Devices, I. Intelligibility, Quality, Speaker Recognizability, Final Report, Contract No. AF19(628)-4195, AFCRL-65-826 (1965).
2. Dolansky, L., Barrs, J. T., Howard, C. R., and Karis, C., Study and Investigation of Speech Data, Final Report, Contract No. AF19(628)-389 AFCRL-63-118 (1965).
3. Voiers, W. D., Performance Evaluation of Speech Processing Devices, III. Diagnostic Evaluation of Speech Intelligibility, Final Report, Contract No. AF19(628)-4987, AFCRL-67-0101 (1967).



5500 TRACOR LANE, AUSTIN, TEXAS 78721

APPENDIX II

MINOR EXPERIMENTAL STUDIES



6560 TRACOR LANE, AUSTIN, TEXAS 78721

SUMMARY OF EXPERIMENTAL STUDY NO. FAV I

TITLE: Intelligibility of Periodically Interrupted Speech.

RESPONSIBLE SCIENTIST(S): Lonnie Wilson, Alan Sharpley, and William D. Voiers.

PURPOSE: To investigate the effect of periodic interruption on the discriminability of consonant features at different rates of interruption and duty cycles.

METHODS AND MATERIALS:

Subjects: Eight male high-school students selected on the basis of consistency of performance on the Diagnostic Rhyme Test (DRT).

Location: TRACOR Human Communications Laboratory.

Stimulus Materials: Eight recorded scramblings of the DRT spoken by a medium pitched speaker.

Stimulus Conditions: Test materials were interrupted at intervals of 5 msec., 10 msec., 20 msec., 32 msec., 40 msec., 80 msec., and 160 msec. for duty cycles of 50%, 25%, and 12.5%. The DRT was otherwise undistorted and administered at a comfortable pre-interruption level of roughly 70-80dB SPL for vowel peaks.

Equipment: Materials were presented diotically via an Ampex 602.2 tape recorder, a high quality custom built amplifier and Koss PRO-600 head phones. Interruption was produced by means of a Grason-Stadler Series 829 Electronic Switch between recorder and amplifier.

EXPERIMENTAL DESIGN: The 21 conditions were each presented once with order of presentation randomized for rate of interruption. The order of the eight DRT scrambling was also randomized over conditions.



6500 TRACOR LANE, AUSTIN, TEXAS 78721

RESULTS AND DISCUSSION: Total DRT scores for the seven interruption rates at the three duty cycles are given in Figure 1. At the 50% duty cycle consonant features are largely unaffected except at the slowest interruption rate. Correct responses for both the positive and negative states of graveness are markedly lowered and biased in favor of the absent state with further decreases in the duty cycle. Scores for graveness range as low as .30 above chance for the gravestate and .50 for the acute state. Scores for the positive state of sustention are hardly affected by any of the manipulations, while those for the negative state fall substantially at 25% and 12.5% duty cycles to as low as .43 above chance. Similarly, the negative states of voicing and nasality are noticeably affected by reduced duty cycles, dropping as low as .48 and .35 respectively above chance, while their positive states remain relatively unaffected. Compactness is somewhat less degraded exhibiting a slight bias in favor of the absent state at the 12.5% duty cycle. Both states of sibilation are almost unaffected by any of the manipulations.

SUMMARY AND CONCLUSIONS: At duty cycles of 50% and interruption intervals between 5 msec. and 80 msec. consonant features are largely unaffected. Interruption periods beyond 80 msec., at a duty cycle of 50% generally, produce decrements in intelligibility.

Compactness and especially sibilation are highly resistant to degradation by interruption. Nasality and sustension and to a lesser extent voicing are biased in favor of the positive states at lower duty cycles. Both states of graveness are degraded at lower duty cycles with a noticeable bias in favor of the acute state.

At 25% and 12.5% duty cycles results are comparable with those obtained for speech masked speech and 2400 bit/s digital vocoders, with exceptions, in the case of vocoders, of nasality and sibilation.

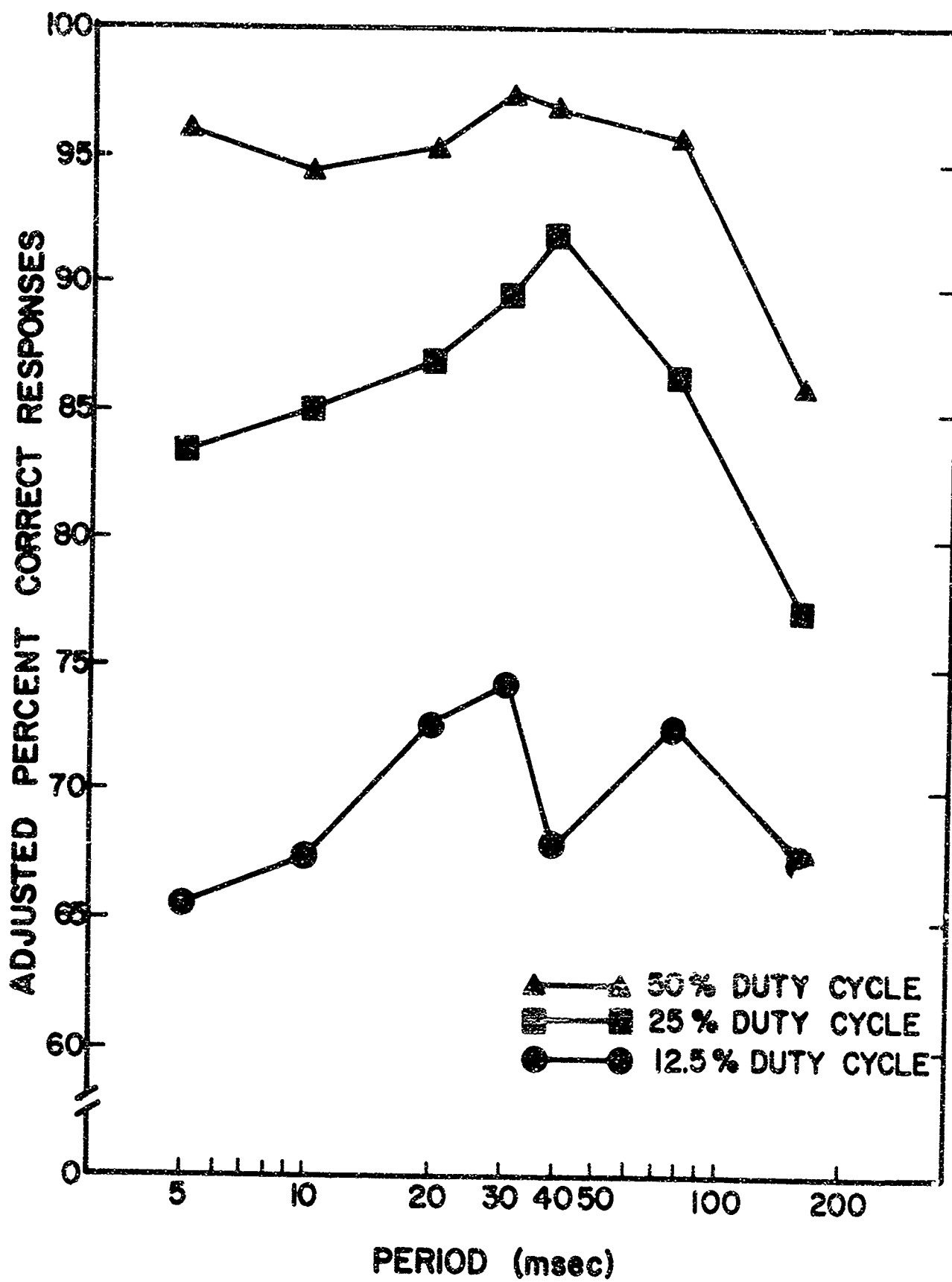


Fig. 1. Effects of periodic interruption on gross DRT score.



6500 TRACOR LANE, AUSTIN, TEXAS 78721

REFERENCES

1. Voiers, Wm. D., The Present State of Digital Vocoding Techniques: A Diagnostic Evaluation, 1968a, IEEE Transactions on Audio and Electroacoustics. Vol. AU-16, No. 2, 275-279 (1968) (See Appendix II for similar report.)
2. Voiers, Wm. D., Effects of Masking Speech on the Discriminability of Elementary Consonant Attributes, 1968b. Paper delivered at the Seventy-fifth Meetings of the Acoustical Society of America. May 1968.



6500 TRACOR LANE, AUSTIN, TEXAS 78721

SUMMARY OF EXPERIMENTAL STUDY: FRT I

TITLE: Effects of familiarity on listener response to Fairbanks Rhyme Test (FRT) Materials.

RESPONSIBLE SCIENTIST(S): Alan Sharpley, Lonnie Wilson, and William D. Voiers.

PURPOSE: To determine the effects of the listeners familiarity with the FRT "percent correct" responses and upon the qualitative character of incorrect responses.

METHODS AND MATERIALS:

Subjects: Five male high-school students selected on the basis of consistency of performance on the Diagnostic Rhyme Test.

Location: TRACOR Human Communications Laboratory.

Stimulus Materials: Six scramblings of the full scale (250 items) FRT materials (two by each of three medium pitched speakers.)

Stimulus Conditions: FRT and Gaussian noise were band-passed (200-4000 Hz). Speech (vowel peaks) to noise ratio was 0dB.

Equipment: Testing facilities included two double-walled IAC rooms, each partitioned into four listening booths. Test materials were presented diotically via an Ampex 602.2 tape recorder, a high quality custom built amplifier and Koss PRO-600 head phones at a comfortable listening level (roughly estimated at 75-80dB SPL for vowel peaks).

EXPERIMENTAL DESIGN: FRT material was presented in two sessions of 16 presentations each. Two scramblings (250 items each) for each of three speakers were used. Speakers and scramblings were cycled as follows: sp. CH-sc. 1, sp. SL-sc. 1, sp. RD-sc. 1, sp. CH-sc. 2, sp. SL-sc. 2, sp. RD-sc. 2. This basic speaker-scrambling set of six presentations was repeated throughout the study.



6500 TRACOR LANE, AUSTIN, TEXAS 78721

RESULTS AND DISCUSSION:

On the first few exposures to FRT materials both percent correct (PC) and within corpus (WC) FRT scores (i.e., correct or incorrect responses utilizing words contained in the corpus of FRT words) rapidly increase (Figures 1 & 2). This increment becomes smaller through further exposures. The early, rapid increase can be attributed to increasing familiarity with the test task, rather than the test materials, whereas familiarization with the corpus of test items may presumably account for the subsequent slower increase. Since the thirty-two presentations were divided into two sessions, fatigue may have operated somewhat to offset the effects of familiarity so that it remains to be determined where the practice effect ultimately becomes asymptotic.

SUMMARY AND CONCLUSIONS:

1. The FRT is extremely speaker sensitive.
2. Results of thirty-two exposures to the FRT failed to substantiate the hypothesis that listeners will tend increasingly to draw their erroneous responses from the corpus of FRT materials as they acquire greater familiarity with this corpus. But, it is possible that the limited number of scramblings utilized here accounts for this trend.
3. The greatest effect of repeated exposure is evident early in the course of familiarization training and can probably be attributed primarily to the learning of the task involved in taking the FRT. More gradual changes, later in the course of training, possibly result from greater familiarity with the corpus of FRT test items.
4. Decreased performance may be the effect of fatigue, becoming evident toward the latter portions of prolonged exposure to FRT materials.

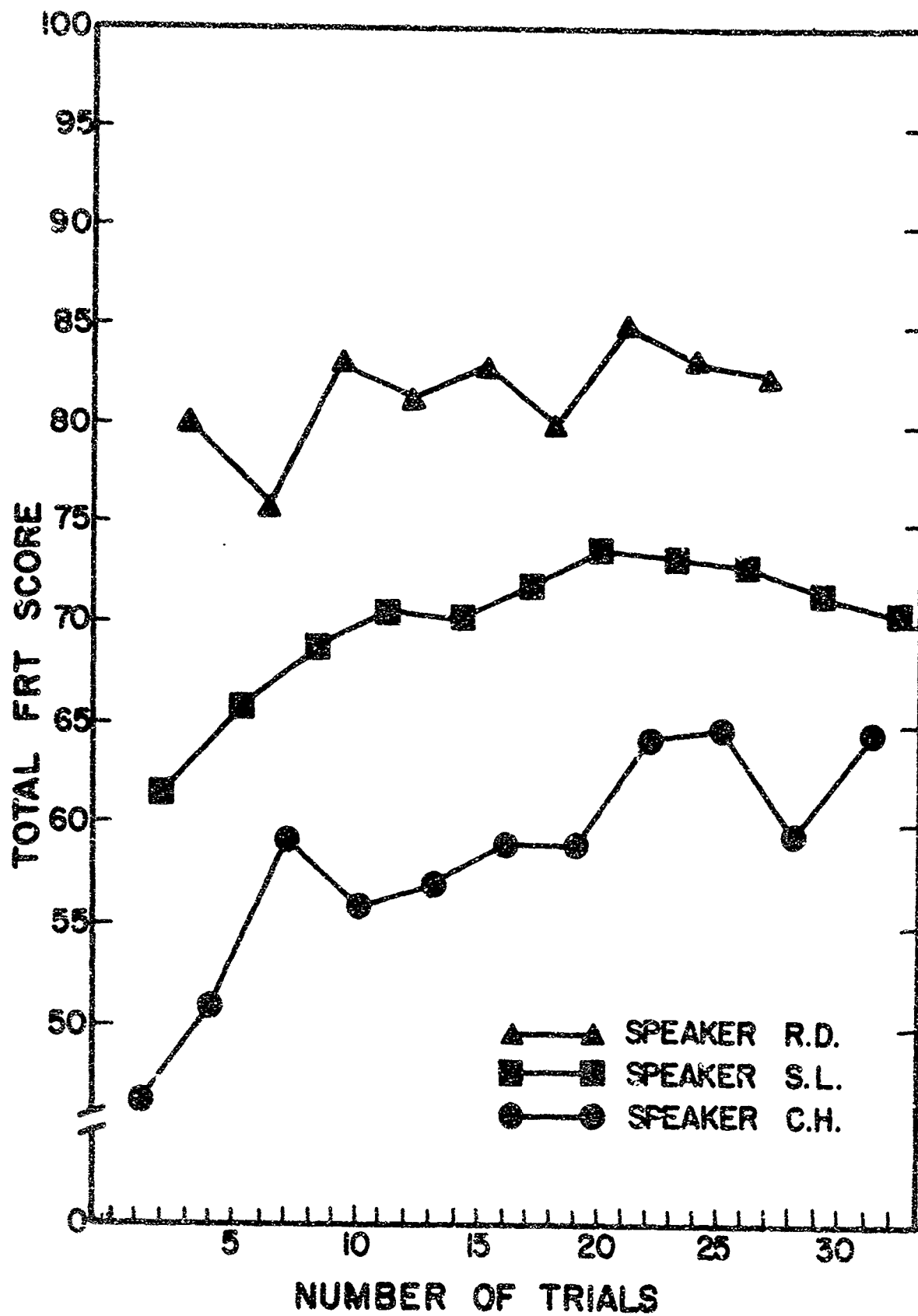


Fig. 1. Effects of practice on percent correct Fairbanks Rhyme Test scores.

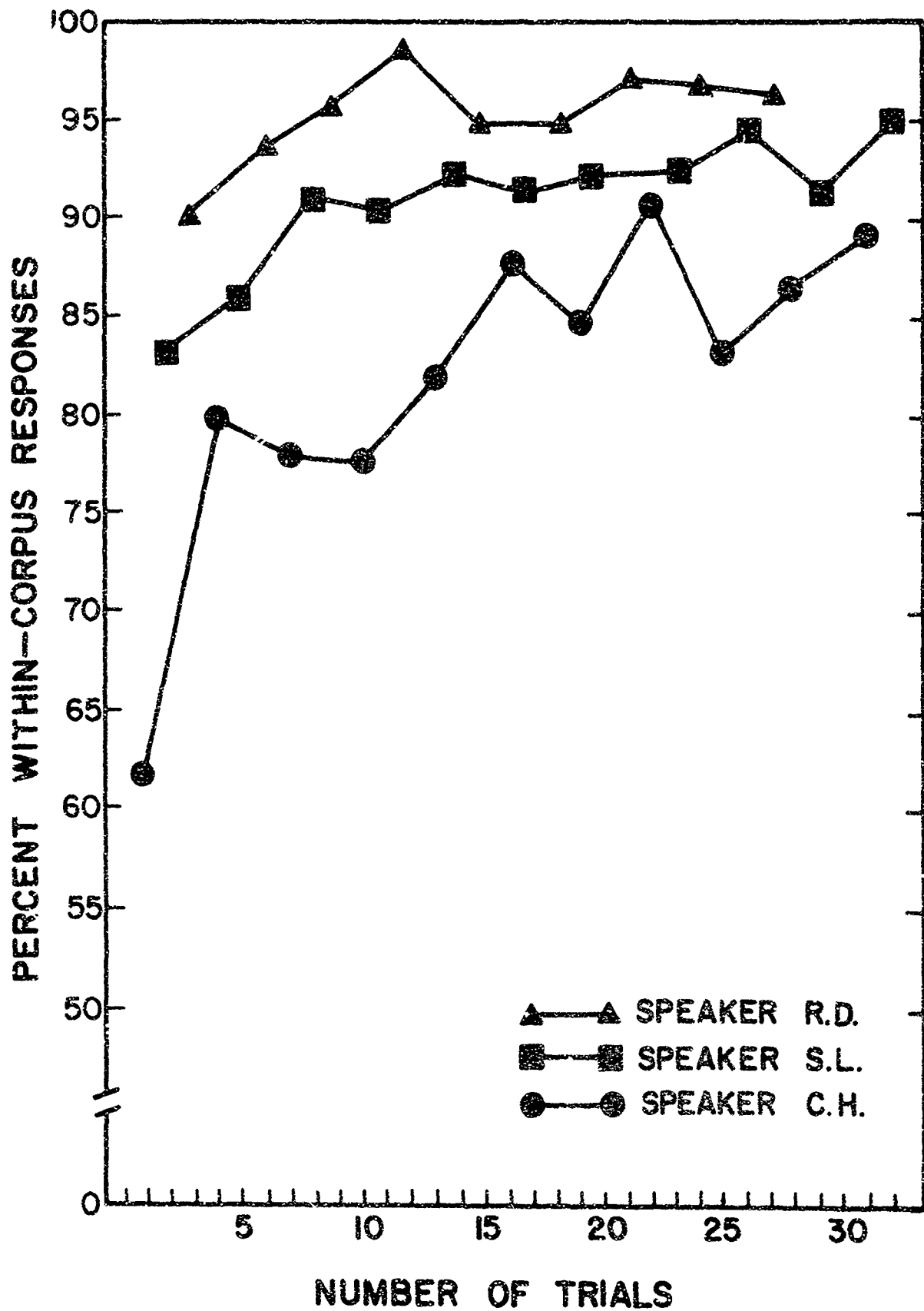


Fig. 2. Effects of practice on within-corpus Fairbanks Rhyme Test score.



6560 TRACOR LANE, AUSTIN, TEXAS 78721

SUMMARY OF EXPERIMENTAL STUDY NO. LEV I

TITLE: Effects of presentation level on listener response to the Diagnostic Rhyme Test (DRT) materials.

RESPONSIBLE SCIENTIST(S): Lonnie Wilson, Alan D. Sharpley, and William D. Voiers.

PURPOSE: To determine the effects of level of presentation of DRT materials on listener performance.

METHODS AND MATERIALS:

Subjects: Six male high-school students selected on the basis on consistency of performance on the DRT.

Location: TRACOR Human Communications Laboratory.

Stimulus Materials: Four scramblings of the full scale Diagnostic Rhyme Test (384 items) by one medium pitched speaker.

Stimulus Conditions: Clear speech at level ranging from 30 to 80dB SPL (vowel peaks) in 5dB steps.

Equipment: Test materials were presented diotically via an Ampex 602.2 tape recorder, a high quality custom built amplifier, and Koss PRO-600 head phones.

EXPERIMENTAL DESIGN: Eleven presentation levels ranging from 30 to 80 dB SPL were randomly ordered and this order presented once. Then, the same order was reversed and repeated. The study was completed in three presentation sessions.

RESULTS AND DISCUSSION: Two presentations were made for each of the eleven level conditions. The diagnostic scores for each of the six consonant attributes for corresponding presentation pairs were then averaged and are shown in Figure 1.



6500 TRACOR LANE. AUSTIN, TEXAS 78721

Since reduction of speech level is essentially equivalent to the addition of Gaussian noise, a comparison of results for the two cases is of possible interest.

Figure 2 presents the results for the case of Gaussian masking noise (Voiers), where it may be observed that the order in which the various consonant attributes affected is similar to that observed under conditions of reduced presentation level. An exception is found in the case of sibilant, however. Although, reduction in presentation level severely decreases the discriminability of this attribute, additive noise effects it only moderately.

SUMMARY AND CONCLUSIONS:

1. Reduction of speech level and addition of Gaussian noise have generally similar effect on the diagnostic scores for consonant attributes. However, the former has substantially greater effect on the discriminability of sibilant. It is possible that this discrepancy is attributable to speaker factors.
2. Variation in the total DRT score and in the various diagnostic scores is negligible over the range from 60 to 80dB SPL.

Ref.: Voiers, Wm. D., "A Diagnostic Rhyme Test," in preparation.

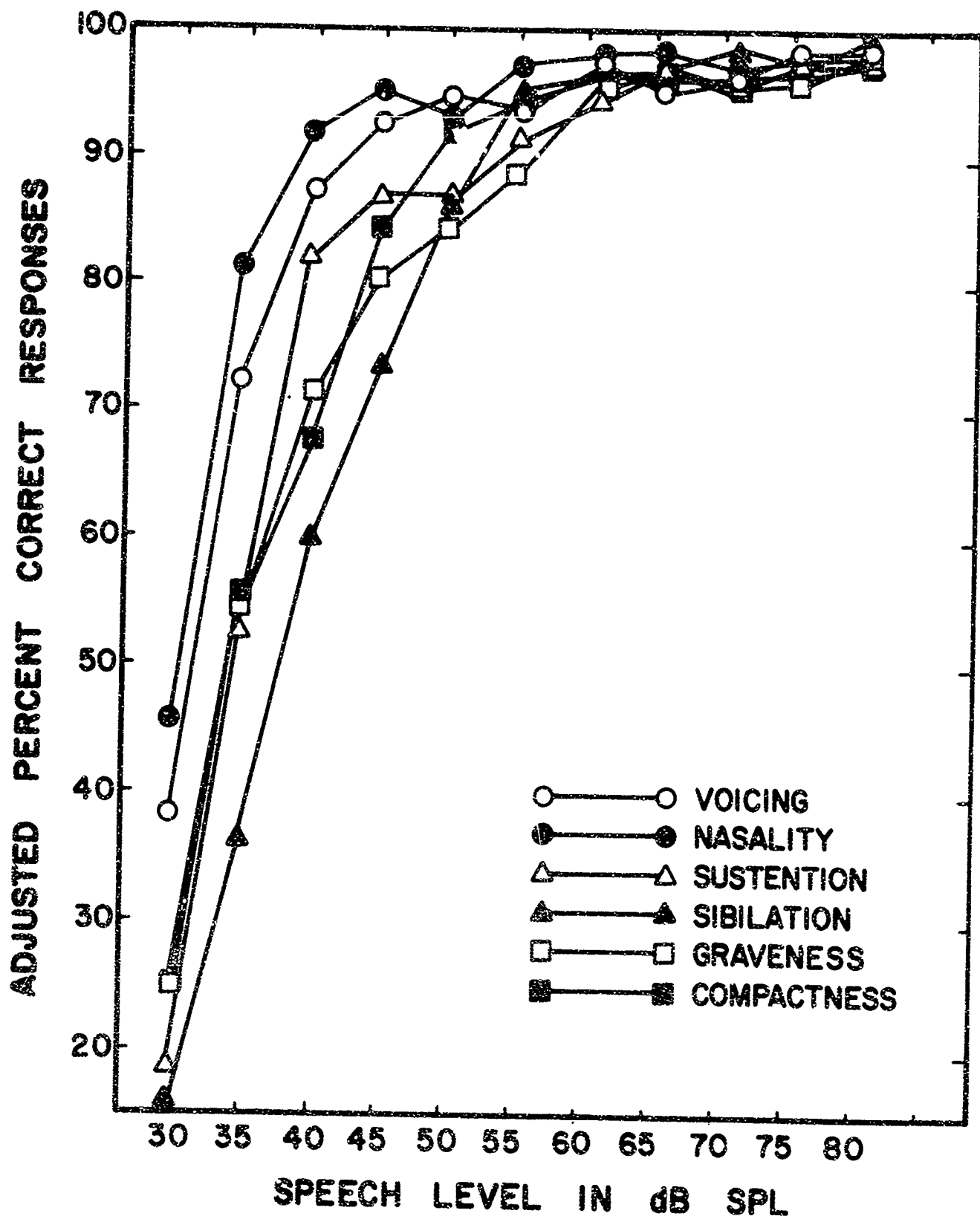


Fig. 1. Effects of speech level on the discriminability of consonant attributes.

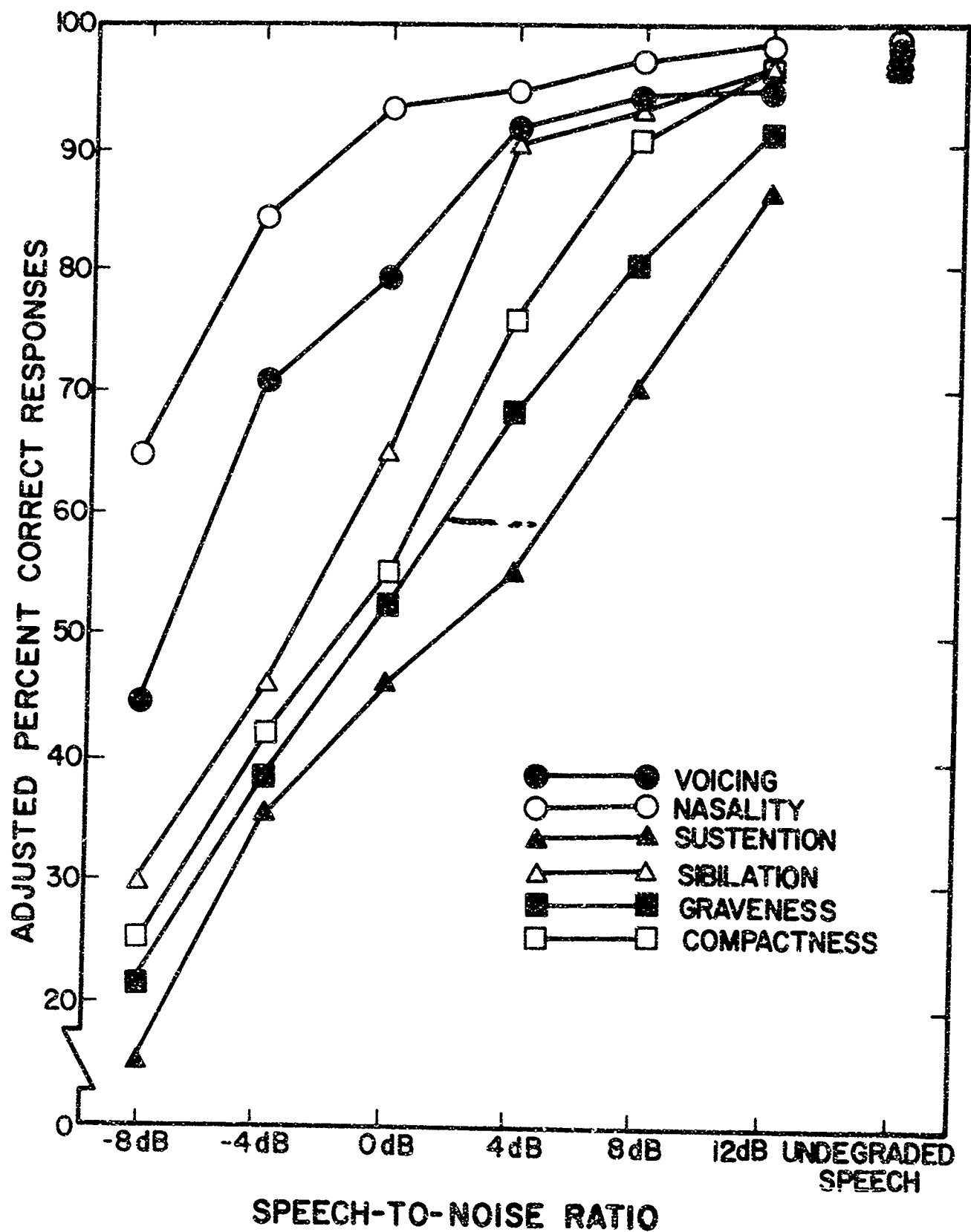


Fig. 2. Effects of band limited Gaussian noise on the discriminability of consonant attributes.



8500 TRACOR LANE, AUSTIN, TEXAS 78721

SUMMARY OF EXPERIMENTAL STUDY NO. VOWEL I

TITLE: Effects of vowel interaction on listener response to the Diagnostic Rhyme Test (DRT) materials.

RESPONSIBLE SCIENTIST(S): Alan D. Sharpley, Lonnie Wilson, and William D. Voiers.

PURPOSE: To determine the effects of following vowels on the intelligibility of initial consonants as shown by "percent correct" responses for the basic consonant attributes affecting speech intelligibility.

METHODS AND MATERIALS:

Subjects: Eight male high-school students selected on the basis of consistency of performance on the DRT.

Location: TRACOR Human Communications Laboratory

Stimulus Materials: Nine scramblings of the full scale DRT (384 items), three scramblings by each of three speakers (one low, one medium, and one high pitched).

Stimulus Conditions: Clear speech was presented at a comfortable listening level (roughly estimated at 75 to 80dB SPL for vowel peaks).

Equipment: Testing facilities included two double-walled IAC rooms, each partitioned into four listening booths. Test materials were presented diotically via an Ampex 602.2 tape recorder, a high quality custom built amplifier and Koss PRO-600 head phones.

EXPERIMENTAL DESIGN: DRT material was presented in three sessions of nine presentations each. Each session included three scramblings by each of three speakers. The nine scramblings were randomly ordered for each of the three presentation sessions.



6500 TRACOR LANE, AUSTIN, TEXAS 78721

RESULTS AND DISCUSSION:

Discrimination errors for each of the elementary consonant attributes were subjected to analysis of variance with factorial design. Three main effects were distinguished: place of maximum constriction in vowel formation (front or back), tongue elevation (high or low), and speakers (high pitched, medium pitched, and low pitched). Trials (or scramblings) "within speakers" constituted the only source of random variation, since all results were averaged across listeners.

Place of maximum constriction proved highly significant ($p < .01$) in the cases of voicing and compactness. Tongue elevation was significant ($p < .05$) only for the attribute voicing, while speakers yielded no significant F ratios. Place of constriction x speakers was significant ($p < .05$) for both graveness and compactness. Several second order interactions were significant but not easily interpreted at this point.

SUMMARY AND CONCLUSIONS:

There is evidence that discriminability of certain elementary consonant features is affected by vowel context.



6500 TRACOR LANE, AUSTIN, TEXAS 78721

SUMMARY OF EXPERIMENTAL STUDY NO. SHORT FORM DRT

TITLE: Validation of abbreviated forms of the Diagnostic Rhyme Test (DRT).

RESPONSIBLE SCIENTIST(S): Alan D. Sharpley, Lonnie Wilson, and William D. Voiers.

PURPOSE: To evaluate four abbreviated forms (48 items) of the DRT from the standpoint of gross difficulty.

METHODS AND MATERIALS:

Subjects: Seven male high-school students selected on the basis of consistency of performance on the DRT.

Location: TRACOR Human Communications Laboratory.

Stimulus Materials: Five scramblings each of four abbreviated forms.

Stimulus Conditions: DRT speech materials were presented at a constant level (vowel peaks) of 70dB SPL while band-limited Gaussian noise was varied from 70 to 88dB SPL.

Equipment: Test materials were presented diotically via an Ampex 602.2 tape recorder, a high quality custom built amplifier, and Koss PRO-600 head phones.

EXPERIMENTAL DESIGN: Five scramblings for each of the four abbreviated forms of the DRT were used. All twenty scramblings were presented twice in the testing session. Each of the four forms was presented at a constant speech level of 70dB SPL for each of ten noise levels. A random presentation order for the ten noise conditions was employed.

RESULTS AND DISCUSSION: The results of this study are summarized in Figure 1., where it may be observed that the four abbreviated versions of the DRT tend, generally, to yield very nearly equivalent results. While there are occasional exceptions to this equivalence, these are evidently attributable to random factors



6500 TRACOR LANE, AUSTIN, TEXAS 78721

rather than to systematic differences among the four forms. In view of the fact that each form utilizes a unique segment of the total corpus of DRT materials (and the fact that each segment, involves only one eighth of the number of test words and test time normally employed with the DRT) the present results are altogether remarkable. They serve in addition to provide important, if indirect validation, of the underlying philosophy of the DRT and of the consonant taxonomy on which it is based.

SUMMARY AND CONCLUSIONS: The use of abbreviated versions of the DRT to obtain gross measures of consonant discriminability under circumstances of limited testing and/or transmission time appears to be both feasible and justifiable in light of the evidence available thus far.

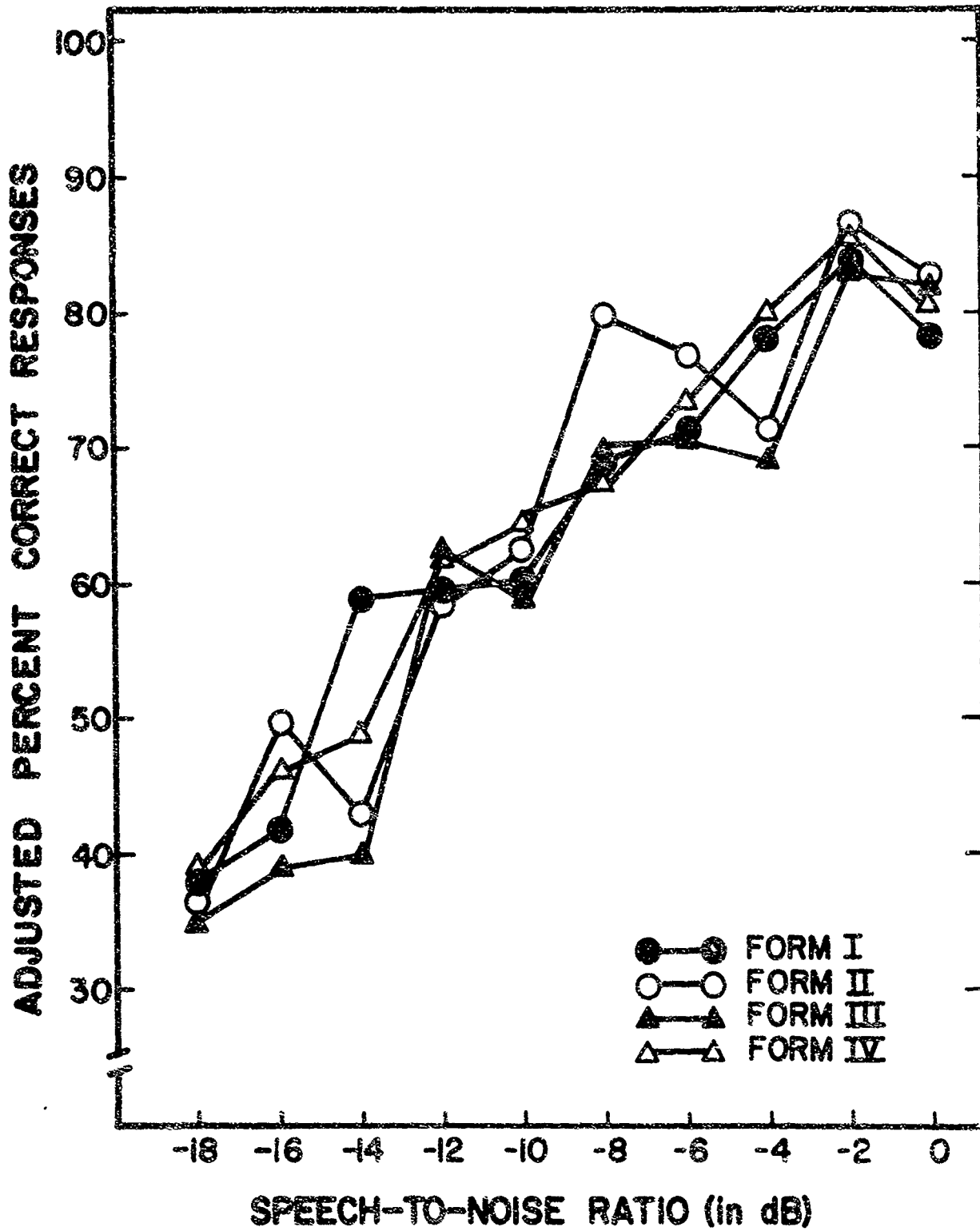


Fig. 1. Effects of band limited Gaussian noise on the abbreviated forms of the Diagnostic Rhyme Test.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D		
<i>(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)</i>		
1. ORIGINATING ACTIVITY (Corporate author) TRACOR, Inc. 6500 Tracor Lane Austin, Texas 78721	2a. REPORT SECURITY CLASSIFICATION Unclassified	2b. GROUP
3. REPORT TITLE DIAGNOSTIC EVALUATION OF SPEECH PROCESSING SYSTEMS		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Scientific. Final. 15 September 1967 - 14 September 1968		
5. AUTHOR(S) (First name, middle initial, last name) William D. Voiers		
6. REPORT DATE December 1968	7a. TOTAL NO. OF PAGES 41	7b. NO. OF REFS 6
8a. CONTRACT OR GRANT NO. F19628-68-C-0068	8b. ORIGINATOR'S REPORT NUMBER(S)	
9. PROJECT NO. 4610-02-01	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
10. DOD Element 674610	AFCRL-68-0568	
11. DOD Subelement 62702F		
10. DISTRIBUTION STATEMENT Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.		
11. SUPPLEMENTARY NOTES TECH, OTHER	12. SPONSORING MILITARY ACTIVITY Air Force Cambridge Research Laboratories (CRB) L. G. Hanscom Field Bedford, Massachusetts 01730	
13. ABSTRACT This report presents the results of a survey in which the Diagnostic Rhyme Test was used to evaluate the present state of digital technique for speech processing and communication. Also presented are the results of a series of minor studies concerned with the methodology of intelligibility evaluation.		

DD FORM 1473
1 NOV 65

Unclassified

Security Classification

Unclassified
Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Diagnostic Rhyme Test Vocoder Performance						

Unclassified
Security Classification