

Temporal Order Judgments in Speech: Are Individuals Language-Bound or Stimulus-Bound?*

Ruth S. Day[†]
Haskins Laboratories, New Haven

Abstract. Speech stimuli, such as BANKET and LANKET, were presented dichotically, with relative onset time varying over trials from 0 to [±]100 msec. When asked to report which phoneme led, subjects fell into two groups: those who performed well, and those who were misled by the temporal constraints on English. A tentative model of temporal order judgment is presented that suggests that there are two modes of listening: a linguistic mode and a nonlinguistic mode.

Introduction

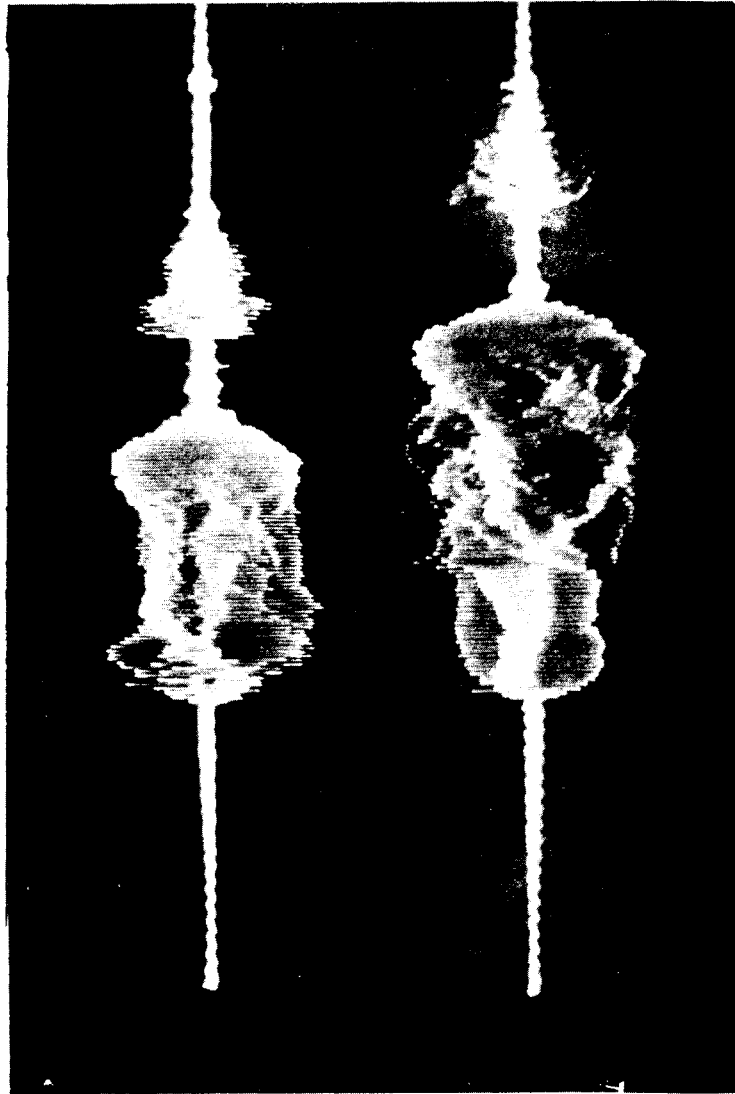
Previous studies of dichotic listening have emphasized the rivalry between the two ears. For example, when the digit TWO is presented to one ear over earphones, while at the same time the digit THREE is presented to the other ear, subjects typically report hearing TWO, or THREE, or both TWO and THREE. Fusion does not occur: no one reports hearing THRU or TEE. However, a study in the present series (Day, 1968) has shown that fusions can occur when the proper psycholinguistic variables are taken into account. For example, given BACK to one ear and LACK to the other, subjects typically report hearing the fusion, BLACK.

Method

The present experiment was designed to study the role of time cues in facilitating or retarding the fusion effect. Figure 1 shows a dual-beam oscilloscope photograph of a sample item. The top channel represents BANKET and the bottom channel represents LANKET. Both are real-speech samples that have been edited using the pulse code modulation system at the Haskins Laboratories (Cooper and Mattingly, 1969). This system permits the experimenter to do four things: 1) he can determine where an item begins and discard all that

*Paper presented at the Ninth Annual Meeting of the Psychonomic Society, St. Louis, November 1969.

[†]Also, Yale University, New Haven.



BANKET

LANCKET

FIG. 1

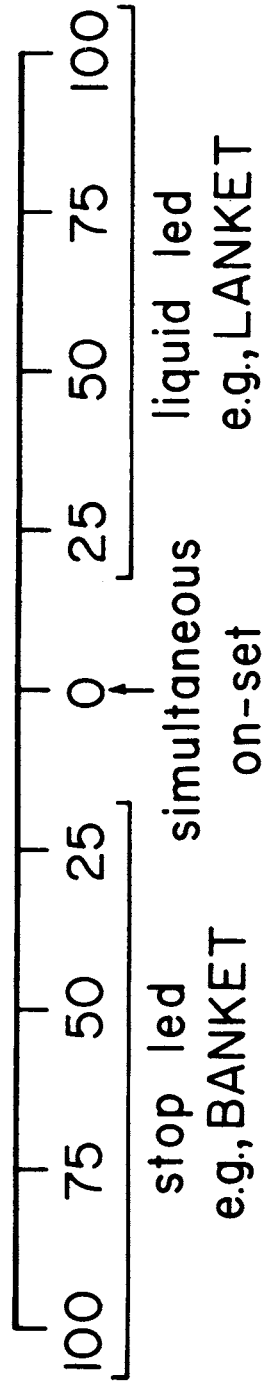
precedes that point; 2) he can determine where an item ends and discard all that follows; 3) he can equalize the over-all intensities of the two items so that they are equally loud; 4) finally, he can line up the onsets of the two utterances with accuracy on the order of 500 microseconds. Note that in this particular example of BANKET/LANKET both utterances begin at the same point in time. We will refer to this situation as the simultaneous onset case, or the 0-lead time case.

Figure 2 shows the general paradigm of the experiment. On one set of trials, BANKET began first by 25, 50, 75, or 100 msec. On another set of trials, LANKET began first by the same intervals. And on the final set, both utterances began at the same point in time. There were ten items, as shown in Figure 3. All involved initial stop and liquid consonants. Each stop consonant (/p,t,k, b,d,g/) was paired with the liquids /r,l/.¹ Thus, for the /pr/ cluster, the inputs PAHDUCT/RAHDUCT can be fused to yield PRODUCT; for the /pl/ cluster, PANET/LANET yields PLANET; for the /tr/ cluster, TEETMENT/REETMENT yields TREATMENT; and so on. All possible fusion responses were acceptable English words, although other experiments (e.g., Day, 1968, Exp. II) have shown that subjects will report fusions that are nonwords, e.g., GORIGIN/LORIGIN yields GLORIGIN. In addition, all inputs were nonwords. (While "wordness" does correlate with meaningfulness, the notion should not be taken too seriously: although BANKET is not an acceptable word, it does answer the question, "What shall I do with the money?" and hence, it is in some sense meaningful.)

Results

The Effect of Relative Onset Time on Fusion Probability. We want to determine what the probability of fusion response is for each of the lead-time conditions. Will fusions occur more readily when the stop consonant (e.g., /b/) leads than when the liquid (e.g., /l/) leads? If so, we would expect that fusion response probability will be higher on the left side of the display in Figure 2 than on the right side. As shown in Figure 4, the obtained function was more or less a straight line. Perhaps fusion was somewhat more probable when the stop led by 75 msec, but in any event, it looks as if time cues per se were not affecting fusion levels. People were about as likely to hear BLANKET when LANKET led.

¹/t1-/ and /d1-/ were excluded since these clusters do not occur in initial position in English.



Lead Times (in msec)

FIG. 2

Experimental Items

	r	l
p	PRODUCT	PLANET
t	TREATMENT	
k	CRACKER	CLOSET
b	BREAKFAST	BLANKET
d	DREADFUL	
g	GREEDY	GLEAMING

FIG. 3

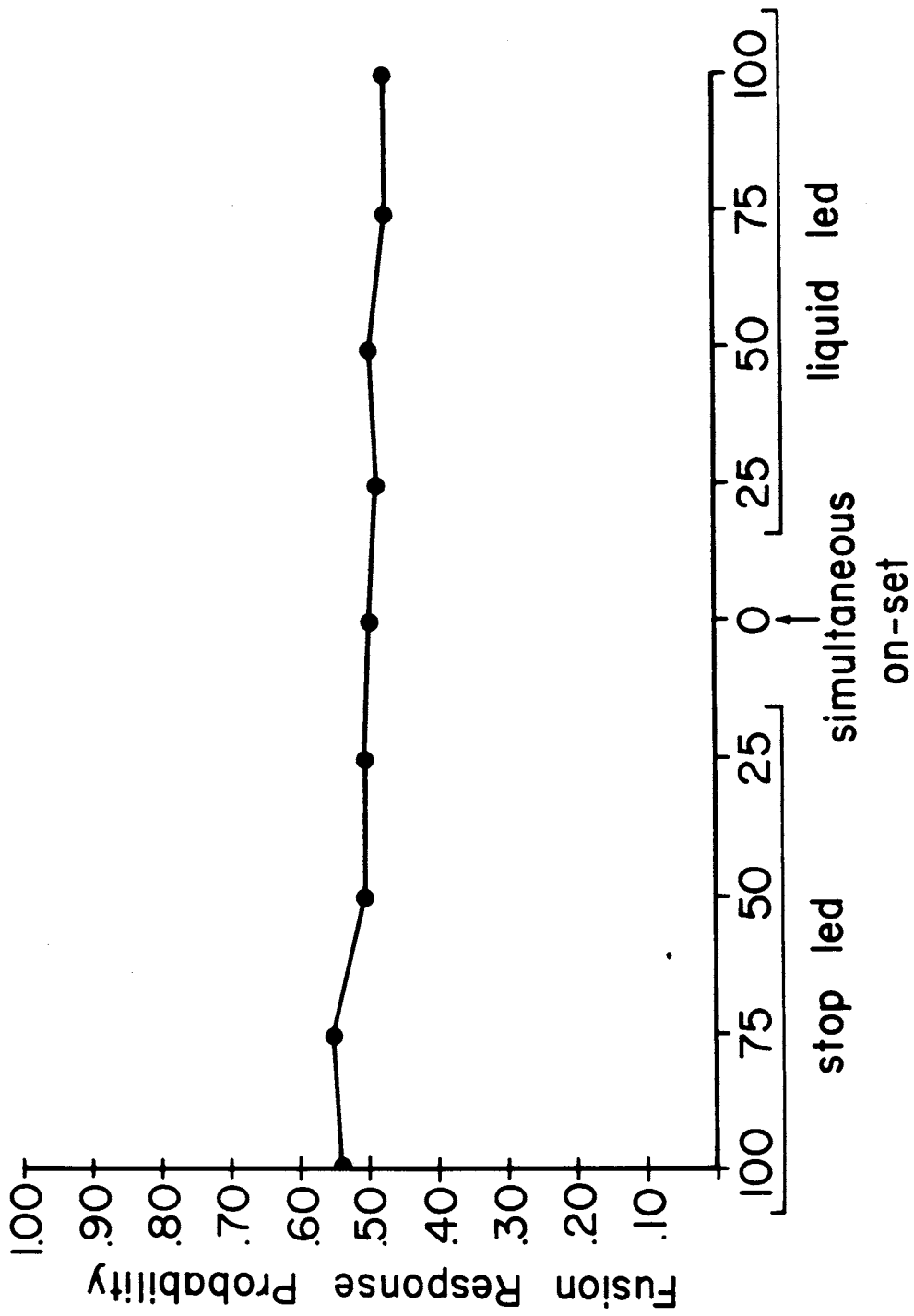


FIG. 4
Phoneme Lead Times (in msec)

Fusion Rates over Subjects. A surprising set of findings emerged from the data based on the performance of individual subjects. Each subject was given a score that reflected how often he fused. This was simply the proportion of times he fused over all trials (180). Contrary to scores on most psychological tasks, fusion rates were not normally distributed (Figure 5). Instead, subjects fell into two groups: those who fused most of the time ("high fusers") and those who fused relatively infrequently ("low fusers"). This experiment with sixteen right-handed subjects has been repeated with sixteen left-handers, and the fusion results are comparable. So the addition of more subjects makes the bimodal distribution even more striking.

Temporal Order Judgments (TOJ). Up to this point, we have been discussing the fusion task. In this task, subjects were asked to report out loud whatever they heard: one word, two words, real words, or nonsense words. There was a second task: temporal order judgment (TOJ). Here, subjects were asked to write down the first sound they heard on every trial. For example, if the first sound they heard was /b/, as in BOY, they wrote down the letter B. In this task we wanted to determine how well subjects could determine which phoneme led as a function of the lead conditions. Consider an individual subject as shown in the top display of Figure 6. When the stop (e.g., /b/) led by 100 msec, he performed perfectly: he always said that the stop led. As the stops lead decreased down to 25 msec, he always said that the stop led. Now consider trials where the liquid (e.g., /l/) led. When the liquid led by 25 msec, the subject performed miserably, that is, he always said that the stop led. As the liquid's lead increased, this subject's performance did not improve at all. He simply reported hearing the stop first, independent of the stimulus conditions. There were twenty observations per point, so the data for each subject are fairly stable. I should also point out that the point representing the 0-lead case is a special case: since neither item led, there is no "correct" temporal order judgment. The open circle at 0-lead indicates the percent of stop consonant responses so that we can assess the overall level of the subject's bias.

Now let's look at another subject as shown in the lower part of Figure 6. He looks very much like the first subject. Note that neither subject improved with increased lead time on either side of the continuum. There were about four more subjects who performed like those of Figure 6. There were other subjects who showed the same over-all effects, but did show a slight increase in performance at the longer leads; nevertheless, their performance on liquid leads never rose above chance.

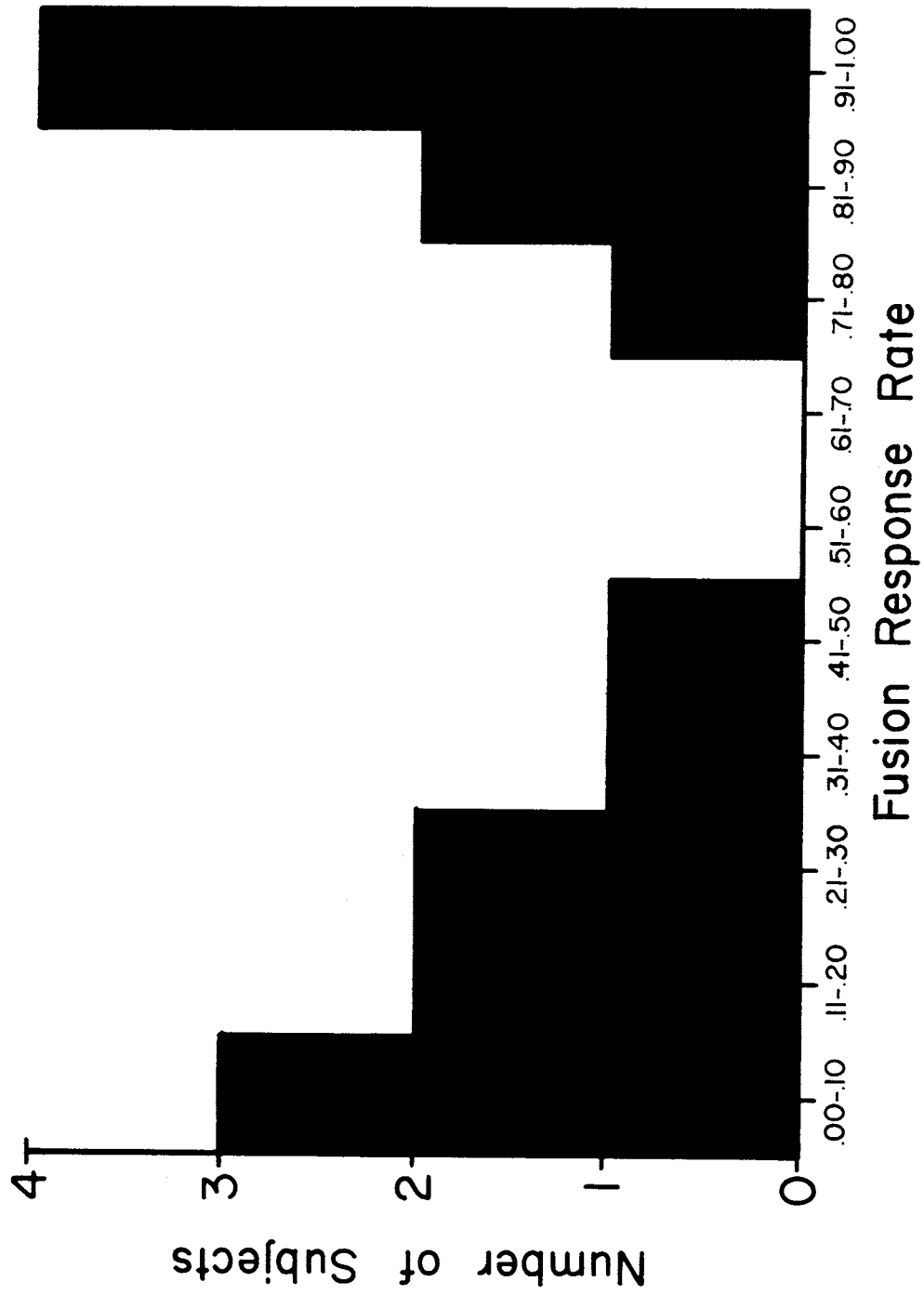


FIG. 5

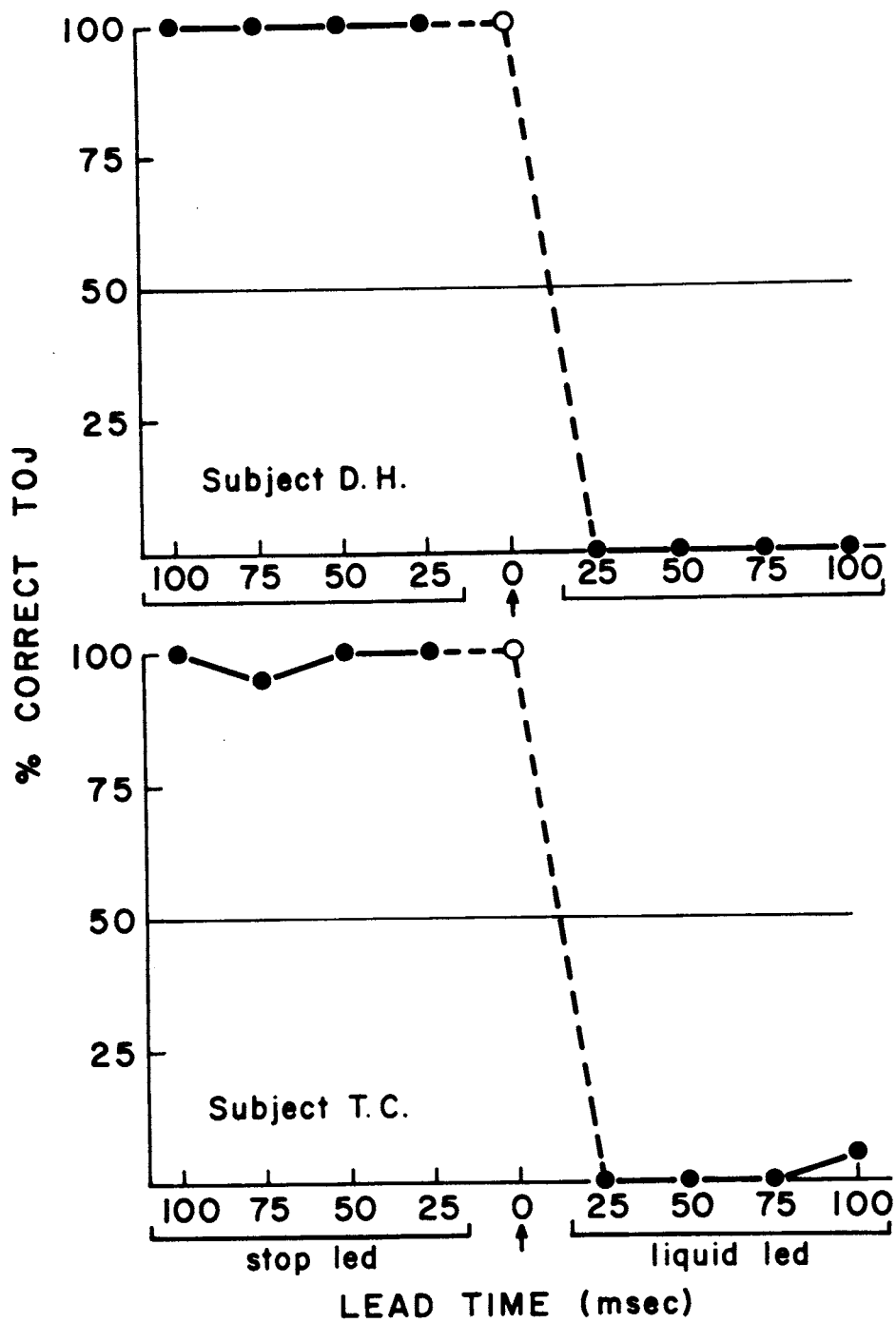


FIG. 6

There were some radically different subjects. Consider those shown in Figure 7. When the stop led, they correctly identified it as leading. When the liquid led, they also correctly identified it as leading. Note that both subjects were sensitive to increased lead times on both sides of the continuum.

Figure 8 shows a schematic diagram of the two types of TOJ performance. The top display shows overall performance that is poor. Subjects here perceive the stop as leading, independent of the stimulus conditions. Further, they show no improvement with increased time leads. These subjects are wholly bound by the facts of the language: in English, (stop + liquid) can occur in initial position, but (liquid + stop) cannot. These facts bias subjects against hearing the phonemes in their given order. We will call them "language-bound" subjects (for lack of a better term). In the bottom display of Figure 8, overall performance is good. Subjects here can tell which stimulus led, and they are sensitive to increased time leads. Since their responses do reflect the stimulus condition, we will call them "stimulus-bound."

Relation of the Fusion and TOJ Tasks. A brief review is in order. On the fusion task, we found two groups of subjects: high fusers and low fusers. On the TOJ task with the same subjects, we found two groups of subjects: those who performed poorly (language-bound) and those who performed well (stimulus-bound). Question: Is there any way to predict how a subject will do on the TOJ task given that we know he is a high fuser or low fuser? Thus we want to correlate performance on the two tasks for the same subjects. Figure 9 shows the scatter diagram for this relationship. Along the ordinate is each subject's TOJ accuracy.² Not only is there a negative correlation, but subjects tend to cluster into two groups: those who are high fusers and poor temporal order judges and those who are low fusers and good temporal order judges.

Discussion

Ordinarily, when we talk about "individual differences," we are trying to account for noisy data. Here, however, the individual difference data suggest that there may be two different types of language perceivers. We have

²The score used here was percent correct on liquid-leading items. Several other scores have been used, and all give essentially the same results.

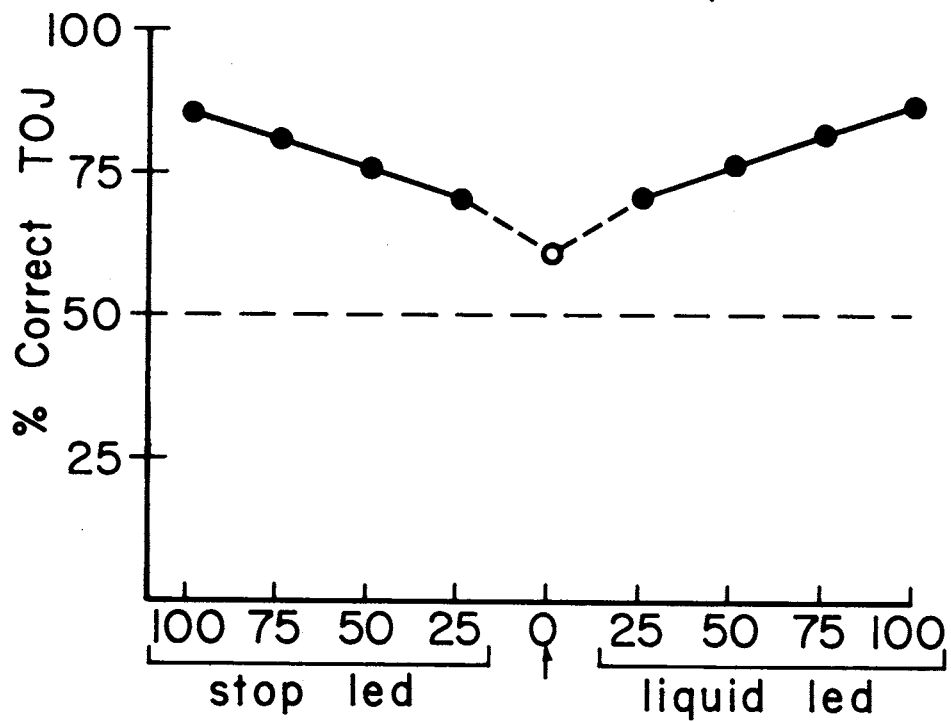
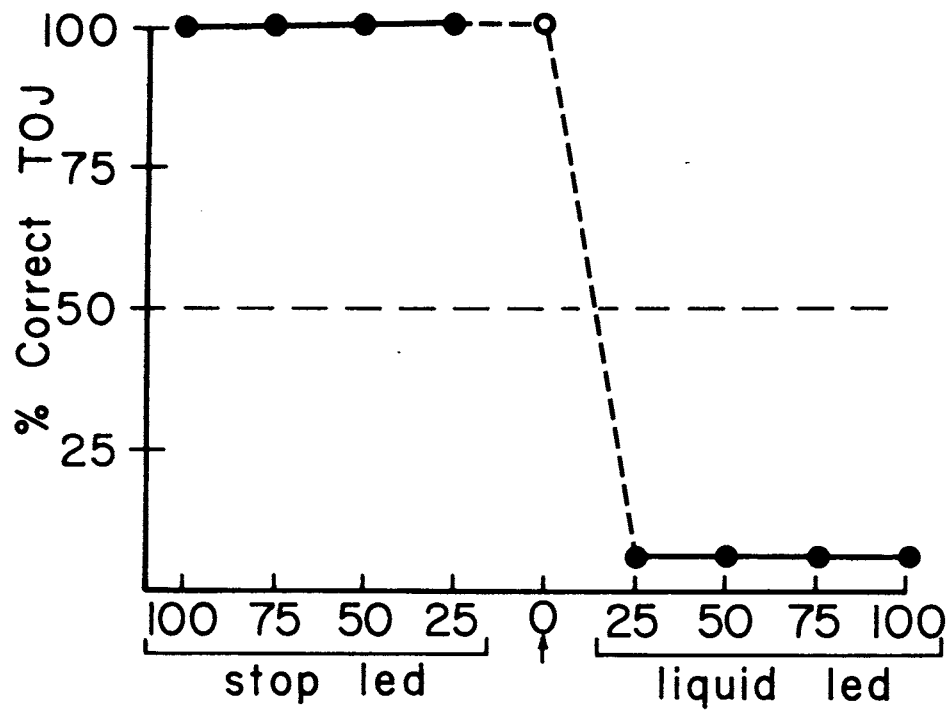


FIG. 8

noticed other differences about the two groups. At the end of the experiment, language-bound subjects are often surprised to learn about the nature of the stimuli and still hear fused clusters even when they are told that there are none on the tape. On the other hand, stimulus-bound subjects can usually tell the experimenter exactly what is on the tape. There are further questions to ask: Will these two groups retain their identities on other speech perception tasks? What about auditory short-term memory tests? We are currently identifying subjects in each category and giving them a battery of relevant tests.

A preliminary and tentative model to describe how subjects make temporal judgments in the present experiment is given in Figure 10. The model is to be used only as a point of departure: I do not know how many boxes there should be, nor how they should be arranged, nor which way all the arrows should go. However, the model does serve as a way to begin thinking about the processes involved. Consider first the analysis stage. At some point, subjects do analyze the stimuli into phonemes. They know that they are deciding between /b/ and /l/, or between /p/ and /r/. At a later stage, synthesis work must be done. That is, the phonemes must be arranged into some order. Before a subject can give a response, the results must be related to past experience with the language, perhaps by way of a linguistic filter or similar device. The filter operates on the basis of the sequential dependencies of phonemes in the language.³ For example, if LBANKET emerges from the synthesis stage, it has difficulty in passing through the linguistic filter and is therefore returned to synthesis for new ordering. If the output is BLANKET, it can pass through the filter, and hence the subject reports hearing /b/ first. The filtering system may have different bias levels across subjects, which would account for the obtained individual differences.

The discussion of temporal order judgment thus far has involved processing of a linguistic nature. However, before the analysis work is done, certain acoustic decisions can be made. For example, there may be a simple detection, a decision that a signal is on, and further, a decision concerning which ear received the signal. A subsequent experiment has been performed

³Perhaps the linguistic filter does not come after synthesis is completed; instead, the synthesis stage itself may have preset probabilities for various sequential dependencies. But this distinction is not crucial for the present discussion.

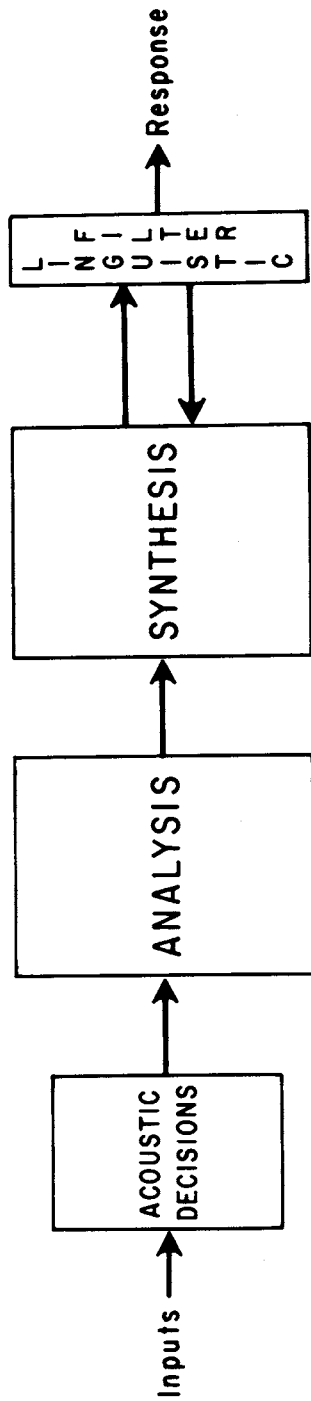


FIG. 10

(Day and Cutting, 1970) in which subjects indicated which ear led. Performance on the ear task was much better: subjects were highly accurate, even though they were language-bound on the phoneme task.

At present, we cannot be sure that acoustic decisions of this sort are made primarily at an early stage. Nor can we assert that they necessarily require less information processing. The claim at this point is simply that they do not require linguistic processing. When subjects judge temporal order at this nonlinguistic level, they perform well. It is only when they must do some linguistic processing, that is, analysis into phonemes, that they get into trouble. Thus, the model suggests that there are two types of processing that speech signals can undergo: linguistic processing and non-linguistic processing. Further, given that a subject's performance does not reflect the stimulus events when asked to identify the leading phoneme, the model suggests where the information is lost: namely during linguistic processing.

Two complementary research strategies have emerged from this work. The first, as described above, involves presenting speech stimuli and asking for temporal order judgments that require linguistic vs. nonlinguistic processing. The second involves presenting speech stimuli and analogous nonspeech stimuli, such as complex tones, and asking for temporal order judgments in the two situations. The results thus far are promising: the data support the notion that there are two general modes of auditory perception, a linguistic mode and a nonlinguistic mode.

Another approach involves investigation of critical cases within the linguistic system. Reversible clusters are of particular interest here (Day, 1970). Given the dichotic item TASS/TACK, subjects can easily determine which phoneme came last since both orders are permissible: TASK and TACKS.

Conclusion

In conclusion, we have seen that: 1) the effect of time cues on fusion and temporal order judgment is surprisingly negligible, and 2) individuals perform in two very different ways, some appear to be language-bound, while others accurately reflect the stimulus conditions. These results suggest that there may be two types of language perceivers in the population at large. A preliminary and tentative model of temporal order judgment was presented. It suggests that there are two modes of listening: a linguistic mode and a nonlinguistic mode.

The dichotic fusion technique is also useful in studying the role of the two cerebral hemispheres in the perception of speech. Therefore, we have extended these studies to other populations: left-handers, temporal lobe patients, and split-brain patients. But those accounts can wait for another occasion.

References

- Cooper, F.S. and Mattingly, I.G. (in press) Computer-controlled PCM system for investigation of dichotic speech perception. Paper presented at the 77th meeting of the Acoustical Society of America, Philadelphia, April, 1969. J. Acoust. Soc. Amer.
- Day, R.S. (1968) Fusion in dichotic listening. Unpublished Ph.D. thesis, Stanford University.
- Day, R.S. (1970) Temporal order perception of reversible phoneme cluster. Paper presented at the 79th meeting of the Acoustical Society of America, Atlantic City, April.
- Day, R.S. and Cutting, James E. (1970) Levels of processing in speech perception. Paper presented to the Tenth Annual Meeting of the Psychonomic Society, San Antonio, November.

