Received 18 March 1965

Distinctive Features and Errors in Short-Term Memory for English Vowels

WAYNE A. WICKELGREN

Department of Psychology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Errors in short-term recall of six English vowels (I, ε, æ, U, ʌ, ɑ) were tabulated and related to several distinctive-feature systems. Vowels were embedded in two contexts: /l[ ]k/ and /z[ ]k/. Subjects were instructed to copy items as they were presented, followed by recall of the entire list of (six) items. Perceptual errors were excluded from the recall error matrix by scoring for recall only correctly copied items. The rank-order frequency of different intrusions in recall of each presented vowel was almost perfectly predicted by a conventional phonetic analysis in two dimensions: place of articulation (front, back) and openness of the vocal tract (narrow, medium, and wide). The error matrix also supported the assumptions that the values of openness are ordered in short-term memory and that the correct value on the openness dimension is more likely to be forgotten than the correct value on the place dimension. The study suggests that a vowel is coded in short-term memory, not as a unit, but as a set of two distinctive features, each of which may be forgotten independently.

INTRODUCTION

CODING is generally considered to be a perceptual or motor problem, but it is equally important in the study of memory. In addition to axioms concerned with the consolidation, decay, and interference of associations between internal representatives in memory, one also wants to know the units of internal representation in memory. Recent findings suggest that a verbal item (word, letter, digit, CV syllable, etc.) is coded in short-term memory (STM) as a set of phonemes, each of which may be forgotten independently. The evidence for this hypothesis is that the errors in short-term recall of correctly perceived verbal items tend to have a phoneme in common with the correct item.[1–3]

The present study attempts to extend the phonemic-coding hypothesis to determine if a vowel phoneme is coded in STM as a set of distinctive features, each of which may be forgotten independently. The general method is similar to that used in determining if "words" are coded in terms of phonemes: namely, examination of the error matrix. However, there are more unresolved issues in the distinctive-feature analysis of English phonemes than there are in the phonemic analysis of English words, and it is not obvious which of the existing distinctive-feature systems one should use. Thus, in addition to testing the hypothesis that phonemes are coded in STM as sets of features, the error matrix should indicate which existing distinctive-feature system best describes the coding of English vowels in STM.

Research on the errors in STM has an obvious methodological similarity to previous research on the errors in auditory perception. Peterson and Barney[4] and Miller[5] have established that similarity in the first two formants and perhaps duration is highly correlated with the errors in auditory perception of vowels. The importance of different features clearly depends on noise and filtering conditions, and the exact nature of the feature system for auditory perception has not been established. For example: How many values are there on each dimension? What is the relative weighting of each dimension? How does this relate to noise and filtering conditions?

The relationship between errors in auditory perception of vowels and errors in STM for vowels is a matter

[1] R. Conrad, "Acoustic Confusions in Immediate Memory," Brit. J. Psychol. 55, 75–84 (1964).
[2] W. A. Wickelgren, "Acoustic Similarity and Intrusion Errors in Short-Term Memory," J. Exptl. Psychol. 70, 102–108 (1965).
[3] W. A. Wickelgren, "Similarity and Intrusions in Short-Term Memory for Consonant–Vowel Digrams," Quart. J. Exptl. Psychol. (to be published).

[4] G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Am. 24, 175–184 (1952).
[5] G. A. Miller, "The Perception of Speech," in *For Roman Jakobson*, M. Halle, Ed. (Mouton & Co., The Hague, 1956), pp. 353–359.

for empirical investigation, and this study provides some evidence on the question. However, this is not the primary goal of the study. The primary goal is to determine, insofar as possible, what feature system works best for STM. This effort has been greatly assisted by feature systems previously developed in articulatory phonetics, acoustic phonetics, and linguistics, but the comparative adequacy of these systems for predicting STM errors is suggestive only with respect to their adequacy in the realm for which they were developed.

Ultimately, it will be possible to decide how many different verbal feature systems exist in the human nervous system. Perhaps, there is only one. Perhaps, there is an acoustic system and an articulatory system. Perhaps, there is also an "abstract" system, as postulated by recent linguistic theories of sound structure.[6,7] This question will probably be answered only when feature systems have been definitely established for many different types of verbal behavior. In the final analysis, a complete theory of verbal behavior must account for all its perceptual, memory, and productive aspects. But this complete theory need not have only one feature system for all the aspects. On the other hand, one feature system that worked best for perception, memory, and production would be the most appealing theoretical possibility.

I. DISTINCTIVE-FEATURE SYSTEMS FOR VOWELS

Let us consider four previously proposed systems for English vowels. All the systems are discrete—by which is meant that there is a finite number of values for each dimension. In fact, for the six vowels in the present experiment—ɪ, ε, æ, ʊ, ʌ, ɑ—there are only two or three values for each dimension.

The conventional phonetic analysis (CPA) of these six vowels in terms of *place* of maximum constriction of the vocal tract (front, back) and *openness* of the vocal tract (narrow, medium, wide) is shown in Table I. Table I also describes the three binary dimensions of Chomsky and Halle's systematic phonetic level of analysis (Pt).[6,8] Although there are some differences in the predictions made by CPA and Pt, there are no differences in the predictions made for the present experiment. Notice that central and back vowels have been lumped together as back vowels in CPA and that the unnecessary flatness feature has been dropped from Pt.

The systematic phonetic level is the final phonological level of the Chomsky–Halle system. The features that characterize a vowel at this level can be understood in terms of the vowel's articulatory and acoustic characteristics. Pt can therefore be interpreted (by others) as a theory of the sensory or articulatory coding of vowels. In order to derive the phonological representation of an utterance from a higher-order syntactic representation of the utterance, Chomsky and Halle have found it useful to interpolate another phonological level between the syntactic level and the systematic phonetic level. This level is called the systematic phonemic level (Pm).[6,7] Chomsky and Halle justify the more abstract Pm level in terms of its simplicity: many regularities of English sound structure appear to be more economically described by their two-level representation than by any existing one-level representation.

Table I. Conventional phonetic analysis (CPA) and the systematic phonetic level (Pt) of the Chomsky–Halle feature system.

	Front	Back	
+Diffuse	ɪ	ʊ	Narrow
−Diffuse			
−Compact	ε	ʌ	Medium
+Compact			
	æ	ɑ	Wide
	−Grave	+Grave	

Table II. Relationships between concrete vowels at the systematic-phonemic level (Pm) of the Chomsky–Halle feature system.

ɪ	ʌ (← ʊ)
ε	ʊ (← ō)
æ	ɑ (← ɔ)

Many of the same features are used at both Pt and Pm levels, but the representation of a Pt-level vowel may be quite different at the Pm level. For our purposes, the principal difference between the two levels is the "vowel shift" that occurs in certain contexts. A particular vowel at the Pm level may be transformed into the same vowel or a different one at the Pt level, depending on the context of the vowel. In the consonant contexts used in the present experiment, /ʊ/ at the Pt level is /ō/ at the Pm level; /ʌ/ at the Pt level is /ʊ/ at the Pm level; /ɑ/ at the Pt level is /ɔ/ at the Pm level; /ɪ/, /ε/, and /æ/ are unchanged. Stated in terms of the Pt level vowels, which are the responses observed in the present experiment, the dimensional structure at the Pm level is as shown in Table II. Table II also shows in parentheses the Pm vowels from which the Pt vowels were derived (according to Chomsky and Halle). Since the dimensional structure is different for the two levels, the two levels predict rather different rank orderings of errors in STM.

6 N. Chomsky and M. Halle, *Sound Pattern of English* (to be published).
7 N. Chomsky and M. Halle, "Some Controversial Issues in Phonological Theory," J. Linguistics (to be published).
8 M. Halle, "Phonology in Generative Grammar," in *The Structure of Language*, J. A. Fodor and J. J. Katz, Eds. (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1964), pp. 334–352.

TABLE III. Jakobson–Fant–Halle feature system.

Diffuse	ɪ	ʌ		ʊ
Compact	ɛ	æ		ɑ
	Acute	Grave	Plain	Flat

The Chomsky–Halle system developed out of the distinctive-feature system of Jakobson, Fant, and Halle (JFH), described in Table III (Ref. 9). JFH is a one-level system like CPA but with a rather different dimensional structure. Although JFH was not intended to be more than a preliminary hypothesis about the feature system, it is interesting to test it against the error data because it makes very different predictions from CPA, Pt, and Pm concerning the rank ordering of errors in STM.

To avoid confusion, it should be noted that many of the terms used to designate features are common to both JFH and the Chomsky–Halle system, but the application of the terms to vowels is rather different in the two systems. Although there is some similarity in the use of the terms "compact," "diffuse," and "grave" in the two systems, the similarity is far from identity.

One final point: Jakobson, Fant, and Halle were careful to eliminate redundancy in the definition of vowels by their features, not only by eliminating redundant dimensions, but also by failing to classify a vowel on a dimension if the dimension was unnecessary to distinguish that particular vowel from other vowels. The latter practice is not observed in the description of JFH in Table III. The STM data validate this decision in every case.

II. METHOD

On each trial, subjects listened to a list of six items, copying the items as they were being presented. As soon as a subject finished copying the list, he covered what he had copied and then attempted to recall the list by filling in six boxes with the correct items in the correct positions. Each trial began with a ready signal lasting about 1 sec, followed by a 1-sec pause, followed by the list presented at the rate of about 2 sec per item, followed by 16 sec in which to recall the list.

The lists were random permutations of the six items in one of the two following populations: (l items) lick, leck, lack, look, luck, lock and (z items) zick, zeck, zack, zook, zuck, zock. Subjects copied and recalled entire four-letter items, but only the vowels distinguished between the items in different positions on any one trial. Thus, it is possible to study the errors in short-term recall of six English vowels, (ɪ, ɛ, æ, ʊ, ʌ, ɑ). The error matrix for these vowels in l items (consisting mostly of English words) can be compared to the error matrix for these vowels in z items (consisting entirely of nonwords). By scoring for ordered recall *only* those vowels that were copied correctly, we can eliminate perceptual errors from the STM errors.

There were 100 trials in the experiment with 50 l lists and 50 z lists. Two l lists and two z lists occurred in each nonoverlapping block of four lists. The experiment was recorded on tape and played back over a loudspeaker. The speaker was a female who had spent the first 11 years of her life in Colorado and who went to high school in Long Island, New York. Subjects were 44 Massachusetts Institute of Technology undergraduates who were taking psychology courses and participating in the experiment as part of their course requirements. These students constituted a rather broad regional sampling of the United States of America. Subjects were run in three approximately equal groups. Instruction was given in the pronunciation of the items and there was a short practice period. The subjects must be presumed to be relatively unpracticed in the pronunciation of the z items. However, the exact parallel between the z items and the l items was designed to produce heavy positive transfer in the direction of "correct" pronunciation of the z items.

III. RECALL ERRORS

Only vowels correctly copied were scored for ordered recall. The recall-error matrices for vowels in l items and z items are presented in Tables IV and V. Except for the omission rate, which is higher for z items than for l items, the matrices are virtually identical. Therefore, they are combined in Table VI and transformed to the conditional probability of each intrusion in recall

TABLE IV. Recall-error matrix for vowels in /l[ ]k/ environments.

		Recalled						
		ɪ	ɛ	æ	ʊ	ʌ	ɑ	Omit
	ɪ	1521	74	81	75	56	45	206
Presented	ɛ	103	1458	110	61	68	43	209
and	æ	80	112	1493	40	59	76	202
correctly	ʊ	66	54	48	1505	99	82	178
copied	ʌ	78	105	74	124	1376	99	206
	ɑ	71	62	88	91	103	1383	203

TABLE V. Recall-error matrix for vowels in /z[ ]k/ environments.

		Recalled						
		ɪ	ɛ	æ	ʊ	ʌ	ɑ	Omit
	ɪ	1453	93	85	69	54	36	235
Presented	ɛ	116	1302	127	64	99	38	254
and	æ	88	101	1449	56	55	53	219
correctly	ʊ	67	58	45	1227	92	73	219
copied	ʌ	75	122	77	112	1236	101	241
	ɑ	57	50	85	97	117	1253	229

[9] R. Jakobson, C. G. M. Fant, and M. Halle, *Preliminaries to Speech Analysis* (MIT Press, Cambridge, Mass., 1952).

TABLE VI. Recall-probability matrix for vowels in both environments.

		Recalled							
		ɪ	ɛ	æ	ʊ	ʌ	ɑ	Omit	N
Presented and correctly copied	ɪ	0.7284	0.0409	0.0407	0.0353	0.0269	0.0198	0.1080	4083
	ɛ	0.0540	0.6811	0.0585	0.0308	0.0412	0.0200	0.1143	4052
	æ	0.0411	0.0522	0.7205	0.0235	0.0279	0.0316	0.1031	4083
	ʊ	0.0349	0.0294	0.0244	0.7165	0.0501	0.0407	0.1041	3813
	ʌ	0.0380	0.0564	0.0375	0.0586	0.6488	0.0497	0.1110	4026
	ɑ	0.0329	0.0288	0.0445	0.0483	0.0566	0.6778	0.1111	3889

of a correctly copied item. Only the combined matrix will be considered in subsequent analyses.

The frequency of correctly ordered recall after correct copying averages 70%. Omissions in recall after correct copying occur about 6% of the time. Intrusions (confusions) in recall make up the remaining 24%. Front vowels appear to be remembered better than back vowels, and within each of these two categories the medium-opening vowels are least well remembered. The differences, however, are not great.

Intrusions in short-term recall of vowels are highly systematic and completely consistent with the CPA or Pt feature system. CPA makes 36 binary (greater than) predictions about the frequency of different intrusions in Table VI. For each of the six presented vowels, CPA predicts that the three incorrect vowels having one feature in common with the correct vowel will occur more frequently than the two incorrect vowels that have no feature in common with the correct vowel. For example, ɛ, æ, and ʊ will each occur more frequently than ʌ or ɑ as an error in the recall of ɪ. Thus, there are six binary predictions for each presented vowel, or 36 altogether. All 36 predictions are confirmed by the data. If the data are analyzed by columns instead of by rows, thus controlling any response bias, we obtain 35 correct predictions out of 36.

It is also perfectly clear that the correct value on one of the dimensions—namely, openness— is more likely to be forgotten than the correct value on the other dimension—namely, place. For example, ɛ and æ each occur more frequently than ʊ as an error in the recall of ɪ. This assumption allows us to make two additional binary predictions for each presented vowel, or 12 altogether. Eleven of the 12 come out as predicted. Analyzing by columns, 9 of 12 are as predicted.

So far, we have not assumed the three values of the openness dimension to be ordered. Since they are ordered by a physical scale, it is reasonable to guess that they are ordered in STM. For example, ɛ occurs more often than æ, and ʌ occurs more often than ɑ, as an error in the recall of ɪ. No prediction is made for medium-opening vowels. For narrow or wide vowels, there are two such predictions for each presented vowel, or 8 altogether. Seven of the 8 come out as predicted, although some of the differences are very small. Analyzing by columns, all 8 comparisons are in the predicted direction.

The Pt system works equally well in predicting the rank order of intrusions for each presented vowel, provided that one makes the assumption that the probability of forgetting gravity is less than the product of the probabilities of forgetting compactness and diffuseness. The Pm and JFH systems are clearly less adequate than the Pt and CPA systems for predicting the rank order of intrusions in STM.

IV. COPYING ERRORS

The copying-error matrices for l items and z items are presented in Tables VII and VIII. The combined matrix is presented in Table IX. As would be expected, more errors are made in copying the less familiar z items than in copying the more familiar l items. Copying "zuck" instead of "zook" and copying "zack" instead of "zock" are particularly large sources of error. These two confusions in the z items are just the ones that would be expected from the relationship between written and

TABLE VII. Copying-error matrix for vowels in /l[ ]k/ environments.

		Copied						
		ɪ	ɛ	æ	ʊ	ʌ	ɑ	Omit
Presented	ɪ	2058	3	1	2	3	2	43
	ɛ	6	2052	9	1	8	1	35
	æ	2	5	2062	1	1	8	33
	ʊ	2	1	···	2032	18	11	48
	ʌ	2	6	2	10	2062	6	24
	ɑ	···	···	23	24	17	2001	47

TABLE VIII. Copying-error matrix for vowels in /z[ ]k/ environments.

		Copied						
		ɪ	ɛ	æ	ʊ	ʌ	ɑ	Omit
Presented	ɪ	2025	18	12	6	1	1	49
	ɛ	11	2000	22	3	10	···	66
	æ	6	9	2021	4	5	12	55
	ʊ	8	3	9	1781	212	27	72
	ʌ	7	11	21	26	1964	20	63
	ɑ	2	1	101	43	22	1888	55

TABLE IX. Copying-probability matrix for vowels in both environments (N=4224 for each vowel).

		Copied						
		ɪ	ɛ	æ	ʊ	ʌ	ɑ	Omit
Presented	ɪ	0.96662	0.00497	0.00308	0.00189	0.00095	0.00071	0.02178
	ɛ	0.00402	0.95928	0.00734	0.00095	0.00426	0.00024	0.02391
	æ	0.00189	0.00331	0.96662	0.00118	0.00142	0.00473	0.02083
	ʊ	0.00237	0.00095	0.00213	0.90269	0.05445	0.00900	0.02841
	ʌ	0.00213	0.00402	0.00545	0.00852	0.95312	0.00616	0.02060
	ɑ	0.00047	0.00024	0.02936	0.01586	0.00923	0.92069	0.02415

spoken English, since the phoneme /ʊ/ is often written as "u" in such words as *put*, *pull*, *full*, etc., and the phoneme /ɑ/ is often written as "a" in such words as *father*, *ah*, *car*, etc. Attributing some part of the total frequency of these two confusions in the z items to orthographic confusion, rather than perceptual confusion, is supported by the much smaller relative frequency of these errors (compared to other errors) in the l items. Since all but one of the l items are words, their spelling is highly overlearned and no orthographic confusion should be expected.

Thus, copying errors with z items undoubtedly include cases both where the item would have been pronounced incorrectly (had this been tested), and where the item would have been pronounced correctly but was written incorrectly. Copying errors with l items very likely include only cases where the item would also have been pronounced incorrectly. Despite this difference, the copying-error matrices for z items and l items demonstrate a high degree of similarity in the rank order of different types of errors in response to each presented item.

The probability of copying a vowel correctly was quite high in the present experiment; 97% for l items, 92% for z items. This indicates that the limited instruction concerning the pronunciation of z items was almost completely effective. Comparison with the data of Peterson and Barney[4] suggests that copying accuracy is not very seriously affected by presenting 6 items in sequence at the rate of 2 sec per item. The fact that recognition of l[ ]k items in the present experiment was actually a little better than recognition of h[ ]d items in the Peterson and Barney experiment can most plausibly be attributed to the fact that only 1 (female) speaker was used in the present study, while 76 speakers were used by Peterson and Barney. The greater number of vowels in the Peterson and Barney experiment also may have reduced copying accuracy, but the difference persists even after application of the "constant-ratio rule."[10] Of course, the subjects and the consonant environments of the vowels were different and this may have had some effect. In short, it is not possible to make precise comparisons between the copying data of the present study and those of Peterson and Barney.

[10] F. R. Clarke, "Constant-Ratio Rule for Confusion Matrices in Speech Communication," J. Acoust. Soc. Am. 29, 715–720 (1957).

The female speaker's first two formant frequencies for each vowel used in the experiment are presented in Table X. Formants were determined to the nearest 50 cps from a spectrogram, and 7 instances of each vowel (3 from l items and 4 from z items) were averaged to yield the figures in Table X. The relative acoustic distances between vowels for the present speaker are very similar to those reported by Peterson and Barney. Also, the distribution of copying errors in the present study shows the same positive correlation with acoustic distance in the space formed by the logs of the first two formants. However, the decline in error frequency as a function of acoustic distance is much less rapid in the present study. Presumably, this results from the greater demand on the subject in the present experiment and the consequent decrease in ability to attend to each item. However, the less rapid decline in copying errors with increasing acoustic distance must not be attributed to a "floor" of random errors produced by complete inattention. The most distant errors are almost as infrequent in this study as in the Peterson and Barney study, but the errors are distributed more evenly among the closest vowels in the acoustic space.

TABLE X. Average formant frequencies (in cps) of the female speaker used in this experiment.

	ɪ	ɛ	æ	ʊ	ʌ	ɑ
F1	450	800	950	650	850	1000
F2	2250	2000	1950	1200	1500	1350

The various discrete-feature systems can also be tested on the copying data in Table IX. Again, the CPA and Pt systems make more-accurate predictions than the Pm and JFH systems. Of the 36 predictions made by the CPA feature system, 35 are correct analyzing by rows and 34 are correct analyzing by columns. However, the assumption that openness is ordered does not fare so well, making correct predictions in only 5 of 8 cases by rows and 4 of 8 cases by columns. There is also no evidence that place is more likely to be perceived correctly than openness. Since the Pt system *must* predict that ɪ, ɛ, æ, are ordered and that ʊ, ʌ, and ɑ are ordered, it is slightly less accurate than the unordered CPA system. However, this is of little significance, since there are so few copying errors and some of the differences are very small.

V. DISCUSSION

Although the assumptions about ranking of dimensions and ordering of values on openness do not appear to hold for copying, the feature system (CPA or Pt) that predicts the recall errors best also predicts the copying errors best. This finding suggests that perception and STM for vowels may use the same feature system, but the conclusion must be considered as tentative. Copying six "words" at the rate of 1 every 2 sec is not the optimal way to study perceptual errors, free of other factors. For example, copying errors in the present experiment could well include STM errors, if the subjects ever fall more than one word behind the speaker. Even if the subjects always keep up, there is the possibility of interference between words being rehearsed in STM and words currently being presented. Copying errors that occur when several words are presented are not necessarily the same as copying errors that occur when only one word is presented at a time.

The present study shows that vowel phonemes are not atomic units in STM, but instead are coded in terms of distinctive features. Furthermore, one feature may be forgotten while another is not. Thus, forgetting in verbal STM is not all-or-none. Whether the forgetting of one feature is completely independent of the forgetting of another feature is a question that requires a more complete theoretical analysis. To test the independence assumption, it appears to be necessary to eliminate omissions and require recall of only one item in a list. It is not possible to test the assumption of complete independence with the present data. However, it is obvious that there is at least partial independence, or else there would be no systematic errors.

Finally, the present study indicates that the feature system that is used to code vowels in STM is CPA, Pt, or a similar system. Whether this places STM in the articulatory system or the acoustic system is impossible to determine at the present time, because of the rather close correspondence between openness and formant 1 and between place and formant 2. However, the plausibility of interpretation in either acoustic or articulatory terms does tend to argue against placing STM in some "abstract" system that does not use a sensory or motor code. Certainly, the systematic phonemic level (Pm), which is postulated to account for regularities in English sound structure, appears to play little or no rôle in STM for unstructured lists of vowels.

ACKNOWLEDGMENTS

This research was supported primarily by a grant from the National Institute of Mental Health, National Institutes of Health, Public Health Service, U. S. Department of Health, Education, and Welfare. Further aid was received from the National Aeronautics and Space Administration.