The Use of Compressed Speech in Selecting Morse Code Operators

[redacted]

The problem of selecting people for morse code training has perplexed researchers for nearly a century. This paper documents an experiment in which the ability to comprehend time-compressed speech was investigated as a possible screening technique for identifying individuals who possess an aptitude for copying morse code signals.

INTRODUCTION

The ability to learn International Morse Code (IMC) is apparently a special aptitude unrelated to other aptitudes or skills (Goffard, 1960). Goffard said:

> For some men code skill seems to be impossible to learn, while for others it presents no problem. Although methods of selecting men with a high aptitude for learning IMC have been the object of research for a number of years, they are still only moderately satisfactory. With the least apt men eliminated by the Army Radio Code Test of the Army Classification Battery, the range of aptitude for code among men in courses which include IMC is still wide. [P. 3]

In one of the first studies of its kind, Thurstone (1919) used "mental tests" in an attempt to predict ability in telegraphy. He concluded that

> the general intelligence tests are not as valuable for diagnosing ability to learn telegraphy as for measuring general intelligence. Ability in telegraphy is probably a special ability. . . . The fact that years of schooling does not agree with ability to learn telegraphy indicates that this is a special ability. College graduates usually do better on general intelligence tests than those who have only finished grammar school. But college graduates do not necessarily excel in learning telegraphy. [P. 117]

Low correlations have repeatedly been found between morse code achievement and intelligence, educational level, mechanical ability, and knowledge of subject matter. Woehlke (1956) summarized the research on attempts to identify selection devices for morse trainees. He reviewed special tests of morse aptitude including the "code learning" types of tests such as tests of code discrimination, learning, and speed of response. He found that attempts to link general ability, achievement, and aptitude tests as well as nontest factors such as age and sex to code ability were unfruitful. Specific aptitude tests included auditory factor tests and clerical, musical, and mechanical aptitude tests as well as many others. When compared to other areas of testing, Woehlke concluded that morse code aptitude validities are inadequate.

The present attrition rate among morse trainees is high. Attrition ranges from 26% to 42% of which "most" (i.e., 20% to 30%) is due to academic failure. Of the academic failures, nearly 100% are because of the inability of the students to copy code sent at a speed of 20 GPM (CODEZ[1]) — the training standard. Historically, attrition has been

[1]CODEZ is the standard now used in the military for determining code speed. It represents the speed that would transmit this 58-baud group with the standard 1–3–7 spacing twenty times in one minute. The previous standard, PARIS, had fewer bauds. Therefore, 20 GPM CODEZ is the same rate as 25 GPM that operators were formerly required to meet.

reported to range from 18% to 60% in both military code training and in the "early" days of training railroad telegraphers (Woehlke, 1956). The Federal Communications Commission indicated that during the last six months of 1956, approximately 38% of the applicants for the General Class amateur radio operators license were rejected for failure to pass the code receiving test (Porter, 1957). The latter group would have possessed the motivation sometimes felt lacking in the military trainee.

Attempts to adjust the training or teaching methods have had only limited success in reducing the total number of hours required to train a morse code operator and have not impacted significantly on reducing attrition. Lengthening the training period to reduce the attrition rate enabled more students to complete the school, but often these same students were rated unsatisfactory on the job by their supervisors.

COMPRESSED SPEECH

Time-compressed speech is defined (Foulke and Sticht, 1969) as speech which has been reproduced in less than the original production time. Compression is most frequently accomplished by electromechanically abutting periodic samples of the original recording. The end product is a tape with an accelerated word rate and a minimum of the frequency distortion associated with simple temporal alteration (e.g., playing a 33⅓ rpm record at 78 rpms).

Large individual differences with regard to the ability to comprehend compressed speech have been observed in the literature but have virtually been ignored in previous research.[2] This is understandable since most of the previous research has emphasized the use of compressed speech as a communications or educational medium in which individual differences were interpreted as error variance. The ability to comprehend compressed speech does not seem to improve significantly with listening exposure to compressed voice tapes. Persons receiving training designed to improve comprehension of compressed speech have frequently been found to show little or no significant differences over neophite listeners (e.g., Foulke and Sticht, 1969).

Morse code can be thought of as the first language to undergo rate compression. It is unique when compared to compressed speech in that up to a rate of 20 GPM (CODEZ) the integrity of the morse code characters themselves does not change. That is, in compressed speech, small bits of words are usually randomly discarded in order to increase the word rate, thereby decreasing somewhat the intelligibility of the words. In speeding morse code rate up to 20 GPM, the character sound is not changed, only the spacing between characters and groups is decreased to increase the rate. Therefore, the ability to comprehend time-compressed speech can be interpreted as "aptitude" related to the speed with which one can accurately process auditory stimuli. If so, it may provide an efficient technique to identify students for morse code training.

METHOD

Subjects

The experiment was planned to include 120 service members enrolled in morse code training at three service schools. At the time of this analysis, data for 92 students was available. The students ranged in age from 17 to 31 (average = 20.5). Service grade ranged from E1 to E4 (average rank = E2). All subjects had normal hearing bilaterally as

---

[2] For a thorough review of the compressed speech literature, see Duker (1974).

determined by a pure-tone audiometric screening test administered as part of the service selection battery for morse operators. Each student had achieved an acceptable score on the radio code subtest of the Armed Services Vocational Aptitude Battery (ASVAB).

Test Instruments

Portions of the STEP (Sequential Tests of Educational Progress) listening test were recorded and compressed[3,4] to create four listening comprehension test audio tapes (A, B, C, and D) of equal length and difficulty. Each test contained three selections whose content ranged from junior high through high school level in difficulty. The recording time for each selection ranged from 62 to 91 seconds at an average rate of 187 words per minute (WPM). Each selection was followed by five multiple choice items. All items and the four response choices were read on tape. The response choices were also printed on the student answer sheets. Therefore, tests A, B, C, and D each contained 15 items.

The selections in tapes B, C, and D were time compressed at 1.5, 2.0, and 2.5 times normal, respectively. The test questions and distractors were not compressed. Thus, four levels of compressed speech (normal, 1.5, 2.0, and 2.5 times normal) were available as the repeated measures dimension in the experiment. The compressed tapes are referred to as BX, CX, and DX and the WPM rates for each is 283, 340, and 434, respectively. Embedded in these tapes are tests B, C, and D which were not compressed.

A "questions only" tape was prepared to assess the prose dependency of the test items. All tapes were presented using a Califone Model 3530 cassette recorder and MPC Model MX-200 headset.

Procedures

The experimental design incorporated two control and two experimental groups. For further control, all subjects were tested individually by a test proctor. An introductory tape was used to present test instructions. Further instructions were read to the subjects by the proctors. The presentation order of the tapes changed for each student to control for possible practice effects. Control groups 1 and 2 (C_1 and C_2) each were made up of 20 students who had just entered code school. Assignment was random. C_1 was administered test tapes A, B, C, and D to assess the extent to which the four tests were equal in difficulty. C_2 was the prose-dependent control group. These students answered the same questions as the other groups in the experiment, but without benefit of listening to stories on which the items were based.

The experimental groups were administered tapes A, BX, CX, and DX. Experimental group 1 (E_1) was made up of 30 randomly selected trainees who had met the 20 GPM (CODEZ) code copying criterion. E_2 was comprised of 22 students who were being dropped from training because they were not making adequate progress toward achieving the criterion of 20 GPM.

After listening to the four tapes and answering test items, each student from the experimental groups listened to selections from tests BX, CX, and DX again in order to assess speech intelligibility at the three compression levels. The students made judgments on a scale of 0% to 100% to estimate what percentage of words from the selection that they thought they heard (independent of how they thought they performed on the earlier tests).

[3]Reproduced by permission from Educational Testing Service, Princeton, N.J.
[4]Recordings were made at the Center for Rate-Controlled Recordings, University of Louisville.

RESULTS

A univariate *t*-test compared the experimental groups on the sum of the four tests to see if there was an overall effect on the tests without regard to repeated measures. The difference was not significant ($t = 1.22$; $df = 1{,}50$; $p = .23$). A second test of interest was a one-sample multivariate Hotelling T^2 to answer the question of whether there is a trend over the repeated measures dimension (levels A, BX, CX, and DX). That is, do the scores drop off as speed increases? To do this, linear, quadratic, and cubic contrast based scores were computed. The scores were transformed scores for each individual. The result of the T^2 test comparing A, BX, CX, and DX for E_1 and E_2 (combined) was significant ($T^2 = 132.73$; $df = 3{,}49$; $p < .001$). The relationship of test performance (comprehension) to rate of compression is best expressed as a straight line or linear trend ($t = -11.15$; $df = 1{,}51$; $p < .01$). The quadratic and cubic trends were not significantly better over and above linearity.

Figure 1 shows the means of the two experimental groups across levels of compression on each test. A two-sample Hotelling T^2 was computed to test the interaction between the experimental groups across the repeated measures dimension. Once again transformed scores were used and contrast scores for the four treatment levels were computed to represent linear, quadratic, and cubic trends. The results of the T^2 approached significance

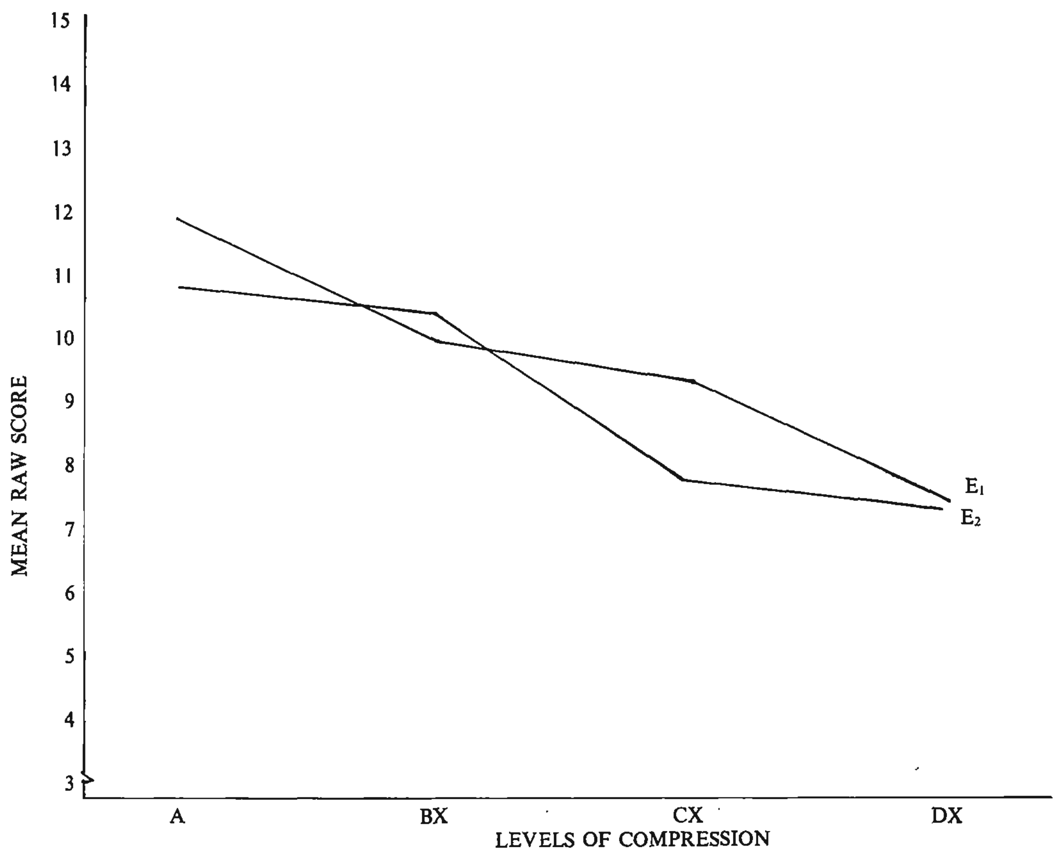

Fig. 1. Average Test Scores Across Levels of Compression for the Experimental Groups

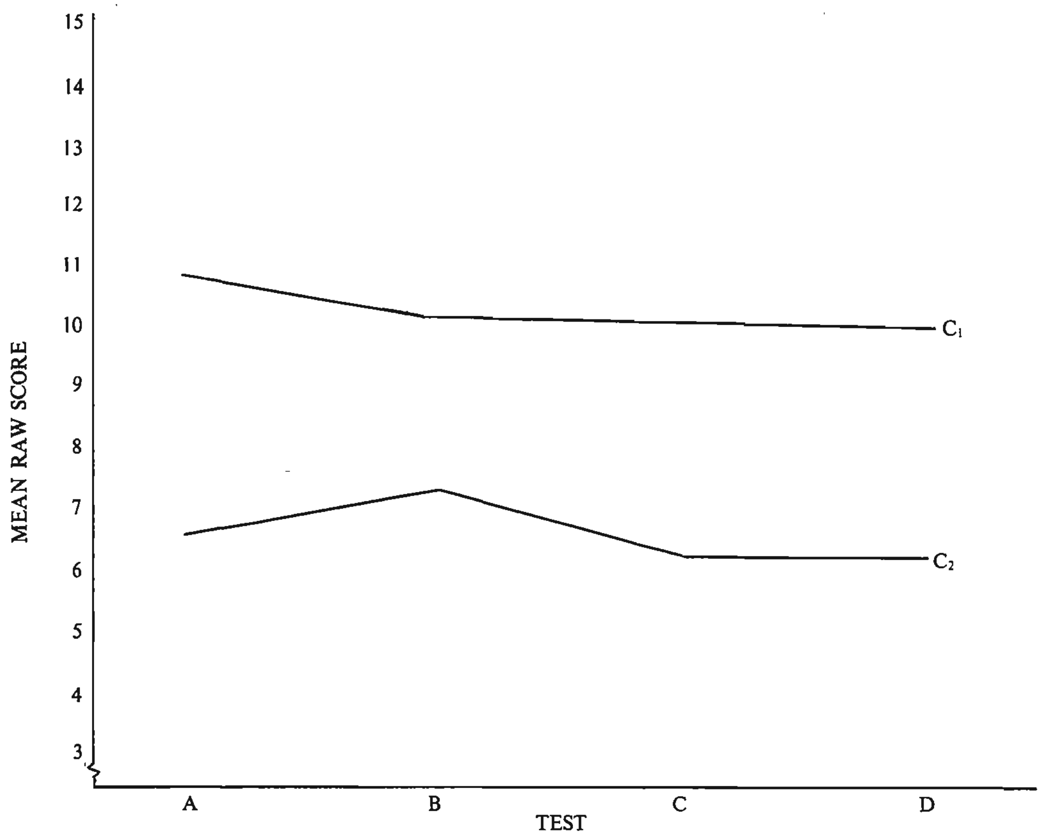

Fig. 2. Average Test Scores for the Control Groups

($T^2 = 7.35$; $df = 3,48$; $p = .08$), but the null hypothesis that successful morse trainees (E_1) and failures (E_2) do not differ, on the average, in their centroids on the comprehension tests compressed at three levels could not be rejected. Using the multivariate T^2 as an omnibus test to control for familywise testing, similar to the Fisher *lsd* approach, would dictate that the analysis stop here. The decision would be that there is no significant interaction between E_1 and E_2 across compression levels. A less conservative approach, however, is to interpret the univariate *t*-tests based on the trend contrasts which are independent of the multivariate T^2. This analysis reveals a significant cubic trend ($t = 2.65$; $df = 1,50$; $p = .01$) and means that two significantly different curved (cubic) lines express an interaction across levels by group (see fig. 1). For reasons to be presented in the discussion, the latter analysis is preferred.

A univariate *t*-test showed that operators who achieved 20 GPM scored significantly higher on level CX than did the failures ($t = 2.06$; $df = 1,50$; $p < .05$). No differences were found between the experimental groups on the other levels. One-way analysis of variance tests were computed to measure whether significant mean differences exist between groups on each test and on the sum of all tests. Significant *Fs*, each with $p < .001$ and $df = 3,88$, were obtained for all tests. The Newman-Keuls procedure was run to determine where the differences occurred. On tests A and B group C_2 was significantly lower than E_1, E_2, and C_1, which were not different from one another. C_1 was significantly higher than C_2, E_1, and E_2 on test D. Again the latter groups were not significantly different from each

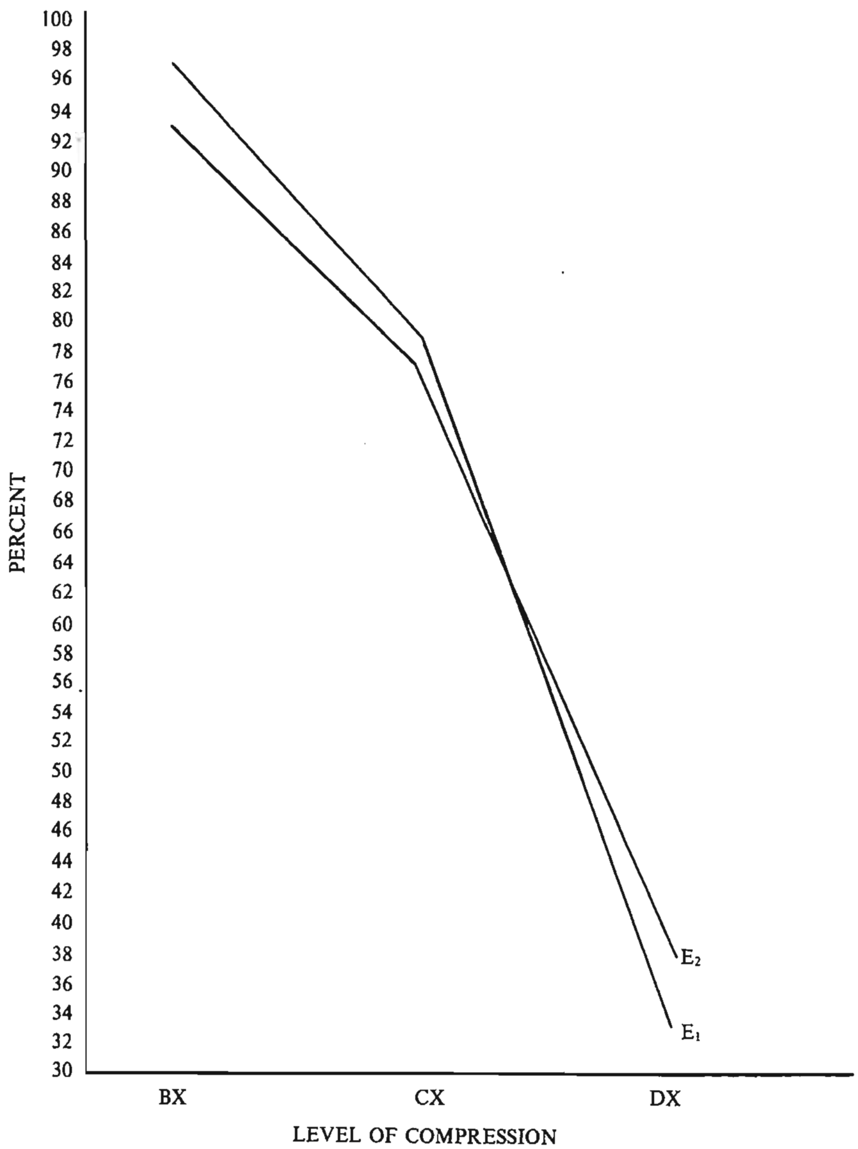

Fig. 3. Group Mean Ratings of Intelligibility Across Levels of Compression

other. Of most interest is the analysis of test C where E_1 was not significantly below C_1 and E_2 did not differ from C_2. However, E_1 and C_1 were significantly above E_2 and C_2.

An analysis of the sum of the test scores revealed that C_2 differs from all the other groups and leads to the intransitive decision that E_1 and E_2 do not differ, nor do E_1 and C_1, but that E_2 is significantly lower than C_1.

A one-sample T^2 was used to test the hypothesis that performance on tests A, B, C, and D was equal in the no-prose control group (C_2). Another test checked the equivalence of the difficulty level for the four tests for C_1 where only the normal level was presented. The null hypothesis was not rejected (as predicted) in either analysis, indicating that the deviations in mean values of each test were not significantly different. In other words, the lines in figure 2 representing C_1 and C_2 are not significantly different from horizontal lines. The T^2 values were 1.88 ($df = 3,17$; $p = .65$) and 4.64 ($df = 3,17$; $p = .28$) for C_1 and C_2, respectively.

As shown in figure 3, student perceived intelligibility of the prose also decreased significantly as a function of speed. The one sample T^2 was equal to 359.78 ($df = 2,50$; $p < .001$). The univariate t for linear trend was significant ($p < .01$); however, the quadratic trend provided a significant improvement over and above linearity ($p < .001$). Therefore, a curved (quadratic) line best represents the relationship between perceived intelligibility and speed. The two-sample multivariate test for interaction between E_1 and E_2 across levels was not significant ($T^2 = 1.64$; $df = 2,49$; $p = .45$). The independent trend tests were also nonsignificant.

The correlation for the experimental groups between final code speed achieved in the course and the radio code (RC) subtest of ASVAB was .35 ($df = 1,45$; $p < .05$) Test C for the experimental groups did not correlate significantly with code speed ($r = .24$; $df = 1,50$; $p > .05$). Additionally, RC and C did not correlate with one another ($r = -.01$; $df = 1,46$; $p > .05$).

DISCUSSION AND CONCLUSIONS

There are several practical as well as theoretical inferences which can be based on the results of this study. Listening comprehension scores at normal speed did not discriminate successful code students from those who failed, supporting earlier findings as reviewed by Woehlke (1956). However, at word rates of 345 WPM (twice normal), successful code trainees were less seriously hampered in understanding the context of the prose selections than were the group of students who failed to meet the code speed requirements. In fact, the successful trainees' performance was not significantly different from the control group that listened to the uncompressed tapes. Furthermore, the unsuccessful students' test scores were not significantly different from the control group that did not hear the prose selections. The study has shown that the use of compressed speech has potential as a screening technique for selecting morse code trainees. It would certainly seem to warrant the expense of further research to develop a specialized test of compressed speech for this purpose.

The four tests used in the study can be regarded as equivalent, based on the performance of the control groups. However, the group mean performance scores for C_2 were well above chance scores, indicating high information load (general knowledge) within the test items. In effect, the extent to which the tests were not prose dependent reduces their usefulness in studying the relationship of comprehension to compression. Because of the high information load of the test questions, rather than having four tests each with 15 items, the tests in effect have 8 or 9 items. To discriminate between the experimental groups using tests this short is difficult. Therefore, the less conservative statistical approach in interpreting the group differences seems warranted.

In order to make these findings more conclusive and to have a test of practical value in selecting recruits for code school, a longer comprehension test using prose selections of speeds ranging from approximately 320 to 380 WPM should be developed. The longer test then should be put to empirical scrutiny to assess the utility of this method.

Student perceived intelligibility of compressed speech was a function of speed. However, this technique of measuring the effect of compression was not useful for discriminating between successful and unsuccessful trainees.

The current selection device (RC subtest of ASVAB) for morse training is a significant predictor of the final code speed students achieved. The lack of correlation between RC and test scores at level CX raises the possibility that two tests might be used in combination (multiple correlation) to produce a better selection procedure than could be obtained using either test alone.

Further research using the ability to comprehend compressed speech as an aptitude may help in selecting students for other jobs which require auditory information processing, i.e., foreign language training. Individuals who are capable of processing auditory stimuli more rapidly (or have larger auditory channel capacity) may be more apt in carrying out the cognitive functions of "translating" the second language back to their first while auditory input continues.