"Posner's Paradigm" and Categorical Perception: A Negative Study

Bruno H. Repp*

ABSTRACT

A reaction-time study was conducted with four synthetic syllables from a "place continuum" (/bae/-/dae/-/dae/-/gae/). A special counterbalanced design was used to assess the effect of acoustic similarity on reaction time. The study included a "same-different" and a classification task, two different temporal delays between the syllables, and binaural versus dichotic (i.e., alternating monaural) presentation. However, no effects of auditory similarity were found, which contradicts a recent study by Eimas and Miller (1975) that used similar stimuli.

INTRODUCTION

Posner and Mitchell (1967) introduced an experimental paradigm that has led to some of the most elegant and successful research in visual information processing (e.g., Posner, 1969; Posner, Boies, Eichelman, and Taylor, 1969). The task consists in judging whether two letters are the same or different, with reaction time as the dependent variable. The two letters can be either identical (AA) or different (AB); in addition, they can have the same name but be physically different (Aa). The subjects are instructed to respond "same" when the two letters have the same name, and "different" otherwise. The principal finding is that "same" reaction times are faster for pairs that are physically identical (AA) than for pairs that are physically different (Aa). This suggests that physically identical letters can be matched at an earlier "node" in processing, which uses purely visual information, while name matches in the absence of physical identity take place at a later processing stage. Posner and Keele (1967) introduced temporal delays between the two stimuli, in order to find out whether the visual information that leads to the relative advantage for physical matches is subject to decay. They found a steady decline of the reaction-time difference over the first 2 sec, suggesting that the visual information is held in a relatively short-lived store.

Similar paradigms have been profitably applied in speech perception (e.g., Springer, 1973; Cole, Coltheart, and Allard, 1974; Repp, 1976a). Perhaps the

*Also University of Connecticut Health Center, Farmington.

Acknowledgment: This research was made possible by the generous hospitality of Haskins Laboratories and its director, Alvin Liberman. The author was supported by NIH Grant T22 DEO0202 to the University of Connecticut Health Center.

[HASKINS LABORATORIES: Status Report on Speech Research SR-45/46 (1976)]
most interesting of these studies is that of Pisoni and Tash (1974). They applied Posner's paradigm to the classical problem of categorical perception. It is well-known that initial stop consonants are easy to discriminate as long as they are perceived as belonging to different categories, but that acoustic differences within these categories are almost impossible to detect (e.g., Pisoni, 1971). It has been suggested that this phenomenon may be due to the rapid loss of auditory information from memory (Fujisaki and Kawashima, 1970; Pisoni, 1971, 1973). Pisoni and Tash (1974) presented two synthetic syllables in close succession, which could be either physically identical (e.g., /ba/₁-, /ba/₁), different acoustically but belonging to the same category (/ba/₁-/ba/₂), or belonging to different categories (/ba/-/pa/). The acoustic variable was voice onset time (VOT), the most important cue for the distinction between /ba/ and /pa/. The listeners were not aware of these acoustic variations and simply made "same-different" judgments with respect to the categories of the syllables. Pisoni and Tash found significantly shorter "same" reaction times for "physical matches" than for mere "name matches," just as Posner did. In addition, they found "different" reaction times to decrease with the acoustic difference between two syllables from different categories, which constitutes additional evidence for the availability of auditory information. (The corresponding finding in vision would be faster "different" latencies for Ab pairs than for AB pairs, a condition that has rarely been included and then has not yielded a positive effect—e.g., Besner and Coltheart, 1975). Pisoni and Tash suggested a two-stage processing model that allows for fast auditory matches to be conducted before slower phonetic (name) matches. Stimuli that are either identical or very different from each other may permit lower-level auditory decisions, while more ambiguous cases are decided at the phonetic level.

The Pisoni and Tash findings are especially interesting because, in contrast to other Posner-type tasks, the subjects are not aware of the physical differences within name categories; that is, no special "name match" instructions are necessary, as in the letter-matching task. Again, the question arises whether and how fast the auditory information is lost from memory. This may be investigated by varying the interval between the two syllables that are to be compared. I conducted such a study two years ago at the University of Chicago. Pairs of syllables from a VOT continuum (ranging from /ba/ to /pa/), as in Pisoni and Tash (1974) were presented at stimulus onset asynchronies (SOAs) between 0 and 3.3 sec. There was a clear effect of acoustic differences on "different" reaction times, which, moreover, did not decrease as the delay between the syllables increased. However, in contrast to the findings of Pisoni and Tash (1974), there was no clear evidence of any effect on "same" reaction times, which is the primary evidence for the availability of auditory information.

The effect on "different" reaction times could have a different explanation. It is well-known that it takes longer to classify stimuli that lie close to a category boundary than stimuli that are far from the boundary (Studdert-Kennedy, Liberman, and Stevens, 1963; Pisoni and Tash, 1974; Eimas and Miller, 1975; Repp, 1975). Pairs of acoustically very discrepant stimuli necessarily contain stimuli from the ends of the acoustic continuum, while pairs of acoustically

---

1Repp, B. H. (1974) Categorical perception, auditory memory, and dichotic interference. Unpublished manuscript. Copies of this paper are available from the author upon request. Some of the results were presented at the 89th meeting of the Acoustical Society of America in Austin, Texas (Repp, 1975).
more similar syllables (from different categories) contain at least one stimulus that is close to the category boundary. Therefore, the differences in categorization time for individual stimuli are confounded with the degree of acoustic discrepancy in between-category comparisons, and the effect on "different" reaction times could simply arise from the successive categorization and phonetic comparison of the two syllables. This explanation would also predict that the effect does not decrease with increasing SOA (Repp, 1975). This methodological objection does not apply to the "same" reaction times, since the individual stimuli contained in within-category comparisons can be properly counterbalanced (as in the experiment of Pisoni and Tash, 1974). One reason my study did not replicate theirs might have been the presentation of the two syllables to different ears, while Pisoni and Tash had presented them binaurally.

The present study asked the following questions:

1. Is the Pisoni-Tash effect obtained with syllables from a "place continuum," that is, with syllables whose acoustic differences lie in the initial formant transitions (and which are also perceived in a highly categorical fashion—see Pisoni, 1971)?

2. If so, does this effect decrease as the temporal separation between the syllables is increased?

3. Is there a difference in the magnitude of the effect when the syllables are presented to different ears rather than binaurally?

4. Does the Pisoni-Tash effect really reflect auditory comparisons between the two syllables, or does it perhaps consist in an influence of the first syllable in a pair on the categorization time of the second syllable? Entus and Bindra (1970) and Eichelman (1970), among others, have provided evidence that "same-different" reaction times and sequential effects in simple choice-reaction time are related and may reflect the same underlying processes. This was investigated here by including a condition in which the subjects had to classify the second syllable in each pair, ignoring the first syllable.

The study used a design that avoids the methodological problem with "different" reaction times discussed above. This design requires three categories on a single acoustic continuum, which is the case with a place continuum (/b/-/d/-/g/). Only four stimuli were used: /b/, /d/₁, /d/₂, and /g/. (The vocalic context, /æ/, was constant.) /d/₁ was acoustically closer to /b/ and /d/₂ was closer to /g/. The predictions were that "same" reaction times should be faster for /d/₁-/d/₁ and /d/₂-/d/₂ than for /d/₁-/d/₂ and /d/₂-/d/₁, and "different" reaction times should be faster for /b/-/d/₂ and /g/-/d/₁ (and their reverse orders) than for /b/-/d/₁ and /g/-/d/₂ (and their reverse orders). It can be seen that this design is completely balanced and therefore leads to unconfounded results for both "same" and "different" reaction times.

**METHOD**

**Subjects**

Eight paid volunteers (five women and three men) from the Haskins-Yale summer subject pool participated. Two of the men were left-handed. All had normal hearing and were relatively inexperienced.
Stimuli

Four synthetic syllables were produced on the Haskins Laboratories parallel resonance synthesizer. Two stimuli were supposedly good instances of /bae/ and /gae/, respectively, while the other two both sounded like /dae/ (cf. Repp, 1976b; the constant vowel will be omitted in referring to the stimuli). The two /d/ s, /d/₁ and /d/₂, differed only in the onset frequencies of the second formant (F₂), which were 1620 and 1772 Hz, respectively. Since the steady-state vowel had its F₂ at 1620 Hz, /d/₁ had a flat F₂, while /d/₂ had a falling transition. The third formant (F₃) fell from 3026 to 2862 Hz in both /d/ s. Likewise, /b/ and /g/ differed only in their F₂-transitions (starting at 1232 and 2156 Hz, respectively) and had identical F₃-transitions (starting at 2180 Hz). All syllables were of 280-msec duration, with 15 msec of prevoicing, no bursts, and a constant fundamental frequency (114 Hz).

Of the sixteen possible ordered pairs of the four syllables, /b/-/g/ and /g/-/b/ were omitted and /b/-/b/ and /g/-/g/ were duplicated instead. This resulted in an equal number of "same" and "different" pairs. Four stimulus lists were recorded. Each contained 80 syllable pairs, viz. 5 blocked replications of the 16 pairs, randomized within blocks. In the first and fourth lists, the SOA was 500 msec; in the second and third lists, the SOA was 2 sec. Each stimulus pair was preceded by a 100-msec warning buzz that came on 500 msec before the first syllable. The two syllables in a pair were recorded on separate channels. The interpair interval was 3 sec.

Procedure

Each subject participated in two 90-minute sessions on different days. The sequence of the two tasks was counterbalanced across subjects. In one session, the subject was instructed to judge whether the two syllables in a pair were the same or different by pressing the response key with the appropriate label (same-different task). In the other session, the instructions were to ignore the first syllable and to classify the second syllable as either "B" or "non-B," that is, "B or G" (classification task). The subjects were told that there were three syllables, /bae/, /dae/, and /gae/. In the classification task, they were informed that /b/ and /g/ never occurred together in a pair but that, apart from this, the first syllable provided no clue about the second syllable. The subjects were encouraged to be as fast and as accurate as possible. A practice series of 32 pairs at SOA=500 was presented at the beginning of each session.

In each session, a subject listened to the experimental tape twice, once binaurally and once dichotically (i.e., with the warning tone and the first syllable in one ear and the second syllable in the other ear; "dichotic" is used here in the wider sense of "different—but not necessarily simultaneous—inputs to the two ears"). The sequence of the two presentation modes was counterbalanced across subjects, but it was the same in both sessions for a given subject. Which ear received the first syllable in the dichotic condition was also counterbalanced across subjects, but fixed for each individual subject.

The tape was played back from an Ampex AG-500 tape recorder through an amplifier/attenuator to Telephonics TDH-39 earphones. Playback intensity was approximately 88 dB SPL (peak deflections on a voltmeter). Dichotic and binaural presentation modes were established by means of electronic switches. Reaction times were recorded on a Hewlett-Packard 522B electronic counter, which
was started by the onset of the warning tone and stopped by depressing either response key. Appropriate constants were subsequently subtracted from all latencies, so that they were measured with reference to the onset of the second syllable in a pair. The subject used both hands for responding, one for each key. Hand-response assignment was again counterbalanced across subjects.

At the end of the second session, each subject was questioned whether he or she had noticed anything about the stimuli that had not been mentioned in the instructions, and subsequently, given that there were two different versions of one syllable, which of the three syllables this might have been. Of the eight subjects, three showed no awareness whatsoever (they also had the lowest error rates), three stated that /d/ and /g/ were more difficult to discriminate than /d/ and /b/, and the remaining two claimed hearing /bla:/ on occasion (these two had the highest error rates). No subject guessed correctly that two /d/s were involved; all guesses were either "two /b/s" or "two /g/s."

The first step in the data analysis was to calculate the medians of the reaction times for the five replications of each stimulus pair in each list, omitting errors. Further analysis was in terms of the means of these medians.

RESULTS

Errors

Since latencies cannot be fully understood without taking the error pattern into consideration, the errors shall be presented first. There was great individual variation: average error rates ranged from 2.0 to 17.7 percent. As pointed out above, they were positively related to the degree of awareness the subject had of the presence of four stimuli. However, no subject made consistent misclassifications or misjudgments of certain stimuli; several changed their error trends during the course of a session.

The overall error rates in the two tasks were similar (same-different: 9.3 percent; classification: 9.2 percent). There was a tendency to commit more errors at the shorter SOA (10.2 percent) than at the longer one (8.3 percent). The most striking difference was between the dichotic and binaural conditions, with almost twice as many errors in the former (11.8 percent) than in the latter (6.7 percent). As might be expected, this difference was more pronounced in the same-different task, but it was also present in the classification task.

In the classification task, /d/ stimuli were misclassified more often than /b/ and /g/ stimuli (14.2 vs. 4.2 percent). Most of the errors on /b/ and /g/ were probably due to inattention and/or hand-response confusions that were not separately identified in this study (i.e., subjects were not asked to "correct" their own errors). /d/ was misclassified more often than /d/ (18.8 vs. 9.6 percent). Misclassifications of /d/ as /b/ or as /g/ were not distinguishable in this task, but it seems likely that /d/ was mostly confused with /b/, and /d/ with /g/. The nature of the preceding stimulus seemed not to make any difference.

In the same-different task, two interactions similar to those predicted for the latencies were expected, since errors and latencies tend to be positively correlated in same-different tasks. Incorrect "same" judgments should have been more frequent in /d/-/b/ and /d/-/g/ (and reverse) pairs than in /d/-/g/ and
/d/₂-/b/ (and reverse) pairs, and incorrect "different" judgments should have been more frequent in /d₁-/d₂ (and reverse) pairs than in /d₁/-/d₁ and /d₂-/d₂ pairs. Both trends were present but not very pronounced (13.3 vs. 9.9 percent, and 9.5 vs. 8.1 percent, respectively). Most surprising was the fact that /b/-/g/ (and reverse) pairs did not show a substantially lower error rate than other pairs (9.2 percent). Clearly, then, the same-different judgment errors could not be predicted from the classification errors, which were more than three times higher for /d/ stimuli than for /b/ and /g/ stimuli. This indicated either that the two stimuli in a pair were matched before complete classification, or that the classification of the second syllable was not independent of the preceding syllable. No such dependence was evident in the classification errors, however.

**Latencies**

It was anticipated that the latencies of subjects with high and low error rates might have to be considered separately, because of the positive correlation between errors and latencies that is usually found. However, this proved to be unnecessary, since the results were completely negative, overall and for each individual subject. While some effects may not have reached significance because of the small number of subjects, the differences of principal interest were clearly not obtained.

Consider first the same-different task. The results for "same" judgments are shown in the first three columns of Table 1. In three of the four conditions, the predicted interaction (the difference between the second and third columns) was in the expected direction (positive) but small; in the fourth condition, binaural at SOA = 2000, it was in the opposite direction. No difference reached significance, and no individual subject showed a clear pattern. The more consistent trend toward longer reaction times at SOA = 2000 than at SOA = 500 also fell short of significance.

<table>
<thead>
<tr>
<th>Mode</th>
<th>SOA</th>
<th>/b+/b/</th>
<th>/d₁+/d₁</th>
<th>/d₂+/d₂</th>
<th>/d₁+/d₂</th>
<th>/b+/d₁</th>
<th>/b+/d₂</th>
<th>/g+/d₁</th>
<th>/g+/d₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dichotic</td>
<td>500</td>
<td>523</td>
<td>536</td>
<td>552</td>
<td>547</td>
<td>547</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>560</td>
<td>565</td>
<td>581</td>
<td>582</td>
<td>589</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binaural</td>
<td>500</td>
<td>526</td>
<td>521</td>
<td>542</td>
<td>562</td>
<td>564</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>550</td>
<td>569</td>
<td>546</td>
<td>586</td>
<td>581</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The last two columns of Table 1 show the "different" latencies. Here, it was predicted that the latencies in column 4 would be shorter than those in column 5. Clearly, there was no difference at all. The only consistent tendency seems to be again longer latencies at the longer SOA, but it did not reach significance. It will also be noted that "same" latencies were somewhat faster than "different" latencies, a difference that is commonly found and was not tested for significance.

158
There were two effects that did reach significance: the Mode × Order and Mode × "B vs. G" interactions \( (p < .01 \text{ and } p < .05, \text{ respectively}) \). They are shown in Table 2.

<table>
<thead>
<tr>
<th>Mode</th>
<th>/b/-/d/</th>
<th>/g/-/d/</th>
<th>/d/-/b/</th>
<th>/d/-/g/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dichotic</td>
<td>564</td>
<td>553</td>
<td>584</td>
<td>566</td>
</tr>
<tr>
<td>Binaural</td>
<td>567</td>
<td>596</td>
<td>554</td>
<td>574</td>
</tr>
</tbody>
</table>

It can be seen that, in the dichotic condition, pairs in which /d/ occurred first tended to have longer "different" latencies than pairs in which /d/ occurred second, and pairs containing /b/ tended to have longer latencies than pairs containing /g/. The opposite was true in the binaural condition. These effects are difficult to interpret.

We turn now to the classification condition. The results for those syllables that were preceded by a syllable from the same category are shown in the first three columns of Table 3.

<table>
<thead>
<tr>
<th>Mode</th>
<th>SOA</th>
<th>/b/-/b/</th>
<th>/g/-/g/</th>
<th>/d/-/d/</th>
<th>/d/-/d/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dichotic</td>
<td>500</td>
<td>543</td>
<td>547</td>
<td>518</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>600</td>
<td>570</td>
<td>594</td>
<td></td>
</tr>
<tr>
<td>Binaural</td>
<td>500</td>
<td>533</td>
<td>544</td>
<td>534</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>562</td>
<td>532</td>
<td>537</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mode</th>
<th>SOA</th>
<th>/b/-/d/</th>
<th>/g/-/d/</th>
<th>/d/-/b/</th>
<th>/d/-/g/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dichotic</td>
<td>500</td>
<td>567</td>
<td>552</td>
<td>505</td>
<td>504</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>619</td>
<td>608</td>
<td>595</td>
<td>570</td>
</tr>
<tr>
<td>Binaural</td>
<td>500</td>
<td>564</td>
<td>563</td>
<td>527</td>
<td>535</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>578</td>
<td>587</td>
<td>530</td>
<td>540</td>
</tr>
</tbody>
</table>

Again, there is no clear evidence for the expected effect (faster latencies in column 2 than in column 3). In the dichotic mode, there was a notable tendency to be slower at SOA = 2000 (not significant), which indicates that the preceding syllable was not completely ignored. The results for syllables preceded by a syllable from a different category are shown in the remaining columns of Table 3. Again, there is no obvious difference between columns 4 and 5, and
columns 6 and 7. However, /b/ and /g/ classification was faster than /d/ classification, and the latencies were again longer at SOA=2000. No effect reached significance. A facilitating effect of a preceding stimulus from the same category may be noted, but only for /d/ classification.

**DISCUSSION**

This experiment provided no evidence for the availability of auditory information in the comparison of syllables from a "place continuum." Although only eight subjects were tested, their results make it quite unlikely that any significant effects would emerge in a larger sample, except for the trivial findings that latencies increase with SOA and that "same" latencies are faster than "different" latencies. Note that, although the data for "same" latencies in Table 1 may be suggestive of a small effect, no individual subject showed a clear pattern of results, despite reasonably stable data (10 replications of each stimulus pair).

Of course, it is entirely possible that the results of Pisoni and Tash (1974) pertain only to differences in VOT, a temporal variable, while differences in formant transitions are not retained in auditory memory. However, a study conducted independently at about the same time as the present experiment by Eimas and Miller (1975) did find a positive effect.

Their study is the more remarkable because it used stimuli from the identical place continuum, originally prepared at Haskins Laboratories by David Pisoni (see Pisoni, 1971). (The present /b/, /d/, and /d/ were their stimuli 1, 6, and 8, respectively—see their Table 1. Their continuum did not include /g/ stimuli.) They used a design similar to that of Pisoni and Tash (1974), counterbalanced for "same" pairs but not for "different" pairs. Miller and Eimas were aware of the alternative explanation for effects on "different" reaction times and emphasized the comparison of "same" reaction times for identical and nonidentical pairs. There were three SOAs (310, 460, and 1000 msec) that were randomized. At the intermediate SOA, which approximates the shorter SOA in the present study, they found a 44-msec difference in "same" reaction times and a 73-msec difference in "different" reaction times, both in the predicted direction. Moreover, the effect on "same" reaction times, but not that on "different" reaction times, decreased as SOA increased. This provides convincing evidence for the involvement of some auditory memory at short SOAs and for its decay over time. It also suggests that the effect on "different" reaction times probably does not reflect auditory memory but differences in categorization time for the component stimuli.

Eimas and Miller's study is elegant and well-designed, and their results must be taken seriously. It will require further research to clarify why the present study did not obtain the same effects, in the absence of any obvious flaws in design. Of course, if the effect of "different" reaction times is due to differences in categorization time alone, no effect should have been obtained in the present balanced design because such differences cancel out. Seen in this way, this portion of the present results even supports Eimas and Miller. However, the reason for the present failure to obtain an effect of acoustic differences on "same" reaction times remains obscure.
REFERENCES


