Auditory information processing and its method of presentation

CHIZUKO IZAWA, DOUGLAS BELL, and ROBERT G. HAYDEN
Tulane University, New Orleans, Louisiana 70118 U.S.A.

The encoding or learning difficulty variable was found to differentiate two theories concerning the relative efficacy of auditory item information presentation methods, anticipation vs. study-test. Barch and Levine (1967) expect superiority of the anticipation method when stimulus encoding is difficult, but that of the study-test method when it is easy (the encoding position). The retention interval model (Izawa, 1981c), however, predicts that the advantage for the study-test method is likely to be negligible when learning is either extremely difficult or extremely easy, but may be large when learning is intermediately difficult, with some possibility of the advantage being a small one. Data from 78 college students were analysed via group means, with either highest or lowest performers, from various perspectives. No situation, throughout all data analyses on three levels of difficulty, produced a significant superiority for the anticipation method vis-à-vis the study-test method. Overall, support for the encoding position was less than meager. The retention interval model, in contrast, seems to have received substantial support from auditory data on all difficulty levels, be they defined a priori (objectively) by the experimenter, or a posteriori (subjectively), by the individual learner.

Key words: auditory information processing, auditory lag effects, retention interval model, encoding difficulty, freed back, anticipation method, study-test method, short-term memory.

Auditory information processing has been modeled and debated in various ways (e.g., Aaronson, 1974; Broadbent, 1958, 1971; Crowder, 1978; Fujisaki & Kawashima, 1970; Massaro, 1976; Neisser, 1967; Pisoni, 1975; Watkins & Watkins, 1982). While some of these positions are based on relatively unique properties of auditory stimuli, others maintain common attributes for both auditory and visual information processing.

By utilizing primarily the visual modality, a puzzle endured for decades concerning the relative efficacy of information presentation methods in learning. For example, when the cue-target combinations are learned in the form of paired-associate learning (PAL), or in verbal discrimination learning (VDL) situations, is the traditional anticipation method that shares the same principle as the "teaching machine" (Equation 1) more advantageous than the study-test method (Equation 2)?

The two item-information presentation methods can be expressed as in Equations 1 and 2, respectively:

\begin{align*}
S_{1*}, S_{1*}R_{1}, S_{2*}, S_{2*}R_{2}, \ldots, \\
S_{j*}, \ldots, S_{n*}R_{n};&
\end{align*}

(Intercycle Interval); \hspace{1cm} (1)

\begin{align*}
S_{1*}, S_{1*}R_{1}, S_{2*}, S_{2*}R_{2}, \ldots, \\
S_{j*}, \ldots, S_{j*}R_{j}, \ldots, S_{t*}, S_{t*}R_{t};
\end{align*}

versus

\begin{align*}
S_{1*}R_{1}, \ldots, S_{j*}R_{j}, \ldots, S_{n*}R_{n}; \\
\text{(Intercycle Interval)}; \\
S_{1*}, S_{2*}, \ldots, S_{2*}, S_{1*};&
\end{align*}

(Intercycle Interval); \hspace{1cm} (2)

1 The present research was supported in part by BRSG Grant S07 RR07040, awarded by the Biomedical Research Support Grant Program, Division of Research Resources, National Institutes of Health. The authors are grateful to Kim Vaz and Nelda Clements for their general clerical assistance.

Requests for reprints should be directed to Chizuko Izawa at Psychology Department.
Auditory information presentation

\[ S_0-R_0, S_1-R_1, ..., S_n-R_n, ..., S_{r2} \]

(Intercycle Interval);
\[ S_{t1}, S_{t2}, ..., S_{tj}, ..., S_{tj}; \]

where \( S \) and \( R \), respectively, stand for the stimulus and response terms of a pair for an \( n \)-pair list in PAL with a randomized item presentation order from cycle to cycle, and where subscripts identify individual items (\( 1 \leq j \leq n \)). The study \( (S_{t1}, R_j) \), a presentation of both terms of the pair) and test \( (S_{t1}, \text{a presentation of the stimulus term alone}) \) events of Item \( j \) are boxed with solid and broken lines, respectively, in Equations 1 and 2.

A puzzling aspect of the comparisons of these two methods was that the results are discouragingly mixed: the study-test method often excels over the anticipation method significantly, but frequently the two methods also differ little. It is clear, nonetheless, that the feedback position based on the Skinnerian operant conditioning principle has no grounds for support. The feedback position predicts the superiority for the anticipation method (Eq. 1) which enjoys immediate feedback, as compared with the study-test method (Eq. 2) which does not. Such a prediction, however, did not materialize in data.

In contrast, the task alternation viewpoint (e.g., Battig & Brackett, 1961) expects inferior performance for the anticipation method that suffers from constantly alternating studying and testing tasks and may be confusing to the learner, as compared with the study-test method which does not do so. The differential acquisition position (e.g., Kanak & Neuner, 1970) expects similar outcomes, assuming better learning for the study-test method. These two latter positions are supported by data, but only partially. They cannot accommodate equal performance levels for both methods.

The puzzle of inconsistent results, large (significant) and small (non-significant) superiority of the study-test method over the anticipation method, seems finally to be accounted for by the retention interval (RI) model (Izawa, 1981a, 1981b, 1981c) both qualitatively and quantitatively. The rationale for the model is based on the fact that individual retention intervals of the \( n \)-items (e.g. the interval between the two boxes with solid and broken lines for Item \( j \) in Eqs. 1 & 2) distribute triangularly under a random item presentation order from cycle to cycle, common in the field, from 0 (minimum) to \( 4n-4 \) (maximum) intervening events with a mean of \( 2n-2 \) under the anticipation method, while the same intervals range from 0 to \( 2n-2 \) with a mean of \( n-1 \) intervening events under the study-test method, as in Fig. 1 (Izawa, 1972). It is essential to note that the two distribution curves do overlap by virtue of sharing the same lower limits, the differential upper limits notwithstanding. Items falling into the overlap area in the distribution are likely to have the same retention interval lengths for both methods, whereas those falling into the nonoverlap areas correspond to the probability that the retention interval is longer with the anticipation method than with the study-test method.

![Number of intervening events between an S and its subsequent T](image)

Fig. 1. Triangular distributions of individual retention intervals within the \( n \)-item list in terms of intervening events for both anticipation and study-test methods (from Izawa, 1972; where \( S \) and \( T \) are, respectively, study and test events).
At the end of any study event, an item may be classified into one of the three types: (1) unlearned, (2) just or recently learned and residing in the unstable short-term memory (STM) store, or (3) well learned and established in the stable long-term memory (LTM) store. The items belonging to the first (1) or third (3) type are noncritical items, each residing outside STM, while those of the second (2) type are critical ones, residing in STM. The noncritical items do not differentiate between the two methods, since the unlearned items (1) are likely to be incorrect under both methods, while the well learned items (3) are likely to be correct under both methods, independent of differential retention interval lengths on the order of \( n - 1 \) intervening events on the average, respectively.

In stark contrast, critical items in unstable STM are likely to survive when retention intervals (lags) are short as is the case frequently with the study-test method, but unlikely to do so when they are long (delays) as often is the case with the anticipation method, i.e., in a situation where critical items fall into the nonoverlap areas (Fig. 1). Here, if the number of such critical items in the nonoverlap areas is sufficient, their differential survival between the two methods is likely to lead to a distinctly better performance with the study-test method.

However, (a) if a sufficient amount of critical items fall into the overlap area of equal retention intervals for both methods, or (b) if only an insufficient number of critical items are generated (independent of where they may fall in the distribution curves, Fig. 1); differential performances between the two methods are unlikely.

Thus, both significant and nonsignificant superiority of the study-test method over the anticipation method seems to be accommodated logically within the single theoretical framework of the retention interval (RI) model, by means of the amount of critical items generated together with the distribution of their retention intervals (Fig. 1).

However, the above is based on data involving almost exclusively visual information processing (in approximately 95% of the studies). A question to be attacked in the present study is: can the theory based on visual data account for results obtained under the auditory modality as well?

While some theories are constructed in modality specifics, others are not. For example, a continuum between visual and auditory modalities is inferred by Craik (1969), Crowder and Morton (1969), and Sperling (1967); two separate processes are assumed by e.g., Murdock (1967, 1968), Murdock and Walker (1969), and Penney (1975). Models such as Neisser (1967), Massaro (1976), or Pisoni (1975) were specifically addressed to auditory information processing.

The RI model is not modality specific. Temporal relationships between study and test events under the anticipation and study-test methods in Equations 1 and 2 remain the same, and so do the retention interval distributions (Fig. 1), independent of the modality utilized. Therefore, the RI model should, if valid, stand empirical tests with the auditory stimuli.

Surprisingly, studies comparing anticipation and study-test methods auditorily are quite small in number. To make matters worse, the inconsistency in the results of the few extant auditory comparisons is gross. Most notably, Barch and Levine (1967) produced significant superiority with the anticipation method vis-à-vis the study-test method when Morse code signals (stimulus terms) and double-digits (response terms) were paired and presented auditorily. The findings directly challenge the validity of the retention interval model.

Nevertheless, when Morse code signals were substituted by words, the complete opposite resulted. This superiority of the study-test method by Barch and Levine
was in agreement with auditory data by Battig and Wu (1965) in PAL, and Kanak, Cole, and Eckert (1972) in VDL. However, in the case of Izawa, Hayden, and Isham (1980), there were no significant differences between the methods. Thus, all possibilities have empirical support in comparative studies with the auditory mode.

In an attempt to explicate inconsistent results within their study, Barch and Levine formulated a position where the anticipation method should excel over the study-test method when stimuli are difficult to encode, and that such a prediction should not be limited to Morse code signals only (1967, p. 287, e.g.). When, however, stimuli are easy, they maintain that the study-test method should generate a distinct advantage. Barch and Levine’s position henceforth will be referred to as the encoding position.

It is to be noted here that via the encoding position, the learning difficulty level entered here as a salient factor, theorized as controlling the relative efficiency of the two methods. Coincidentally, learning difficulty is relevant to the RI model as well. According to the model, when learning is extremely difficult, not very many items are learned, and therefore the gross total number of critical items in STM is likely to be small. Given an insufficient amount of critical items, the differential lengths of the retention intervals are unlikely to be large. This would lead to little performance differences between the two methods. Similarly, when learning is extremely easy, most items are easily overlearned and go to LTM quickly, leaving few items in STM as critical items. This may also lead to relatively small performance differences between the two methods.

However, when learning is only intermediately difficult, a substantial number of critical items may be generated and reside in STM. (a) If a sufficient number of them falls into the nonoverlap areas in the distribution, the retention intervals under the study-test method will be shorter than under the anticipation method, and therefore a large advantage for the former is likely. (b) If, however, a sufficient number of the critical items should fall into the overlap area instead, the lengths of retention intervals of these items are likely to be the same for both anticipation and study-test methods, and therefore, little performance differences between them are to be expected. For a more detailed discussion, see Izawa (1983).

Hence, the two theoretical positions make quite different predictions in reference to encoding and learning difficulty. When learning is extremely difficult, the encoding position predicts superior performances for the anticipation method, while the RI model predicts little difference between the two methods. When learning is easy, the former expects superiority of the study-test method, while the latter predicts the same superiority for the study-test method to be likely when learning is moderately easy (or moderately difficult), with some reservations for possibly no difference at all. However, for the RI model, small differences are most likely when learning is extremely easy.

Thus, learning difficulty provides an ideal variable to test the two rival theories empirically. The present study was planned to achieve such a test, by manipulating the encoding difficulty levels via word frequency by Thorndike and Lorge (1944).

Method

The experimental design planned to fulfill our goals was a 2 x 3 factorial one, with two information processing methods: anticipation vs. study-test method, and three levels of learning difficulty: easy, medium, and difficult. Each of the three lists was composed of 20 pairs, derived from 40 words of two or three syllables each. The easy list consisted of words oc-
curring 100 times or more per one million words, while the intermediate one, of five times per four millions, and the difficult list, of words of four occurrences per 18 millions, respectively, selected from Thorn-dike and Lorge (1944). No initial letter was repeated among the stimulus words, or among the response words, respectively, in each list. Pairs were constructed so as to avoid apparent connections between the two terms of each pair. One hundred twenty words thus selected were nouns, with two minor deviations (one adjective and one verb) forced by the restrictions to medium and difficult lists.

Auditory item presentations were made by a male voice (Douglas Bell), pre-recorded on a Realistic Tape Recorder, Model CTR-42, with a 2 s rate. One pair was presented every 2 s during the study cycle, and on a test event of 2 s. The subject had to respond during that period inclusive of the cue (stimulus) presentation which signaled a test event. Each subject was run individually, and responded orally.

Item presentation order was randomized from cycle to cycle for both methods. To be comparable with the anticipation method which begins with a test event prior to any study opportunity, the study-test conditions started with the test cycle.

Seventy-eight Introductory Psychology students at Tulane University volunteered to participate in the experiment, 13 per condition. The subjects were run one at a time, and were assigned to one of the six conditions under a semi-Latin square method, in order of their appearance in the laboratory. A practice task consisting of both anticipation and study-test arrangements, the same for all subjects, preceded the main task, so as not to bias them differentially.

Results and Discussion

Tested by the practice task performance, there were no group differences in the subject variable, ensuring the comparability of the six conditions.

Figure 2 presents major results in terms of incorrect responses including both overt errors and no responses (omissions). There were large differences among difficulty levels in a 2 × 3 factorial analysis of variance: \( F(2, 72) = 45.417, \ p < .001 \), with performance levels lining up neatly in accordance with those of learning difficulty. The findings testify that the difficulty levels we manipulated via word frequency did indeed produce expected effects, to legitimize empirical tests of the two different theories that predict differential performances for the anticipation and study-test methods with differential encoding or learning difficulty.

Our main interest in the present study was the performance difference between the two item information presentation methods under comparison. Barch and Levine's encoding position predicts significantly better performances with the anticipation method than with the study-test method for the difficult list, but a large superiority for the study-test method instead with the easy one. There is no explicit statement on what the encoding position predicts for the medium list. It may be fair, however, to infer that the medium list should result in some compromise of the two opposite extremes: small and large differences between the two item presentation methods. Independent of the accuracy of such an inference, the position clearly expects large interactions between methods and learning difficulty levels. Quite contrary to this theoretical prediction, there were no traces of such interactions: \( F(1, 72) = 0.942, \ MS_e = 946.88 \).

Furthermore, overall performance differences between anticipation and study-test methods were too small to be significant: \( F(1, 72) = 1.361, \ p > .20 \). Thus, neither significant superiority for the anticipation method with the difficult list, nor for the study-test method with the
easy list, as predicted by the encoding position, materialized. The only non-significance with the medium list, which was never explicitly expressed but inferable, seems to be supported. No difference between the two methods is apparent for the difficult list (Fig. 2, right). There seems, nonetheless, a slight edge (less errors) for the study-test method over the anticipation method with the easy and medium lists. However, even if analyses were limited to these two lists, performances between the two methods were still too small to be reliable: $F(1, 48)=2.574, \text{MS}_{\epsilon}=1.119, p>.10$, with no interactions at all $F(1, 48)=0.002$. The only significance for that analysis was the verification of the learning difficulty levels, $F(1, 48)=14.138, p<.001$.

In clear contrast, the present results seem to provide considerable support for the retention interval (RI) model. The model predicts little difference between anticipation and study-test methods when encoding, therefore learning is difficult, on the grounds that only a few items are learned and enter STM. With the absence of sufficient critical items in STM, the differential loss over differential retention intervals between the two methods would not lead to any tangible differences thereof. Indeed, the two methods were practically the same when learning was difficult (Fig. 2, right). Similarly, when learning is very easy, a great majority of them are quickly learned and possibly over-learned, and established in LTM in no time, leaving few critical items behind in STM. Here again, an insufficient critical item situation results, and thus, little performance differences between the methods ensue. This prediction with the easy list was also well borne out, while the advantage of the study-test was not significant.

When learning is only intermediately difficult, the RI model expects the advantage for the study-test method, but the degree of that advantage may be either (a) significant, or (b) nonsignificant. With a moderately difficult list, substantial critical items are expected to be generated in STM. If a great majority of them fall into nonoverlap areas of the retention interval distributions (Fig. 1), the former, or (a) should result. If, however, a sufficient number of them fall into the overlap area instead, the latter, i.e. (b), is likely. Thus, the present nonsignificance can be interpreted as resulting from the latter or (b) situation.

The present data, then, seem to be in agreement with the RI model, but in disagreement with the encoding position. In search of possible support for the latter position that may be lost in analyses so far, we carried out elaborate data analyses from several additional perspectives, relevant to encoding difficulty, by examining every item throughout all subjects in each condition. First, we obtained proportions correct for the first time, given incorrect on all previous trials; the statistics give us a fair estimate for the conditioning probability or learning rate for every test subsequent to the study trial, together with the weighted mean, as seen in Table 1 for each method for each list. Consistent with the earlier analyses, there were large differences between the three lists; the conditioning probabilities lined
Table 1
Proportion correct for the first time given incorrect on all previous trials (conditioning probability), mean test trials to the first correct response and last incorrect response, and number of items never learned in each condition

<table>
<thead>
<tr>
<th>Difficulty level</th>
<th>Method</th>
<th>Proportion correct for the first time given incorrect on all previous trials (conditioning probability)</th>
<th>Weighted mean</th>
<th>Mean T trials to the first correct (SD)</th>
<th>Mean T trials to the last error (SD)</th>
<th>No. of items never learned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Preceding study trials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Easy</td>
<td>Anticipation</td>
<td>.208</td>
<td>.155</td>
<td>.207</td>
<td>.246</td>
<td>.173</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>260</td>
<td>206</td>
<td>174</td>
<td>138</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Study-test</td>
<td>.181</td>
<td>.244</td>
<td>.317</td>
<td>.255</td>
<td>.195</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>260</td>
<td>213</td>
<td>161</td>
<td>110</td>
<td>82</td>
</tr>
<tr>
<td>Medium</td>
<td>Anticipation</td>
<td>.092</td>
<td>.144</td>
<td>.124</td>
<td>.073</td>
<td>.110</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>260</td>
<td>236</td>
<td>202</td>
<td>177</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>Study-test</td>
<td>.092</td>
<td>.169</td>
<td>.163</td>
<td>.189</td>
<td>.105</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>260</td>
<td>236</td>
<td>196</td>
<td>164</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>260</td>
<td>255</td>
<td>244</td>
<td>234</td>
<td>218</td>
</tr>
<tr>
<td></td>
<td>Study-test</td>
<td>.008</td>
<td>.053</td>
<td>.080</td>
<td>.057</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>260</td>
<td>258</td>
<td>249</td>
<td>229</td>
<td>216</td>
</tr>
</tbody>
</table>

N = the number of cases.
Auditory information presentation

Table 2
Overt errors, total errors, and propositions of overt errors in each condition

<table>
<thead>
<tr>
<th>Difficulty level</th>
<th>Method</th>
<th>Overt errors</th>
<th>All errors</th>
<th>Overt error proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td>Mean</td>
</tr>
<tr>
<td>Easy</td>
<td>Anticipation</td>
<td>0.954 1.541 248</td>
<td>5.400 3.464 1.404 1.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study-test</td>
<td>0.588 1.071 133</td>
<td>4.685 3.079 1.218 1.26</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>Anticipation</td>
<td>1.050 1.684 273</td>
<td>7.173 3.484 1.865 1.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study-test</td>
<td>0.919 1.506 239</td>
<td>6.400 3.475 1.664 1.44</td>
<td></td>
</tr>
<tr>
<td>Difficult</td>
<td>Anticipation</td>
<td>0.642 1.335 167</td>
<td>8.962 2.801 2.330 0.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study-test</td>
<td>0.788 1.402 205</td>
<td>9.231 2.671 2.400 0.85</td>
<td></td>
</tr>
</tbody>
</table>

up according to the learning difficulty dimension, largest for the easy list, followed by medium, and difficult lists, in this order. Learning rates differed little between the methods for the difficult list; the study-test method showed trends to be better than the anticipation method, but only slightly so.

Next, we obtained the mean test trials for the first correct response per item within the list, and entered it in the third column from the right in Table 1, with the standard deviations in parentheses. The difficult list required a large number of trials, while the easy one needed the least number in order to score a correct response for the first time, with the medium placing in between. The same state of affairs was apparent in the mean test trials to the last error per item (see the second to the last column). Another consistent evidence is the number of items never learned throughout all tests by any of the subjects (last column in Table 1). Throughout all data analyses, the same consistent message emerged: (a) learning was a function of the difficulty level, and (b) little performance differences between the two item information presentation methods. The latter was particularly remarkable with the difficult list, with nonsignificant advantages for the study-test method with the easy and medium lists.

For the reason that auditory events occur sequentially in the time dimension, it is intuitively appealing to assume that the anticipation method that enjoys immediate feedback (Eq. 1) should have some sort of advantage. Skinnerian feedback position and Barch and Levine's encoding position notwithstanding, we failed to see such advantages so far. Are there any aspects of our data that may show that such is the case? Note that incorrect responses consist of two kinds, overt errors (wrong responses) and omissions (failure to respond). The analyses so far were based on both types combined. To uncover possible effects buried in the combined data, overt errors were separated in the first three data columns in Table 2. Total errors minus overt errors give us omissions. Entered in the last column, the proportion of overt errors was quite small, demonstrating that a great majority of errors were omissions. There were small, undecisive trends that the subject ventured more guesses, thereby committing more overt errors, when lists were easier. Differences in overt error proportions between anticipation and study-test methods were practically none for both medium and difficult lists, but they were somewhat more numerous for the anticipation method when the learning list was easy.

Given that learning difficulty is a key issue in this paper, we analysed data from still another vantage point. The difficulty level in the present study was objectively manipulated by the experimenter via a large-scale word frequency count by
Thorndike and Lorge. That difficulty level may, however, not correspond subjectively to a particular learner. For any difficulty level, selected a priori, some subjects may find it extremely difficult, while others may find it very easy. It is highly instructive, then, to consider individual differences and to examine data in terms of subjective difficulty levels from the standpoint of the subject.

To this end, data from the two best (lowest incorrect responses) and the two poorest (highest incorrect responses) performers were examined, and entered in the lower, or upper panels, respectively, in Fig. 3. As expected, learning difficulty levels were reconfirmed to have generated large effects, for both types of learners: \(F(2, 12)=86.201, p<.001\). Several other aspects of the Fig. 3 analyses enlighten us in a fascinating manner.

First, as evident from Fig. 3, by far the largest differences in this study occurred between fastest and slowest learners: \(F(1, 12)=279.184, MS_e=43.042, p<.001\), supporting the necessity of separate consideration of the subjects in addition to their mean scores.

Second, intriguingly enough, when the learners' ability was considered, overall performance differences between study-test and anticipation methods were significant: \(F(1, 12)=18.351, p<.01\). The present findings contrast sharply with nonsignificance when based on the group means in Fig. 2. The present significance, however, seems attributable almost exclusively to the best performers, for whom no list in the present study was extremely difficult; even for the difficult list (Fig. 3, bottom right), their error proportions were only on the order of .400 or less at the end of the experiment, which was equivalent to, or better than, the medium list for all subjects (Fig. 2, center). The highest performers learned the other lists even faster, particularly so with the easy list.

Making a dramatic contrast, no list was easy for the poorest learners, achieving only .350 and .600 in error probability at the end of the experiment even with the easy list (Fig. 3, top left). Reflecting these contrasting effects, interactions (methods x subjects) were significant: \(F(1, 12)=8.910, p<.05\). The subject variable also interacted significantly with the difficulty level: \(F(2, 12)=7.776, p<.05\), confirming qualitative differences between highest and lowest performers. The triple interactions, however, were not large: \(F(2, 12)=3.061, .05<p<.10\).

Consistent with the foregoing, when separate analyses were made for highest and lowest performers, respectively, the two item presentation methods did not differ among the poorest learners: \(F(1, 6)=3.139, MS_e=38.333, p>.10\); but did differ significantly among the best learners: \(F(1, 6)=15.260, MS_e=247.750, p<.01\). For both types of performers, again the
Auditory information presentation 21

difficulty levels produced significant main effects; $F(1,6)$s were, respectively, 89.383 and 39.852, each $p<.001$. Interactions (methods × difficulty) were large: $F(2, 6) = 5.148, p<.05$ among the poorest learners, suggesting that the degree of differences between the methods varied depending on the difficulty levels, but not noticeably among the best performers: $F(2, 6)$ being only 1.285, $p>.20$.

All the above data analyses clearly indicate powerful effects of the learning difficulty variable, be they based on the mean (Fig. 2), the best performers only (Fig. 3, bottom), or on the poorest ones only (Fig. 3, top). In Fig. 3, in particular, we observe fine grained information regarding performance differences between anticipation and study-test methods. Independent of the experimenter-determined difficulty level, any given list was very difficult for the poorest learners, while very easy for the best learners, albeit the difficulty variable had its orderly effects, with most errors in the difficult list, followed by medium and easy lists in this order.

When learning or encoding is easy, the encoding position expects better performances for the study-test method than for the anticipation method. An ideal situation to test this position is to examine the easy list data by the highest performers (Fig. 3, bottom left). This expectation was indeed borne out, but the degree of superiority with the study-test method seems unimpressive, as compared with medium and difficult lists (Fig. 3, bottom center and right), where subject-learning difficulty may, in a sense, be regarded as only either intermediately easy or difficult.

Conversely, a difficult list learned by the lowest performers created a situation where encoding was extremely difficult, an ideal situation to test another prediction by the encoding position, i.e., the large superiority of the anticipation method over the study-test method. This prediction, however, did not turn out to be the case. Failing in this ideal situation, the other situations were still less encouraging. As seen in the top panels in Fig. 3, no situations, where learning/encoding was difficult, produced significantly better performances for the anticipation method in support of the encoding position.

The total failure to replicate Barch and Levine's findings of the superiority for the anticipation method in any of the present situations which were carefully examined by both objective and subjective learning and encoding difficulty, is a mystery, in spite of extensive data analyses achieved from every conceivable vantage point, unless their group testing situation invited artifacts that were unlikely in the present individual testing situation. It is, nonetheless, quite clear, that in the absence of empirical support, their position concerning the supremacy of the anticipation method seems to lack any persuasive power.

In contrast, the retention interval (R1) model predicts little differences between the two methods when learning is very difficult, as with the poorest learners (Fig. 3, top panels), particularly so when the difficult list was learned (Fig. 3, top right, extremely difficult). As the severity of difficulty lessens, as with medium and easy lists for the lowest performers, the study-test method is predicted to increase its degree of advantage. When learning is intermediately difficult (or intermediately easy), that advantage over the anticipation method may be either large or small, depending on where a sufficient number of critical items fall into non-overlap or overlap areas, respectively, in the retention interval distributions (Fig. 1). Indeed, significant superiority was produced under the study-test method vis-à-vis the anticipation method with both difficult and medium lists learned by best learners (Fig. 3, bottom right and center, de facto only intermediately difficult). When learning is extremely easy, however, the model predicts only a small
advantage for the study-test method, if any. In support of the model, performance differences were considerably smaller when the easy list was learned by the highest performers (Fig. 3, bottom left).

In a sense, the six panels in Fig. 3 can be regarded as subjective learning difficulty lined up in the consecutive rank-order from the most to the least difficult from top right to top left, and then from bottom right to bottom left, as numbered at the lower left corner of each panel. In such a situation, the RI model expects the advantage of the study-test method over the anticipation method to be negligible in the most difficult situation (1), gradually becoming larger and larger (2 to 3, and then to 4), reaching a maximum somewhere in the intermediate stage (in this case, 5, the medium list learned by best performers), and then again decreasing in magnitude (6, the easy list learned by best performers). Should there be a much easier situation than this (6), the advantage for the study-test method is likely to be even smaller (cf. Izawa, 1974).

The data in Fig. 3 seem to support predictions from the RI model to a minute detail, indeed. Considering all aspects of the present study, it seems fair to conclude that the retention interval model holds up well under auditory information processing situations also.

References


(Received Nov. 4, 1982; accepted Jan. 21, 1984)