TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. FAA-AM-73-13	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle

RECEPTION OF DISTORTED SPEECH

5. Report Date

December 1973

6. Performing Organization Code

7. Author(s)

Jerry V. Tobias, Ph.D.
F. Michael Irons, M.Ed.

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.

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

5. Supplementary Notes

This research was conducted under Tasks No. AM-A-71-PSY-16, AM-A-72-PSY-16, and AM-B-73-PSY-16.

6. Abstract

Noise, either in the form of masking or in the form of distortion products, interferes with speech intelligibility. When the signal-to-noise ratio is bad enough, articulation can drop to unacceptably--even dangerously--low levels. However, listeners are capable of learning to listen to such speech and to improve their comprehension of it. In the experiments described here, the nature of this learning and the necessary amounts of time for maximum improvement are explored. The effects of several types of signal degradation are discussed, as are suggestions for training listeners to understand them. Among the processes investigated are the transfer of listening experience with one kind of signal to the analysis of another kind; the effects of actively participating in the listening task, and the results of manipulating motivation. Some inferences are drawn regarding speech learning by listeners with non-normal ears.

7. Key Words Learning Noise Speech Speech Reception	18. Distribution Statement Availability is unlimited. Document may be released to the National Technical Information Service, Springfield, Virginia 22151, for sale to the public.

9. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 9	22. Price $3.00

RECEPTION OF DISTORTED SPEECH

I. Introduction.

Noise has direct physiological, psychological, and social consequences. It also has indirect consequences that are associated with public health and that certainly are not limited to damage to the auditory physiology, to the psyche, or to the community's acceptance of loud sounds. Consider the effect of a bit too much noise on an airline pilot's reception of an air traffic control message: the physical well-being of hundreds of passengers and of unknown numbers of people on the ground can be changed by the inaccurate understanding of an instruction. A missed warning in a steel mill can produce frightening—even deadly—effects on personnel; the physiological results are not confined to the temporal bone.

We know ways to measure speech interference, and we know something about the acoustic factors that determine how well a listener will be able to understand masked speech. However, there is another kind of influence on speech intelligibility that all of us have had experience with, but that no one has measured before: people can learn to manipulate signal-to-noise ratios mentally as well as acoustically. It only requires that the listener's brain be adequately exposed to the masked signal. This exposure allows the signal-selection mechanisms to search out the best methods for processing the speech-plus-noise, and, after a time, produces greatly improved intelligibility. One of the questions that has not been answered before is how much time it takes to learn that new analyzing process.

Here is a practical illustration of what the phenomenon is. People often whistle or hum or sing while they work. Many nod their heads or tap their fingers to keep time with their music. In offices, you can sometimes see three or four people, each tapping out a different rhythm, oblivious of the tempo being strummed on the next desk. Sometimes, though, in a noisy work environment, a bizarre variation of this behavior appears: a group of employees who could not possibly hear each other's humming because of nearby loud machinery all move in time to the same invisible drummer. The first time we saw such a thing, we asked one of the workers what he was waving at. He said, "It's the music," and we pretended to understand. Of course, when we went into a storeroom a little distance from the machinery, there *was* music. Whether for morale or for entertainment or for setting a working pace, the company pumped recorded music into the factory. The workers heard it even though a visitor could not make it out above the din of the equipment.

Similar stories can be picked up from anyone who measures noise. All the anecdotes lead to the same conclusion: the ability to hear masked signals that are inaudible or unintelligible to the untrained or inexperienced observer can be improved by listening practice. The anecdotal evidence has been overwhelming. The laboratory evidence has been non-existent.

II. Methodology.

A. *Signals.* Masking and distortion are similar kinds of operations; each covers up part of the otherwise available information with extraneous matter to make the signal "noisy." They both decrease the intelligibility of speech signals. The effect of an intelligibility-decreasing distortion can be nearly indistinguishable from that of masking. For example, in a masked-speech experiment, Kryter (1946) 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 non-reverberant speech raised 80 dB. In that study, the reverberation had a masking effect similar to the effect of an extra 20 dB of noise.

A learning process permits man to overcome the change that the noisiness produces. Practicing listening to the speech without the noise (or without the distortion), however, seems not to help intelligibility much. Exposure to the noise alone or to the distortion of non-semantic signals seems not to help. But practice listening to the combination of speech and its intelligibility-destroying noise leads to rapid improvement. The available data cover studies of both masked and distorted speech; the results from experiments with one kind of signal are similar to the results from experiments with the other.

Subjects were all taught to shadow (Cherry, 1953) while listening to recorded speech. In shadowing, the listener immediately repeats every word he hears, even as he is hearing new material. Although the idea may sound difficult, subjects are quite adept at learning it, and intelligibility-test scores measured by shadowing are similar to scores earned in other kinds of tests. Indeed, if anything, shadowing is a particularly sensitive measuring tool (Pierce and Silbiger, 1972). Most subjects reach 95–100% intelligibility scores on clear, continuous speech within a few minutes. Our subjects were trained in shadowing until they had scored higher than 95% in five successive one-minute intervals.

The speech used for the speech-learning experiments was not the same as that used to teach shadowing. The experimental speech is a series of easy-to-understand 120-word passages, read by a male talker who monitored himself during the recording session in order to insure a constant speaking level. Later, slight variations in level were made from passage to passage in order to insure that all would be equally intelligible in a simple masking experiment. Each passage is approximately 50 seconds long, with a 10-second pause between passages. A total of 54 such one-minute segments was available, and the segments were spliced together in many randomized orders.

For some subjects, the passages were masked with a wide-band Gaussian noise; for others, the speech was infinitely peak clipped; for still others, the signal became a pulse train whose spacing was determined by the line-crossings of the speech wave; and finally, in one series of tests, the speech became a carrier that was amplitude-modulated by a band of noise. Subjects selected their own signal levels; for a 1000-Hz tone adjusted to the same peak level as the speech, the sound-pressure level was 75±4 dB, which was near the optimum choice according to preliminary tests of the relation between level and intelligibility. Figure 1 illustrates the kinds of distorted signals that were used. In the masked condition, the speech wave is simply added to the noise. In the modulated condition, though, a multiplication transform is used, with the effect

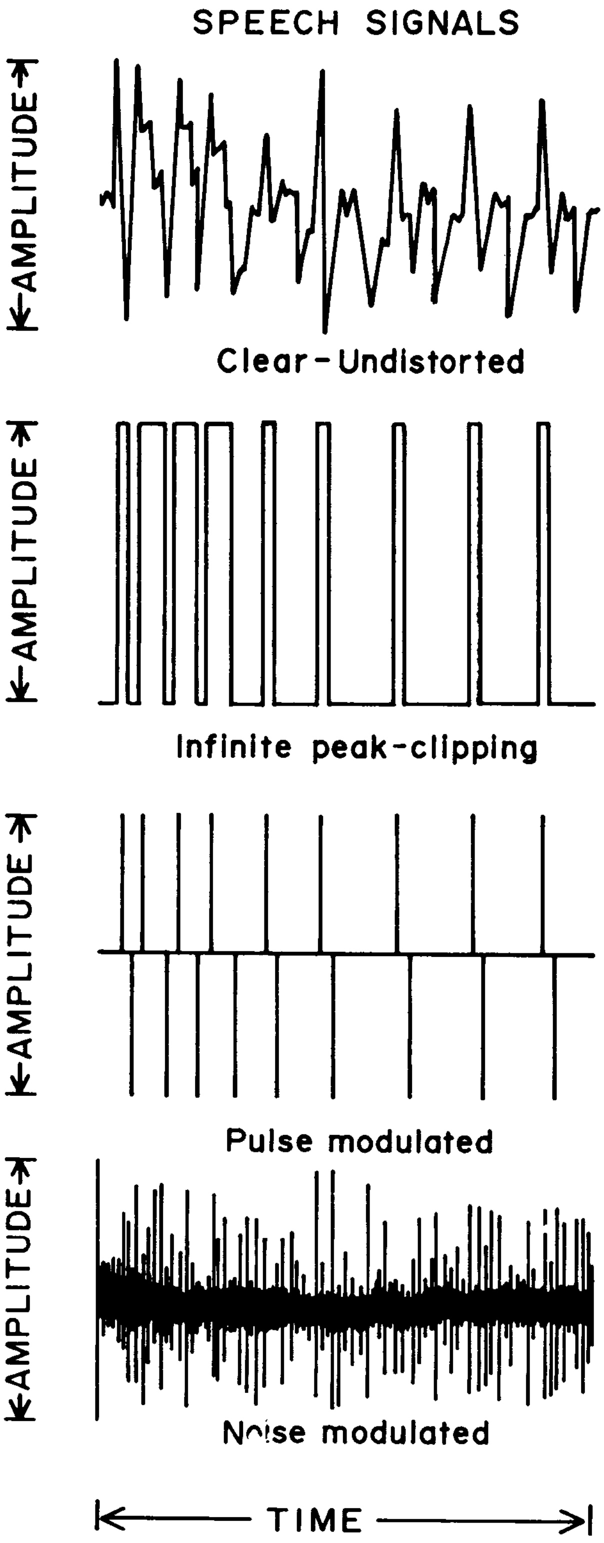

FIGURE 1. Waveforms of test signals used.

that each partial in the original instantaneous speech spectrum is replaced by a steep-skirted band of noise, 1200–Hz wide, centered on the partial. In the pulse-modulated procedure, all that is retained of the original waveform is the time and polarity of axis crossings; infinitely peak-clipped signals look to have only that same information (Licklider and Pollack, 1948), but they are generally much more intelligible (Ainsworth, 1967), even when experienced listeners adjust the levels of both types for maximum intelligibility. Clipped speech sounds harsh; pulsed speech sounds harsher. Noise-modulated speech sounds very noisy, but is generally reported to be much clearer than one would expect with "that much noise" present.

The masking level and the modulator bandwidth were selected to produce approximately the same maximum intelligibility score (80% correct) for highly trained listeners as unmasked pulsed speech does. Clipped speech is a bit easier to understand, and maxima near 90% are common.

B. *Subjects.* Each segment of these studies used six university students as listeners. All subjects had normal hearing, and none had any previous experience with this kind of task. A total of 13 series of experiments used 78 subjects. Everyone was trained in shadowing before being exposed to the distorted or masked signals. Most subjects were then simply instructed to shadow whatever they could hear. Several groups, though, received special treatment: some shadowed for a total of only eight minutes in a 54-passage session; another few shadowed everything, but were informed that they would be given a monetary incentive to do well.

III. Results and Discussion.

A. *Speech Learning.* The basic outcome of all these experiments is perfecly predictable: intelligibility starts at a low level and improves with listening practice up to a plateau value. Figure 2 shows a learning curve for each of the four kinds of signal. The rates of change are fairly similar from one condition to another, although the plateau values vary somewhat. The immediately apparent point to note about all of these data is that learning seems to be complete within 15 or 20 minutes. The auditory system makes its analysis of the signal-plus-noise, determines how to extract the maximum information, and makes whatever modifications are necessary in order to perform the extraction—and it does all that in less than half an hour. The listeners are probably not especially conscious of what they are doing in order to get this analytical processing under way; most of them report no special effort to get better, and generally they have little recollection of how well they performed.

Although the curves are similar in shape, it is inappropriate to try to draw inferences and conclusions about the speech-learning mechanism from that fact. Learning curves simply look alike. That does not necessarily demonstrate anything about similarities or differences in the analysis of modulated and pulsed-speech signals.

The curves that represent what happens in this learning mechanism do have one particularly fascinating aspect, though (Figure 3). They show that subjects returning after one or two weeks away from the task start the first couple of passages with scores slightly lower than their previous maxima, but then, almost immediately, they rise to a higher plateau than the one they attained during their first test session. The change from the first to the second plateau is statistically significant at better than the .05 level; final scores on session two are 12 to 15 percentage units above those on session one, making a total improvement that is equivalent to about an 8–dB shift in signal-to-noise ratio. During their time away from the laboratory, the subjects had no opportunity to listen to the kinds of distortions that were used, so it is unlikely that any conscious rehearsal could have influenced the results. The phenomenon is similar to the psychological entity called reminiscence (Buxton, 1943), in which some previously measured ability improves following an interval of time during which the subject is not permitted to practice. But even in that context, a question remains: why didn't subjects continue to improve throughout their first hour's practice? We may speculate that the improvement does indeed continue, but at a much slower rate—too slow to measure with this method. However, we do not know how to determine if the curves are biphasic—if they have two separate slopes.

Whatever the solution to that question, though, one thing is clear. The brain, once it has organized itself to determine the transformations necessary for analyzing difficult speech messages,

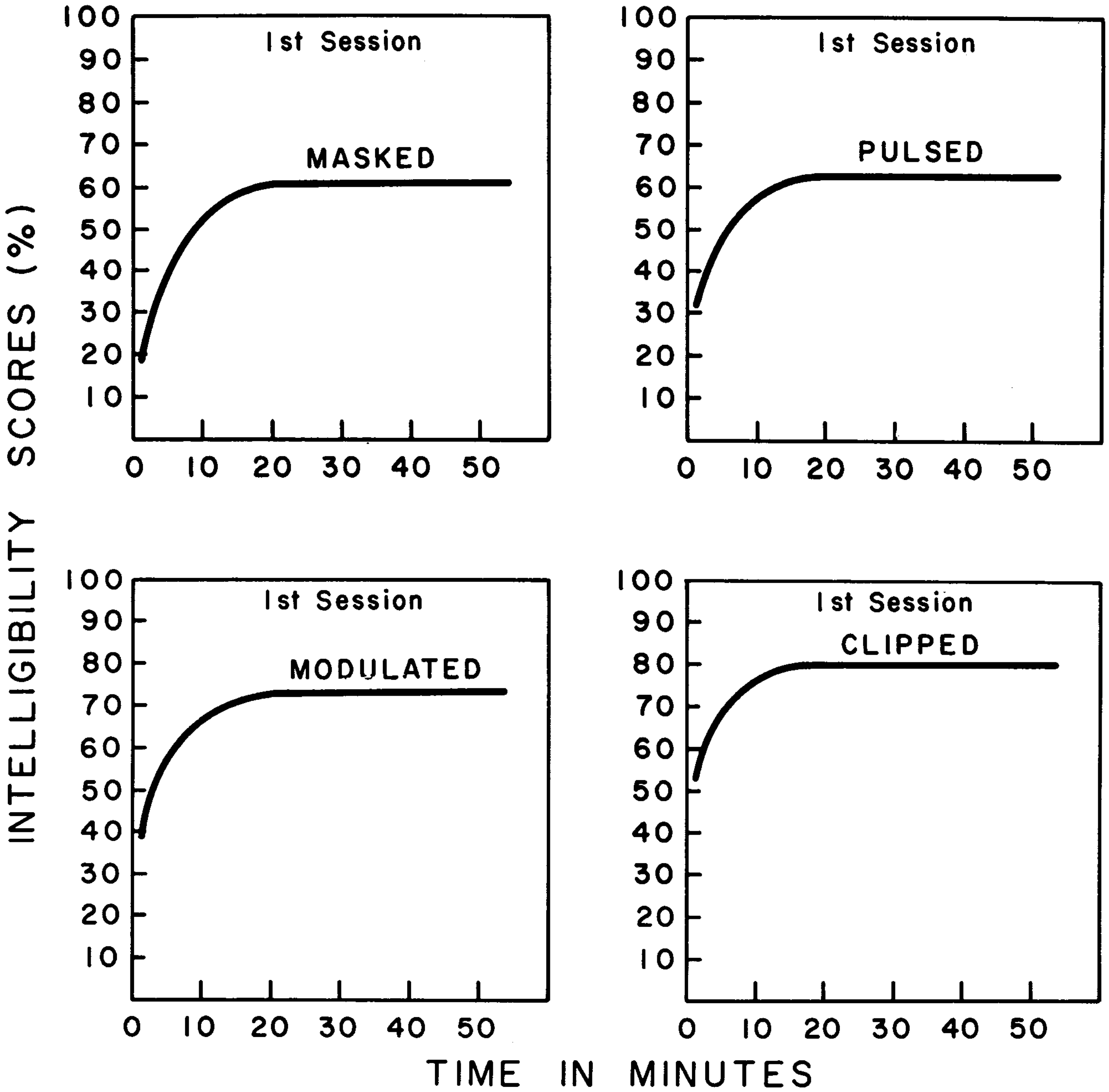

FIGURE 2. Mean learning curves for each type of signal.

continues to refine the analyzing process. These changes and refinements are the kinds that allow the listener to generalize or transfer the techniques of decoding one sort of distortion to the interpretation of other sorts.

B. *Transfer.* In tests of the transfer of speech learning, subjects were trained during the entire first session with material that had been treated with one variety of distortion. For the first half hour of the second session, two weeks later, they continued with that same distortion; by the end of the 30 minutes, it was certain that they had reached their new, higher asymptotic intelligibility score. Then, for the last half hour of that session, they were given material that had been subjected to a different distortion. For example, subjects who spent all of their hour and a half of practice time on modulated speech were tested on pulsed speech (Figure 4). Within three to five minutes, they reached plateaus that were at least as high as those reached by similar subjects who had listened to nothing but pulsed speech during both sessions. Transfer is equally good in the opposite direction.

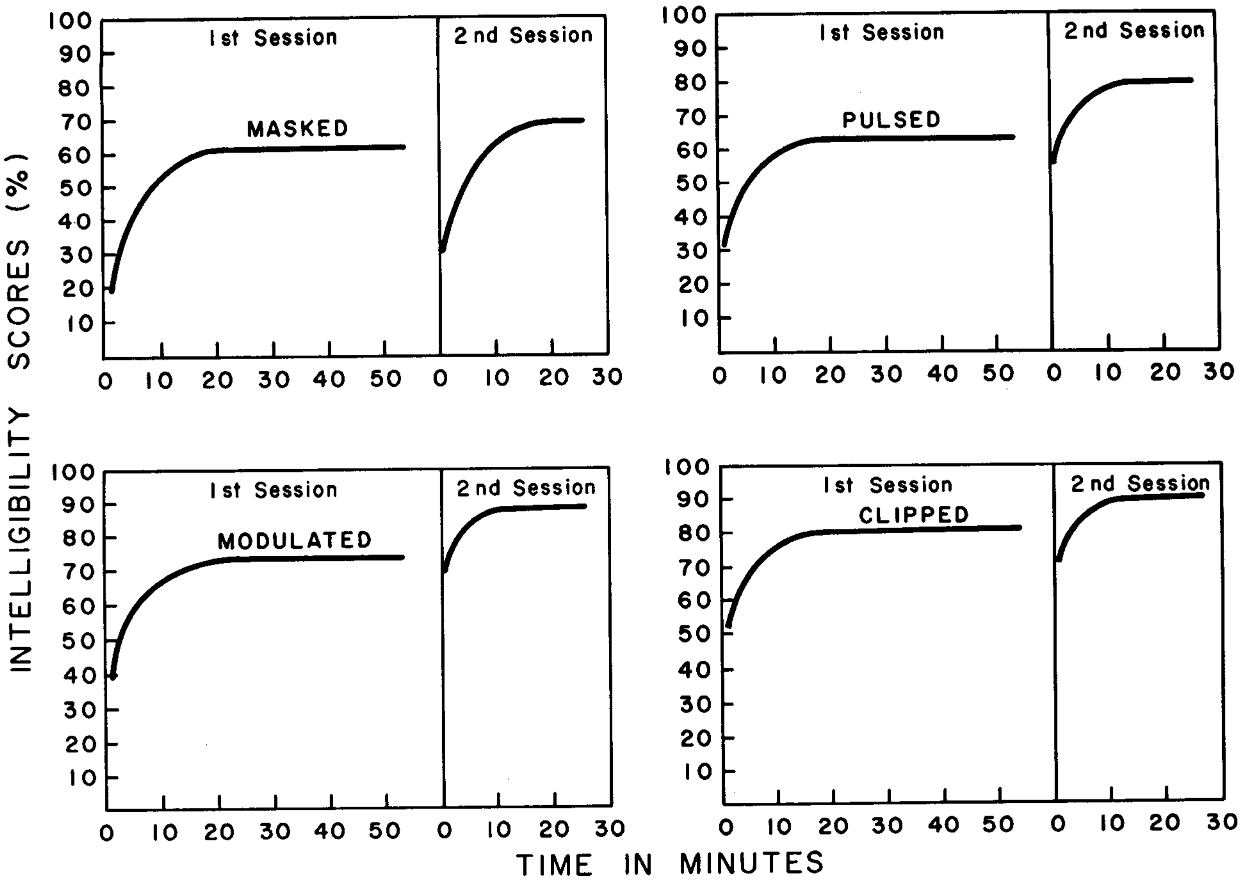

FIGURE 3. Mean learning curves for second experimental session. Note the improvement following a week or more without practice.

Although pulsed speech lacks most of the spectral information of the original waveform, and modulated speech lacks most of the temporal information, the transfer of ability from one to the other seems complete. The suggestion is strong that human observers, once they have learned how to listen to difficult speech, can successfully understand almost any form of it (for another example, see Beadle, 1970). Perhaps that idea helps to account for the fact that some people are able to understand English spoken with many kinds of dialects, but that others cannot get much intelligibility out of what is said to them by talkers with just moderately variant speech.

A practical result of this finding is that a person probably does not need to be trained to listen to distorted or masked signals that are of the precise form that will occur in his work. Once he has mastered some types of speech learning, he will be able to assimilate others rapidly.

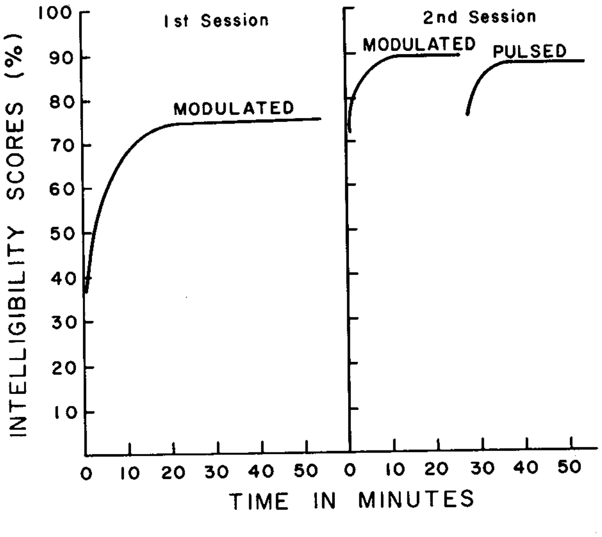

FIGURE 4. Mean learning curves for subjects who received all of their listening experience on modulated speech. When tested on pulsed speech, their scores were almost immediately at the maximum expected for subjects experienced with pulsed-speech listening.

Should we decide to make a set of recordings to train aviators to listen to radio transmissions, those records need not contain precisely the same kinds of signal degeneration that actually arise in the cockpit; anything similar—or maybe even dissimilar—ought to do as well.

C. *Passive Listening.* Must such training involve continuous speech-related activity on the part of the listener? To find out, we ran an experiment on the learning that accrues to subjects who hear the distorted signals, but who do not have to shadow them (Figure 5). If the shadowing activity is contributing to the learning, then levels of performance ought to be higher for those who are thus employed during their listening. Twelve subjects were used to test this idea. Before their first exposure to the distorted speech, they were trained in shadowing, just as all the subjects had been, using clear speech. Then they were asked to shadow the first pair, the last pair, and two equally spaced intermediate pairs of distorted passages. For the rest of the time—46 passages—they were silent. When these passive listeners' responses were compared with the responses of active, continuously shadowing subjects, two differences were apparent from the data and from the subjects' reports. First, the passive group's final first-session scores are consistently higher than the active group's (second-session curves are alike). Second, active shadowers have almost no retention of the material that they listen to, but passive listeners remember a great deal of what they hear. One interpretation of this finding is that our active listeners are giving us scores that do not represent improvements in shadowing technique, but rather real improvements in the ability to understand difficult speech signals. The higher passive scores might be interpreted to mean that shadowing somehow interferes with speech learning, and we cannot refute that possibility. However, it is possible too that the listener who can understand and retain what he listens to is better *motivated* to learn. That possibility can be tested.

D. *Motivation.* Two groups of subjects were tested—one group on masked and one on pulsed speech—but, unlike previous listeners, these were given monetary incentives to do well. After a subject had worked through the first three passages, the three scores were averaged, and he was told that for each 1% by which the 54th passage was better than that beginning average, he would be paid a bonus of $.05. Also, in order to keep him working at a high level during the entire test session, he got an additional $.10 for each passage on which the score was above 90%.

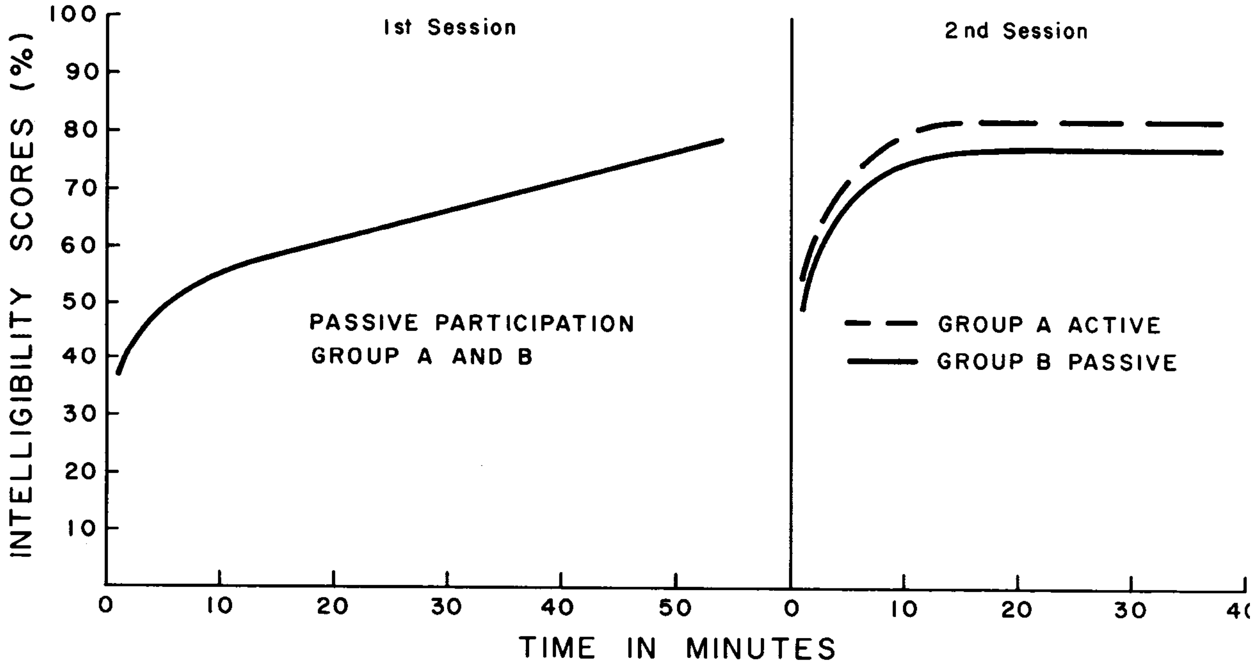

FIGURE 5. Mean learning curves for passively listening subjects. In the second session, half were asked to shadov

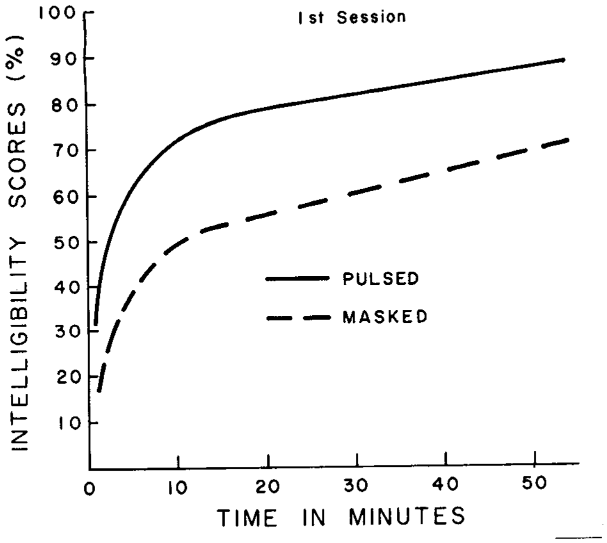

FIGURE 6. Mean learning curves for subjects who received monetary incentives.

The results (Figure 6) are similar to those for the passive listeners: curves continue to rise for a longer period of time during the first session, and, within the first hour, they reach values that are comparable to second-session plateaus. This relation between passive-listening results and motivated-listening results certainly suggests that the passive subject continues to improve because he is more interested in the task than the active subject is. He is able to relax a bit, and he can actually attend to what the talker is saying (remember that his retention is better than the active subjects).

Second-session scores for these subjects are indistinguishable from those of any other subjects. Changes in ultimate peak scores, if they occur as a result of monetary reward, are not large enough for us to measure with these techniques.

E. *Masked Speech.* Experiments with masked speech at a −3 dB signal-to-noise ratio show one kind of quantitative difference from the other experiments: first-session subjects reach two quite different kinds of asymptotes, apparently as a function of their earliest scores. Listeners who do well in the first few minutes are like most listeners; they improve rapidly to plateaus of 80% or so. But those who start with intelligibility scores of approximately 10% reach peaks in the neighborhood of only 50 or 55% (Figure 7). The intention had been to set a signal-to-noise ratio that would give a first-minute intelligibility score of 20 or 30%, but the selection was not right for these subjects; their initial scores actually ranged from 5 to 29%. The low-plateau subjects show greater variability than might be accounted for by the relatively unrestricted range in which they were working. Their learning curves rise comparatively slowly, sometimes taking 35 or 40 minutes to get the asymptotic value. The curves are unlike any others we have seen.

An explanation may lie in an evaluation of the learning experience that each group receives: the people who start off well get exposed to large numbers of correctly heard words; those who start poorly receive relatively little information that helps them in organizing an attack upon the problems of learning to understand difficult speech. Indeed, they hear very little that they recognize as being *any* kind of speech.

We cannot be certain that the apparent separation of subjects into exactly two groups is correct, although it looks to be. And we have no testable explanation for why the original differences in intelligibility exist. But it seems clear that some listeners do start off with higher scores. Are they the sort who have learned to listen to dialects, perhaps? Whatever the initial reason, after a listener has some success in pulling intelligence out of noise, he can magnify that success into higher scores. The listener who starts low may spend most of his training time struggling to hear anything at all. His decoder never gets enough samples of the difficult sounds to permit the

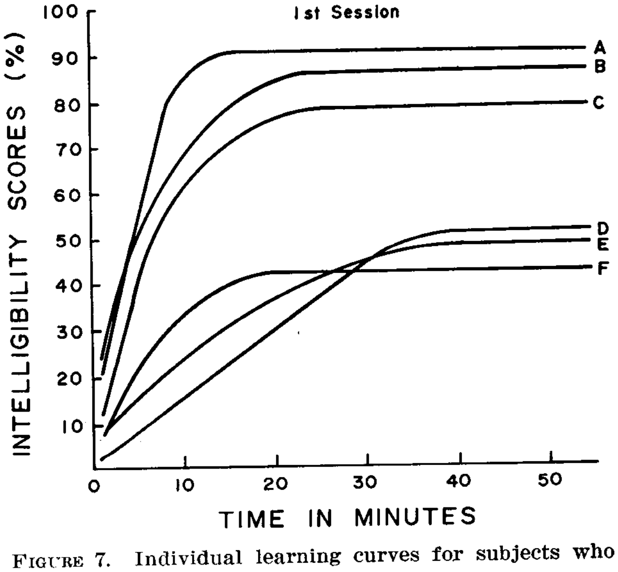

FIGURE 7. Individual learning curves for subjects who listened to masked speech.

formation of useful hypotheses about how to listen.

F. *Retention.* Most subjects were retested in a third session one month following the second (six weeks after the first). Third-session curves and scores are similar to second-session curves and scores. One group was retested after a year had passed with no known intervening practice. Their latest perfomances are similar to their second sessions, too.

G. *Non-Normal Hearing.* Normal-hearing subjects learn to understand badly mangled speech after a short period of practice. There is no evidence that people with pathological hearing do as well, but it is certainly possible—even reasonable—to conclude that they do.

The plateaus of subjects in their second sessions, no matter what the conditions of their training (except those who are able to receive only small amounts of information during the first session), and no matter what kinds of signals they were trained with, all are improved, and all look similar. They crowd together. That fact may partially explain why hearing-aid users are reported to do much better at speech discrimination after they have used an aid for a week than they did when the instrument was first tried. The early listening presents them with a kind of sound that is somehow different to them; they have to learn about it before they can get maximum sense from it. If this learning process works with some kinds of pathological ears as well as it does with normals, we might also expect to find that, for those ears, experience with *any* hearing aid will transfer to any other.

The hard-of-hearing person may have a greater problem learning to understand distorted speech than the normal-hearing listener, though, for the very reason that he cannot hear enough of the signal to work out an appropriate analysis strategy. He could be like the low-plateau subjects in the masked-speech experiments. However, usually, the overall sound pressure level of his work environment will be high enough to overcome much of the problem caused by an elevated threshold, so he can learn as well as his colleagues. This likelihood leads to the interesting possibility that the results of some audiometric tests of the ability to understand speech that is immersed in noise may be more a function of learning than of hearing.

H. *Training.* Even two minutes of listening can improve the ability to understand a talker (Peters, 1955). Six to eight hours may be needed to teach people to understand speech sounds that are transformed by a spectral inversion (Beadle, 1970), and even then, it takes longer to learn from an unfamiliar talker. But for optimum training for the reception of non-inverted speech, about an hour is needed.

How should the time be spent? If you want to improve your reception of distorted speech, it is not enough to listen to the right kind of interfering noise. It probably will not help to be exposed to non-speech sounds that are subjected to the same sorts of distortions that affect the speech; the analyzing activities of the brain are quite different for speech and for non-speech signals (Stevens and House, 1972). Student pilots take far longer than half an hour of flight time to learn to understand air traffic control communications; factory workers do not begin to understand what is said to them in noise until days of listening have passed, not minutes. Both groups commonly hear speech-plus-noise for only short moments at a time, and then return to listening to noise alone. The requirement for rapid learning, though, is that the listener be able to hear the *combination* of signal and noise at a signal-to-noise ratio that is high enough to permit him some success in interpreting the messages that are being transmitted. If his motivation to learn is high (or heightened), he can reach his maximum capacity in an hour. He will probably learn still faster if his training is done by a talker whose voice is familiar to him.* And when he has found the knack of how to listen he will keep it for a long, long time.

Once, after a static-filled, phase-distorted narrow-band, whistling short-wave radio broadcast of a concert, Sibelius is reputed to hav

* Schubert and Parker (1955) reported in a paper o a speech intelligibility study, "A puzzling phenomeno occurs with three of the subjects, who were wives c the talkers. In each case the wife exhibits only a ver slight dip in intelligibility, if any, . . . when her ow husband is the talker, but shows about the averag dip for either of the other two speakers. This ol viously falls beyond what has previously been considere the boundary of auditory theory and the authors, wh were two of the talkers, are relieved of the risk of di cussing it further." So are Tobias and Irons.

ɔointed at the radio receiver and said, "I can't ınderstand how anyone but a musician could ɛnjoy listening to that thing." In the same way, one who is not trained to listen to speech will not enjoy it. Indeed, in many circumstances, he may not even be able to hear it.

REFERENCES

1. Ainsworth, W. A.: Relative Intelligibility of Different Transforms of Clipped Speech, JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 41:1272–1276, 1967.
2. Beadle, K. R.: The Effects of Spectral Inversion on the Perception of Place of Articulation. Doctoral Dissertation, Stanford University, 1970.
3. Buxton, C. E.: The Status of Research in Reminiscence, PSYCHOLOGICAL BULLETIN, 40:313–340, 1943.
4. 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.
5. 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.
6. Licklider, J. C. R., and I. Pollack: Effects of Differentiation, Integration, and Infinite Peak Clipping Upon the Intelligibility of Speech, JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 20:42–51, 1948.
7. Peters, R. W.: The Effect of Length of Exposure to Speaker's Voice Upon Listener Reception. Joint Project Report No. 44, U. S. Naval School of Aviation Medicine, 1955.
8. Pierce, L., and H. R. Silbiger: Use of Shadowing in Speech Quality Evaluation, JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 51:121, 1972.
9. Schubert, E. D., and C. D. Parker: Addition to Cherry's Findings on Switching Speech Between the Two Ears, JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 27:792–794, 1955.
10. Stevens, K. N., and A. S. House: Speech Perception. In J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory. Volume II*, New York, Academic Press, 1972.