Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FAA-AM-76-3		

4. Title and Subtitle	5. Report Date
MASSED VERSUS DISTRIBUTED PRACTICE IN LEARNED IMPROVEMENT OF SPEECH INTELLIGIBILITY	March 1976
	6. Performing Organization Code
7. Author(s) Jerry V. Tobias, Ph.D.	8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, Oklahoma 73125	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D.C. 20591	13. Type of Report and Period Covered OAM Report 14. Sponsoring Agency Code

15. Supplementary Notes

This research was conducted under Tasks AM-D-74-PSY-51 and AM-B-75-PSY-50.

16. Abstract

Student pilots or new air traffic controllers have two ways to learn to understand the noisy and distorted communications common to aircraft operations: they may learn as they are working on other aspects of the activity, or they may learn by devoting a continuous period of time to speech-intelligibility improvement. Work completed at this laboratory indicates that two listening sessions of less than 1 hour each can accomplish the improvement and, under some circumstances, a single session will do. The current experiments are intended to reveal some of the micro-structure of this learning process. For the first experiment, subjects were divided into five groups: One was permitted 2 minutes of listening practice; another, 4; another, 8; another, 16; and the other, 32, Retests 2 weeks later showed intelligibility improvement as a function of the number of minutes of previous experience up to 16; the 2- and 4-minute subjects appeared to be no different from subjects with no previous experience on this task, 8-minute subjects were somewhat better, and 32-minute subjects were indistinguishable from ones with a full hour of experience. Subjects for the second experiment were similarly divided except that each one returned to the laboratory three times a week until every listener had heard 32 minutes of distorted speech. The five groups are nearly indistinguishable from each other, which suggests that the nervous system stores even small amounts of information gained from practice with distorted speech to use later for comparisons and improvements.

17. Key Words	18. Distribution Statement
Communication Learning Noise Speech Speech Reception	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22151

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	6	$3.00 PC .95 MF

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Acknowledgments

This work could not have been successfully completed without the valuable assistance of F. M. Irons, V. Kidd, and W. E. Morris.

MASSED VERSUS DISTRIBUTED PRACTICE IN LEARNED IMPROVEMENT OF SPEECH INTELLIGIBILITY

I. Introduction.

Distorted speech becomes more intelligible following listening experience of various kinds. Until recently, not much else could be said to describe this intelligibility-learning process; certainly no quantification of it was available. Now Tobias and Irons[5] have provided some data. Gathering such information helps both the applied and the theoretical worker; obvious applications exist for student pilots, air traffic controllers, and those who have to communicate with non-native speakers (and a less obvious application may exist for students of foreign languages); some of the theoretical implications are related to reminiscence, rehearsal, attention, memory, and motivation. A survey of all these matters is included in the Tobias and Irons paper, together with descriptions of and data from the previous experiments.

In the Tobias and Irons experiments, subjects listened to highly distorted speech for 1 hour at a time and learning curves were plotted to describe the measured intelligibility of the speech during each minute of the hour. In general, except where additional confounding factors are part of the test condition, intelligibility scores rise to a plateau in about 20 or 30 minutes. Subjects who return to the laboratory after an absence from the task (usually a week or more) start at a slightly lower intelligibility level than the one they reached at the end of the first session, but within a few minutes the scores rise to a level significantly higher than the first-session plateau. This delayed improvement suggests that, since scores increase even in the absence of further experience, subjects might also improve with far less experience in listening to difficult speech material. The present paper is concerned with two experiments designed to investigate this question. As with all the experiments on learned speech intelligibility, these two look at the way in which people can manipulate signal-to-noise ratios mentally. The signal-selection mechanisms of the listener's brain seem to search out the best method for processing speech-plus-noise, and, after a time, they produce improved intelligibility.

The ability to hear or understand signals that are inaudible or unintelligible to the untrained or inexperienced observer can be improved by listening practice. The use of such an approach for distorted as well as for masked sounds is based in the idea that both masking and distortion cover up part of the otherwise available information to make the signals "noisy." Both decrease the intelligibility of speech. The effect of an intelligibility-decreasing distortion can be nearly indistinguishable from that of masking. For example, Kryter[3] showed that the measured intelligibility of highly reverberant speech that is masked by enough noise to raise it to a level 60 dB above threshold is comparable to the intelligibility of masked nonreverberant speech raised 80 dB; the reverberation has a masking effect similar to that of an extra 20 dB of noise.

II. Method.

Sixty college-age subjects with normal hearing were divided into two groups—one to participate in each of the two experiments. Each group was further subdivided into five groups of six subjects each. No subject had previous experience with this kind of task.

Subjects were all taught to shadow[1] while listening to recorded speech. In shadowing, listeners quietly speak what they hear at the same time that they are listening to it. A person who has not tried it might assume that shadowing would be difficult, but it is really quite easy to do and is a particularly sensitive measuring tool.[4] Subjects are trained in shadowing normal, undistorted speech until they reach intelligibility scores of 95 percent or higher in five successive 1 minute intervals; then they continue training—or overtraining—for approximately 50 minutes longer. In the training period, the speech materials used are different from those that are later subjected to distortion for the experimental conditions.

The speech, both practice and experimental, is a series of easy-to-understand 120-word passages, read by a man who monitored himself during the recording sessions to maintain a constant speaking level. Later, small amounts of attenuation were inserted from passage to passage to insure that all passages were essentially equal in intelligibility in a simple masking experiment—that is, they were "homogenized" according to their masked intelligibility. Each passage is approximately 50 seconds long, with a 10–second pause between passages. Sixty passages were available for practice. Another 123 experimental passages were available with the segments spliced together in many randomized orders. A full test session comprised 3 practice passages and 54 test passages.

Because previous experiments had indicated that learning to listen to one kind of distortion is much like learning to listen to any other (in fact, transfer of learning between types of distortion is also nearly perfect[5]), only one variety of distortion was tested. This distortion is produced by replacing the original speech waveform with 0.1-msec pulses, all of the same height, whose spacing and polarity are determined by the instant and direction of axis crossing in the original waveform (see Figure 1). Pulsed speech sounds very harsh. In the earlier experiments, subjects selected their own signal levels: A 1000–Hz tone adjusted to the same peak VU-meter reading as the speech had a sound-pressure level (SPL) of 75 ±4 dB, which is within the optimum range according to preliminary tests of the relation between level and intelligibility. In the current experiment, the level was preset to 75 dB SPL. For highly trained listeners, pulsed speech is 85–90 percent intelligible.

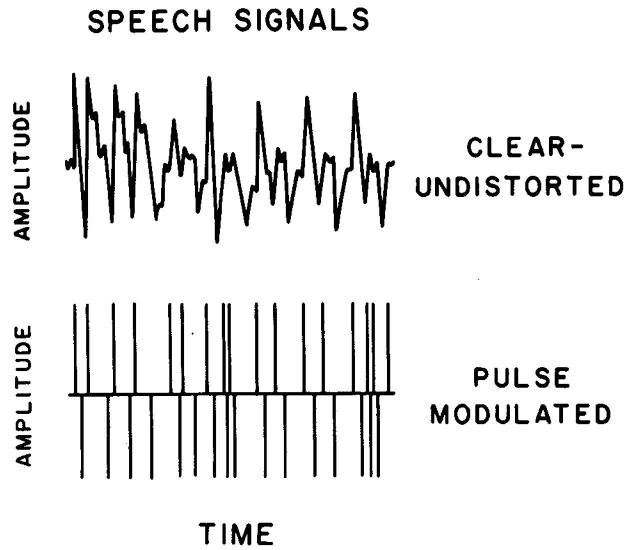

FIGURE 1. The upper curve represents a short sample of an undistorted-speech waveform. The lower curve, drawn from the same sample and subjected to the distortion used in these experiments, illustrates how the waveform information was discarded except for the time and direction of axis crossings. Each pulse has the same amplitude and width as each other pulse.

Intelligibility scores are obtained by having a trained observer score the subjects' shadowing while listening to the same signal before it undergoes the pulsed distortion and while following a written transcription of the text. The last 100 words of each 1–minute segment are scored to produce a "percent correct" number.

III. Results and Discussion.

A. *Experiment I.* Following their training period, subjects in the first experiment went right on to shadow the difficult pulsed speech. One group of six subjects listened to it for 2 minutes; another group listened for 4 minutes; another, for 8; another, for 16; and another, for 32. Then they were dismissed and scheduled to return 2 weeks later. In their second sessions, subjects were given 5 minutes of shadowing practice on clear, undistorted speech and were then tested for a full hour on pulsed speech.

Figure 2, a composite curve for all five groups, shows the first-session results. The subjects in all five groups began with intelligibility scores of about 45 percent. The 2–minute subjects improved a little in the second minute and then were dismissed. Subjects in the other groups continued to improve with additional practice. The 4-minute subjects were dismissed after their fourth minute, the 8-minute subjects were dismissed after their eighth, and so on. Predictions from the previous experiments were confirmed: Only the 32-minute subjects reached any kind of plateau.

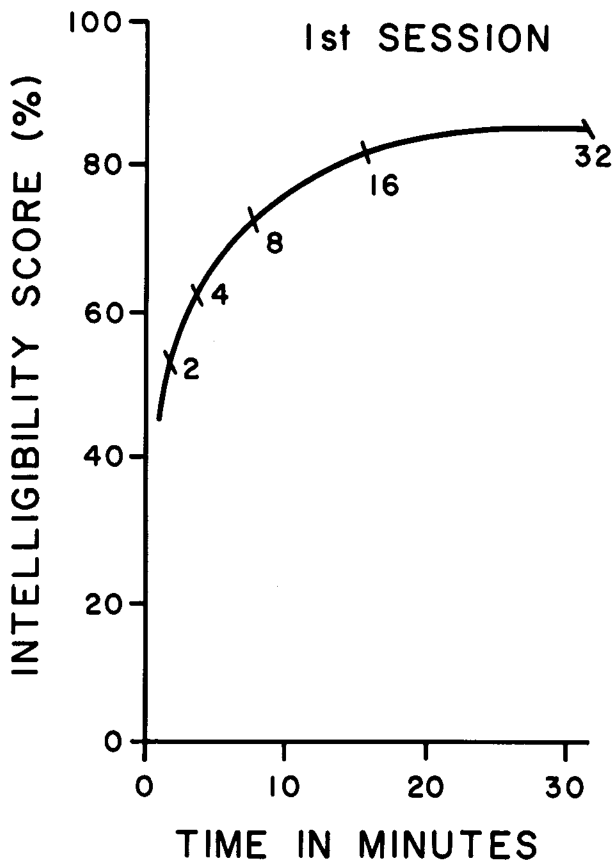

FIGURE 2. Intelligibility of pulsed speech as a function of time spent shadowing. Each mark across the curve represents the final score for one of the five groups of subjects.

Figure 3 illustrates the second-session performance of each group. The small horizontal line crossing each curve represents the average final score reached during the first session by that group.

The curves for 2-minute and 4-minute subjects are not detectably different from similar curves obtained from subjects with no previous

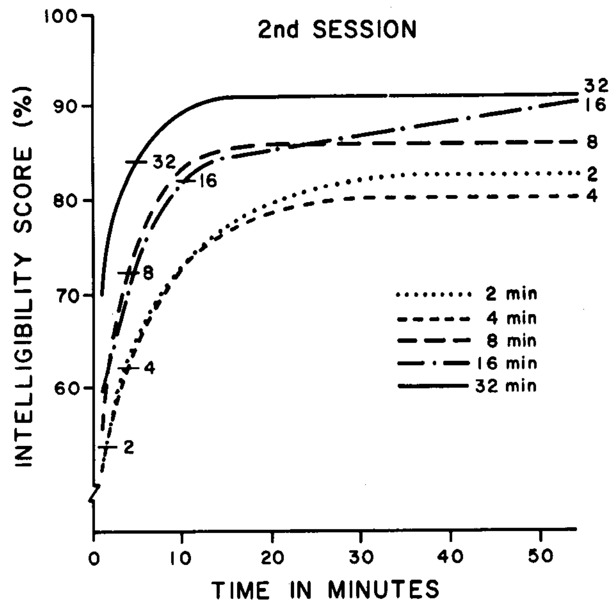

FIGURE 3. Second-session performance of the subjects shown in Figure 2. The mark that crosses each curve indicates the average final score reached by that group during its first session.

experience. Figure 4, from an earlier study, shows mean first-hours and second-hour performance for subjects whose first-session training time was 54 minutes; comparison with Figure 3 shows the similarity between the current 2- and 4-minute subjects' second-hour performance and the fully trained subjects' first-hour performance.

The curve for 8-minute subjects rises more rapidly than curves for inexperienced subjects, but the plateau value at the end of their second session is essentially the same as that for a normal, full-hour, first session. The rising slope of the second-hour curve for these 8-minute subjects is beginning to show differences from the slope for untrained listeners, though.

The 16-minute group's performance is clearly transitional: Like those of the 8-minute subjects, their scores rise for the first 10 minutes or so to an early (but otherwise normal) first-session-plateau value. However, instead of stopping there as the 8-minute curves do, they slowly continue to improve with additional second-session practice until, by the end of 40 or 45 minutes, they have reached the 90-percent second-session-plateau value expected of highly experienced listeners.

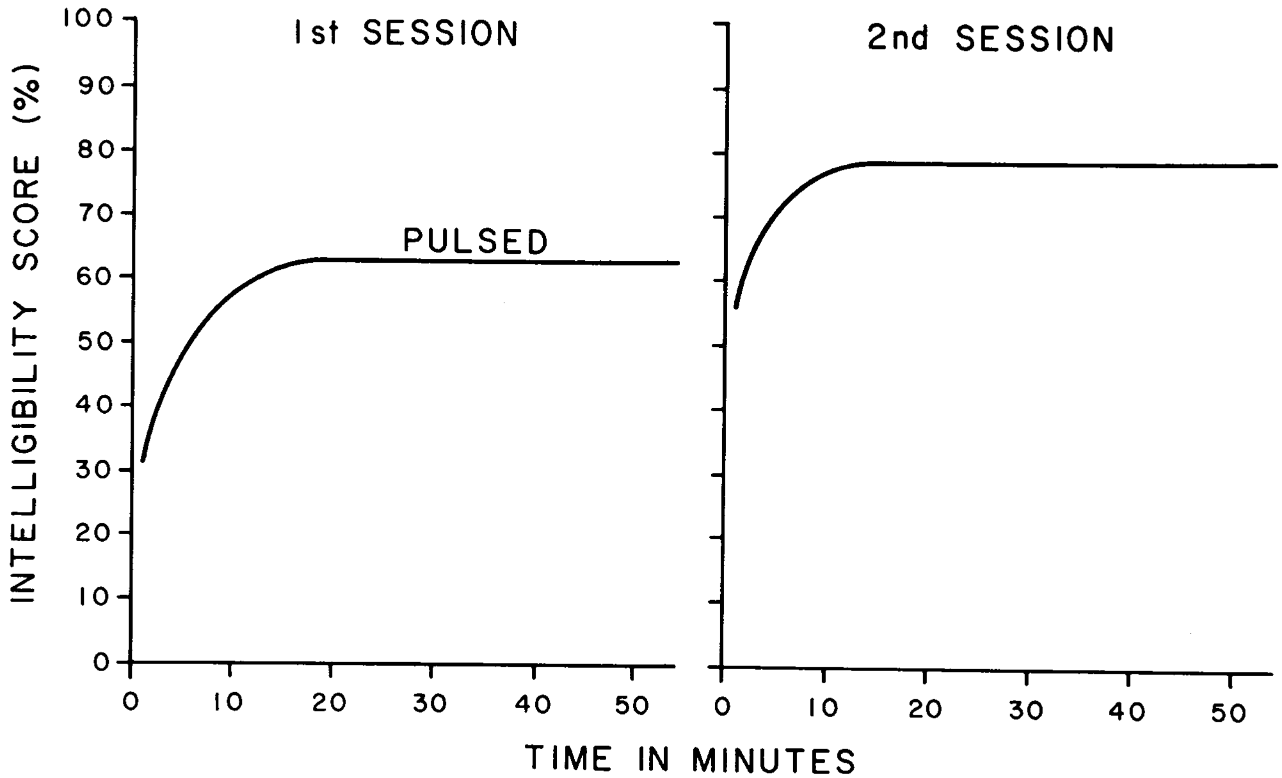

FIGURE 4. Average performance on pulsed speech for subjects who were trained for 54 minutes.

The 32–minute curve, like the 8– and 16–minute curves, rises less rapidly than curves for subjects who have listened for a full hour during the first session (see Figure 4). However, the plateau has increased to the usual second-session value, although it has done so in 10 to 15 minutes rather than 3 to 5.

B. *Experiment I Discussion.* Consider that the brain's distorted-speech analyzer may incorporate a kind of time constant such that unless it receives at least that much input information, it cannot function well. Certainly a qualitative difference in the shapes of learning curves exists for subjects who have been trained for 15 or 20 minutes rather than for much longer or much shorter periods, and the connotation of a discontinunity in the analyzer's response pattern is clear. One might conclude that the process is associated with memory such that permanent storage of the template against which a listener measures incoming distorted speech depends on his receiving a quarter hour or more of rehearsal. This experiment determines the duration of that rehearsal but does not speak to whether it must be continuous or whether it can be broken into shorter segments separated by comparatively long periods of other activity. Experiment II attempts to answer that question.

C. *Experiment II.* Following their training period, subjects in the second experiment went right on to shadow pulsed speech, just as did the subjects in the first experiment. On that first day, again as in Experiment I, one group of six subjects listened to distorted speech for 2 minutes; another group listened for 4 minutes; another, for 8; another, for 16; and another, for 32. However, instead of being sent away for 2 weeks, the first four groups were brought back sooner. Members of the 2–minute group were first tested on a Monday (or a Tuesday) and then returned each Wednesday, Friday, and Monday (or each Thursday, Saturday, and Tuesday) until they had completed 16 sessions for a total of 32 minutes of experience with pulsed speech. Similarly, members of the 4–minute group returned for eight sessions; the

8–minute group, for four sessions; and the 16–minute group, for two. When a subject completed his 32 minutes of listening, he was dismissed to return 2 weeks later for retesting. In second sessions, subjects were given 5 minutes of shadowing practice on clear, undistorted speech and were then tested for a full hour on pulsed speech.

Of the several ways available to graph the results, the most informative is to chart the intelligibility score in percent as a function of the number of minutes of listening experience, no matter how many days it took to get that experience. Figure 5 is such a graph. Except that the curve for the 32–minute group rises more rapidly, the groups are practically indistinguishable from each other during their 32 minutes of practice and during their retest following 2 weeks. The result would probably not be predicted on the basis of Experiment I.

D. *Experiment II Discussion.* Listening to distorted speech for a long enough period of time, whether that time is continuous or intermittent, allows one to improve his ability to understand that kind of speech. Whereas Experiment I demonstrates that short periods of exposure to difficult speech contribute nothing to the ability to analyze and thus to understand it, Experiment II demonstrates that even for short exposure times, a residue of information about the distortion pattern is retained, at least for a few days, to be summed with later samples of similar speech as if no hiatus had occurred. The analyzing system appears to use data from previous days as readily as it uses data from previous minutes.

IV. Conclusions.

Past experience in listening to badly distorted speech is stored by listeners for comparison with

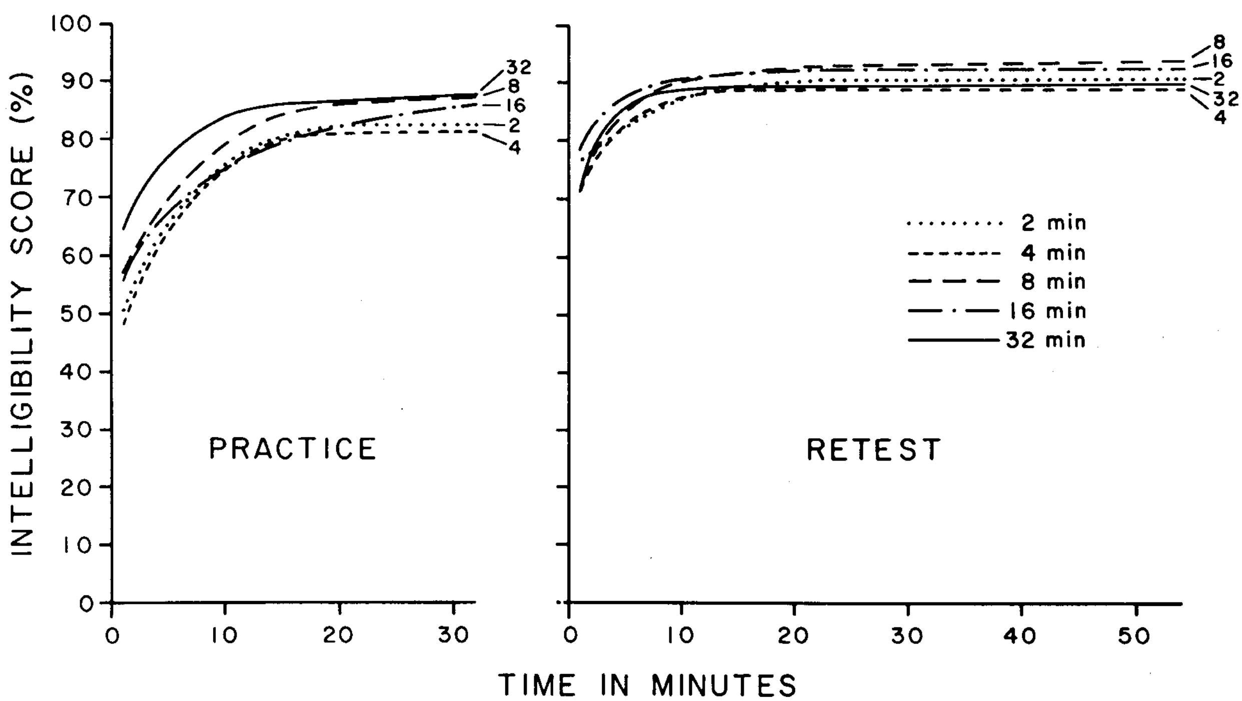

FIGURE 5. Practice and retest results of Experiment II. Each practice curve shows the result of 32 minutes of shadowing pulsed speech; the curve marked "2" represents sixteen 2–minute sessions and required about a month to complete, the curve marked "4" represents eight 4–minute sessions and required about 2 weeks, and so on to the curve marked "32," which represents one 32–minute session. Retest curves for all five groups represent a single, 54–minute session.

other difficult-to-understand speech signals. Just as Durlach[2] theorized (to explain a binaural masking process), a neural equalization and cancellation must be performed to improve the intelligibility of the difficult speech. The nervous system must somehow create a yardstick against which to gage the distortion and a replica of the distortion process that can be reversed to return the signal to its original form (as nearly as possible), once the process is well learned. A critical amount of exposure to distortion is required before the task can be done efficiently; until the listener has heard at least 15 or 20 minutes worth, the stored information on how to transform the speech into a more intelligible version is inadequate. At least 30 minutes are needed before adequate information has been stored, and for the best level of performance, either a period of rest or an especially high level of motivation is necessary.[5] The process is accelerated slightly if the original rehearsal period is an hour rather than a half hour, but the intelligibility level measured 2 weeks later differs between 30-minute and 1-hour exposures only during the first 10 or 15 minutes of testing, suggesting that the extra exposure is not critical to later performance, nor even very efficient.

In practical terms, the experiments reported here show that, although the person who learns to listen to radio communications over noisy and distorted systems generally complains that he is learning to understand speech transmissions very slowly, he really may learn as rapidly as if all his listening time were lumped together. His samples of distorted speech are infrequent, it takes many of them to add up to an hour of exposure, and the actual elapsed time may be weeks. The inverse interpretation may also be valuable: The student pilot or new air traffic controller can overcome this aspect of his communication problem by devoting an hour or two to doing nothing but listening to distorted transmissions (or recordings of them might be even more effective). Should he do so, his learning of radio-associated tasks might well be facilitated significantly.

References

1. Cherry, E. C.: Some Experiments on the Recognition of Speech, With One and With Two Ears, JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 25:975–979, 1953.
2. Durlach, N. I.: Binaural Signal Detection: Equalization and Cancellation Theory. In J. V. Tobias (Ed.), *Foundation of Modern Auditory Theory,* Volume II, New York, Academic Press, 1972.
3. Kryter, K. D.: Effects of Ear Protective Devices on the Intelligibility of Speech in Noise, JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 18:413–417, 1946.
4. Pierce, L., and H. R. Silbiger: Use of Shadowing in Speech Quality Evaluation, JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 51:121, 1972.
5. Tobias, J. V., and F. M. Irons: Reception of Distorted Speech. In W. D. Ward (Ed.), *Proceedings of the International Congress on Noise as a Public Health Problem,* 550/9–73–008, Washington, D.C., U.S. Environmental Protection Agency, 1973.