Sperling, G., & Speelman, R. G. (1970). Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory* (pp. 151-202). New York: Academic Press.

6

Acoustic Similarity and Auditory Short-Term Memory: Experiments and a Model

George Sperling and Roseanne G. Speelman[1]
Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey

---

[1] Now at Seton Hill College, Greensburg, Pennsylvania.

Introduction[2]

A common tenet in current theories of short-term memory is that items to be reported are rehearsed subvocally prior to report (Broadbent, 1958; Brown, 1958; Conrad, 1967; Glanzer and Clark, 1963, 1964; Sperling, 1960, 1963, 1967). If this is so, then one may expect the acoustic properties of these items—their sound—to influence the number of items that can be recalled correctly. For example, Sperling (1960, p. 21) observed that acoustic confusions (e.g., between B and D, D and T) occurred in a task where stimuli were presented visually and reported in writing, a task in which the items were never overtly represented in an acoustic form. Further experiments showed "that subjects required to memorize sequences of the letters B, P, D, T, etc., do not do as well as when confronted with an equivalent sequence of letters which do not sound so much alike" (Sperling, 1963, p. 31).

Since about 1963, the effect of the sound of verbal stimuli on the ability to recall them has been intensively investigated. The experiments fall roughly into three general classes: (1) experiments that compare memory for sequences of items that sound alike with memory for sequences that do not sound so much alike (Baddeley, 1966; Conrad, 1963; Conrad, Baddeley, and Hull, 1966; Conrad and Hull, 1964; Laughery, 1963; Laughery and Pinkus, 1966; Sperling, 1963; Speelman and Sperling, 1964; Wickelgren, 1965e); (2) experiments that study interference with to-be-remembered items by other items which may or may not sound like the to-be-remembered items (Conrad, 1967; Dale, 1964; Dale and Gregory, 1966; Wickelgren, 1965a, 1966a, 1966b); (3) experiments that are concerned primarily with analyzing response

---

[2] An account of these experiments was presented to the Eastern Psychological Association (Speelman and Sperling, 1964). An earlier version of the present article appeared as Sperling and Speelman (1967). The procedure and results sections of the article represent the joint work of both authors; the discussions, predictions, and appendices present Sperling's opinions and calculations. The research was carried out at the Bell Telephone Laboratories. The final report was prepared by the first author, much of it being done when he was a visiting staff member at the University of California, Los Angeles, in 1967–8. The authors wish to express their gratitude to Dr. Saul Sternberg and Dr. Donald Norman who contributed many helpful suggestions. Mrs. Judy T. Budiansky wrote the computer program for the graphs in Fig. 11.

errors in terms of the acoustic confusability between a stimulus item and an incorrect report of that item (Conrad, 1962, 1964; Wickelgren, 1965b, 1965c, 1965d, 1966c).

We know from these studies that a string of acoustically similar items is less well recalled than a string of acoustically different items, both with visual and with auditory stimulus presentations. From the interference studies, we know that items that sound like a to-be-remembered item impair its recall more than do items that sound different from it. From the analysis of errors, we know that a whole item, even a simple monosyllable, is not remembered in an all-or-none fashion; its constituent phonemes may be forgotten separately. We know also that items that have a phoneme in common with the correct response item are more likely to occur as errors than items which do not have a common phoneme, and that perceptual confusions between items heard in noise are similar to—but different from—memory confusions in recall.

On the other hand, there are many unanswered questions. How are results obtained with visual and with auditory presentations related to each other? Does interference act directly in memory or only at the time of retrieval from memory? (Is this a useful distinction to make?) Can the results of recall experiments with acoustically similar items be predicted from results with acoustically different items? To answer some of these questions we proceed as follows.

First, we propose a phonemic model of auditory short-term memory. The model is not sufficiently elaborated to predict performance in general, so we later add a set of empirical rules for predicting performance. We also consider in detail the role of rehearsal in maintaining items in memory.

Second, we present the results of a series of experiments in which we measured the differences in memory for acoustically similar and for acoustically different items. In these experiments, we varied the modality of stimulus presentation, the presentation rate, the number of letters presented, and the type of report required.

Third, we use the data of the experiments to examine differences in performance with acoustically similar and acoustically different items, to determine the effect of scoring for items or for positions, to compare performance with visual and auditory presentations, to estimate the capacity of short-term memory, and to assess the role of rehearsal.

Finally, we consider scoring procedures and measures of performance in greater detail. Interpretations of experimental results depend on the scoring method chosen to measure the results. For example, certain classes of interpretation (e.g., partial vs. all-or-none knowledge) require compatible methods of scoring. In other instances, use of the appropriate measure greatly simplifies interpretation. Certain of these methodological considerations have been neglected in the existing experimental literature. Therefore, in the appendices to this chapter we consider the following problems: correcting for the effect of chance guessing, correcting the observed memory span for truncation due to the span's variability, correcting estimates of capacity for the effect of response interference, how partial knowledge about items should be added into a total score, and the effect of scoring with and without regard for the serial position of the remembered item. These results are of general use, even to those who may disagree with the particular theories presented in the paper.

Phonemic Model of Auditory Short-Term Memory

Ultimately, we propose a set of rules for predicting the auditory recall data. A critical concept of the prediction rules is capacity. To provide some insight into the rules, we present first a preliminary model of memory which was developed to deal with capacity (Sperling, 1968). It was not developed further to deal with performance because performance depends on rehearsal and on response-interference, neither of which has been measured in sufficient detail to justify its use in making predictions. On the other hand, when capacity estimates are corrected for response-interference, they are found to be independent of rehearsal, so that predictions of capacity are possible even without information about these factors. Although the model does not relate directly to data, it makes some parameter predictions that can be used by the prediction rules. From capacity, the model predicts the trial-to-trial variability of capacity and the differences between capacity for acoustically similar and acoustically different items. These predictions reduce the number of parameters that need to be estimated in order to apply the rules. In addition, the model provides insight into a number of problems of short-term memory.

The Phonemic Model

BASIC ASSUMPTIONS

(1) When letters are presented auditorily (or when they are rehearsed), all phonemes of the letters are stored in memory; (2) once in memory, constituent phonemes of letters are retained or lost independently; (3) at recall, if only one of the two phonemes of a letter is available, a guess is made from among those letters of the alphabet that contain the retained phoneme in the same position (initial or terminal).

To illustrate the implications of these assumptions, let us consider specifically the actual alphabets used in the experiments of this paper. There were two alphabets each constructed of eight consonants: the acoustically similar (AS) alphabet consisted of the letters *b, c, d, g, p, t, v,* and *z* (pronounced zē); the acoustically different (AD) alphabet consisted of the letters *f, l, m, x, h, k, q, y*. A stimulus sequence was composed by selecting randomly with replacement within one of the two consonant alphabets. Thus, in the AS alphabet, retention of the final phoneme of a letter (ē) is of no value because it does not help to discriminate among the letters of the alphabet. To express the basic assumptions precisely requires some definitions.

DEFINITIONS

Let the most recent arrived phoneme in memory be designated as phoneme 1, the next-most recent as phoneme 2, etc. When letters each consist of two phonemes, the most recent letter contains phonemes 2 and 1, letter 2 contains phonemes 4 and 3, and letter n contains phonemes $2n$ and $2n-1$. Let $p(i)$ be the probability of correct recall of the ith most recent phoneme. Let f_1 and f_2, respectively, be the conditional probabilities of correctly recalling a letter given recall of only its initial or final phoneme. Let f_{12} and f_{00}, respectively, be the conditional probabilities of correctly recalling a letter given recall of both or of neither of its constituent phonemes.

A letter is reported correctly if and only if one of four mutually exclusive events occurs: (1) only its first phoneme is recalled and this phoneme leads to correct identification of the letter; (2) only its second phoneme is recalled, and it leads to correct identification; (3) both phonemes are recalled; or (4) it is guessed by chance when neither phoneme is recalled. The assumption that the memory for each phoneme is independent of memory for other phonemes means that the probability of correct recall of the nth most recent letter is given by the sum of the probabilities of these four events, namely,

$$p(L_n) = f_1 p(2n)[1 - p(2n-1)] + f_2 p(2n-1)[1 - p(2n)] \\ + f_{12} p(2n) p(2n-1) + f_{00}[1 - p(2n)][1 - p(2n-1)]. \quad (1)$$

Evaluation of Eq. 1 requires knowledge of the f's and $p(i)$s.

EVALUATION OF PHONEMIC EFFICIENCIES

When all letters of an alphabet have an equal chance of occurring, we let f_j (the conditional probability of correctly identifying a letter, given only its jth constituent phoneme) be equal to the number of different phonemes of the alphabet that occur in the jth phoneme position divided by the number of letters in the alphabet. We may consider f_j an index of the phonemic efficiency of the jth phoneme position. When $f_j = 1$, the jth phoneme position is totally efficient.

For the AS alphabet, $f_1 = \frac{8}{8}$ (all initial phonemes are different), and $f_2 = \frac{1}{8}$ (all terminal phonemes are e). For the AD alphabet, $f_1 = \frac{4}{8}$ and $f_2 = \frac{8}{8}$. For both alphabets, $f_{12} = 1$ and $f_{00} = \frac{1}{8}$. For simplicity, the double phonemes ks in x and ju in q are treated as single phonemes. The e^I phonemes in h and in k are considered not to confuse with each other because they occur in different phoneme positions.

EVALUATION OF $p(i)$

If the memory contained complete knowledge of the first n letters, and zero knowledge of the others, then the probability $p(i)$ of recalling a constituent phoneme i would be given by $p(i) = 1$, $1 \leqslant i \leqslant 2n$ for the first n letters, and by $p(i) = 0$, $i > 2n$, for the remaining letters. In this case, obviously, no difference in memory will be observed for stimuli of different alphabets. In the phonemic model, differences between alphabets result only from *partial* recall of letters. To make an exact prediction of the difference between alphabets requires knowledge of how $p(i)$ varies from 1.0 to 0.

While there is considerable basis for assuming exponential decay processes in memory strength (Atkinson and Shiffrin, 1968; Norman, 1966; Waugh and Norman, 1965; Wickelgren and Norman, 1966; and many of the models in this book), it is parsimonious to avoid the usual dual-parameter description of exponential decay, in which strength decays with one parameter and a second parameter relates strength to $p(i)$. Therefore, it is assumed simply that $p(i)$ decays exponentially with i,

$$p(i) = \int_{i-1}^{i} e^{-x/\alpha}\, dx, \tag{2}$$

(see Fig. 1). This choice of $p(i)$ yields a one-parameter description of the contents of memory in terms of the total number of phonemes α [area under $p(i)$].

The reason for defining $p(i)$ in terms of an integral instead of directly as an exponential (which would be simpler and practically equivalent to Eq. 2) is that the integral definition gives a better intuitive picture of the underlying memory. The abscissa of Fig. 1 may be regarded as space in memory, in which each phoneme occupies one unit of space. Ultimately, of course, it may be desirable to consider that different phonemes occupy different amounts of space in memory and to consider the ordinate of Fig. 1 to represent trace strength [or signal-to-noise ratio (S/N)], rather than probability directly. The advantage of the

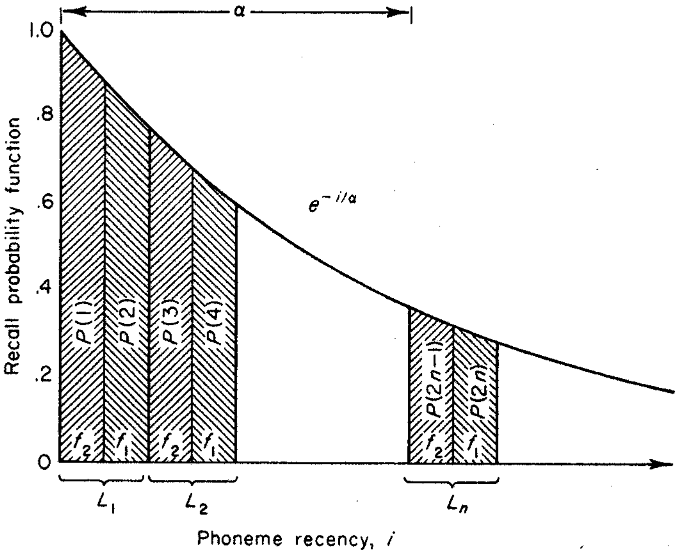

FIG. 1. Theoretical phonemic memory. Abscissa: phoneme recency. The most recently arrived phoneme is numbered 1, the next most recent is 2, etc. Ordinate: recall probability generating function. Shaded areas indicate the probability of recall $p(i)$ of the ith phoneme. The f's indicate the conditional probability of correct recall of the nth letter given correct recall of only its first (f_1) or second (f_2) phoneme. The phonemic capacity α illustrated here ($\alpha \approx 8$) is only $\frac{1}{3}$ of the observed value ($\alpha = 24$).

integral definition of $p(i)$ by Eq. 2 is that it is immediately adaptable to such complications. In any event, the exact form of the decay function is only of second-order importance; the major features of the model are phonemic representation and partial recall.

In the model, there is a strict one-to-one relation between phoneme capacity α and predicted letter capacity C. To derive C for an n-letter stimulus, it is assumed that the n letters occupy phoneme positions 1 to $2n$ of the memory. Application of Eqs. 1 and 2 to each phoneme pair and summing $p(L_n)$ gives the total predicted number of correctly retrieved letters. To compare this number with data, it must then be corrected for chance guessing (to be discussed later); the result is the predicted letter-capacity, C. For example, to anticipate some of the experimental results, if a subject has an apparent capacity of 7.31 AD letters (from a stimulus of 12 AD letters), the model requires a phoneme capacity of $\alpha = 15$. [If the post-stimulus cue (which instructed the subject) also is considered to be in memory, the required phoneme capacity is $\alpha = 24$. The post-stimulus cue is about six phonemes long, on the average, and, being presented last, would occupy the first three letter-positions in memory.]

Once the phoneme capacity α has been determined for one alphabet composed of two-phoneme letters, the model makes an exact prediction of letter capacity for any other alphabet composed of two-phoneme letters. (Generalizing the prediction to more complex stimulus sets than sets composed exclusively of two-phoneme letters would require new assumptions and is not attempted here.) The phoneme capacity, $\alpha = 24$, derived from a capacity of 7.31 AD letters implies an AS letter-capacity of 5.52 letters, i.e., a difference between the capacity for AD and AS letters of 1.79 letters. The predicted difference is virtually the same (1.72 letters) when the cue is ignored.

Further Properties of the Phonemic Model

This section deals with various interesting details and elaborations of the phonemic model. The reader who wishes to follow only the main points should skip directly to the next section.

The model is concerned only with gross similarities among *letters,* whether they have identical phonemes or not. It trivially accounts for Conrad's (1964) observation that an incorrectly reported letter tends to be from the same class as the stimulus letter (i.e., to share a common phoneme). The model is not concerned with whether *different phonemes* are similar or not, although phonemic similarity (as well as letter similarity) plays a role in short-term recall (Wickelgren, 1965d, 1966c). For example, the model makes no distinction between the alphabet (*m, n*) and the alphabet (*m, s*). On the other hand, by generalizing the model so that recall of a phoneme is based on correct recall of the relevant articulatory features, the model could, in principle, handle such data also.

The model does distinguish between alphabets on grounds other than phoneme redundancy. For example, it predicts that *b, c, d* letters are slightly more difficult (about five percent) than *f, l, m* letters because the useful terminal consonant of the *f, l, m* letters is more recent in memory than is the useful initial consonant of the *b, c, d* letters. But it probably underestimates this effect (cf. Wickelgren, 1965e).

REDUNDANCY, INTELLIGIBILITY, MEMORABILITY

There is a nonlinear relation between phonemic redundancy (number of efficient phonemes per item) and item retrievability. That is, retrieval of one efficient phoneme suffices for correct recall of an item. Additional efficient phonemes are redundant, but not necessarily useless. When one efficient phoneme is forgotten, a redundant phoneme still can support correct recall. The phonemic redundancy of digits suggests why digits heard in noise are more readily forgotten than high-quality digits (Dallett, 1964). Digits (most of which contain three phonemes) are highly redundant stimuli, and theoretically should remain intelligible even in noise that greatly reduces the probability of correctly identifying individual phonemes in isolation. In fact, an intelligibility test may be considered as the simplest memory task—recall of one item. In general, then, in simple memory tasks, redundant phonemes provide a margin of safety, a margin which is lacking at low S/N ratios. In a difficult memory task, the redundant phonemes ultimately are needed. Thus, noise digits can suffer in recall, even if originally they were fully intelligible.

More generally, the phonemic model suggests an avenue of approach to an important problem of auditory communications; namely, what character sets optimize performance at particular noise levels? The approach is to find characters such that (1) their constituent phonemes have high phonemic efficiencies and (2) the number of constituent phonemes maximizes a criterion function which weights the values of intelligibility in noise, memorability, and length of message (brevity). As the number of efficient constituent phonemes increases, intelligibility increases, brevity decreases, and memorability first increases to a maximum and then decreases. For example, *A, B, C* is an inefficient brief code; *Able, Baker, Charlie* is an inefficient intelligible code. The optimum code obviously depends on the particular performance desired.

REHEARSAL IN THE MEMORY MODEL

This section investigates some consequences of the following assumption about rehearsal: as letters are rehearsed, their constituent phonemes enter into phonemic memory in just the same way as do the phonemes of unrehearsed letters; the sequence of items in memory is determined by the sequence of external and internal inputs; recall is based on reconstruction from all images of a letter in memory; when an item is correctly reconstructed from partial information (e.g., a letter from only one of its constituent phonemes) and then rehearsed, the rehearsed copy contains the full information. Rehearsal is thus particularly effective with redundant codes. The net effect of rehearsal is to make a rehearsed item more resistant to loss, since it has to be lost at two independent locations in order to disappear from memory.

The phonemic model assumes that the only cause of loss of items from phonemic memory is the entry of new items; these push the old items out into the tail of the exponential curve. New items may be new test items or deliberate interference items (retroactive interference). Passive decay (i.e., the effect of time) may be considered as a special case of retroactive interference in which the new items are blank (silent) phonemes. Response interference may be considered a special case in which the new items are the responses. The proposition that rehearsed material in phonemic memory is more resistant to loss thus translates into the hypothesis that rehearsed items are more resistant to retroactive interference, passive decay, and response-interference.

Another consequence of rehearsal is that it is itself a source of retroactive interference. Consider, for example, a rehearsal of the three most recent letters in phonemic memory, designated as A, B, C in order of their arrival. Rehearsal of the 3rd most recent (A, first in order of arrival) causes it to occupy the first letter position. The C and B move down one position, so that B now occupies the 3rd position. Rehearsal of B moves

C to the 3rd position; rehearsal of C restores A to the 3rd position and the cycle is complete. The rehearsed letters now occupy the first three letter positions, the original letters have been moved to positions 4, 5, 6.

Let ϵ be the probability of incorrect retrieval from the 3rd position. The probability that a letter will have been rehearsed correctly then is $1 - \epsilon$. For each of the first three positions, rehearsal reduces by $1 - \epsilon$ the probability that the letter stored there is correct. Because the rehearsed copy has probability ϵ of being incorrect, rehearsal is useful only when ϵ is very small, e.g., for very recent material. It follows that rehearsing long sequences is undesirable; even rehearsal of short sequences is undesirable except when resistance to interfering stimuli is of overriding importance. These considerations are qualitatively in good agreement with our intuitions about rehearsal.

REHEARSAL AND CAPACITY OF THE MODEL

The model gives a good account of the data when the model is assumed to have a mean capacity of 15 phonemes; that is, when the area under the curve of Fig. 1 (the α of Eq. 2) equals 15. In actuality, as illustrated in Fig. 1, there are an infinite number of slots. Thus, if we were to present a stimulus ensemble of 12 letters (as we do, later on), the phonemes of these letters would only occupy the first 24 slots of the model: approximately 80% of the total area under the curve of Fig. 1. Might not rehearsal improve the efficiency of memory by filling some of the remaining slots, thereby using to good advantage some of the remaining 20% of the memory capacity? To anticipate, the answer is *no*; the probability of retrieving information from that part of memory is so low that letter storage would not be improved by more than about 5%, about $\frac{1}{3}$ of a letter.

Maximum use of capacity is achieved when the 12 stimulus letters occupy the 24 most favorable phoneme slots of memory. In general, as we have already noted, it is undesirable to rehearse long strings. Were a subject to rehearse the first group of four letters, we would find that these letters would occupy only about an additional 6.7% of the area under the model's exponential curve. After we correct for the errors that occur in the rehearsal itself, plus correcting for the fact that some of the stored letters are now duplicated in memory, we find that letter storage is increased by less than 5%.

The Experiments

Subjects

Fourteen adult employees of Bell Telephone Laboratories, seven males and seven females, served as subjects. Each subject served for five sessions of approximately one hour each, plus some shorter additional sessions for the control procedures.

Procedures

AUDITORY STIMULI

A stimulus was composed of a sequence of letters selected randomly (with replacement) within either one of two consonant alphabets: an acoustically similar (AS) alphabet (*b,c,d,g,p,t,v,z*) and an acoustically different (AD) alphabet (*f,l,m,x,h,k,q,y*). Every letter occurred in each position of the stimulus an equal number of times in each procedure. The inclusion of different letters having a common phoneme in the AD alphabet (ɛf, ɛl, ɛm, ɛks; e$^{\mathrm{I}}$tʃ, ke$^{\mathrm{I}}$, kju) is unfortunate but impossible to avoid given the restrictions of nonoverlapping sets of items and of exclusion of vowels.

The letters were spoken by one of us (RGS) in a female voice with American pronunciation without any significant regional accent. The letters were spoken in time to a metronome heard through earphones. They were recorded on magnetic tape and presented to subjects through a loudspeaker at approximately conversational speech intensity. The two alphabets were never mixed within a single stimulus. Three presentation rates were used—one letter per second (1/sec), two letters per second (2/sec) and four letters per second (4/sec). The presentation conditions of rate and alphabet were randomly varied from trial to trial.

INSTRUCTIONS

In all procedures, the task of the subject was to write, immediately after termination of the stimulus, as many correct letters as possible in their correct positions on a prepared answer grid-sheet. Subjects were told that in the main scoring procedure only correct letters in the correct positions would be scored as correct, but they were encouraged to guess because there was no penalty for guessing. The temporal sequence in which subjects wrote the letters was not controlled. At the end of each procedure, subjects were asked informally to report any particular tricks they used in performing the tasks.

PROCEDURE I: PARTIAL REPORT

Stimuli were composed of three groups of four letters (all from the same alphabet) followed by the spoken instruction "write one" or "write two" or "write three." The task of the subject was to write the particular group of four letters indicated by the post-stimulus cue. The subjects did not know until the post-stimulus cue which group of letters would be called for (cf. Anderson, 1960; Sperling, 1960). The pause between groups of four letters within a stimulus (and between the last letter and the cue) was equal to the time interval between letters at (1/sec and 2/sec) or to the time interval between three letters (at 4/sec). The sound-duration of letters was less than ¼ sec at the 4/sec rate, and increased to almost ½ sec at the 1/sec rate, due to the increased persistence of terminal vowels. Thus, the silent interval between groups of letters was equal to about ¾ sec at the 4/sec and 2/sec rates, and to about 1½ sec at the 1/sec rate. Eighteen conditions were tested: 2 alphabets × 3 rates × 3 recall instructions, giving 180 trials in all.

PROCEDURE II: RUNNING MEMORY SPAN

The number of letters in a stimulus was varied haphazardly from trial to trial, i.e., between 10 and 37 letters. Subjects were given no prior indication of the length of a stimulus and there were no intentional intonation cues to stimulus length. Subjects were asked to record the last *n* letters (as many as possible) in their correct positions on a prepared answer grid-sheet (cf. Pollack, Johnson, and Knaff, 1959). Six conditions were tested: 2 alphabets × 3 rates, 60 trials in all.

PROCEDURE IIIA: WHOLE REPORT 12

Stimuli consisted of 12-letter sequence spoken evenly without grouping by stress or intonation. Subjects were told to write all the letters of the stimulus. Six conditions were tested: 2 alphabets × 3 rates, 30 trials in all.

PROCEDURE IIIB: WHOLE REPORT 4–10

This procedure was conducted about two years after all the others. Only seven of the original subjects were still available; the unavailable seven subjects were replaced by new subjects who were matched to them for sex, approximate age, and job classification within the laboratory.

Stimuli were recorded and presented as in procedure IIIa. A stimulus consisted of a sequence of 4, 6, 8, or 10 letters spoken without grouping. A condition consisted of 12 consecutive trials with stimuli of the same length, of the same alphabet, and spoken at the same rate. Stimuli of length 10 were tested only at the 1/sec rate. Twenty conditions were tested: 2 alphabets × 3 rates × 3 lengths (4, 6, 8) + 2 alphabets × 1 rate × 1 length (10), 240 trials in all. The trials were divided into two sessions of approximately one hour each. The sequence of conditions was chosen randomly. Two different random sequences of conditions were used (different recordings) and half the subjects were tested with each sequence.

PROCEDURE IV: SIMULTANEOUS VISUAL PRESENTATION

Letters (about .30 × .35 in.) were drawn in India ink on white cardboard in a 3 × 4 array (Fig. 2). They were viewed binocularly at a distance of 20 in. in a three-field tachistoscope (Sperling, 1965). The viewing sequence was preexposure field, stimulus exposure, postexposure field. The preexposure field was white with a small fixation dot in the center. It was terminated .5 sec after the subject pressed a button to initiate the trial. The preexposure was followed immediately by the stimulus field containing 12 letters. There were three exposure conditions: the stimulus was terminated after .2, 2.0, or 12.0 sec, and immediately replaced by a postexposure field consisting of randomly scattered letter fragments (visual noise). Duration of visual noise was 2.0 sec. The luminance of the white portions of all fields was 23 ft·L. Subjects were instructed not to begin writing until after the letters were turned off.

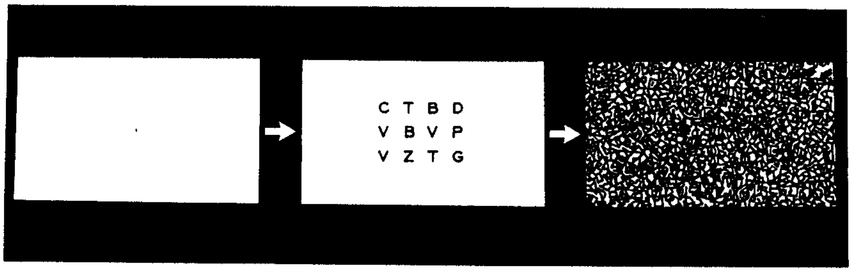

FIG. 2. From left to right, the sequence of stimuli presented to subjects in procedure IV (visual whole report). An AS stimulus is shown.

Trials were ordered in groups of five of the same condition, each group preceded by a sample trial. Six conditions were tested: 2 alphabets × 3 exposure durations, 60 trials in all.

It should be noted that in procedures I, II, and IIIa, trials with different alphabets and rates were intermixed; in procedures IIIb and IV, trials with stimuli composed of the same alphabet spoken at the same rate were grouped together. In all procedures, the various kinds of trials were counterbalanced.

Control Procedures

INTELLIGIBILITY TEST 1

Subjects were tested individually with the stimuli previously used in the partial report procedure. Each subject was instructed in advance to report one particular position from each group of four letters, e.g., the second letter. This yielded three reported letters per stimulus (of 12 letters). Every letter was tested by at least three subjects, and each subject reported 540 letters.

RESULT OF TEST 1

Errors in report occurred only with acoustically confusable (AS) stimuli, and then only when presented at the rate of 4/sec. These were therefore retested by the next procedure.

INTELLIGIBILITY TEST 2

The same stimuli were used as in intelligibility test 1, except that they were rerecorded to increase to several seconds the interval between groups of four letters. Subjects were told in advance to report only one particular letter (e.g., the second) from each group of four letters. Different subjects reported different letters, each letter being reported by at least three subjects. The subjects were told they would be given $1.00 for a perfect score and $.50 for a score with only one error.

RESULTS OF TEST 2

The subjects made a total of three errors in the 1260 letters reported. Thus, even with the most difficult stimuli, the AS alphabet presented at the most rapid rate (4/sec), intelligibility exceeds 99%. We conclude that the ability of subjects to identify any of our letters in its context is not one of the limiting factors in these experiments.

Scoring

POSITION SCORING

Except when stated otherwise, all data will be given as corrected position scores. In position scoring, an item must be reported in its correct serial position to be scored as correct. From the raw position score $\hat{R}$, a position score corrected for random guessing $\hat{S}$ is computed as follows:

$$\hat{S} = \hat{R} - \tfrac{1}{7}\,(N - \hat{R}), \tag{3}$$

where N is the total number of letters reported, $\hat{R}$ is the number reported correctly, and $\hat{S}$ is the corrected score. The cap above $\hat{R}$ and $\hat{S}$ designates that these are sample estimates of underlying population values of R and S. (That $\hat{S}$ is an unbiased estimate of S is proved in Appendix 1.) The corrected score $\hat{S}$ is derived by assuming that an observed score R is composed of S letters, which the subject knows the position of with absolute certainty, plus $R - S$ letters, which the subject writes correctly by pure chance from an alphabet of eight equiprobable letters (cf. Woodworth and Schlosberg, 1954, p. 700). In the partial report procedure, $\hat{S}$ (obtained as above) was multiplied by 3 (because the subject reported only $\frac{1}{3}$ of the stimulus letters) to obtain $\hat{A}$ the estimated numbers of "letters available."[3]

ITEM SCORING

The item score of a response may be defined as a position score obtained after the response letters have been permuted (i.e., rearranged) so as to maximize the position score. In item scoring, inversions of stimulus letters, e.g., writing YX for XY) are counted as correct. The difficulty with item scoring is that chance guessing inflates the item score far more than the position score. (See *Results* and *Appendix* 2.)

Results

MAIN RESULTS, AUDITORY MEMORY PROCEDURES

The results of procedures I, II, and IIIa are summarized in Table 1 and Fig. 3. The main result is that, in all tasks, stimuli composed of

---

[3] When N letters are presented and j are called for, the estimated number of "available letters" is $\hat{A} = \hat{S}(N/j)$, where $\hat{S}$ is the corrected score Eq. 3 for the response of length j. This calculation assumes each of the N presented letters has an equal probability of being called for in the required response.

TABLE 1
Number of Letters Reported Correctly, Auditory Memory Span, Procedures I, II, IIIa

	Procedure	1/sec	2/sec	4/sec	Mean
	Acoustically different				
I.	Partial report	7.16	7.72	7.38	7.42
IIIa.	Whole report-12	6.00	6.71	4.78	5.83
II.	Running span	4.32	3.04	2.88	3.42
	Mean	5.83	5.82	5.01	5.56
	Acoustically similar				
I.	Partial report	4.75	5.01	4.20	4.65
IIIa.	Whole report-12	4.73	4.34	3.36[a]	4.14
II.	Running span	2.81	2.85	2.08	2.58
	Mean	4.10	4.07	3.21	3.79
	Mean *AD&AS*	4.96	4.94	4.11	4.67

[a] This value represents the mean obtained after deleting data from one stimulus in which 7 of 12 letters were "*b*."

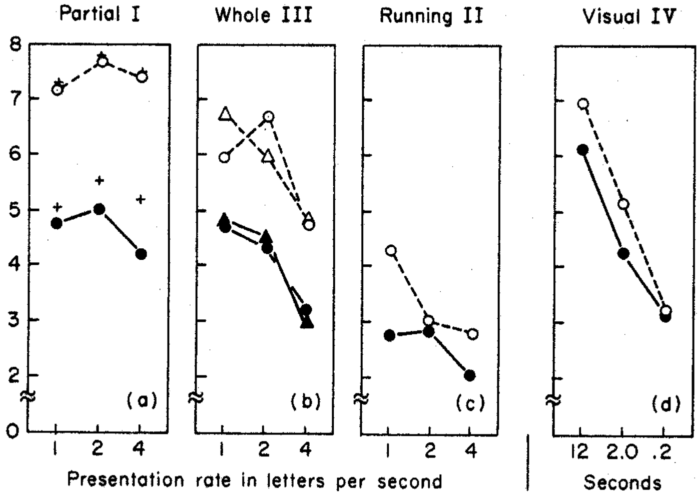

FIG. 3. Mean number of letters reported correctly in various memory span procedures. Abscissa is presentation rate in letters per second for auditory presentations (a, b, c) and exposure duration in seconds for visual presentations (d). Upper curves illustrate data for acoustically different AD letters, lower curves illustrate data for acoustically similar AS letters. The triangles in (b) illustrate averaged scores for list lengths 6 and 8, the circles illustrate scores for list length 12. The + symbols represent the memory capacity as estimated by the method described in the discussion section.

TABLE 2
Number of Letters Reported Correctly, Auditory Whole Report, Procedure IIIb

Acoustically different				
	rate			
List length	1/second	2/second	4/second	Mean
4	3.95	3.99	3.93	3.95
6	5.63	5.50	5.08	5.40
8	6.79	6.58	4.63	6.00
10	6.46	–	–	–
Acoustically similar				
4	3.76	3.63	3.24	3.54
6	4.62	4.69	3.11	4.08
8	5.25	4.43	2.98	4.21
10	5.19	–	–	–

AS letters are reported less accurately than comparable stimuli composed of AD letters.[4]

The results of the auditory experiments with lists of length 4–10 with whole report (procedure IIIb) are shown in Table 2. In procedure IIIb, unlike procedures I, II, IIIa, scores were limited in some cases by the total number of letters in the stimulus.

To see whether the change in subjects or in procedure between experiments IIIa and IIIb affected the results, list length 8 in procedure IIIb may be compared with length 12 in procedure IIIa; both are whole-report auditory tests. The mean number of letters reported correctly in the two AD conditions are 6.00 and 5.83, and in the AS conditions, 4.21 and 4.32, respectively (see also Fig. 3b). These comparisons indicate that the level of performance remained substantially the same after replacement of seven of the original 14 subjects, and the effect of alphabets is substantially the same in both conditions.

ITEM SCORES

The scoring procedure used above was *position scoring:* a response letter was correct if and only if it was the same as the letter in the cor-

---

[4] An analysis of variance showed all main effects and three interactions to be highly significant. For details, see Sperling and Speelman (1967).

responding position of the stimulus. In *item scoring*, all correspondences between stimulus and response items are counted, irrespective of their position. Whole report data of procedure IIIb were scored by both methods, and an analysis of the difference between the observed position and item scores is given below. The differences between the item scores and position scores are illustrated in Fig. 4. The abscissa represents stimulus length minus the mean corrected position score for each condition. The abscissa thus indicates the mean number of opportunities for item guesses, i.e., the mean number of letters by which the stimulus exceeds the corrected position score. The ordinate represents the mean difference between the uncorrected item score and corrected position score for each type of stimulus, i.e., the guessing success.

The data of Fig. 4 clearly indicate that the difference between item scores and position scores increases almost linearly as a function of the mean number of item guesses, and that the same relation obtains for all presentation rates, stimulus lengths, and alphabets. For example in almost every task, data for AD and AS stimuli differ less from each

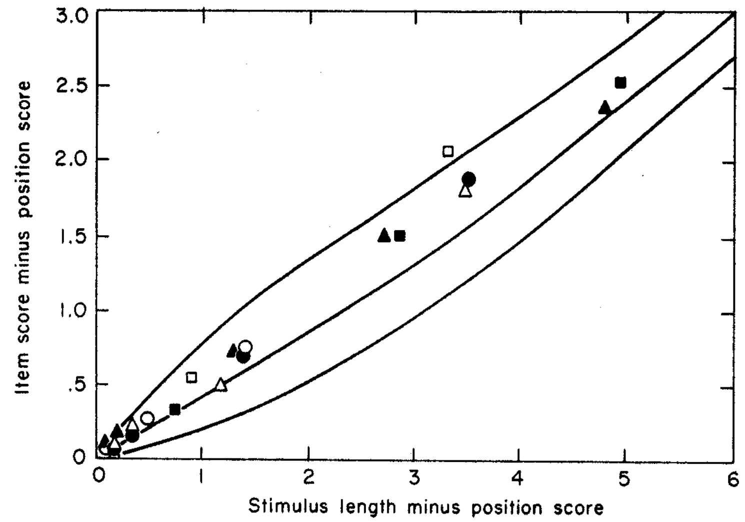

FIG. 4. Difference between item scores (independent of position) and position scores in the whole report procedure. The abscissa represents the opportunity for guessing items, i.e., the difference between number of presented letters and the position score. Unfilled and filled points indicate AD and AS alphabets, respectively; triangles, circles and squares indicate rates of 1, 2, and 4/sec, respectively. Theoretical curves were obtained by assuming that the true difference I between the item-span and the position-span was $I = 0$, 1, and 2 items (lowest, middle, and highest curves, respectively).

other in their item score than in their position score. Figure 4 illustrates that the differences between the item scores for the two alphabets is predicted by the difference between their position scores and does not reflect any new information.

THE EFFECT OF ITEM KNOWLEDGE

Let the mean position memory span $\bar{m}$ be defined as the *position* score that would be observed in the absence of random guessing and the item memory span $\bar{m} + I$ be defined as the *item* score that would be observed in the absence of chance guessing. (Note: $\hat{S}$ in Eq. 3 estimates $\bar{m}$.) The lowest curve drawn in Fig. 4 represents theoretical predictions based on the null hypothesis; namely, it represents the difference between item scores and position scores that would be expected (due to chance guessing) if the item memory span were identical to the position memory span. (The position memory span is assumed here to be distributed normally across subjects with a standard deviation of one letter—see Discussion section.) Although the lowest curve indicates that chance guessing accounts for much of the observed difference between item and position scores, it also indicates that the null hypothesis is untenable. The two curves that bound the data of Fig. 4 were derived by assuming that the amount I by which the item span exceeds the position span for the same task is 1.0 letters and 2.0 letters.[5] Nineteen of the 20 data points fall between these two hypotheses; the one exception is the AD, 4/sec, length-8 condition, where the item span appears to exceed the position span by slightly over 2.0 letters.

If it is assumed that the item span exceeds the position span by $1\frac{1}{2}$ letters (on the average), then inspection of Fig. 4 shows that no predicted item score would be in error by more than about $\pm\frac{1}{4}$ letters—most predictions would be considerably better. We conclude that in these experiments item scores can be predicted from position scores. Because item scores can be derived from position scores, we confine our subsequent analysis to position scores. The relatively frequent presence in memory of two or more items whose positions are not known also provides a simple account (perhaps more plausible than

---

[5] When a subject's true item span exceeds his true position span by I items, the observed position score—even corrected for chance guessing—may be greater than the true position span. This occurs because the probability of the I items being written in their correct positions (and thus being scored correct by position scoring as well as item scoring) is greater than the probability of a purely random guess being correct. It can be proved that the inflation of the position score by item knowledge does not exceed 1.0 letters. A correction for the inflation of observed position scores by item knowledge is incorporated in the theoretical curves of Fig. 3 (see Appendix 2).

that of Conrad, 1965) for the frequently observed transposition errors (e.g., reporting YX for XY).

MAIN RESULTS, VISUAL MEMORY SPAN (PROCEDURE IV)

At the conclusion of the visual test, each subject was questioned about the strategy he used to remember letters, i.e., whether he rehearsed them in a rote manner or whether he looked for patterns among the letters, formed associations, or tried some other form of coding. For the .2- and 2.0-sec exposures all subjects concurred; they reported rehearsing the letters in a rote, sequential way. During 12-sec exposures, however, subjects differed in their strategy. Based on their replies, subjects were divided into three groups for separate analysis: five subjects who reported using rote rehearsal exclusively (**R**), six subjects who reported using some more complicated form of grouping or coding (**C**), and three subjects who reported both, i.e., a mixed strategy. Typical coding strategies were looking for and noting the location of repeated letters and of familiar letter combinations, e.g., BCD (binary coded decimal), BVD (a brand name), etc.

Table 3 gives the results of the visual memory experiment. Scores of rehearsing, mixed-strategy, and coding subjects are roughly similar at .2- and 2.0-sec exposure durations, and at the 12-sec exposure for

TABLE 3
Number of Letters Reported Correctly, Visual Whole Report, Procedure IV

		Acoustically different Exposure duration			
Subjects	*n*	12.0	2.0	.2	Mean
Rehearsing	5	7.26	5.07	2.99	5.11
Mixed	3	6.28	4.83	3.05	4.72
Coding	6	7.13	5.53	3.63	5.43
Mean	14	6.99	5.21	3.27	5.16
		Acoustically similar			
Rehearsing	5	5.18	4.28	2.76	4.07
Mixed	3	5.51	3.79	3.06	4.12
Coding	6	7.33	4.74	3.67	5.25
Mean	14	6.17	4.37	3.21	4.59

AD letters. At the 12-sec exposure duration, rehearsing subjects show a 2.1 letter deficit for AS letters, while coding subjects actually score higher with the AS than the AD stimuli. [A *t*-test for the difference between coding and rehearsing subjects at 12-sec exposures gives a t of 2.92 ($p < .01$).]

Discussion of the Experimental Results

Comparisons of Performance with Visual and Auditory Stimuli

Results of visual and of auditory whole report procedures are compared in Fig. 5. The abscissa represents the average number of letters the 14 subjects reported correctly with AD stimuli. The ordinate represents the AS deficit in each task; it is the AD score minus the AS score. In the visual data in Fig. 5, brief exposures are represented at the far left and long exposures at the far right. Figure 5 segregates visual data, indicating separately the visual data of coding and of rehearsing subjects. (Visual data of three mixed-strategy subjects are not represented in Fig. 5; they can be computed from Table 3. The performance of mixed subjects is intermediate between that of rehearsing and of coding subjects.)

We make three general observations (with the exceptions noted below) from Fig. 5. (1) AD scores and AS deficits are strongly correlated. (2) Visual AS deficits are similar to auditory AS deficits at rates of 1/sec and 2/sec but different from auditory AS deficits at rates of 4/sec. (3) Increasing the duration of visual exposures increases visual AS deficits in the same way that increasing the number of letters presented in an auditory stimuli (at rates of 1/sec and 2/sec) increases auditory AS deficits. Varying visual exposure duration does not effect AS deficits in the same way as varying auditory presentation rate. These observations are fully consistent with the visual-to-auditory encoding hypothesis (Sperling, 1963, 1967). The encoding hypothesis states (1) that subjects rehearse letters from a visual presentation at a rate of less than 4/sec and (2) that subsequent memory for the rehearsed letters ultimately is limited by the same factors as if the letters originally had been presented to the ears instead of to the eyes.[6] This hypothesis is considered in more detail later.

---

[6] Additional support for the encoding hypothesis is obtained from data of some individual subjects. Individual data usually exhibit regularities when they are graphed as in Fig. 5, although subjects vary widely. For some subjects, the data for the visual whole report fall exactly between (*a*) data of the auditory whole report at the rate of 2/sec and (*b*) auditory data at the rate of 4/sec.

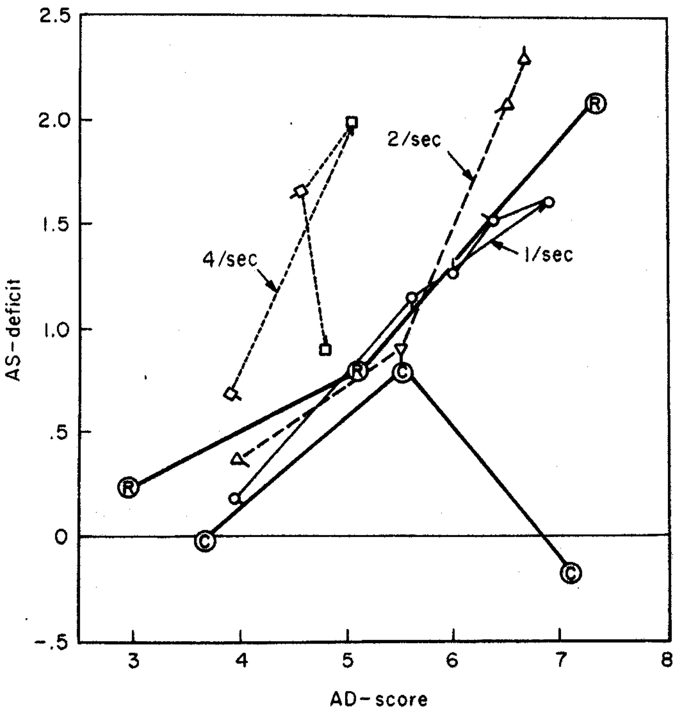

FIG. 5. Comparison of visual and auditory procedures. AD minus AS scores (AS deficits) are graphed against the AD score. Letters (inside circles connected by heavy solid lines) indicate data from the visual whole-report procedure; all other symbols represent data from the auditory whole-report procedure. R and C indicate scores of rote-rehearsing and of pattern-coding subjects in the visual tasks. The exposure duration is shortest for points at left, longest for points at right. Auditory data are the averages of all subjects; the circles, triangles, and squares represent rates of 1/sec, 2/sec, and 4/sec, respectively. Points representing the same auditory presentation rate are connected, and the rate is indicated. The small pointer emanating from each point indicates the length of the presented list. (The pointer should be interpreted like the hour hand of a clock.)

MEMORY WITHOUT AS DEFICIT

The coding-subject group is a remarkable exception to the equivalence of auditory and visual AS deficits. At visual exposures of .2 sec and 2.0 sec their performance is similar to the rehearsing subjects. However, at 12-sec exposures (where rehearsing subjects show the greatest AS deficit) coding subjects show no AS deficit. Obviously, their mnemonic process of recalling patterns and distinctive combinations of letters does not depend on the sound of stimulus letters in the same way that a rote rehearsal of stimulus letters does. (The AS letters are more frequent in the English language than the AD letters and this may facilitate a coding process based on complex associations.)

That rote-rehearsing and pattern-coding subjects perform similarly at 2.0-sec exposure durations indicates that pattern-coding, a time-consuming mnemonic aid, is not attempted at brief exposure durations. It should be noted here that coding and rehearsing subjects were not segregated on the basis of their data, but on the basis of statements they made about their coding strategy at the conclusion of the experiment, prior to data analysis. (For a similar segregation of subjects on the basis of reported strategies, with substantial observed differences in performance, see also Harris and Haber, 1963.)

ANALYSIS OF OTHER DATA

The analysis illustrated in Fig. 5 (AS-deficits versus AD-scores) may be extended to a study by Laughery and Pinkus (1966). They presented subjects with auditory stimuli of length 6 and 8 as in our procedure III. The main difference was in their selection of presentation rates (⅓/sec, 1/sec, 3/sec). Laughery and Pinkus also presented stimuli visually at the same rates. Unlike our procedure IV (simultaneous visual presentation), in their procedure each new letter was superimposed upon and replaced its predecessor. Our analysis of their published data is illustrated in Fig. 6.

Laughery and Pinkus's subjects have lower average scores than ours. To facilitate comparison between their study and ours, we selected the seven subjects with the lowest scores in one of our whole-report experiments (list length 8, presentation rate of 2/sec, AD alphabet) and added their data to Fig. 6.

Inspection of Fig. 6 shows that Laughery and Pinkus's data obtained both visual and with auditory presentations at rates of 1/sec and 3/sec are comparable to the data obtained from our low-scoring subjects at rates of 1/sec and 2/sec. This means that to a first approximation in their experiments, as in ours, AS deficits are predictable from AD scores, independently of the modality of presentation, of the presentation rate (1/sec to 3/sec), or of the length of the presented list (4–12 letters).

Results and Conclusions about Presentation Rate

SLOW PRESENTATION RATES

At slow rates, presentation times can be strikingly long. For example, at the ⅓/sec rate, a six-letter stimulus requires more than 15 sec for

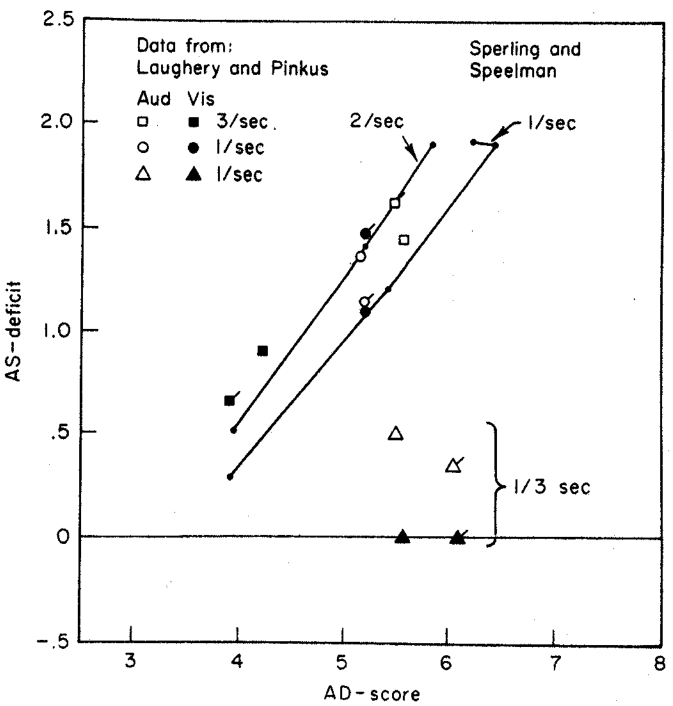

FIG. 6. Data from Laughery and Pinkus (1966). Coordinates same as Fig. 5. Stimuli were of length 6 and 8; length 8 is indicated by short diagonal strokes. Unfilled and filled points represent AD and AS alphabets, respectively; squares, circles, and triangles represent rates of 3, 1, and ⅓/sec, respectively. The solid lines represent Sperling and Speelman's data obtained with lists of lengths 4, 6, 8, and 10. These data are the same as those of Fig. 5, except that they are taken only from the lowest-scoring 7 (of 14) subjects.

presentation, and an eight-letter stimulus requires more than 21 sec The eight-letter stimulus at 3/sec requires only 2.6 sec. It is not so surprising, then, that Laughery and Pinkus's data for the ⅓/sec presentation rate are quite different from their other data, showing no AS deficit for visual presentations and only a slight AS deficit for acoustic presentations.

In our procedures, recall without AS deficits occurred only at very brief visual exposures (where recall presumably was limited by visibility rather than mnemonic factors) and at long visual exposures for coding subjects.

From Laughery and Pinkus's auditory and visual data with ⅓/sec stimuli, and from our visual data with the 12-sec stimulus, we conclude that subjects can recall without AS deficits, but that such a mode of recall requires more time per stimulus item. As a rule of thumb, the time required is about 2 sec per letter in successive visual or auditory presentations, and about 2 sec per correctly reported letter in simultaneous visual presentations. From the coupling of recall-without-AS-deficit to pattern coding in our visual task, we infer that the time-consuming process in tasks that fail to show AS deficits is some form of pattern and/or association coding. The mechanisms underlying the selection of patterns to code, the formation of associations, the storage and retrieval of patterns and associations are far more complex than the mechanisms underlying time-limited rote rehearsal. We are content merely to characterize the conditions under which pattern coding can occur and to note that it is beyond the scope of this chapter.

FAST PRESENTATION RATES

In the graph of AS deficits vs. AD scores (Fig. 5), the 4/sec data fall about one letter to the left of 1/sec and 2/sec data. Our interpretation of this finding is that the 4/sec rate produces an overall performance decrement of about one letter (compared to 1/sec and 2/sec) and that this decrement is independent of acoustic similarity. Later, this observation is given a quantitative interpretation.

MODERATE PRESENTATION RATES

For auditory presentation rates of 1/sec to 3/sec (list lengths of 4–12 letters), for successive visual presentation at 1/sec to 3/sec, and for simultaneous visual presentation at exposure durations of .2 sec—2.0 sec (12.0 sec for rote-rehearsing subjects), the data fall approximately on the same straight line on a graph of AS deficits vs. AD scores (Figs. 5 and 6). This colinearity is interpreted to mean that only one significant underlying factor varies as the many presentation parameters vary. Because the data derived from lists of different lengths, particularly short lists, fall in such a regular order on this line, we propose that the underlying factor is the effective length of a list, i.e., the number of items from the list that the subjects attempt to maintain in memory. For example, in simultaneous visual tasks, the number of effective items is the number of items the subject rehearses. In short auditory lists, all items are effective items. In long auditory lists, we suppose that the effective items are those which are rehearsed plus those at the end of the list.

Although we cannot yet measure the number of effective items directly, we assume that the AD score is a monotonic indication of their number. The high correlation of AS deficits with AD scores is assumed to result from an underlying dependence of both AS deficits and AD scores on the number of effective items.

The stimulus parameters of list length, presentation rate, modality, (but not alphabet), are assumed to determine the number of effective items in the list, as indicated by the AD scores. Because AS deficits depend mainly on AD scores (effective items) and do not show any strong additional dependence on list length, modality, exposure duration, or rate, it follows that one need postulate only one kind of memory to account for all these results. The effective items from all these kinds of lists ultimately are stored in this memory which, because of its large AS deficits, may reasonably be called auditory short-term memory.

Capacity of Auditory Short-Term Memory

In this section, we draw a distinction between the *capacity* of a memory—its maximum possible contents—and the observed *performance* (e.g., whole-report score) under particular conditions. Partial-report procedures give the best estimates of the capacity (see Sperling, 1960). The main advantage of the partial- over the whole-report procedure is that only a few items are called for in a partial report. If reporting these few items interferes with retention of the remaining items, it is no matter, the remaining items are not called for. A second advantage of the partial report is that it samples memory at a more precisely defined point in time than does a whole report, simply because it takes less time to retrieve and to report a few items from memory than it takes to retrieve and to report many items.

When the partial report requires more than one item, then retrieval and reporting of the first item or items, will impair recall of remaining items within the report group. Unless the partial report be limited to a single item, its advantages are only relative, not absolute. By recognizing the inadequacy of the partial report procedure, we can make a correction for it and obtain an improved estimate of memory capacity.

A procedure for estimating a correction factor f is stated and justified in Appendix 1. Basically, f is derived from data of whole reports of four-letters lists. With such short lists, the limiting factor on performance is response-interference, not memory capacity, and therefore f corrects for response-interference. In fact, f is never very different from one here ($1.005 \leq f \leq 1.236$), so that any possible errors in f produce only second-order errors in estimates of capacity.

TABLE 4
Estimates of the Capacity for Letters of Auditory Short-Term Memory

	Rate			
Score	1/sec	2/sec	4/sec	Mean
AD partial report score	7.16	7.71	7.38	7.42
Correction factor	1.014	1.005	1.019	
Est. *AD* memory capacity	7.27	7.75	7.51	7.51
AS partial report score	4.75	5.01	4.20	4.65
Correction factor	1.064	1.105	1.236	
Est. *AS* memory capacity	5.05	5.54	5.19	5.26
Partial report deficit (*AD–AS*)	2.41	2.70	3.18	2.76
Capacity deficit (*AD–AS*)	2.22	2.21	2.32	2.25

ESTIMATES OF MEMORY CAPACITY

Table 4 gives: (a) the partial report scores (numbers of letters available corrected for chance guessing, *A*, from Procedure I, Table 1); (b) the estimated correction factors *f* for response-interference; and (c) the estimates of memory capacity obtained by multiplying (a) and (b). Table 4 also gives the AS deficits of partial reports (AD score minus AS score) and the AS-capacity-deficits (AD capacity minus AS capacity).

The most significant aspect of Table 4 is that estimates of capacity (even uncorrected partial report scores) are relatively independent of presentation rate (whereas measures of whole-report and running-span performance depend strongly on rate). In estimating the total capacity of auditory short-term memory, it must be remembered that the post-stimulus cue was an auditory stimulus, e.g., "write two." The cue was correctly interpreted in every instance, indicating perfect memory for it. We may conclude that short-term memory can contain 7.5 AD letters plus one post-stimulus cue; this capacity varies ±¼ letter depending on presentation rate. Memory capacity for AS letters is smaller by 2.25 letters, independent of rate.

The Role of Rehearsal

When stimulus letters are spoken at a slow rate (e.g., slower than 1/sec), or when delays are interposed before report, subjects rehearse letters during the silent intervals. According to the phonemic model, rehearsal occurs because memory decays with time (silent phonemes)—

even when there are no external disturbances—and rehearsal can replace the decaying trace. Rehearsal may have functions for storage in long-term memory, but these do not concern us now. The present questions are: (1) Are rehearsed letters stored in the same memory as unrehearsed auditory letters? and (2) Does rehearsal increase the capacity of short-term memory?

If the answer to the first question were no, then almost certainly the answer to the second would be yes: that is, if rehearsed letters could utilize some new memory—different from the memory for unrehearsed letters—then rehearsal would place at the subject's disposal the combined capacity of two memories. Rehearsal would increase capacity.

If rehearsed letters were stored in the *same* memory as unrehearsed letters, rehearsal still might increase the capacity of memory. For example, imagine a world where spoken letters occur only in conjunction with noise, and at low S/N ratios. If the noise were stored in memory along with the signal, memory for unrehearsed letters would be poor. If rehearsed letters did not contain the noise, the opportunity to rehearse would appear to increase the capacity of memory.

We shall anticipate the conclusion of this section by stating that rehearsed letters are stored in the same memory as unrehearsed letters and that capacity is not increased. To reach this conclusion, we use information about capacity obtained in the last section, and information about rehearsal obtained from introspective reports.

WHAT WAS REHEARSED? INTROSPECTIVE EVIDENCE

Introspective reports of rehearsal in the partial report task are persuasively clear at low presentation rates and paradoxically ambiguous at high rates. At the 1/sec presentation rate, subjects say they rehearse each group of letters in the interval between groups. The silent interval is about 1½/sec, easily long enough to permit rehearsal of the entire previous group. Some subjects say they rehearse the previous group twice (or two previous groups once each) during the intergroup pause. When rehearsal is measured, its rate is found to be three letters per second (Landauer, 1962) and to vary from three to ten items per second (Sperling, 1963). The highest rates are possible only for highly practiced, repeated sequences; three letters per second is a typical rate for new material. Assuming a rate of rehearsal of three letters/second (giving a 2½/sec rehearsal period) implies that the rehearsal overlaps the first letter of the following group. Alternatively, a rehearsal rate of 5/sec would fit the eight letters entirely into the silent interval.

At the 2/sec and 4/sec rates, subjects report that active rehearsing interferes with recall, and that an optimum strategy is to listen passively to the entire stimulus. Whether rehearsal ever occurs in these cases is ambiguous. Particularly at 4/sec rates, subjects report trial-to-trial variations in the "attention" given to particular groups of letters, but it is not clear whether such attention can be equated with the active rehearsal that occurs at the 1/sec rate. From introspective reports, therefore, it is abundantly clear that there is massive rehearsal of stimuli presented at the 1/sec rate, and minimal rehearsal at the higher rates. A rough estimate of the average number of single-letter rehearsals would be 12 to 24 for the 1/sec presentations and less than 4 for the 2/sec and 4/sec presentations.

REHEARSAL, TIME, AND SERIAL POSITION EFFECTS IN PARTIAL REPORTS

Figure 7 shows the accuracy of recall of each report group in the partial report experiment. The result to which we wish to call attention is that the accuracy of report of the first group is virtually independent of

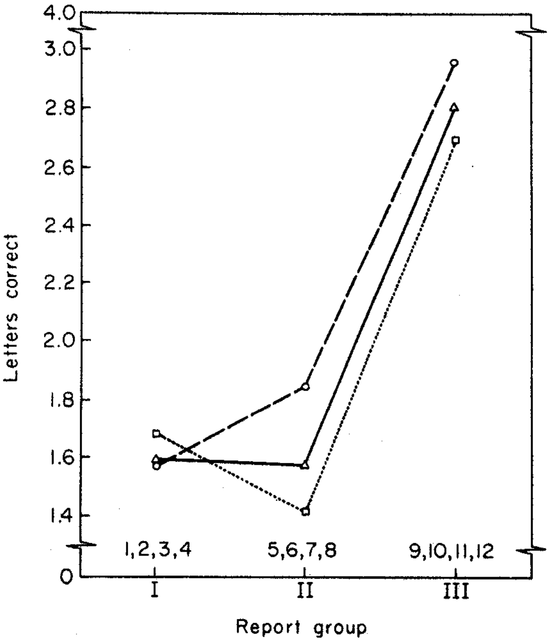

FIG. 7. Mean number of letters reported correctly in each report group averaged over alphabets. The only correction applied to these data is the correction for guessing (Eq. 3).

presentation rate. Calculation of the time from stimulus to response yields some interesting comparisons. At the rate of 1/sec, the first letter of the first group occurs 15½ sec before the cue to report it, as it is followed by 11 letters, three pauses, and the cue. On the basis of Anderson's (1960) data, delays of this length would be expected to interfere substantially with accuracy of report. At the 2/sec rate, the delay of the first item is 8 sec, at the 4/sec rate it is 5½ sec. Therefore, on the basis of time delay from stimulus to report, one would expect a large decrease in accuracy of the first group as a function of increasing delay (decreasing presentation rate). The virtual absence of this expected effect of delay may be explained by the observation we made above; namely, subjects rehearse the first group during the delay. The introspective accounts of rehearsal and the other serial position data thus tend to corroborate each other.[7]

REHEARSAL AND CAPACITY

For both alphabets, the 1/sec rate (most rehearsal) yielded the lowest estimates of memory capacity and the 4/sec rate yielded the second lowest estimates. In fact, these are minor effects, as the capacity estimates are nearly independent of rate. From this observation, and from the observation of massive rehearsal at 1/sec, it can be concluded that when rehearsal occurs, the total amount of material stored in auditory short-term memory is not increased.

Though rehearsal may compensate for passive decay at slow presentation rates, it does not increase estimated capacity beyond that which is observed when rehearsal is minimal or absent (2/sec rate). This conclusion rests on the simple facts that the amount of rehearsal depends greatly on presentation rate, and that memory capacity depends only slightly on presentation rate.

Although 1/sec rates yield the lowest estimates of capacity [based on their lowest partial report score (AD) and second-lowest partial report score (AS)], they yield the highest scores in nearly all other tasks. We explain these results by supposing that rehearsal duplicates and thereby rearranges the contents of memory so that its capacity is optimally utilized for the task at hand. For example, we assume that rehearsing a letter causes it to be entered into memory again. This makes the rehearsed letter more recent in memory (at the expense of the other letters) and means it is stored twice in memory.

---

[7] For a more detailed analysis of the data of Fig. 7, see Sperling and Speelman (1967).

Dual storage is particularly important when a letter originally is stored in a slot of memory which loses information between the time of storage and the time of retrieval. When a letter has been rehearsed, the net partial information remaining in the original plus the rehearsal storage locations may suffice for correct recall, whereas either location alone might not contain sufficient information. Without rehearsal of short lists, there would be unused capacity—empty locations which are capable of retaining partial information.

When the stimulus list is so long that it occupies the entire useful space of memory, rehearsal of some stimulus letters inevitably displaces others. This may improve performance by increasing resistance of the remaining letters to response-interference. But, even after correction for response-interference in estimates of capacity, rehearsal still will appear to have reduced capacity because duplication occasionally is needless.

Recency and the dual storage of rehearsed letters explain why they are less susceptible to interference by subsequent stimulus letters and less susceptible to response-interference. A similar conclusion (rehearsal protects against interference) is reached by Howe (1965). He presents indirect evidence, very similar to that reported above, to indicate that subvocal rehearsal protects items stored in memory from interference at the time of a vocal report; Howe (1967) reports comparable direct evidence that overt rehearsal protects items in memory from the self-produced interference of an interpolated task.

The effect of rehearsal on utilizing the full capacity of auditory short-term memory must not be confused with the possible role of rehearsal in utilizing other memories. When 3 sec were available to rehearse each stimulus item (or 12 sec for a simultaneous visual stimulus) the data indicated that a different kind of memory became effective (a memory for patterns and associations) with a different dependence on the acoustic similarity of items.

THE LOCUS OF MEMORY FOR REHEARSAL MATERIAL

Are letters which have been silently rehearsed remembered in a different memory system than letters which are heard but not rehearsed? Specifically, is there a memory of the muscular movements (or presumptive muscular movements) involved in rehearsing a letter, a memory which is different from the auditory memory for letters that are acoustically produced at the ears?

Hintzman (1965, 1967) proposes muscular memory because he notes that confusion errors in memory tasks differ systematically from confusion errors in listening tasks. His argument rests on the implicit assumption that the frequency spectrum of internal memory-noise is the same as the flat frequency spectrum of the noise which was used to create listening confusions.

Because the spectrum of speech itself is not flat—it is deficient in high frequencies—a masking noise with a flat frequency spectrum selectively destroys speech discriminations based on high frequencies (e.g., on 2nd and 3rd formants) and leaves discriminations based on low frequencies (e.g., on the 1st formant) relatively intact. In memory tasks, subjects make both kinds of confusions, which implies that the spectrum of memory noise approximates the spectrum of speech. Differences observed between the errors made in listening and in memory tasks (Conrad, 1964) can be explained by the arbitrary selection of the flat-noise stimulus to produce the listening errors.

While there is not yet any reason to suppose that significant muscular memory of rehearsal exists, there are some reasons to doubt its existence.

1. If there were a special memory for rehearsed materials, then rehearsal should increase memory capacity. That is, by remembering rehearsed letters in another memory, the space they occupied in auditory memory would be freed, or at least supplemented, thereby providing greater overall memory capacity. In fact, the opposite was observed: when there was evidence for rehearsal, memory capacity was slightly reduced. This is consistent with the assumption (Sperling, 1963, 1967) that rehearsed letters are remembered in the same memory as unrehearsed, acoustically-produced letters.

2. If memory storage for rehearsed letters were different from memory for unrehearsed letters, then it might be expected to have different properties with respect to the qualities of the letters being remembered. In particular, it might be expected to be less sensitive to acoustic similarity then auditory memory. In fact, a remarkable finding of the partial report procedure was that memory capacity was reduced by almost exactly the same amount (2.25 letters) at all rates of presentation when AS letters were substituted for AD letters, i.e., capacity deficits are independent of rate and therefore of rehearsal. This means that acoustic similarity impairs capacity for rehearsed letters to just the same degree that acoustic similarity impairs capacity for unrehearsed auditory letters. In this respect, at least, memory for rehearsed material is identical to memory for unrehearsed auditory stimuli.

3. It might here be argued that the vocal efforts required to produce various sounds are similar to each other to the exact extent that the produced sounds are similar to each other. Insofar as this is true, motor and sensory memory cannot be distinguished on the basis of sound. It is correct to conclude, however, that memory for rehearsed and unrehearsed letters is qualitatively the same and quantitatively interchangeable. This memory is properly called auditory memory (or auditory information storage) because it depends critically on the sound of the stimuli, even for groups of letters, which according to reasonable inference and introspection, are not rehearsed. There is no need at present to attribute verbal or linguistic attributes to this memory (e.g., Atkinson and Shiffrin, 1968); the data of these experiments can be explained on the basis of acoustic confusability (see below, also Conrad, 1964).

Predictions of Performance

Rules for Predicting Performance

To predict performance directly from the phonemic model, it would be necessary to make more detailed and accurate assumptions about the shape of the phoneme recall function, about the subjects' patterns and rules of rehearsal, about how rehearsed letters are stored in memory, about whether reporting letters is equivalent to rehearsing them, about the effects of time (e.g., silence is a sequence of blank phonemes), and so on. Even if we made all these assumptions, implementing them would require more parameters than could possibly be estimated from the data at hand.

Instead, we have taken an entirely different approach, and attempted to predict the data from the minimum possible number of parameters. In making predictions, we change from the partial-knowledge model to a most nearly equivalent all-or-none model because it offers simpler, more direct predictions. (However, we retain the partial-knowledge phonemic model of capacity to generate two parameters.)

PREDICTION PARAMETERS

The model contains seven estimated parameters. To predict the data of the experiments, the relevant parameters are combined according to a simple set of rules. Many predictions require less than the full set of seven parameters; indeed, it is possible to reduce the number of estimated parameters to five without much loss of accuracy. The seven parameters (and, for convenience, the final estimations of their values) are:

1. letter capacity
    a. AD alphabet (7.31 letters)
    b. AS capacity deficit (1.79 letters)
2. Response-interference parameters for presentation rates of
    a. 1/sec (.70 letters)
    b. 2/sec (1.04 letters)
    c. 4/sec (2.34 letters)
3. Standard deviation of memory capacity (1.22 letters)
4. Running span factor (.59)

PREDICTING WHOLE REPORT SCORES

Subtract from capacity, the AS capacity deficit (if appropriate) and the appropriate response-interference parameter. The result is the predicted mean memory span, $\bar{m}$. For a stimulus of length N, the predicted whole-report score $S(N)$ basically is the smaller of N and $\bar{m}$.

A correction for truncation must be subtracted in computing $S(N)$ whenever $|N - \bar{m}| < 2$ letters. The correction is needed because we assume memory span is distributed normally with a mean $\bar{m}$ and a standard deviation of 1.22 letters.[8] Subjects cannot report more letters than are presented to them, even though their memory span m may exceed N on some of the trials. Hence, we need to truncate the normal distribution at the value of N and to correct for this truncation. The correction is .5 letter when $|N - m| = 0$ and smaller otherwise. (See Appendix 1 for details.)

PREDICTING PARTIAL REPORT SCORES

When a subject makes a partial report of j letters from among N presented letters, the predicted mean number of letters available A is $A = C \cdot S(j)/j$, where C is the memory capacity for the appropriate alphabet and $S(j)$ is the predicted whole-report score for a stimulus of length

---

[8] Individual (average) memory spans tend to be distributed normally (cf. Pollack *et al.*, 1959). No matter what the distribution of one individual's memory span over i successive trials may be, the overall population distribution (over j subjects × i trials) is more closely normal than are the individual distributions. The approximation to normality improves as j increases.

j. (This prediction procedure simply reverses the procedure for estimating memory capacity from observed partial report scores. See Appendix 1 for more details.)

PREDICTING RUNNING MEMORY SCORES

The predicted score is simply .59 $\bar{m}$.

SAMPLE CALCULATIONS

We shall illustrate the rules by predicting scores for 4/sec AS letters. (1) Letter capacity for AS letters is 5.52 letters (AD letter-capacity of 7.31 letters minus an AS capacity deficit of 1.79 letters). (2) Mean memory span $\bar{m}$ at 4/sec equals 5.52 letters minus the rate-performance factor of 2.34 letters = 3.18 letters. (Note, the observed memory span for list lengths 6, 8, and 12 are, respectively, 3.11, 2.98, 3.36 letters.) (3) The predicted running memory span is $\bar{m} = 3.18$ letters times $.59 =$ 1.87 letters (2.08 observed). (4) The prediction for list length-4 required a correction for truncation of .17 letters giving a prediction of 3.01 letters (3.24 observed). To obtain the predicted partial report score, the memory capacity 5.52 must be multiplied by 3.01/4.00, giving 4.17 available letters (4.20 observed).

Discussion of the Prediction Rules

CAPACITY

There basically is only one capacity parameter in the rules, the total number of phonemes. The phonemic model derives letter capacities for various alphabets from the phonemic capacity. For convenience, the prediction rules give two parameters for capacity, one for each alphabet. As prediction of the AS capacity deficit by the phonemic model is virtually perfect, only one of these parameters actually needs to be estimated.

Another parameter is saved by predicting the standard deviation of capacity from the phonemic model, rather than estimating it. Predicted values of σ are: 1.43 (AD), 1.56 (AS), compared to 1.22 letters (AD and AS) estimated from data. It happens that the prediction rules are insensitive to variations of σ in this range and goodness of fit is little affected by using one of the predicted values rather than the estimated value.

RESPONSE INTERFERENCE PARAMETER (r)

The phonemic model predicted two properties of response-interference: (1) its dependence on rate (because of the dependence of rehearsal on rate), and (2) its dependence on list length, response-interference being greater for long lists. The first of these properties is explicit in the prediction rules: the parameter r explicitly depends on rate. The second property is implicit: the effect of r depends implicitly on the length N of a presented list. To see this, consider a graph of predicted mean score $S(N)$ vs. N. Figure 8 illustrates $S(N)$ for two values of r, differing by $\Delta r = 1$, e.g., $r = 0$ and $r = 1$. The bottom curve of Fig. 8 gives the difference, which is an indication of how much r subtracts from $S(N)$ as a function of stimulus length. Although r is a constant, its effect depends on stimulus length, becoming important only as N approaches the memory span. The model predicts that partial reports are

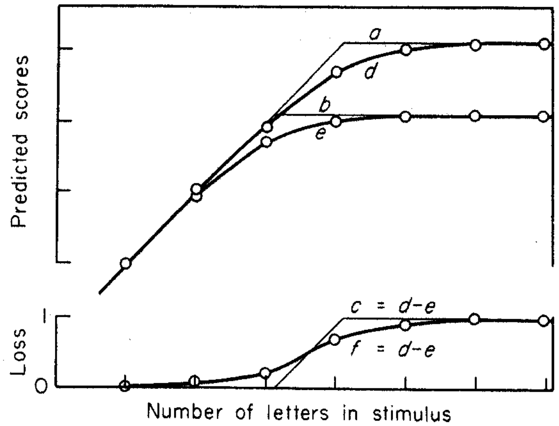

FIG. 8. Examples of theoretical predictions of whole report scores as a function of stimulus length. Curves a,b represent predictions that assume a negligibly small capacity-variance; curves d,e assume a capacity-variance of 1.0 letter^2. The difference in memory span (i.e., difference in response-interference parameter, r) between curves a and b is one letter ($\Delta r = 1.0$). Curves c and f show how the predicted loss in response accuracy depends on the number of letters presented. The ordinate for c and f is the predicted loss. Because the predictions are based on linear equations, the shape of the curve does not depend on the number of letters presented; therefore this number need not be indicated. However, a stimulus contains only integral numbers of letters, therefore predictions are made only at these specific points, indicated by circles. Coordinate markings are spaced at 1-letter distances.

more accurate than whole reports because the effect of r is smaller in the shorter partial report.

EFFECT OF ALPHABET

All differences in performance between AD and AS stimuli derive from the initial difference in letter capacity for these alphabets, and these differences are derived *a priori* from the phonemic efficiencies, f_1 and f_2. The rules translate the initial capacity difference into predicted AS deficits. A large capacity difference does not necessarily imply a large AS deficit in performance. For example, predicted and observed performance on lists of length-3 is virtually perfect for AD and AS alphabets. In the rules, the effect of alphabet is treated functionally in the same additive way as rate. For example, Fig. 8 can be interpreted as referring to two different alphabets.

RUNNING SPAN

In making running-memory reports, subjects typically write the last string of three–four letters first (beginning with the oldest), and then add a string of two to three letters in front of these, and occasionally yet a third string in front of the second. This recall strategy (which may be the only one possible) is inefficient because it unloads the most recent items first. It fails to balance response-interference and recency. The factor of .59 is a *post hoc* empirical estimate of the efficiency of the strategy.

Evaluation of the Predictions

GOODNESS OF FIT

The parameters used in the rules were chosen by means of an iterative optimization procedure which minimized the summed squares of the differences between each of the 38 predicted and observed pairs of values. Figure 9 illustrates a scatter plot of the predicted and observed results. The correlation of the predictions with the observed results of the 38 experiments is $r = .98$. The regression slope is 1.00 and the prediction accounts for .962 of the variance in the data, with a residual error of .27 letters per prediction (expressed in standard deviation terms).

These same parameters are not optimal for predicting the 19 observed AS deficits; nevertheless, they account for .78 of the variance of these

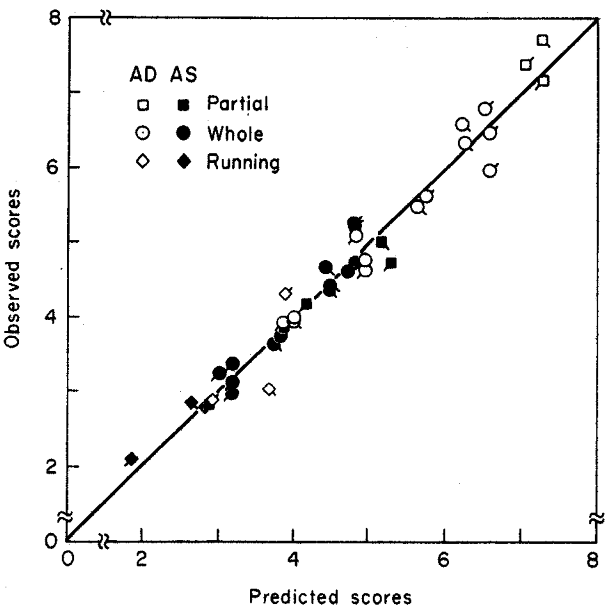

FIG. 9. Scatter plot of the observed scores in the 38 auditory experiments vs. the scores predicted by the calculation algorithm. Presentation rate is indicated by the direction of the diagonal strokes: up-right = 1/sec, down-right = 2/sec, down-left = 4/sec. The constants of the model are: AD capacity (7.31 letters), AS capacity deficit (1.79 letters), response-interference parameter (.70, 1.04, 2.34 letters for 1/sec, 2/sec, 4/sec), running span factor (.59), and capacity variability ($\sigma = 1.22$ letters).

data. Considering the small range of observed AS deficits and their greater error (each is the difference of two scores), this also is satisfactory prediction.

ECONOMIZING THE PREDICTIONS

The model can predict AS scores directly, i.e., without using any data whatsoever from experiments with an AS alphabet. To predict AS scores, six parameters of the rules are estimated only from AD scores, and the AS capacity deficit is calculated from the phonemic model. (The estimated and calculated values are virtually identical.) This prediction accounts for .90 of the observed variance of the 19 AS scores.

The absolute minimum basis for prediction is five experiments; one partial report condition (e.g., 2/sec), three whole reports (e.g., length-8, 1/sec, 2/sec, 4/sec), and one running memory condition (e.g., 2/sec).

The variability may be estimated from the whole reports. The six parameters estimated from such sets of five scores typically predict about .90 of the observed variance of the remaining 33 scores.

The predictions are not sensitive to small changes in parameters. For example, by choosing the memory capacity as 7.5 letters, the AS capacity deficit as 2.0 letters, the effect of rates as 1.0, 1.0, 2.0 letters (1/sec, 2/sec, 4/sec), the running memory factor as .6 and the standard deviation of the memory span as 1.0 letters, the rules still predict .94 of the variance of the original data. This set of seven parameters contains only four simple numbers (7.5, 2.0, 1.0, .6), which certainly brings it within the memory capacity of the reader.

Summary and Conclusions

Recall was tested in 19 different auditory and three visual recall tasks. In all 22 tasks, subjects reported fewer letters correctly when stimuli were composed of acoustically similar (AS) letters than when they were composed of acoustically different (AD) letters.

Subvocal rehearsal was shown not to increase the *capacity* of auditory short-term memory, but to increase the efficiency of utilizing the capacity. Scoring for items recalled, independent of their position, indicated an additional mean capacity of about 1.5 letters (i.e., without knowledge of their position), which was independent of the alphabet, presentation rate, and list length. The relatively frequent presence in memory of two items whose position is not known provided a plausible explanation for transposition errors. Results from visual presentations could be predicted from auditory results by assuming that subjects rehearsed letters of visual stimuli subvocally at rates of less than 4/sec. In presentations that allowed about 2 or more sec per letter, more complex memory mechanisms came into play.

A phonemic model of auditory short-term memory was proposed in which individual phonemes were retained or forgotten independently, a letter being reconstructed from the retained phonemes. The model fully accounted for the effect of acoustic similarity on recall. A set of prediction rules based on the model predicted .96 of the variance of the results of the 2 × 19 auditory conditions.

Appendix 1

Corrections for Guessing, Truncation, Partial Knowledge, and Response-Interference

Correction for Guessing

In this section, we show that the scores, as corrected for guessing by Eq. 3 of the text, are unbiased estimates of an underlying performance factor, the truncated memory span.

DEFINITIONS

Let the length of the presented stimulus be N letters, each chosen randomly without replacement from an alphabet of L equiprobable different characters. Let the memory span, an integer k, have associated with it a distribution function $\varphi(x)$, where $\varphi(x) \geqslant 0$ for all x and

$$\int_{-\infty}^{\infty} \varphi(x)\, dx = 1$$

(see Fig. 10). Let the probability $p(k)$ of x being equal to exactly k letters be given by $p(k) = \int_{k-1/2}^{k+1/2} \varphi(x)\, dx$. The definition of $p(k)$ as the integral of $\varphi(x)$ between $k - \frac{1}{2}$ and $k + \frac{1}{2}$ is in accordance with the intuitive notion of memory span. For example, when $\varphi(x)$ is a symmetric distribution about its mean m, and we say $m = 5.5$, we mean that scores of 5.0 and 6.0 are observed approximately equally often (for $N >> 6.0$ and scores corrected for guessing).

Given that the memory span equals k, the expected score $r(k)$ has two components: a performance factor (the truncated span), $s(k)$

$$s(k) = \min[N, \max(0, k)] \qquad \text{(A1)}$$

and a guessing factor $g(k)$

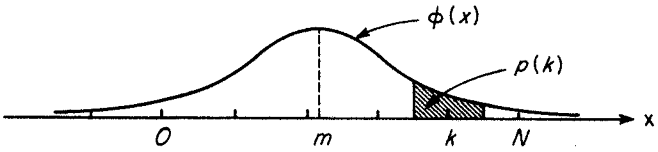

FIG. 10. A memory-span distribution function.

$$g(k) = [N - s(k)]/L, \qquad \text{(A2)}$$

which results from $[N - s(k)]$ guesses, each with probability $1/L$ of success. We say $s(k)$ is truncated because it equals k only for $0 \leqslant k \leqslant N$, and it equals 0 or N outside this range. The predicted score R is given by the sum over k,

$$R = \sum_{k=-\infty}^{k=\infty} p(k)r(k) = \sum_{k=-\infty}^{k=\infty} p(k)[s(k) + g(k)] = S + G, \qquad \text{(A3)}$$

where S and G represent the sums of the performance and guessing terms, respectively.

CORRECTION FOR GUESSING

We wish to show that the scores corrected for guessing $\hat{S}$ (Eq. 3 of text) are unbiased estimates of the performance factor S. Given that the memory span equals k, we may write the corrected score $\hat{s}(k) = \hat{r}(k) - [N - \hat{r}(k)]/(L - 1)$. Substitution of $r(k)$ from Eq. A3 for $\hat{r}(k)$, and algebraic reduction gives $\hat{s}(k) = s(k)$; thus, the corrected score $\hat{s}(k)$ is an unbiased estimate of the performance factor $s(k)$. Similarly, by summing over k, it is shown that $\hat{S} = \hat{R} - (N - \hat{R})/(L - 1)$ is an unbiased estimate of S. The corrected scores $\hat{S}$ estimate performance S by "removing" the factor $G = (N - S)/L$, which results from pure chance guessing.

Correction for Truncation

The score corrected for guessing $\hat{S}$ estimates a truncated span S, which depends not only on $\bar{k}$, the mean memory span, but also on N, the number of letters presented. In this section, we show how to compute S from $\bar{k}$. The reader should note that the mean memory span $\bar{k}$ is not exactly equal to the mean m of the memory distribution function because of the quantization of k, but the difference between $\bar{k}$ and m is negligible for $\sigma \geqslant 1$.

The truncated span S is less than $\bar{k}$ whenever there are trials on which the memory span k exceeds N, the number of letters presented. (Some truncation also occurs when $k < 0$, but, with the parameters of the experiments, this effect is entirely negligible.) In the text, $\varphi(x)$ was taken to be normal with mean m and standard deviation σ, so that some truncation always occurs. Given a normal distribution, the truncated span S in Eq. A3 can be given explicitly by

$$S = \sum_{k=n_1}^{n_2} \frac{1}{2} \left[\operatorname{erf}\left(\frac{k + \frac{1}{2}m}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{k - \frac{1}{2}m}{\sigma\sqrt{2}}\right) \right] \cdot s(k), \tag{A4}$$

where $n_1 = -\infty$ and $n_2 = +\infty$. There is no appreciable loss of accuracy in Eq. A4 if we take n_1 as the largest integer $< m - 5\sigma$ and n_2 is the smallest integer $> m + 5\sigma$.

The correction for truncation T is used in the text is simply

$$T = \max(m,N) - S. \tag{A5}$$

A useful feature of this definition of T is that it depends only on $d = |m-N|$, being greatest when $d = 0$. Some values of (d,T) are: $(.0\sigma, .381\sigma)$; $(.25\sigma, .270\sigma)$; $(.5\sigma, .183\sigma)$; $(1.0\sigma, .073\sigma)$; $(2.0\sigma, .006\sigma)$. For $d > 2\sigma$, T can be neglected; therefore for $N > \bar{k} + 2\sigma$, $\hat{S}$ estimates $\bar{k}$ directly.

The Effect of Partial Knowledge on Scores

THE ALL-OR-NONE MODEL

The correction for guessing assumed that a subject had all-or-none knowledge, i.e., perfect knowledge of exactly k stimulus letters and zero knowledge of the remainder. We have just shown that the corrected score $\hat{S}$ is given in Eq. 3 of the text is an unbiased estimate of $\bar{k}$. The corrected scores thus are appropriate whenever subjects have all-or-none knowledge.

While subjects may, in fact, have virtually perfect knowledge of the letters in some serial positions of the stimulus, and virtually zero knowledge of letters in other serial positions, there usually are letters at still other serial positions about which subjects have partial knowledge. Partial knowledge violates the assumption of all-or-none knowledge. The aim of this section is to show that the corrected scores $\hat{S}$ given by Eq. 3 also are reasonable to use when subjects have partial knowledge. The strategy in establishing reasonableness will be to show the consistency of the guessing correction with measures of transmitted information.

THE INFORMATION-TRANSMITTED MODEL

The only appropriate additive measure of partial knowledge known to the authors is the amount of information transmitted between stimulus and response (Shannon, 1948). Unfortunately, this measure is impractical. For example, to measure the information transmitted by stimuli of length equal to ten letters (each letter being chosen from an alphabet of eight letters), the experimenter must be prepared to present and to receive any of the $8^{10} = 1{,}073{,}741{,}824$ possible stimuli or responses, and to do so repeatedly in order to obtain the distributions of incorrect responses.

To simplify the problem, each serial position of the stimulus is treated as an independent channel. The information transmitted by a serial position is measurable; the information transmitted by the whole response is the sum of the amounts of information transmitted by each component serial position. This estimate of total information is slightly too low because subjects inadvertently use codes that involve several serial positions simultaneously (e.g., inversions); nevertheless, it is the best measure we have. For the remainder of this section, therefore, we confine ourselves to the analysis of the information transmitted about a single serial position; total information is given by summing over positions.

BOUNDS ON INFORMATION TRANSMITTED

Partial knowledge is revealed in the information-transmitted measure as the amount of information transmitted by incorrect responses. For example, when there is no partial knowledge, all incorrect responses are equally probable and they transmit no information. In this case, the information transmitted $T_{\min}$ is given by

$$T_{\min} = \log_2 L - p \log_2(1/p) - (1-p) \log_2[(L-1)/(1-p)], \quad \text{(A6)}$$

where L is the number of equiprobable stimulus alternatives and p is the uncorrected probability of a correct response.

For the case of maximum information transmitted by incorrect responses, we add the restriction that no incorrect response is more probable than the correct response. The maximum information transmitted, $T_{\max}$ then is

$$T_{\max} = \log_2 L - jp \log_2(1/p) - (1-jp) \log_2[1/(1-jp)], \quad \text{(A7)}$$

where j is the greatest integer $\leqslant 1/p$.

Let s be the probability of a correct response in the ith serial position as corrected for chance guessing by Eq. 3 of the text,

$$s = p - (1-p)/(L-1); \quad \text{(A8)}$$

s varies between zero and one. To compare the T's with s, we normalize the T's by dividing them by the stimulus information, $H_S = \log_2 L$. (H_S is the limiting amount of information transmitted when recall is perfect.) The normalization gives $P_{\min} = T_{\min}/H_S$ and $P_{\max} = T_{\max}/H_S$. $P_{\min}$ and $P_{\max}$ are the minimum and maximum possible *fractions* of information transmitted assuming, respectively, the minimum and maximum amounts and information to be transmitted by incorrect responses. Figure 11 shows s, $P_{\max}$, and $P_{\min}$, as functions of p.

Figure 11 illustrates a desirable "betweenness" property of s, namely that s mostly lies between $P_{\max}$ and $P_{\min}$. In fact, the following conditions of "betweenness" easily can be proved: (1) $P_{\min} < s$ for all p, $1/n < p < 1$, and for all $n \geqslant 2$; (2) $P_{\max} > s$ for the cusps of $P_{\max}$, except the first and last [i.e., for all p, $1/(n - 1) \leqslant p \leqslant \frac{1}{2}$]; and (3) in the limit as $n \rightarrow \infty$, $P_{\max} > s$ for all p, $1/n < p < 1$.

The betweenness property of s means that if s is interpreted as a fraction of information transmitted, the fraction usually lies between its theoretical lower and upper bounds. Betweenness means that, when it is not desired to assume all-or-none knowledge, S can be given an alternate interpretation as a reasonable estimate of partial knowledge, i.e., of the fraction of information transmitted.

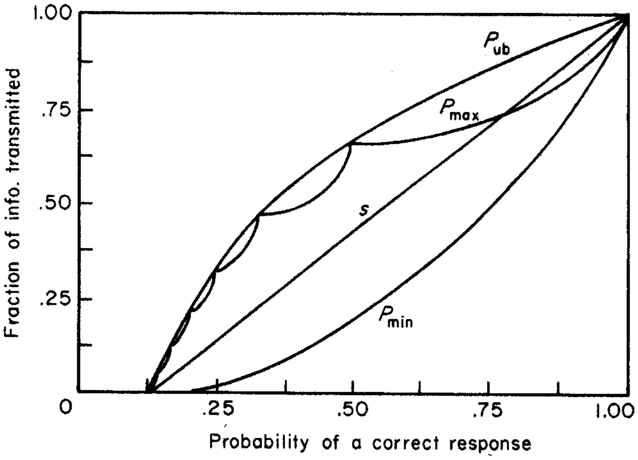

FIG. 11. Fraction of information transmitted vs. p, the (observed) probability of a correct response in a single serial position. The number of equiprobable stimulus alternatives is $N = 8$. $P_{\max}$ is the theoretical maximum possible fraction of information transmitted; $P_{\min}$ the theoretical minimum; s the score corrected for chance guessing by Eq. 3; $P_{ub} = \log_2 NP$, a simple upper bound that yields correct values of $P_{\max}$ whenever $1/p$ is an integer.

In summary, the justifications for using Eq. 3 to correct scores are: (1) there is no feasible alternative computation; (2) interpreted as an estimator in the all-or-none model, it leads to scores appropriate to the mathematical analysis used; (3) it is a reasonable estimate of partial knowledge for partial-knowledge models. This last advantage gives us freedom to use all-or-none and partial-knowledge models interchangeably, depending on which is more convenient in a particular situation.

Correction of Partial Report Scores to Give Estimates of Capacity

THEORY UNDERLYING THE CORRECTION

Consider the case of a subject required to make partial reports of length equal to j letters from a stimulus of N letters. The subject's score A may be less than N for several basic reasons.

A. Some fraction $1 - p$ of the stimulus letters no longer are in memory at the time the report is called for. Perhaps the letters never were stored or perhaps they were stored and subsequently lost. We seek to estimate p, the proportion of letters remaining in memory at the time report is called for. When there are no other losses, capacity $C = pN$.

B. The act of reporting the first and subsequent letters, produces response-interference with the memory of the remaining called-for letters (within-group interference), so that ultimately recall is reduced by some additional fraction $1 - q$. The number of letters available A, i.e., the apparent number of letters in memory, is reduced to qpN (see footnote 3, page 166). The aim of the following section is to estimate q from whole-report data, and then to estimate capacity $C = pN$ from $\hat{A} = \hat{q}\hat{p}N$ by multiplying $\hat{A}$ by a correction factor of $1/\hat{q}$.

C. Various factors related to the cue-to-report reduce recall performance; for example, difficulty in interpreting the cue and possible cue-confusions. Such factors conceivably might be estimated from experiments in which the number and compatibility of cues is varied. Hopefully, they are negligible when the number of cues is small. They are omitted here in estimating a correction factor.

ESTIMATING THE CORRECTION FACTOR

The correction factor is estimated from experiments in which response self-interference is the primary cause of recall errors. This situation obtains in whole report experiments with short list lengths because here the factor of overloaded capacity is negligible. In the ideal case, we would know *a priori* that memory capacity is exactly C letters, and we would obtain whole reports for lists of length j, where $j < C$. Recall errors could not be attributed to capacity overload because $C > j$; therefore, if errors occurred, they would be attributable entirely to response-interference.

Suppose now that C is a statistical quantity which varies somewhat from trial-to-trial. Then, the argument becomes quantitative rather than all-or-none; that is, the more unlikely it is that $C < j$, the more confidence we have in attributing errors to response-interference. (The prediction rules in the text give a full quantitative expression for these arguments.)

The partial reports were of length $j = 4$, so that the response-interference factor is measured from whole reports of length 4. This demands that the memory capacity exceed 4. Memory capacity has not yet been calculated, but we can use the partial report data directly (number of letters available, $\hat{A}$, see footnote 3, page 166) as a conservative lower bound of capacity. Table 1 showed that in all six conditions (2 alphabets × 3 rates) $\hat{A}$ exceeds 4, i.e., $C > j$. Let $\hat{S}(j)$ be the whole report score or lists of length j. The estimated fraction $\hat{q}$ of letters lost due to self-interference is $\hat{q} = 1 - \hat{S}(j)/j$, so that the estimated correction factor $\hat{f}$ $(\hat{f} = 1/\hat{q})$ is

$$\hat{f} = j/\hat{S}(j).$$

The two major kinds of error in estimating f are: (1) partial rather than all-or-none knowledge of items, (2) not all items have equal susceptibility $(1 - q)$ to response-interference. Happily, these two kinds of error tend to cancel each other.[9]

Appendix 2: Predicted Scores When Items Are Remembered without Knowledge of Position

Definitions

Let $\varphi(x)$ be the distribution function associated with position memory span. Let the distribution function associated with the item memory span be $(x - I)$, i.e., the item span exceeds the position span by I

---

[9] For more details, see Sperling and Speelman (1967).

items, where I is a nonnegative integer. We wish to calculate R_1, the truncated position score as defined in Appendix 1; R_2, a position score which is higher than R_1 because of an increment resulting from memory of I items (but without memory of their position); and R_3, the item score (i.e., the score obtained after response items have been permuted so as to maximize the position score).

For example, suppose on a particular trial six letters are presented and the memory span is five. This would usually lead to the correct report of only five items. Suppose $I = 1$; one additional stimulus item can be remembered but without any memory of its position. Because only one position on the score sheet remains to be filled after the five letters whose position is known have been written, the additional item always will be written in its correct position. Thus, memory of items even without memory of their position inflates the position score R_2, as well as the item score R_3. This Appendix shows how R_1, R_2, and R_3 are calculated and outlines (without mathematical rigor) the basis of the calculations.

Position Score, R_1

From Appendix 1, we have

$$R_1 = \sum_{k=-\infty}^{\infty} p(k)[s(k,N) + g_1(k,N,L)], \qquad \text{(B1)}$$

where

$$s(k,N) = \min[N,\max(0,k)], \qquad \text{(B2)}$$

and

$$g_1(k,N,L) = [N - s(k,N)]/L. \qquad \text{(B3)}$$

Inflated Position Score, R_2

The position score R_2 is inflated by memory of I items whose position is unknown (g_2), and by pure chance guessing (g_1). R_2 is given by

$$R_2 = \sum_{k=-\infty}^{\infty} p(k)[s(k,N) + g_2(I,k,N,L) + g_1(k + g_2,N,L)], \qquad \text{(B4)}$$

where s and g_1 are as defined above and $g_2(I,k,N,L)$ represents the expected number of the I items, which are written in their correct position.

The use of g_1 as an estimate of guessing in Eq. B4 assumes that, when items are remembered but written in an incorrect position, their probability of being correct is equal to that of randomly selected items.[10]

To derive an expression for $g_2(I,k,N,L)$, the following definitions are needed. Let B represent the number of positions by which the stimulus length N exceeds the memory span. B is the number of positions the subject fills with guesses or partial guesses. Let A represent the number of items the subject knows (without knowledge of their position) on a particular trial. Then

$$B = \max[0,\min(N,N-k)]$$
$$A = \min(I,B).$$

The response (beyond the position memory span k) contains A items from the stimulus $(x_1, x_2 \ldots x_A)$ and B-A items $(y_1, y_2 \ldots y_{B-A})$ guessed randomly (with or without replacement). The order of the x_i and y_i is random. The probability that an item x_i finds itself in its originally correct position is simply $1/B$. (Of course, x_i may find itself in a position belonging originally to x_j, but still be correct there by chance whenever $x_j = x_i$. This eventuality is subsumed in the guessing term g_1.) The expected number g_2 of the A items in their originally correct position is

$$g_2 = \begin{cases} 0 & B = 0 \\ A/B & 0 \leqslant A \leqslant B \\ 1 & 1 \leqslant B \leqslant A. \end{cases} \tag{B5}$$

Note that g_2 is not a function of the total number of possible letters L, but only of A/B.

Let the position scores, corrected for random guessing by Eq. 3 of text, be designated as S_1 and S_2, respectively. It is easily proved that $R_2 - R_1 = S_2 - S_1 = g_2$ defined by Eq. B5. A remarkable corollary is that a subject cannot inflate his corrected position score by more than one item, no matter how many nonpositional items he knows, i.e., no matter how much item knowledge he has.

The theoretical predictions of Eqs. B4 and B5 were tested by a Monte Carlo simulation. The empirical results of sampling with replacement from an alphabet of size $L = 8$ differed insignificantly from the predicted results.

---

[10] This assumption has been proved by S. Johnson (personal communication).

Item Scores, R_3

The predicted item scores are given by

$$R_3 = \sum_{k=-\infty}^{\infty} p(k) \cdot [s_1(k+I) + g_3(k,N,L)],$$

where

$$g_3(k,N,L) = F(A,L) \quad \text{and} \quad A = \min(\max(N - k - I,0),\ N).$$

Here A is the number of guesses at stimulus items, and $F(A,L)$ is the expected number of correct items, scored irrespective of position. It is assumed here, as in the previous derivations, that the number of response items equals the number of stimulus items.

Derivations of $F(A,L)$ have been carried out by S. Johnson (unpublished) and Riordan (1967). The authors are indebted to Dr. Johnson for the values of $F(0,8)$ to $F(8,8)$ which are: .0; .1250; .4414; .8864; 1.4201; 2.0167; 2.6593; 3.3365; and 4.0404.

Figure 4 of the text illustrates the expected difference between item scores and correct position scores as a function of the difference between stimulus length and position score, for three values of I. In treating the data, position scores were corrected for chance guessing by Eq. 3 of the text. For the theoretical curves, therefore, the ordinate of Fig. 4 was $\hat{R}_3 - \hat{S}_2$ and the abscissa was $N - \hat{S}_2$ where $\hat{S}_2$ is $\hat{R}_2$ corrected for chance guessing, i.e., $\hat{S}_2 = \hat{R}_2 - (N - \hat{R}_2)/(L - 1)$. The function $\varphi(x)$ was taken to be Normal with mean m and $\sigma = 1$; the curves were generated by assuming $N = 100$ and varying m in steps of .1 from 90 to 105.

References

ANDERSON, N. S. Poststimulus cuing in immediate memory. *Journal of Experimental Psychology,* 1960, **60**, 216–221.

ATKINSON, R. C., AND SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *Advances in the psychology of learning and motivation research and theory.* Vol. II. New York: Academic Press, 1968. Pp. 89–195.

BADDELEY, A. D. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity in short-term memory. *Psychonomic Science,* 1966, **5**, 233–234.

BROADBENT, D. E. *Perception and communication.* New York: MacMillan (Pergamon), 1958.

BROWN, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology,* 1958, **10**, 12–21.

CONRAD, R. An association between memory errors and errors due to acoustic masking of speech. *Nature*, 1962, **193**, 1314–1315.

CONRAD, R. Acoustic confusions and memory span for words. *Nature*. 1963, **197**, 1029–1030.

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

CONRAD, R. Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 161–169.

CONRAD, R. Interference or decay over short retention intervals? *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 49–54.

CONRAD, R., BADDELEY, A. D., AND HULL, A. J. Rate of presentation and the acoustic similarity effect in short-term memory. *Psychonomic Science*, 1966, **5**, 233–234.

CONRAD, R., AND HULL, A. J. Information, acoustic confusion and memory span. *British Journal of Psychology*, 1964, **55**, 492–432.

DALE, H. C. A. Retroactive interference in short-term memory. *Nature*. 1964, **203**, 1408.

DALE, H. C. A., AND GREGORY, M. Evidence of semantic coding in short-term memory. *Psychonomic Science*, 1966, **5**, 75–76.

DALLETT, K. M. Intelligibility and short-term memory in the repetition of digit strings. *Journal of Speech and Hearing Research*, 1964, **7**, 362–368.

GLANZER, M., AND CLARK, W. H. Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 289–299.

GLANZER, M., AND CLARK, W. H. The verbal loop hypothesis: Binary numbers. *Journal of Verbal Learning and Verbal Behavior*, 1964, **2**, 301–309.

HARRIS, C. S., AND HABER, R. N. Selective attention and coding in visual perception. *Journal of Experimental Psychology*, 1963, **65**, 328–333.

HINTZMAN, D. L. Classification and aural coding in short-term memory. *Psychonomic Science*, 1965, 3, 161–162.

HINTZMAN, D. L. Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 312–316.

HOWE, M. J. A. Intra-list differences in short-term memory. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 338–342.

HOWE, M. J. A. Consolidation in short-term memory as a function of rehearsal. *Psychonomic Science*, 1967, **7**, 355–356.

LANDAUER, T. K. Rate of implicit speech. *Perceptual and Motor Skills*, 1962, **15**, 646.

LAUGHERY, K. R. Effects of symbol size on immediate memory. *American Psychologist*, 1963, **18**, 415. (Abstract)

LAUGHERY, K. R., AND PINKUS, A. L. Short-term memory: Effects of acoustic similarity, presentation rate and presentation mode. *Psychonomic Science*, 1966, **6**, 285–286.

NORMAN, D. A. Acquisition and retention in short-term memory. *Journal of Experimental Psychology*, 1966, **72**, 369–381.

POLLACK, I., JOHNSON, L. B., AND KNAFF, P. R. Running memory span. *Journal of Experimental Psychology*, 1959, **57**, 137–146.

RIORDAN, J. A matching problem. *Bell Telephone Laboratories Technical Memorandum*, 1967.

SHANNON, C. E. A mathematical theory of communication. *Bell System Technical Journal*, 1948, **27**, 379–423, 623–656.

SPEELMAN, R. G., AND SPERLING, G. Effect of the sound of stimuli upon short-term memory. Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, April 18, 1964.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74**, No. 11 (Whole No. 498).

SPERLING, G. A model for visual memory tasks. *Human Factors*, 1963, **5**, 19–31.

SPERLING, G. Spatial and temporal visual masking. I. Masking by impulse flashes. *Journal of the Optical Society of America*, 1965, **55**, 541–559.

SPERLING, G. Successive approximations to a model for short-term memory. *Acta Psychologica*, 1967, **27**, 285–292.

SPERLING, G. Phonemic model for short-term auditory memory. *Proceedings of the American Psychological Association*, 1968, **4**, 63–64.

SPERLING, G., AND SPEELMAN, R. G. The effect of sound of stimuli on short-term auditory memory. *Bell Telephone Laboratories Technical Memorandum*, 1967.

WAUGH, N. C., AND NORMAN, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

WICKELGREN, W. A. Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 54–61. (a)

WICKELGREN, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, **70**, 102–108. (b)

WICKELGREN, W. A. Similarity and intrusions in short-term memory for consonant-vowel digrams. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 241–246. (c)

WICKELGREN, W. A. Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, 1965, **38**, 583–588. (d)

WICKELGREN, W. A. Short term memory for phonemically similar lists. *American Journal of Psychology*, 1965, **78**, 567–574. (e)

WICKELGREN, W. A. Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology*, 1966, **71**, 396–404. (a)

WICKELGREN, W. A. Short-term recognition memory for single letters and phonemic similarity of retroactive interference. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 55–62. (b)

WICKELGREN, W. A. Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, 1966, **39**, 388–398. (c)

WICKELGREN, W. A., AND NORMAN, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, **3**, 316–347.

WOODWORTH, R. S., AND SCHLOSBERG, H. *Experimental psychology*. (Rev. ed.) New York: Holt, 1954.