COGNITIVE PSYCHOLOGY 6, 41–60 (1974)

Acoustic Memory and the Perception of Speech[1]

C. J. DARWIN AND A. D. BADDELEY[2]
University of Sussex

The nature of acoustic memory and its relationship to the categorizing process in speech perception is investigated in three experiments on the serial recall of lists of syllables. The first study confirms previous reports that sequences comprising the syllables, *bah, dah,* and *gah* show neither enhanced retention when presented auditorily rather than visually, nor a recency effect—both occurred with sequences in which vowel sounds differed (*bee, bih, boo*). This was found not to be a simple vowel–consonant difference since acoustic memory effects *did* occur with consonant sequences that were acoustically more discriminable (*sha, ma, ga* and *ash, am, ag*). Further experiments used the stimulus suffix effect to provide evidence of acoustic memory, and showed (1), increasing the acoustic similarity of the set grossly impairs acoustic memory effects for vowels as well as consonants, and (2) such memory effects are no greater for steady-state vowels than for continuously changing diphthongs. It is concluded that the usefulness of the information that can be retrieved from acoustic memory depends on the acoustic similarity of the items in the list rather than on their phonetic class or whether or not they have "encoded" acoustic cues. These results question whether there is any psychological evidence for "encoded" speech sounds being categorized in ways different from other speech sounds.

The perception of speech clearly presents specific problems, not encountered in other types of auditory perception. The phonetic message is carried in a complex code whose constraints are not those of any other class of sound (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). These are grounds enough for postulating a specialized speech perceiving mechanism. However, experimental evidence has been cited in support of a special processing mechanism which is required more for the perception of some speech sounds than for others. In particular it has been claimed (Liberman *et al.*, 1967) that stop consonants need

[1] We would like to acknowledge the help of S. Rudlin, A. Adlard, and M. Johnson in collecting some of the data. The synthesis program used was derived from one originally written by Penny Pickbourne; we are grateful to Professor A. M. Uttley for access to this program. Please address reprint requests to: C. J. Darwin, Laboratory of Experimental Psychology, The University of Sussex, Brighton, England BN1 9QY.

[2] Now at University of Stirling.

more of the services of this special processing mechanism than do vowels. These two phonetic classes are taken as extremes of a continuum of "encodedness", a term which describes the complexity of the relationship between the acoustic signal and its phonetic category.

Thus, stop consonants, which are described as "encoded", are cued by a very different acoustic signal depending on the context in which they appear. A /d/ before /i/ can be cued by a rising second formant transition, whereas a /d/ before /u/ can be cued by a falling second formant transition. This variety in the cues is much less pronounced for vowels if one considers the case of a single speaker, speaking carefully, but the distinction between stops and vowels becomes less clear-cut in rapid speech, where the cues for vowels become dependent on the surrounding consonants (Lindblom & Studdert-Kennedy, 1967) or where there is a difference in the acoustic signal when two different speakers articulate the same vowel (Ladefoged & Broadbent, 1957).

The experimental evidence for suggesting that this dimension of encodedness is perceptually significant rests on a number of paradigms. Stop consonants, for example, show more clearly the ideal of categorical perception than do isolated vowels (Liberman, Harris, Hoffman, & Griffith, 1957; Fry, Abramson, Eimas, & Liberman, 1962). They show a greater tendency than vowels to be reported more accurately from the right than the left ear (Shankweiler & Studdert-Kennedy, 1967) and are more susceptible to dichotic backward masking than vowels (Studdert-Kennedy, Shankweiler, & Schulman, 1970; Darwin, 1971b). Experiments on other classes of speech sound also indicate that encodedness may be perceptually significant. Dichotically presented fricatives only show a right-ear advantage for recall of place of articulation when the encoded formant transitions are present as cues as well as the less encoded steady-state friction (Darwin, 1971a). Similarly, Cutting (1973) has shown that laterals (/r,l/) which lie between stops and vowels on the encodedness dimension show a right-ear advantage which is also intermediate between the two. Thus, there appears to be a correlation between the amount of acoustic restructuring or encoding that a class of phonemes undergoes in the speech of a single, careful speaker, and the degree to which that class of phonemes shows various behavioral effects such as categorical perception, dichotic masking, and a right-ear advantage. This relation seems to hold even in experiments where, within the experiment itself, there is in fact a simple one-to-one relationship between the phoneme and the acoustic signal.

This emphasis on encodedness as the significant variable in experiments which show different results for different phonetic classes has not gone unchallenged. Fujisaki and Kawashima (1968) suggested that the

different results obtained for stops and for vowels in discrimination experiments (where stops show more nearly categorical perception) can be attributed to the greater persistence of vowels in an auditory short-term memory accessible to the discriminating process. For vowels a discrimination can thus be based both on the results of a phonetic classification and on some cruder representation in auditory memory, while the auditory memory for stops is of little use and the discrimination can be based only on on the phonetic classification. Their evidence for this hypothesis was that merely shortening the duration of vowels gave more nearly categorical perception. Their results have recently been amplified by Pisoni (1973), who used an ingenious variant of the traditional discrimination paradigm with success. The nub of their argument is that the different perceptual results obtained for stops and vowels are due, not so much to differences in the categorizing process as Liberman *et al.* maintain, but rather to differences in the availability from acoustic memory of the acoustic cues underlying the categorization. They supposed that the relatively long duration steady-state cues for the vowels were preserved better than the brief transitions which cued the stop consonants.

The issue of the role of auditory memory in speech perception has been raised in yet another paradigm due to Crowder (1971), but derived from a technique for studying acoustic memory first used by Crowder and Morton (1969). This technique requires the subject to recall immediately and in the correct order a list of seven digits. If the subject only sees the list, he makes many more recall errors in the final positions than if he had heard the list. This is the modality effect. The relative improvement in performance on the last item for auditory presentation is shown also in the fact that subjects make fewer errors on the last than on the penultimate position of the auditory list. This is called the recency effect and does not occur for visual presentation. A third, related effect can be observed when a redundant suffix, which the subject does not recall, is added to the end of the list. If this suffix is physically similar to the list items (in speaker's voice, amplitude, and location) the recency effect and the modality effect are abolished. But if the suffix is distinct from the list items in physical characteristics (a different voice, or extremely, a different tone) recall is virtually unaffected and the recency and modality effects remain intact. This difference in the effects of different suffixes is called the suffix effect. There is no suffix effect for visual presentation.

Crowder and Morton (1969) interpreted these various effects as due to a form of acoustic memory, called precategorical acoustic storage (PAS), which can hold sounds for a short time and can be used to aid

recall. The modality effect suggested that PAS was an auditory store rather than one shared with vision, while the suffix effect suggested that it was a precategorical store on the grounds that the efficacy of a suffix depended only on its resemblance to the list items along crude physical dimensions. Furthermore, since delaying the suffix decreased the suffix effect they suggested that the store decayed with time. The effect of the suffix appeared to be a displacement of the last list item from PAS.

Crowder's (1971) contribution was to show that the three effects which had given rise to the notion of PAS did not occur, if, instead of digits, syllables differing only in an initial stop consonant (*bah, dah, gah*) were used. He also showed that syllables differing in a final vowel (*bee, bih, boo*) did give the modality, recency, and suffix effects. Broadly similar results have been obtained independently by Smallwood and Tromater (1971) and Cole (1973). The former's data suggest (although there is no statistical support) that both the recency and suffix effects are smaller for lists made up of a vocabulary of letters differing only in a consonant (BCDGPTVZ) than for those differing in vowels as well (HJLNRXQY). Cole's study uses nonsense syllables composed of six or seven consonants and six or seven vowels and finds in general that the consonants show less recency effects than the vowels.

Crowder's (1971) interpretation of his own and these results was that PAS was selective in the material that it retained. Vowels are retained while consonants are not. Such selectivity seemed to Crowder to be incompatible with PAS being a tape recording of the stimulus, since a tape recording would preserve indiscriminately both consonants and vowels.

This conclusion depends very much on whether PAS is subject to any degradation with time. Crowder's own views on this have been mixed, but unless PAS is very different from acoustic memories revealed by other paradigms (e.g. Darwin, Turvey, & Crowder, 1972) it seems likely that some temporal degradation does occur. Given this assumption, the tape-recording analogy gains new life, since one might reasonably suppose that degrading a tape recording would destroy the distinction between say the consonants /b,d,g/ before it destroyed that between the vowels /i,I,u/. The extent to which this is true will naturally depend on the type of degradation, but there is at least evidence from perceptual confusions in white noise that the /b,d,g/ distinction disappears at a much higher S/N ratio than the /i,I,u/ distinction (Miller & Nicely, 1955; Pickett, 1957). Perhaps then, we can use the concept of acoustic discriminability to explain the absence of acoustic memory effects for a vocabulary of three stop consonants. Acoustic memory would thus be viewed as an initially tape-recording-like representation of the stimulus

which became degraded with the passing of time. This degradation might then render the memory less effective for finer auditory discriminations than for coarser.

However, a broader issue than the particular nature of the representation is involved, namely the relationship between this representation and the categorization process in speech perception. As we have already mentioned, the idea of acoustic memory has been used to explain some of the experimental results used as support for a special processing mechanism for certain classes of speech sound (Fujisaki & Kawashima, 1968). This has eroded the experimental evidence for such a processor. However, in a recent paper Liberman, Mattingly, and Turvey (1972) welcome Crowder's results as support for the idea of a special processor, since, they claim, "the special process which decodes the stops strips away all auditory information" (p. 329). They are suggesting that the properties of acoustic memory are in fact *dependent* on the special processor. This proposition, if justified, would clearly prevent acoustic memory from being used to attack experiments supporting this special processor for encoded sounds.

To prevent the special processing mechanism from being defined in an entirely circular way we clearly need to test which of these two accounts of acoustic memory is the more tenable. Is it a memory which holds material irrespective of its phonetic class rather like a tape recording might, but which becomes degraded with the passing of time? Or is it a memory in which only material which is not processed by some special processing mechanism can be held? If the former is the case, then we are free to use acoustic memory as an alternative explanation for those experiments on which a special processor for encoded speech sounds is based. If the latter is the case, then the special processor is by definition immune.

The experiments reported in this paper attempt to determine whether the availability of information in auditory memory is determined by the relative discriminability of the items which have to be distinguished or by some higher level property such as their phonetic class or encodedness.

EXPERIMENT 1

Crowder (1971) and Cole (1973) concluded that consonants show little evidence of being preserved in acoustic memory compared with vowels. Crowder used a very confusable set of consonants (/b,d,g/) and Cole, though he used more discriminable consonants, used a larger set (/d,s,m,θ,n,s,p/). If, as we suppose, these particular vocabularies have failed to show acoustic memory effects because of their acoustic

confusability, we should be able to show such effects by using a small vocabulary of distinct consonants. On the other hand, if acoustic memory effects are determined solely by whether the sounds used are consonants or not, a small vocabulary of distinct consonants should be no more likely to show acoustic memory effects than the vocabularies which Crowder and Cole used.

We use the modality and recency effects in this experiment to find whether a small vocabulary of discriminable consonants shows the effects of acoustic memory more than a similar set of confusable consonants. In all, the experiment has four vocabularies: (1) three stop consonants, (2) three vowels (these replicate Crowder, 1971), (3) three acoustically distinct consonants in syllable-initial position, and (4) the same in syllable-final position. We have varied the position of the distinct consonants in the syllable since this variable was confounded with phonetic class in Crowder's study.

Method

Subjects read, either silently or aloud, typed lists of seven items. Each list was exposed for 2 sec in the window of a Forth Instruments memory drum. The subject then had to write down the list in the correct order, guessing if necessary and without going back to correct mistakes. He had 12 sec for recall when the advancing of the drum warned him that the next trial approached. Four different vocabularies were used for the lists: initial stop consonants (*bah, dah, gah*), final vowels (*bee, bih, boo*), initial dissimilar consonants (*ga, ma, sha*), and final dissimilar consonants (*ag, am, ash*). Fifteen consecutive trials formed a block, within which the vocabulary and the mode of reading (aloud/silent) was constant. Each of 16 undergraduate subjects took all eight blocks in a predecessor-balanced Latin square design (Williams, 1949).

Results

Serial position curves, giving the mean error probability under each of the eight conditions and seven serial positions, are shown in Fig. 1. The stop consonant and the vowel data confirm Crowder's findings: The vowels show both recency ($T = 5$, $n = 12$; $p < .01$) in the last serial position of the read aloud condition and a modality effect in the last serial position ($T = 0$, $n = 14$; $p < .001$), while the stops show neither recency ($T = 33½$, $n = 14$; $p > .25$) nor a modality effect ($T = 47$, $n = 15$; $p > .4$). The acoustically dissimilar consonants, however, fail to confirm Crowder's conclusion that consonants in general show little effect of acoustic memory. When they occur in syllable-final position there is clear evidence both from recency ($T = 1½$, $n = 15$; $p < .001$)

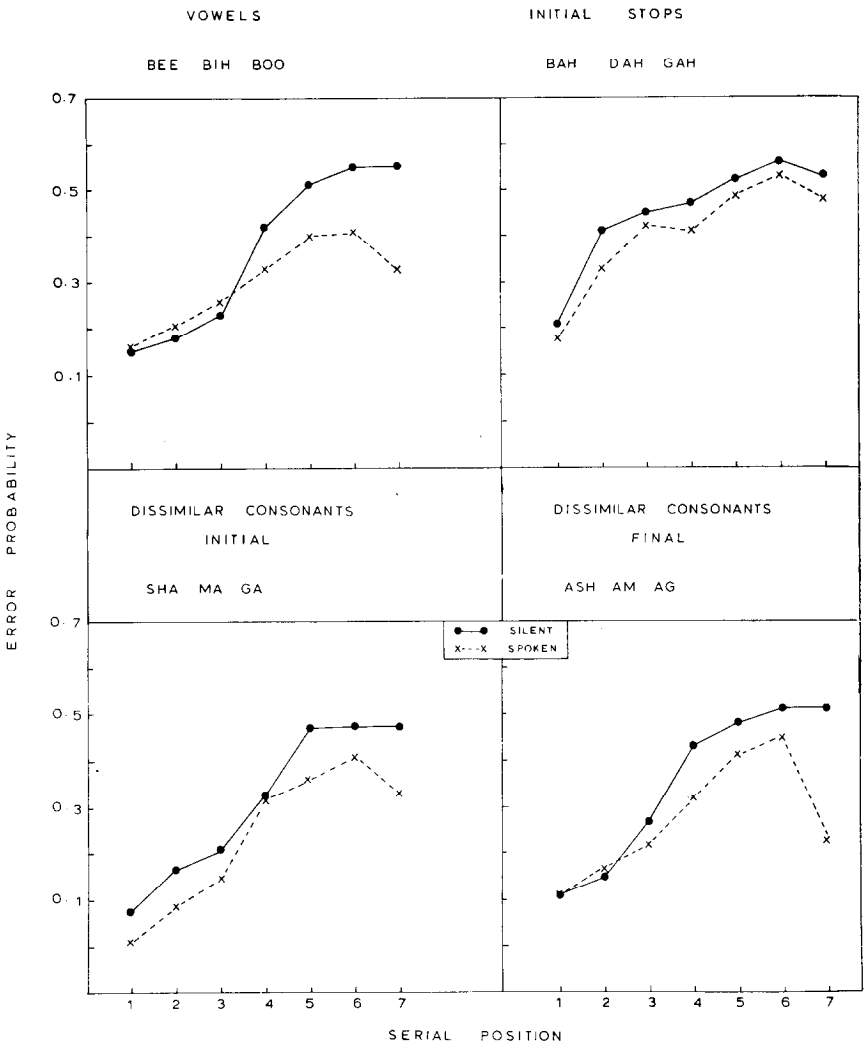

FIG. 1. Mean error probabilities as a function of serial position for lists of seven syllables read silently or aloud (Experiment 1).

and from the modality difference on the last position ($T = 0$, $n = 16$; $p < .001$) that acoustic memory is contributing to their recall. In syllable-initial position the evidence is less clear. Although there is a significant modality effect on the last serial position ($T = 6½$, $n = 14$; $p < .005$) there is only a marginally significant recency effect ($T = 21$, $n = 13$; $p < .1$). Comparing the modality effect in the last serial position across the three consonant conditions showed that the position of the

dissimilar consonant within the syllable was important ($T = 14½$, $n = 16$; $p < .005$) but that which type of consonant occurred in initial position was not important ($T = 36$, $n = 16$; $p > .1$).

Discussion

There are circumstances under which consonants can show a large recency effect for serial recall after auditory presentation. Acoustically distinct consonants in syllable-final position give large and highly significant recency and modality effects—evidence that acoustic memory is contributing to their recall. In a paper published after this experiment was performed Crowder (1973) found no evidence for acoustic memory effects for syllable-final stop consonants, but he did find very small, though significant, acoustic memory effects for a vocabulary of fricative consonants in either syllable-initial position or in isolation. Our experiment taken in conjunction with this more recent finding by Crowder demonstrates that the consonant–vowel distinction is largely irrelevant. Rather, both the nature of the consonants used and their position within the syllable are important. The small recency effect for syllable-initial consonants may be due to the vowel of the final syllable acting as a suffix itself. The small size of Crowder's fricative results are well in accord with our hypothesis that acoustic memory can be regarded as an adulterated tape recording since the acoustic cues that distinguish the three fricatives that Crowder used (/v,z,ʒ/) are very much more similar than those distinguishing the consonants that were used in our experiment. However, there are at least two other hypotheses which can explain the results so far obtained. One is the encoding hypothesis, which maintains that only those cues which are unencoded are preserved in acoustic memory. The second, which Crowder (1973) favors, is that steady-state sounds are preserved better in acoustic memory than are transient sounds. Both these hypotheses are still alive since among those consonants used in the experiments which have shown evidence for acoustic memory have been ones which are at least partly cued by steady-state sounds. These sounds are also relatively unencoded. They are Crowder's fricatives and our fricative (/ʃ/) and our nasal (/m/).

The next two experiments attempt to test both of these hypotheses. We first ask whether acoustically similar vowels show to the same extent the indicants of acoustic memory as do acoustically dissimilar vowels. We then ask whether sounds which are entirely transient can show acoustic memory effects to the same extent that steady-state sounds can.

EXPERIMENTS 2a & 2b

It is implicit in both alternative hypotheses described above that the size of any effect of acoustic memory depends on whether or not the

sounds used fall into a particular class, such as vowels or steady-state sounds. If they do fall within this class which items are actually used is immaterial. The encoding hypothesis, for example, could not account for any variation in the size of acoustic memory effects for different vocabularies of vowel, since all steady-state vowels within a single speaker are equally "unencoded." By contrast, it is the essence of the tape-recording hypothesis that the particular items chosen for the experiment are crucial in determining the size of the effect obtained. Particularly, we would expect that the acoustic memory effects obtainable from a vocabulary of acoustically similar vowels would be smaller than those obtainable from a vocabulary of dissimilar vowels. This is so because after a certain amount of degradation there might be sufficient information available in acoustic memory to distinguish between acoustically dissimilar items, but not between acoustically similar ones; the gross features of the scene are preserved as on a blurred photograph, but the fine detail is lost. This experiment thus uses two different vocabularies of steady-state vowels, one vocabulary which consists of three acoustically dissimilar vowels (/ɪ,æ,u/) and another of three similar vowels (/ɪ,ɛ,æ/). The two vocabularies have two vowels in common and differ only in the relation between these two vowels and the third. Both the encoding and the steady-state hypotheses must predict no difference between these two vocabularies since all the sounds belong to that class of sounds able to show acoustic memory effects. On the other hand, the degraded tape-recording hypothesis maintains that the acoustically similar vowels should show the effects of acoustic memory less than those vowels which are acoustically dissimilar. The vocabulary for which Crowder (1971) obtained significant recency effects was /i,ɪ,u/ (*bee, bih, boo*); this resembles our acoustically dissimilar condition in having two vowels quite close together in F1/F2 space and a third remote from them. A third condition is also included in this experiment to see whether acoustic memory effects can be obtained for transient sounds as well as for steady-state sounds. To this end, three diphthongs were used.

Method

Auditory presentation was used throughout this experiment to allow control over the particular sounds used. The two acoustic memory effects with which we are concerned are, therefore, the recency effect and the suffix effect. To give precise control over the particular sounds heard by the subject synthetic speech was used throughout. This was synthesized on the University of Sussex's software parallel-formant speech synthesizer. This program produces speech according to a similar philosophy and of a similar quality to that used by Crowder from the Haskins Laboratories.

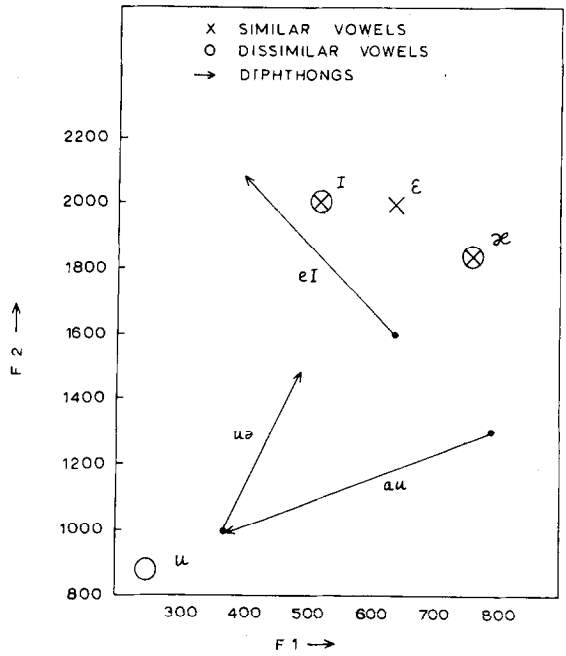

FIG. 2. Formant frequencies for the sounds used in Experiments 2a and 2b.

The first two formants of the stimuli used in these two experiments are shown diagrammatically in Fig. 2. All the sounds had an additional steady third formant at 2500 Hz. In Fig. 2 the crosses mark the similar vocabulary (/ɪ,ɛ,æ/ as in *bit, bet, bat,* respectively) and the circles the dissimilar vowels (replacing /ɛ/ with /u/ as in *boot*). These vowels were preceded by the stop consonant /b/ to give a CV syllable which lasted 60 msec, of which the final 30 msec was steady-state. The diphthongs, represented by arrows in Fig. 2, were tokens of /eɪ,au,uə/ preceded by the stop /d/ to give the three words *day, dhow, dour.* These sounds were synthesized to be entirely transient and lasted 190 msec. For the vowels, the two suffixes used were a steady tone of 1000 Hz lasting 60 msec and the syllable /b/ɜ (as in *burr*) also lasting 60 msec and at the same pitch as the vowels of the vocabulary. For the diphthongs the tone suffix lasted 190 msec and the speech suffix was another diphthong /gou/ (*go*) with the same intonation contour as the main vocabulary. These were the sounds used for Experiment 2a. In Experiment 2b the diphthong condition was compared with the dissimilar vowels whose duration was extended to be the same as the diphthongs. This afforded a more appropriate comparison than with the shorter vowels used in Experiment 2a.

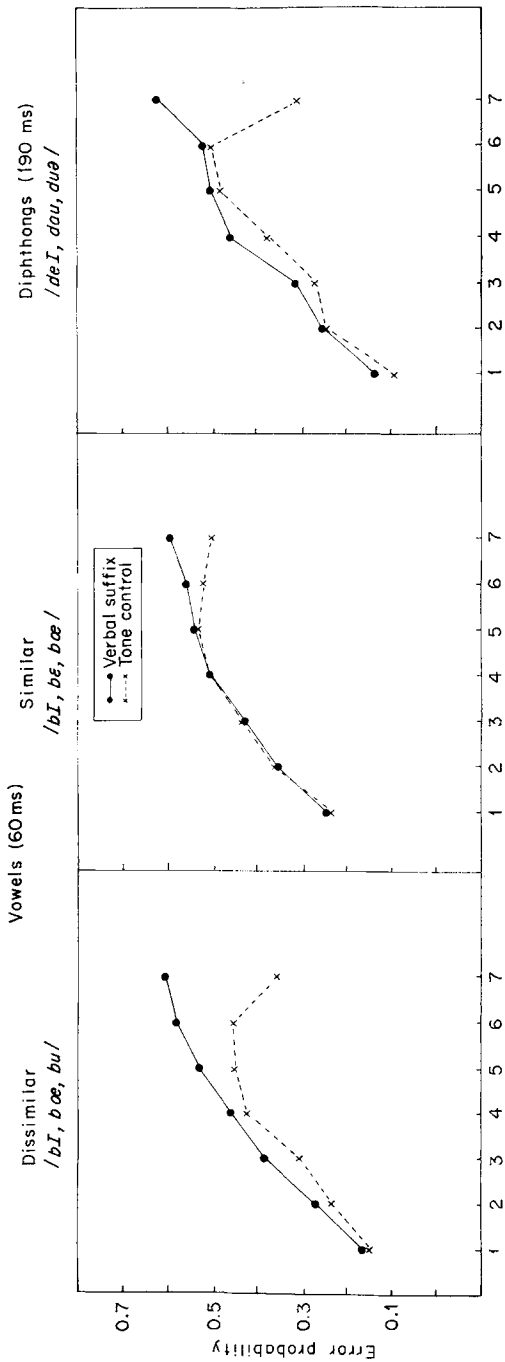

FIG. 3. Mean error probabilities as a function of serial position for auditorially presented lists of seven syllables with either a speech or a tone suffix (Experiment 2a).

In both experiments undergraduate subjects were tested in groups of between one and four. They listened to eight-item lists (seven memoranda plus a suffix) played binaurally over headphones, at a rate of 2 items/sec with 12 sec allowed for written recall. A warning tone was played 2 sec before each trial started. As in Experiment 1 the experiment was divided into blocks in which the vocabulary and the suffix remained constant. Because of the more complicated nature of the vocabularies used in these experiments each block this time consisted of 20 trials, the first five of which were not scored but were used to acquaint the subjects with the new vocabulary. In Experiment 2a each of the 30 subjects took all six possible blocks (two suffixes combined with three vocabularies—similar vowels, dissimilar vowels, diphthongs). In Experiment 2b each of 20 subjects took all four blocks (two suffixes times two vocabularies—long dissimilar vowels, diphthongs). As in Experiment 1 the order of the blocks followed a Latin-square.

Results

The serial position curves for mean errors are shown in Figs. 3 and 4 for Experiments 2a and 2b, respectively. All the vocabularies used, except the similar vowels, show significant effects attributable to acoustic memory. For the dissimilar vowels there is both recency with the tone suffix ($T = 27$, $n = 29$; $p < .001$) and a suffix effect ($T = 63$, $n = 25$; $p < .01$); for the diphthongs this is also true ($T = 24½$; $n = 29$; $p < .001$ for

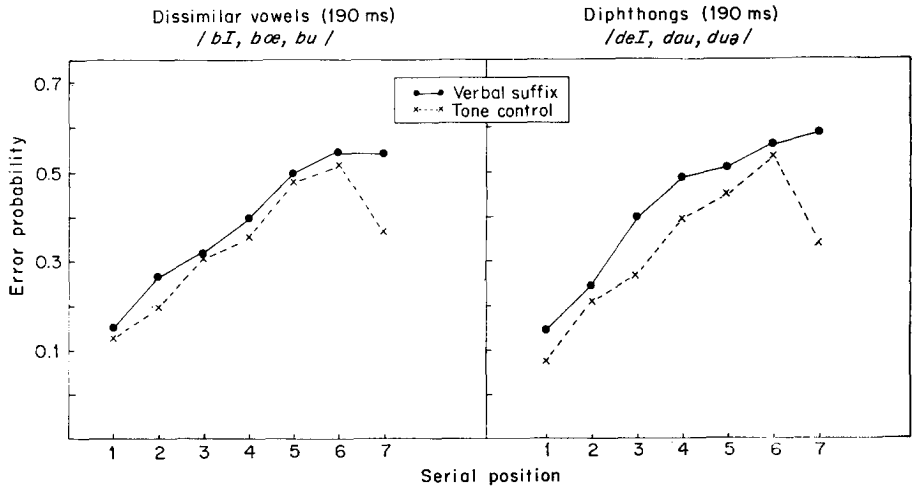

FIG. 4. Mean error probabilities as a function of serial position for auditorially presented lists of seven syllables with either a speech or a tone suffix (Experiment 2b).

recency; $T = 0$, $n = 28$; $p < .001$ for suffix effect). For the similar vowels, though, there is no significant recency effect ($T = 133$, $n = 24$; $p > .5$) and only marginal evidence for a suffix effect ($T = 78½$, $n = 25$; $p < .05$). Moreover, if we compare the size of the suffix effect for the similar vowels with that for the dissimilar vowels we find that the size of the effect is significantly smaller for the similar vowels ($T = 75$, $n = 27$; $p < .01$), as is the difference in the recency effect ($T = 62$, $n = 25$; $p < .01$).

Experiment 2b confirms that dipthongs show both a recency effect ($T = 2½$, $n = 19$; $p < .001$) and a suffix effect ($T = 9½$, $n = 20$; $p < .001$) as again do the long duration dissimilar vowels ($T = 8$, $n = 15$; $p < .001$ for recency; $T = 10$; $n = 15$; $p < .001$ for suffix effect). However, there is no significant difference between these two vocabularies either in recency effect ($T = 66½$, $n = 18$; $p > .4$) or in the magnitude of the suffix effect ($T = 70½$, $n = 19$; $p > .3$)—the diphthongs in fact show slightly larger effects.

Discussion

The absence of any convincing auditory memory effects for the similar vowels, and the significant change in the size of these effects when the vowel vocabulary is made acoustically more distinct clearly indicate that whether a sound appears to be preserved or not in auditory memory has little to do with its phonetic class. It depends on the acoustic similarity of the items used in the vocabulary. Any correlation with phonetic class (or with the encodedness dimension) can, consequently, be explained more parsimoniously by appealing to the greater acoustic confusability between members of some phonetic classes than others.

The hypothesis that acoustic memory holds transient sounds worse than it holds steady-state sounds is not borne out by the results of Experiment 2b. This shows no difference between the acoustic memory effects for transient diphthongs and those for steady-state vowels of the same overall duration. In fact the diphthongs showed slightly larger acoustic memory effects, but as indicated above the difference was not significant. Transience itself is not a sufficient condition to preclude sounds from being preserved in echoic memory. However, it is possible that transience may, under some conditions, contribute to the acoustic confusability of sets of sounds.

Although for ease of exposition we have talked of particular sounds being or not being preserved in acoustic memory, we believe this to be an inappropriate description. We would rather talk of the information necessary for the distinction between particular sets of sounds being preserved or not. For example, in Experiment 2a the vowels /ɪ, æ/ do or

do not appear to be preserved in acoustic memory, depending on the other item from which they have to be distinguished.

We can see no way of explaining the results of Experiment 2a without recourse to some notion such as acoustic confusability. However, as must be painfully obvious to the reader, we have yet offered no definition of how acoustic confusability can be independently determined. In fact this objection is not as serious as might first appear since it is now a matter for empirical investigation what particular form of distortion the acoustic signal undergoes. It is very unlikely that the type of errors would reflect accurately those obtained in white noise. It is possible that acoustic memory is in fact somewhat more sophisticated than a tape recording in that some crude analysis of the speech signal may have already occurred. A better analogy might be a three-dimensional F1/F2/F3 plot similar to Fig. 2, with the effects of distortion equivalent to blurring a particular vowel's position in this space. This simple representation is obviously unable to account for differences in the rate of change of formant transitions, but time represented as a fourth dimension might give a reasonable analogy.

General Discussion

Properties of Acoustic Memory

Previous work has suggested that acoustic memory holds some relatively crude representation of the auditory input for a short time during which the representation becomes degraded (Darwin, Turvey, & Crowder, 1972). The experiments reported here suggest that the result of this degradation is in some way to blur the information held in acoustic memory. After some degradation has taken place there may be sufficient information left to distinguish between a number of very different items, but perhaps not enough information left to distinguish between the same number of more similar ones. Thus we find, in otherwise similar experiments, that acoustic memory can contribute substantially to the distinction between the vowels /ɪ,æ,u/ and between the consonants /g,ʃ,m/, but only minimally to the distinction between /ɪ,ɛ,æ/ and not at all to that between /b,d,g/. It may be that the different estimates of the duration of acoustic memory, that previous workers have found, are due to the different auditory resolution that their tasks required.

Our experiments give no new evidence that the system responsible for the recency, modality, and suffix effects is an auditory one rather than an articulatory one. They could be accounted for in terms of articulatory similarity as readily as acoustic similarity. Massaro (1972) has criticized interpretations of recency effects which attribute them to acoustic mem-

ory, suggesting that the effects are due to postcategorical interference in primary memory. However, as has been pointed out before (Crowder & Morton, 1969), it is hard to see why, if the effects are due to operations at the articulatory level, they are confined to the auditory input modality. There is no clear evidence to distinguish the form of the acoustic memories revealed by short-term memory experiments such as these from those using other paradigms such as partial report (Darwin, Turvey, & Crowder, 1972) or backward masking (Studdert-Kennedy, Shankweiller, & Schulman, 1970; Massaro, 1972; Darwin, 1971b). Massaro claims that the store that he identifies from backward masking experiments has a duration of only about 250 msec on the grounds that his subjects' performance asymptotes at that interstimulus interval. One can, however, interpret this result as well by saying that within 250 msec the subject has extracted all the available information from the store. Whether or not the store continues to preserve this information would not affect performance beyond this point.

Massaro also claims that there is better evidence for his store being precategorical than that revealed by the recency experiments. His reason is that the effectiveness of the suffix is sensitive to some physical properties of the stimulus, whereas in his backward masking experiments the frequency of the mask has little effect. This argument would only hold if he could show backward masking effects for the discrimination of two sounds, one of which was impervious to the suffix effect of the other in a recency experiment. He could then justly claim that PAS came after the store involved in the backward masking experiments. Lacking this evidence we are entitled to suppose that those gross features which determine the efficacy of a suffix are extracted before Massaro's store is reached. Massaro's argument is further weakened by Darwin's (1971b) finding that the extent of backward masking depended crucially on the type of mask used.

Acoustic Memory in Speech Perception

A number of experimental paradigms have been used to support the claim that different phonetic classes are perceived by different mechanisms. In the introduction we raised the possibility that these differences might be explained in terms of auditory memory rather than in terms of different mechanisms of categorization. This final section explores how this might be.

(i) *Discrimination experiments.* The role of auditory memory in experiments on the discrimination of vowels and stop consonants has already been given considerable discussion by Fujisaki and Kawashima (1968) and Pisoni (1973). Their main finding is that the discrimination

of vowels can be influenced by factors which are irrelevant to the categorization process thought by Liberman *et al.* to be responsible for the different perceptual functions of vowels and stop consonants. They explain their results in terms of a phonetic classification plus an additional mechanism for discrimination, based on acoustic memory. There is no unequivocal evidence here for different categorization processes for vowels and stop consonants.

(ii) *Ear differences.* Other evidence cited in support of special categorization processes for stop consonants, in particular, and encoded sounds, in general, is the right-ear advantage under dichotic listening. It is much harder to show a convincing right-ear advantage for vowels than for stop consonants. This has been interpreted in terms of the stop consonants requiring the special abilities of a left-hemisphere speech processing mechanism more (Liberman *et al.*, 1972). Darwin (1973) has already discussed the difficulties of interpreting the dichotic listening experiments in this way, and concluded that auditory memory can explain the differences between stops and vowels in this paradigm also. Essentially, the argument is that a convincing right-ear advantage will only appear if the left-hemisphere has a sufficiently privileged access to the opposite ear. Now it may be that sounds which are still discriminable after a relatively long stay in acoustic memory give the left-hemisphere more time in which to categorize the sounds that came to the inappropriate (left) ear. This would reduce the size of the ear difference for those sounds. A number of results support this interpretation. First, for vowels an ear difference can be found under two different circumstances. In one case a right-ear advantage can occur when more than one speaker is used in the experiment (Haggard, 1971; Darwin, 1971a). In the second case, a right-ear advantage has been found when vowels are presented dichotically at a very unfavorable signal-to-noise ratio (-12 dB) (Weiss & House, 1973). In both these cases the utility of the representation in acoustic memory might be less than in the usual case where one speaker is used and the signals played at a very high S/N ratio (Shankweiler & Studdert-Kennedy, 1968; Darwin, 1971a). The presence of two speakers means that greater auditory resolution is needed in order to identify the vowel correctly, and an originally unfavorable S/N ratio in the stimulus could enhance the effect of whatever distortion occurs in acoustic memory. Although the case of the two speakers can also be accounted for by the encoding interpretation, the Weiss and House experiment cannot be. Other dichotic experiments which are open to the echoic memory interpretation are due to Darwin (1971a) and Cutting (1972). In Darwin's experiment fricative consonants only gave a right-ear advantage when they contained formant transitions as cues to place of articulation.

They failed to give any ear difference when only steady-state sounds gave this cue. Crowder (1973) has shown that there are small but significant recency effects and suffix effects for sounds identical to the ones Darwin used irrespective of the presence of formant transitions. This is precisely what we would expect on the hypothesis that the distinction between the rapid transitions is lost from echoic memory while that between the other cues is preserved. Cutting's experiment showed that although stop consonants can give a reliable right-ear advantage in both the syllable-initial and the syllable-final position, the laterals /r,l/ only show a right-ear advantage when they are in initial position. They fail to show any advantage when the formant pattern is reversed in time. Again, this result fits snugly with the recency and suffix data since Crowder (1973) finds no recency or suffix effects for either initial or final stop consonants, while we have shown that acoustically distinct consonants show larger acoustic memory effects in the syllable-final position than the syllable-initial position.

(iii) *Dichotic masking.* The final type of experimental evidence used in support of the encoding dimension as a perceptually significant variable comes from experiments on dichotic backward masking. The result here is that when stop consonants are presented dichotically in C–V syllables with a slight temporal offset (say 60 msec), the second consonant is reported more accurately than the first, (Studdert-Kennedy, Shankweiler, & Schulman, 1971). This result has been attributed to an interruption phenomenon since it depends crucially on the relationship between the two sounds which are dichotically opposed (Darwin, 1971b). When two vowels are similarly opposed, there is much less evidence for superior recall of the second sound over the first (Studdert-Kennedy, Shankweiler, & Schulman, 1971). The suggestion has already been made (Darwin, 1971b) that the effect has an acoustic rather than a linguistic basis since it occurs for nonspeech as well as for speech distinctions, but the properties of acoustic storage allow us to explain the vowel–stop difference in acoustic terms. We might suppose that after the processing of the first sound has been interrupted by the arrival of the second, there remains in acoustic memory some representation of the first sound which can be inspected again by the categorizing process. The state of this representation will naturally determine how well the initial sound is perceived. If, as we might suppose for stop consonants, this representation carries little useful information by the time it is inspected, then the initial interruption will appear to have had a substantial effect and there will be appreciable backward masking. But if, as we might suppose for the vowels, there is still useful information in this representation, then the initial interruption will seem to have little effect

and there will be much less evidence for backward masking. Our prediction then is that those vocabularies of sounds which show large recency effects should be less susceptible to backward masking than those which show small recency effects. Some independent confirmation of this hypothesis has recently appeared. Darwin (1971b) showed that there was much less evidence of backward masking for the recognition of voicing, than for the recognition of place of articulation. We might expect then that voicing should show greater evidence of recency and suffix effects than place of articulation. A. Thomasson (personal communication) has evidence from Dutch listeners that this is indeed the case. The voicing distinction could be made, at least for the stimuli that Darwin used, simply on the detection of a periodic excitation during the first 50 msec or so of the sound. This information might well be more resistant to the distortion of a stay in acoustic memory than that about the trajectories of the second formant which cued place of articulation. Thomasson's subjects spoke the sounds themselves so it is not possible to tie their cues down.

In summary, we would like to claim that the concept of acoustic memory can be used to explain a large number of experimental results from a variety of paradigms which have hitherto been attributed to differences in the categorizing process itself. We have presented experimental evidence that an item's preservation in acoustic memory cannot be regarded as a consequence of a special decoding mechanism for a subset of speech sounds. Indeed the very existence of such a mechanism is called into doubt since the phenomena it was assumed to explain can themselves be explained in terms of acoustic memory. The correlation that has been noted between the dimension of encodedness and performance in various perceptual tasks can perhaps be attributed to the fact that those acoustic cues which are most encoded (i.e., show most acoustic restructuring with context) are the most ephemeral in acoustic memory.

We do not doubt that speech requires different processing mechanisms from other sounds, and indeed would regard the right and left ear advantages for speech and nonspeech sounds as good evidence for this. We do, however, question, in the light of our results, whether there is any psychological evidence for supposing that different speech sounds are perceived in different ways.

REFERENCES

COLE, R. A. Different memory functions for consonants and vowels. *Cognitive Psychology*, 1973, **4**, 39–54.

CROWDER, R. G. The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 587–596.

CROWDER, R. G. Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, 1973, **1**, 14–24.

CROWDER, R. G., & MORTON, J. Pre-categorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, **5**, 365–373.

CUTTING, J. E. Ear advantage for stops and liquids in initial and final position. Haskins Labs: Status Report on Speech Research SR-31/32, 1972, 57–65.

CUTTING, J. E. A parallel between encodedness and the magnitude of the right ear effect. *Journal of the Acoustical Society of America*, 1973, **53**, 358 (A).

DARWIN, C. J. Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, 1971a, **23**, 46–62.

DARWIN, C. J. Dichotic backward masking of complex sounds. *Quarterly Journal of Experimental Psychology*, 1971b, **23**, 386–392.

DARWIN, C. J. Ear differences and hemispheric specialization. In F. O. Schmitt and F. G. Worden (Eds.), *The Neurosciences, Third Study Program*, Cambridge, MA: M.I.T. Press, 1973.

DARWIN, C. J., TURVEY, M. T., & CROWDER, R. G. An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, 1972, **3**, 255–267.

FRY, D. B., ABRAMSON, A. S., EIMAS, P. D., & LIBERMAN, A. M. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, **5**, 171–189.

FUJISAKI, H., & KAWASHIMA, T. The influence of various factors on the identification and discrimination of synthetic speech sounds. Paper read at 6th International Congress on Acoustics, Tokyo, Japan, August, 1968.

HAGGARD, M. P. Encoding and the REA for speech signals. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 34–45.

LADEFOGED, P., & BROADBENT, D. E. Information conveyed by vowels. *Journal of the Acoustical Society of America*, 1957, **29**, 98–104.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431–461.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, **54**, 358–368.

LIEBERMAN, A. M., MATTINGLY, I. G., & TURVEY, M. T. Language codes and memory codes. In A. W. Melton and E. Martin (Eds.), *Coding Processes in Human Memory*, New York: Wiley, 1972.

LINDBLOM, B. E. F., & STUDDERT-KENNEDY, M. On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, 1967, **42**, 830–843.

MASSARO, D. M. Preperceptual auditory images. *Journal of Experimental Psychology*, 1970, **85**, 411–417.

MILLER, G. A., & NICELY, P. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 338–352.

MORTON, J., CROWDER, R. G., & PRUSSIN, H. A. Experiments with the stimulus suffix effect. *Journal of Experimental Psychology Monographs*, 1971, **91**, 169–190.

PICKETT, J. M. Perception of vowels heard in noises of various spectra. *Journal of the Acoustical Society of America*, 1957, **29**, 613–620.

PISONI, D. B. Auditory and phonetic memory codes in speech discrimination. In I. G. Mattingly (Ed.), *The Speech Code: Readings in Speech Perception*, 1973, in press.

SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology,* 1967, **19**, 59–63.

SMALLWOOD, R. A., & TROMATER, L. J. Acoustic interference with redundant elements. *Psychonomic Science,* 1971, **22**, 354–356.

STUDDERT-KENNEDY, M., SHANKWEILER, D., & SCHULMAN, S. Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. *Journal of the Acoustical Society of America,* 1970, **48**, 599–602.

WEISS, M. S., & HOUSE, A. S. Perception of dichotically presented vowels. *Journal of the Acoustical Society of America,* 1973, **53**, 51–58.

WILLIAMS, E. J. Experimental designs balanced for the estimation of the residual effects of treatments. *Australian Journal of Scientific Research,* 1949, **2**, 149–168.

(Accepted August 21, 1973)