Predicting the Judged "Similarity of Sound" of English words.

Article *in* Journal of Verbal Learning and Verbal Behavior · August 1973

Paul C. Vitz
Institute for the Psychological Sciences

JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 12, 373–388 (1973)

Predicting the Judged "Similarity of Sound" of English Words[1]

PAUL C. VITZ AND BRENDA SPIEGEL WINKLER

New York University

In six experiments subjects rated the "similarity of sound" of pairs of words. The words varied from three to nine phonemes, and from one to three syllables. A model is proposed to account for the ratings which assumes the subject codes each word into a string of phonemes and computes a distance between the phoneme strings. The predicted phonemic distances correlated highly with rated similarity. Revisions of the model using distinctive features failed to improve the predictions, suggesting that factors at a level below phonemes are not important in judgments of word sound similarity. However, a revision using consonant and vowel phoneme clusters did improve the predictions. Evidence for rhyme, alliteration, stress, and serial position effects is also presented.

The purpose of this paper is to propose a simple model that can be used to predict the judged similarity of sound of any two English words. Very roughly, the following type of experiment provides the primary data to which the model is addressed. A native English speaking subject hears two isolated words, one word immediately after the other and the subject's task is to judge, on a rating scale, how similar the two words sound. The subject is to attend only to the sound of the words and to ignore any semantic or spelling characteristics. The average rating of the subject is used as a measure of the similarity of the two words. Presumably these ratings reflect how the words are perceived and/or how they are represented in memory while being compared.

Considerable previous research has been done on the perceptual or short-term memory similarities of the English vowels and consonants. See Cole, Sales, and Haber (1969); Sales, Cole, and Haber (1969); Wickelgren (1965b, 1966); Greenberg and Jenkins (1964); and Miller and Nicely (1955). These studies typically were limited to subsets of common vowels or to subsets of common consonants and in all cases they were concerned with the auditory perception or memory of various dimensions or features within phonemes. We are unaware of any studies directly comparing the auditory similarity of complete English words. One partial exception is a study by Nelson and Nelson (1970) in which subjects rated the "acoustic (articulatory)" similarity of three letter words. The stimuli were presented visually, the word pairs being typed on the pages of a response booklet. This allows spelling to influence the ratings, a possibility which the authors argue did occur for one set of words. Apparently, however, for the majority of the word pairs the ratings were primarily determined by acoustic or articulatory factors and the present model is applied to their data in a later section.

THE MODEL

Word Perception

In the subsequent experiments we assume the subject perceives the first word and then transfers the representation of this word to a short-term memory. Next, the subject per-

[1] This work was supported by Grant NS-07636-04 from the Institute of Neurological Diseases and Stroke, National Institutes of Health. Requests for reprints should be sent to Paul C. Vitz, Department of Psychology, New York University, 4 Washington Place, 9th Floor, New York, New York 10003. We thank Martin Braine, Eric Brown, Charles Cairns, and Murray Glanzer who provided useful comments and information at various stages of this work.

ceives the second word, transfers its representation or code to memory and finally the subject compares the two coded words in short-term memory to arrive at a rating of how similar they sound. For evidence that phonemic information is processed in short-term memory (and long-term memory), see Glanzer, Koppenaal, and Nelson (1972). What this code might be, and particularly how the codes are compared so as to arrive at a similarity rating, is described shortly. However, it will first be useful to outline the assumed process involved in word perception, the end result of which is the coded word.

The model of word perception followed here with some modifications is the hierarchical stages model presented by Studdert-Kennedy (1973, in press). The physical or acoustic word stimulus is first processed in stage 1, an auditory stage where the stimulus is translated into the "simple" auditory dimensions of loudness, pitch, quality, and duration. Stage 2 is the first stage in word perception as distinct from sound perception, and is called the phonetic stage. Here, the auditory signal is converted into discrete features and then into bundles of such features, phonetic segments. Stage 3, the phonological stage, codes the phonetic segments into a phonemic representation. It is also likely that phonemes are further grouped into consonant and vowel clusters and the clusters are grouped into syllables. The coded output from the phonological stage passes onto a higher and presumably final stage where lexical, syntactic, and semantic factors are all added to the phonological representation. In the present experiments, in which the words are presented separately and not in sentences, this final stage probably makes a minimal contribution particularly since semantic identification is not required. In short, the large amount of linguistically unnecessary information in the continuously varying auditory signal has been reduced via these hierarchical stages to the single category of a particular word.

The hierarchy is primarily a bottom-up or constructive process. However, linkages in the other directions are also postulated by Studdert-Kennedy. For example, both phonological and higher order hypotheses can feed back so as to strongly influence perception at the phonetic level. Such top-down or deductive influences from higher stages are used to explain the phonemic restoration effect of Warren (1970). We assume that most of the early stages are not consciously perceived, at least not when the subject is in the normal set occurring when listening to speech, and that subjects first become aware of speech at a relatively high level. In the present experiment this is assumed to be the syllable level; thus, the subject is first aware of a within syllable, within cluster string of phonemes.

Recently, Massaro (1972), Savin and Bever (1970), and Warren (1970) have argued that the syllable is the unit of speech perception. We assume that it is the first unit the subject is aware of when listening to speech but as Studdert-Kennedy points out, the proposed earlier processing of features and phonemes at lower levels is very likely rapid, unconscious, and essentially automatic. In the case of visual perception, much evidence and theory support the interpretation that prior to perceptual awareness of a letter or other geometric figure, various visual features, such as lines and angles (or corners), are being detected and organized to form the pattern (Hubel & Wiesel, 1965, McFarland, 1967, Vitz & Todd, 1971). In fact, awareness may be what occurs when the perceptual-cognitive process is terminated. According to this argument, awareness will occur at the highest level of organization that the subject brings or is instructed to bring to the task. At any rate, such a hierarchical process allows one to avoid arguments about *the* unit in word perception or word memory.

Although it would be more realistic to represent a word with a complex code summarizing some of this hierarchical process, only a simple phonemic code will be considered in the model. The hierarchical coding process is not directly involved in the distance measure or directly tested by the experiments; as mentioned, it is proposed to describe how the subject arrives at the phonemic code and to make it clear that the phonemic code is a considerable simplification. The particular phonemes to represent or code English words are shown in Table 1. Examples of phonemic representation are: plant/plænt/; relation /rəlēʃən/.

Predicted Phonemic Distance (PPD)

Two words must be aligned in order to compute the distance between them. To align two words is to place one above the other so that the phonemes in one can be compared with those in the other. The alignment rule is to minimize the distance between words. The distance between two identical phonemes is defined as zero and between two nonidentical phonemes as one. Therefore, minimizing is accomplished by maximizing the number of corresponding phonemes which are identical. Two phonemes correspond if, after alignment, one is directly above the other. Alignment can be accomplished by sliding the phonemes of one word above those of the other until the maximum number of identical corresponding phonemes is observed. Such identical corresponding phonemes are called 'matched' phonemes. Examples:

/sit/
/hit/

here the last two phonemes match;

/sit/*
*/its/,

here the two *i*s and two *t*s match;

/sit/**
**/tēk/,

here only the *t*s match. The asterisks represent the blank space against which the phoneme(s) in the other word is matched. In sliding the segments to maximize matches, it is permissible to break either word between

TABLE 1

PHONEMES USED FOR CODING ENGLISH WORDS

No.	Phoneme	Examples	No.	Phoneme	Examples
1	ī	Bee, leap	22	l	Lad, table
2	i	Pit, physic	23	p	Pat, tap
3	e	Pet, said	24	b	Bat, tab
4	ē	Pay, they	25	m	Mat, ram
5	æ	Pat, add	26	f	Fat, laugh
6	ai	Ride, aisle	27	v	Vat, avail
7	ā	Father, art	28	t	Tap, pat
8	au	Loud, fowl	29	d	Dab, bad
9	ɔ	Lot, what	30	n	Nab, ban
10	ɔ̄	Caught, all	31	s	Sat, scene
11	ɔi	Boil, coin	32	θ	Thatch, faith
12	ō	Low, coal	33	z	Zinc, houses
13	u	Put, full	34	ð	That, the
14	ū	Shoe, rude	35	ʃ	Sham, nation
15	yū	Few, mew	36	č	Chat, rich
16	ʌ	Cut, sun	37	ž	Azure, vision
17	ɚ	Supper, father	38	ǰ	Rage, jet
18	ə	Rodent, telephone	39	k	Cat, ache
19	y	Yam, yes	40	g	Go, ghost
20	w	Wail, await	41	h	Hat, cohere
21	r	Rail, arise	42	ŋ	Bang, sing

phonemes but the order of the phonemes is never changed. For example, *unchangeable* and *unexchangeable* can be aligned:

/ʌn***čēnǰəbəl/
/ʌneksčēnǰəbəl/,

and: *underwritten* and *relation*:

/*****rəlēʃən/
/ʌndərri**tən/.

The predicted phonemic distance, PPD, between two words is the proportion of phoneme positions after alignment which do not match. Thus, /sit/ and /hit/ have a PPD of 1/3; /sit/ and /its/ a PPD of 2/4; /sit/ and /tēk/ a PPD of 4/5; /ʌnčēnǰəbəl/ and /ʌneks-čēnǰəbəl/ have a PPD of 3/13. It is important to note that the asterisk positions count in computing the proportion of segment positions that match. Two words with no phonemes in common have a distance of 1.

The model is obviously greatly simplified and makes a number of assumptions which are known to be incorrect or are not likely to be correct. It assumes that all phonemes contribute equally to the total distance measure; that phonemes are either identical or completely different, that is, no degree of difference between segments is represented; that the distance is a simple average of the separate contributions, that is, possible interactions between adjacent segments are ignored. In spite of these and other assumptions, we believe the model provides a useful initial frameword with which to approach the similarity data. The following experiments explore the accuracy of the PPD scores by varying different phonemic and syllabic properties of the words being compared.

Experiments 1, 2, 3, and 4: Method

In each experiment there is one standard word which occurs in every pair; it is the distance from this standard that the subject rates. The comparison words range from those which have all but one phoneme in common with the standard to those with no phonemes in common. Between these extremes are words which have different numbers of segments in common, that is, the first phoneme or a middle phoneme.

Subjects

Each of the four experiments used 16 New York University students, median age approximately 22, about half of each sex. All subjects spoke native American English.

In Experiments 1 and 4, the subjects were paid volunteers. In Experiments 2 and 3, the subjects were fulfilling a course requirement. The subjects were run in sets of four except where smaller numbers were needed to complete a set.

Stimuli

Experiment 1. The 25 comparison words with their phonemic coding and other characteristics are shown in Table 2. The standard word for this list is: *sit*/sit/. Each comparison word is one syllable and consists of three phonemes and contains either two or three clusters. The words are familiar English words with the exception of the easily pronounced nonsense word: *tass*. Such occasional nonsense words were used to get a set of comparison words that provide systematic variation in the number and position of the phonemes that differ from the standard word. Four words differ from *sit* by three phonemes, nine differ by two, twelve differ by one, and one word, *its*, has the same phonemes in a different order. Those words with phonemes in common have them in the different possible positions and represent the logically possible ways of differing from the standard.

Experiment 2. The 25 comparison words and the standard word, *plant*/plænt/, with their phonemic coding and other characteristics are shown in Table 3. All of these words consist of one syllable, three clusters, and four or five phonemes. Again, there are a few easily pronounced nonsense words, for example, *plint*, *prant*. Five words have no phonemes in common with *plant*; six have one phoneme in common, this phoneme being in five different locations and representing five different phonemes. Five words have two phonemes in common with *plant*, four have three phonemes in common and five have four phonemes in common. It is not practical to create words that differ in all possible ways, however these comparison words cover the major possibilities.

Experiment 3. The 25 comparison words and the standard, *wonder*/wʌndɚ/, are shown in Table 4. These words all consist of two syllables (slight accent or stress on the first) and five or six clusters and five or six phonemes. Five words have no phonemes in common, six have one phoneme in common, four have two in common, five have three in common, and five have four in common. The phonemes which are in common occur in different positions and involve different phonemes.

TABLE 2

COMPARISON WORDS OF EXPERIMENT 1[a]

English spelling	Phonemic coding	Predicted phonemic distance	Obtained average similarity[b]	English spelling	Phonemic coding	Predicted phonemic distance	Obtained average similarity
1. Wage	wēj	1.00	0.15	14. But	bʌt	0.67	1.18
2. Rule	rūl	1.00	0.24	15. Live	liv	0.67	1.20
3. Keys	kīz	1.00	0.45	16. Song	sɔŋ	0.67	1.23
4. End	end	1.00	0.56	17. Tin	tin	0.67	1.27
5. Take	tēk	0.80	0.35	18. This	ðis	0.67	1.49
6. Tall	tāl	0.80	0.39	19. Miss	mis	0.67	1.66
7. Dose	dōs	0.80	0.53	20. Its	its	0.50	2.40
8. Tass	tæs	0.80	0.77	21. Sat	sæt	0.33	2.34
9. Toss	tɔs	0.80	0.83	22. Set	set	0.33	2.86
10. Mess	mes	0.80	1.09	23. Sick	sik	0.33	2.89
11. Imp	imp	0.75	1.18	24. Pit	pit	0.33	3.10
12. Feet	fīt	0.67	0.80	25. Hit	hit	0.33	3.16
13. Chin	čin	0.67	1.07				

[a] Standard word: *sit* / sit/.
[b] Obtained average similarity to standard (N = 16).

TABLE 3

COMPARISON WORDS OF EXPERIMENT 2[a]

English spelling	Phonemic coding	Predicted phonemic distance	Obtained average similarity[b]	English spelling	Phonemic coding	Predicted phonemic distance	Obtained average similarity
1. Screech	scrīč	1.00	0.11	14. Blond	blɔnd	0.60	0.90
2. Bricks	briks	1.00	0.15	15. Pierced	pīrst	0.60	0.98
3. Crowd	craud	1.00	0.18	16. Stand	stænd	0.60	1.51
4. Waves	wēvz	1.00	0.18	17. Plots	plɔts	0.60	1.79
5. Harmed	hārmd	1.00	0.42	18. Split	split	0.50	0.80
6. Limps	limps	0.83	0.32	19. Prank	prænk	0.40	1.85
7. Drink	drink	0.80	0.28	20. Grant	grænt	0.40	3.43
8. Smart	smārt	0.80	0.75	21. Plast	plæst	0.20	2.27
9. Blamed	blēmd	0.80	1.01	22. Plint	plint	0.20	2.28
10. Parks	pārks	0.80	1.27	23. Plans	plænz	0.20	2.84
11. Grams	græmz	0.80	1.79	24. Prant	prænt	0.20	3.39
12. Cramp	kræmp	0.80	2.37	25. Slant	slænt	0.20	3.49
13. Flushed	flʌʃt	0.60	0.29				

[a] Standard word: *plant* / plænt /.
[b] Obtained average similarity to standard (N = 16).

TABLE 4

COMPARISON WORDS OF EXPERIMENT 3[a]

English spelling	Phonemic coding	Predicted phonemic distance	Obtained average similarity[b]	English spelling	Phonemic coding	Predicted phonemic distance	Obtained average similarity
1. Sickles	siklz	1.00	0.04	14. Danger	dēnǰɚ	0.60	1.28
2. Tossing	tɔ̄siŋ	1.00	0.06	15. Widows	widōz	0.60	1.29
3. Locket	lɔket	1.00	0.10	16. Bundle	bʌndl	0.40	1.36
4. Raising	rēziŋ	1.00	0.16	17. Tender	tendɚ	0.40	2.05
5. Rapid	ræpid	1.00	0.18	18. Windle	windl	0.40	2.05
6. Chinning	činiŋ	0.80	0.01	19. Welder	weldɚ	0.40	2.38
7. Bucket	bʌket	0.80	0.14	20. Winter	wintɚ	0.40	2.64
8. Cradel	krēdl	0.80	0.16	21. Wuzder	wuzdɚ	0.20	2.69
9. Willows	wilōz	0.80	1.20	22. Wundle	wʌndl	0.20	2.91
10. Member	membɚ	0.80	1.57	23. Wander	wɔndɚ	0.20	3.56
11. Colors	kʌlɚz	0.67	0.71	24. Sunder	sʌndɚ	0.20	3.59
12. Handle	hændl	0.60	0.53	25. Wunter	wʌntɚ	0.20	3.63
13. Butler	bʌtlɚ	0.60	0.96				

[a] Standard word: *wonder* /wʌndɚ/.
[b] Obtained average similarity to the standard (N = 16).

Experiment 4. The characteristics of the comparison words and the standard, *relation*/rəlēʃən/, are shown in Table 5. The standard word consists of three syllables, with the stress on the second syllable, of seven clusters and seven phonemes. The comparison words range from four-syllable, nine-cluster, nine-phoneme words to one-syllable, three-cluster, three-phoneme words. The stressed syllables are marked with an accent. A one, a two-, and a three-syllable word have no phonemes in common with *relation*. One word has six phonemes and two syllables in common and the other 21 words have intermediate numbers of phonemes, phoneme positions, and syllables in common.

Procedure

The subjects sat at desks, 3–4 feet apart, and heard the tape-recorded word list presented through a high quality loudspeaker approximately 5 feet away. The tape recording presented pairs of words, each pair consisting of the standard and one of the 25 comparison words. The subjects were instructed to mark, on a rating scale, the similarity of the sound of the comparison word to the sound of the standard word. Subjects were specifically instructed to judge on the basis of the sound characteristics of the words and not on the basis of other characteristics such as meaning, associations, or spelling. Subjects used a 5-point scale, with the integers indicating degree of similarity marked as follows: 0, zero similarity, 1, slightly similar; 2, moderately similar; 3, very similar; and 4, extremely similar.

A practice list of word pairs preceded the first test list. This list had a standard word of the same number of segments, clusters, and syllables and the same stress as the standard word in the test list. The warm-up comparison words were also structurally similar to the comparison words in the test list. Rating the warm-up list cleared up any questions which arose about the procedure.

The subjects were first presented with the standard word, for example, *sit*. After 5 sec, the list of 25 pairs was presented in the following manner. "Comparison number one" (2-sec pause), "*sit*" (½ sec pause), "*dose*," and so on through the list. The subject was allowed as much time as needed for a rating, usually about 4 or 5 sec. After the 25 word pairs had been presented, there was a break of 2 min and a retest list of the same word pairs was presented with the standard word in the opposite position from where it occurred in the first list. That is, each pair of words occurred on one list in the order standard-comparison; and in the other in the reverse order. At the end of the experiment, the subjects described what criteria they used for their ratings. These comments were recorded. The entire experiment took 45–60 min. Each group of four subjects received a different and counter-balanced order of presentation for both lists of stimuli. Two groups of four subjects heard a female voice (accent: middle-eastern seaboard) reading the words, two groups heard a male voice (accent: middle-western). The procedure for Experiments 1, 2, 3, and 4 was identical in all respects.

TABLE 5

COMPARISON WORDS OF EXPERIMENT 4[a]

English spelling	Phonemic coding	Predicted phonemic distance	Obtained average similarity[b]	English spelling	Phonemic coding	Predicted phonemic distance	Obtained average similarity
1. Get	get	1.00	0.35	14. Lash	læʃ	0.71	0.80
2. Asthmatic	asmǽtik	1.00	0.36	15. Rush	rʌʃ	0.71	1.12
3. Deters	dītəz	1.00	0.40	16. Repel	rəṕel	0.67	1.36
4. Hesitate	hezitḗt	0.89	0.52	17. Inhibition	iń̃hiḃiʃən	0.67	1.59
5. Fascinating	fǽsinḗtiŋ	0.89	0.75	18. Emotions	imóʃənz	0.63	1.62
6. Sun	sʌn	0.86	0.41	19. Resolution	rézəlū́ʃən	0.55	2.38
7. Effect	əfekt́	0.86	0.48	20. Laceration	lǽsərḗʃən	0.50	2.59
8. Become	bəkúm	0.86	0.73	21. Revision	rəvížən	0.43	2.11
9. Rich	rič	0.86	0.87	22. Relate	rəĺēt	0.43	2.68
10. Legislates	léǰislḗts	0.80	1.23	23. Location	lōkḗʃən	0.43	2.84
11. Mechanic	məkǽnik	0.78	0.66	24. Ration	ŕēʃən	0.29	2.98
12. Underwritten	ʌńdərítən	0.73	1.47	25. Belation	bəĺēʃən	0.14	3.33
13. Pain	pēn	0.71	0.73				

[a] Standard word: *relation* /rəlḗʃən/.
[b] Obtained average similarity to standard (N = 16).

EXPERIMENTS 1, 2, 3, AND 4: RESULTS

The Rating Task

The subjects reported no serious difficulties in making their ratings. Since each comparison was presented in the order AB and BA, the correlation between these two sets of ratings provides a test-retest measure of rating reliability. The median test–retest correlations are .84, .86, .86, and .82 for Experiments 1–4, respectively. Each subjects 25 average similarity ratings were correlated with the ratings of the other 15 subjects. These correlations provide a measure of between subject reliability; they are .84, .76, .85, and .79 for Experiments 1–4.

PPD Accuracy

The major results are shown in Tables 2–5 and especially in Figs. 1–4. The tables present the PPD values and the obtained average similarity rating for each word. The figures show the correlation between PPD values and the ratings. In Experiment 1, shown in Fig. 1, the correlation is −.94; in Experiment 2, Fig. 2, the correlation is −.81; in Experiment 3, Fig. 3, r equals −.92; and in Experiment 4, Fig. 4, r equals −.95. With the exception of the −.81 value, these correlations are high and represent substantial support for the model.

Excluding the r of −.81, the other three scatter diagrams show a small but consistent pattern of discrepancy about the regression line. The similarity ratings of the small or low PPD scores tend to lie above the line and the large PPD scores tend to lie below it. This suggests an underlying S-shaped function is more descriptive.

Subject Reports

The subjects' post experiment reports of factors influencing their ratings provide interesting data, although they can not be relied upon heavily. Four possible structural units were mentioned as being used in making the ratings: words, syllables, clusters, and phonemes. The "word" was mentioned by only 5 subjects and appeared in the least articulate comments, for example, "(my) ratings were based on a global impression of the whole word." Sixteen subjects said they

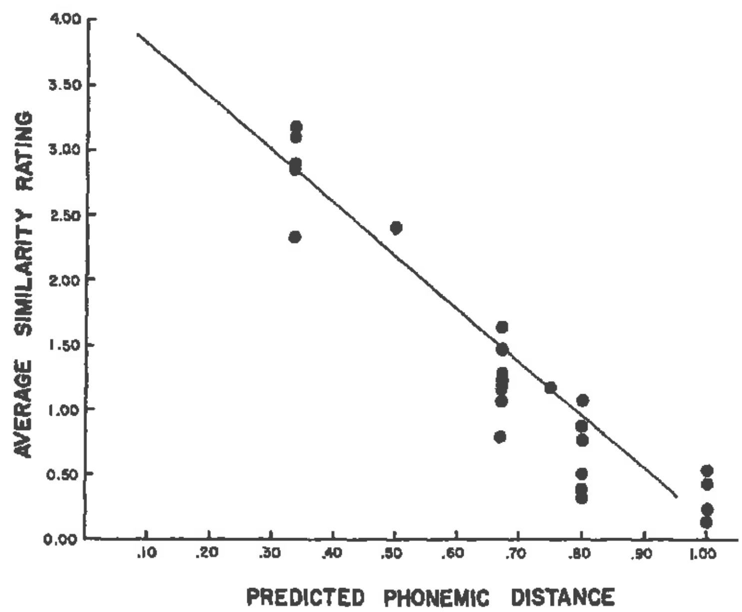

FIG. 1. Average similarity rating of comparison word to standard word as a function of predicted phonemic distance. Experiment 1—standard word: *sit*.

compared syllables and if subjects who did not use the term "syllable" but did refer to using particular syllables are included, then 25 subjects, or about 35%, referred to this unit. All but one subject who mentioned syllables also referred to units at the phoneme level as well. Clusters were never described by a general name but were described explicitly, for example, in Experiment 2 with *plant* as the standard, 5 subjects referred to the clusters

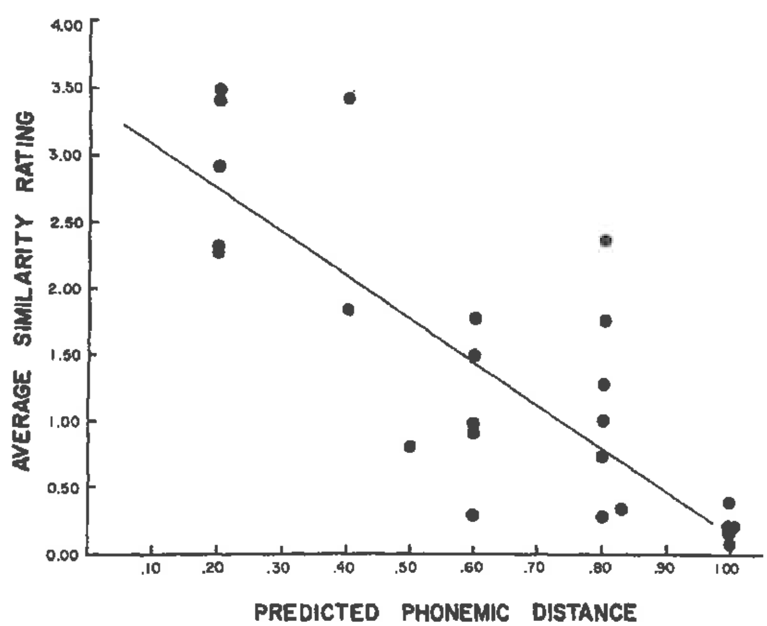

FIG. 2. Average similarity rating of comparison word to standard word as a function of predicted phonemic distance. Experiment 2—standard word: *plant*.

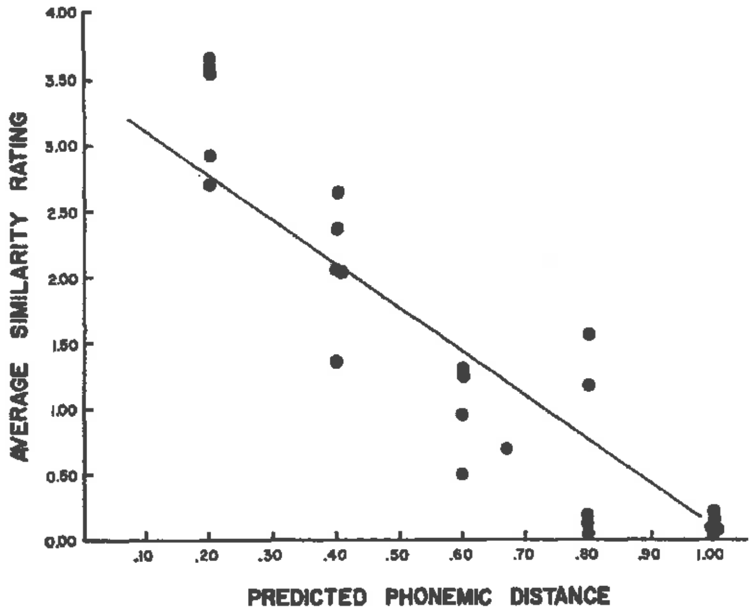

FIG. 3. Average similarity rating of comparison word to standard word as a function of predicted phonemic distance. Experiment 3—standard word: *wonder*.

pl and *cr*. Clusters almost never occurred in the other experiments and were not mentioned by subjects. Various items were interpreted as referring to phonemes, for example, "the individual sounds," "the sounds of the letters," and "phoneme" (this term was used by 2 subjects). Fourteen subjects used such general expressions; if the subjects who referred to particular phonemes, for example, /æ/, /i/, and /t/, are included, then a total of

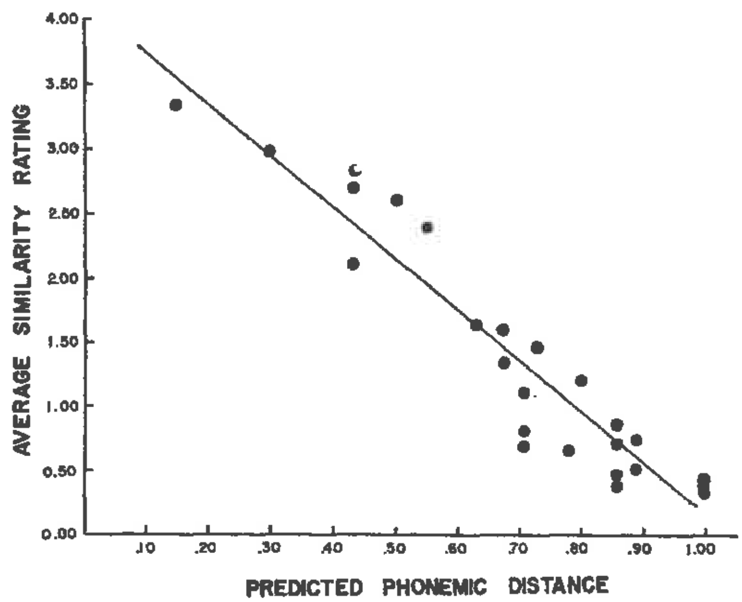

FIG. 4. Average similarity rating of comparison word to standard word as a function of predicted phonemic distance. Experiment 4—standard word: *relation*.

31 subjects, or about 50%, mentioned this level of unit. The subjects gave very little evidence for the significance of distinctive features since no subject referred directly to anything that resembled a distinctive feature, but 2 subjects (see below) did mention phoneme similarity and hence must have been responding to features. One other class of general structural unit was mentioned, vowels and consonants. About 30% mentioned the term "vowel" and only 16% referred to the class "consonants" even though the stimuli contained 3 to 4 times as many consonants as vowels.

We interpret the above data as generally supporting the model's assumption that subjects compare a phoneme code of the two words in order to get their rating. The subjects who referred to comparing whole words or syllables, we assume, compared them primarily on the basis of their phoneme similarities.

Rhyme was mentioned by subjects in each experiment and if one includes the subjects who did not use the term "rhyme" but referred to the importance of the last syllable, then about 25% of the reports showed evidence for rhyme as a factor. Rhythm, stress, and inflection were mentioned by only 4 subjects, three of whom were in experiment 4 involving multisyllable words where these factors were most prominent.

About half the subjects referred to the position of the syllable or phoneme as being important. The last consonant or the last syllable was mentioned as being important twice as often as the first syllable or sound. The middle positions were never mentioned as being most important and one subject said, "the middle positions were not so finely discriminated." This obviously suggests some kind of serial position effect of segment importance across the words.

One subject said the meaning of the words had an influence, 3 subjects reported they pictured the letters and the spelling; 2 subjects said the stimulus words to themselves before making a judgment. All of these comments suggest nonauditory factors as possible influences in the data.

Discrepancies from the PPD Scores

The model fails to reflect the similarity of phonemes but evidence for a possible similarity effect is found in Experiments 1, 2, and 3. In Experiment 1, /sat/ and /set/ each differ by the vowel from /sit/; but /set/ is judged closer to /sit/, presumably, because /e/ is closer to /i/ than /æ/ is to /i/. In Experiment 2, /græmz/ and /kræmp/ are rated much higher than their PPD scores predict. Presumably, this is due, at least in part, to the fact that the /r/ of the cluster /gr/ and /kr/ is similar to the /l/ of /pl/ and because the /m/ s are similar to the /n/ in the cluster /nt/. One subject in Experiment 2 reported explicitly that /pl/ and /bl/ were similar and that the /r/ of /gr/ and /kr/ was close to the /l/ of /pl/. In Experiment 3, /membɚ/ has only one segment in common with /wʌndɚ/ but the other four pairs of segments are similar and this probably accounts for its high rating. One subject referred to the similarity of the elements by saying he compared for similarity but used a "strict standard," for example, the segments had to be very similar for any similarity to be counted. Evidence for a rhyme effect can be seen in Experiment 1 in the five highest ratings, /hit/, /pit/, /sik/, /set/, and /sat/. Each of these is different from /sit/ by one segment but the two rhyming words are the highest and /sik/, a partial rhyme, is next. In Experiment 2 the three highest ratings rhyme with the standard: /slænt/, /prænt/, and /grænt/. The latter gets a very high rating even though it has only three segments in common. Other words which also have four or three segments in common are rated much lower when the common segments do not happen to rhyme, for example, /plint/ and /prænk/. In Experiment 3, the three highest ratings rhyme with the standard; six of the top seven rhyme with the standard in Experiment 4.

There is some evidence suggesting that the first letter was especially important. In Experiment 3 both /wilōz/ and /widōz/ are rated much higher than words with the same number of segments in common with the standard but with the common segments in middle positions, for example, /ciniŋ/, /bʌket/, and /kʌlɚz/.

There is, in Experiment 2, a suggestion that vowels or at least certain vowels tend to be weighted more than consonants. Thus, /kræmp/ and /græmz/ have only the vowel in common which may have resulted in the high ratings of 1.79 and 2.37, while /split/ with three consonants in common with /plænt/ has a low rating of 0.29. That vowels, or at least certain vowels, are given more weight would not be surprising, since vowels tend to have considerably more energy and to last longer than consonants. In addition, there may also be strong and weak consonants, for example, ʃ, s, p, b versus m, n, l, r but there is no case for this in our data.

An alternative interpretation to the general emphasis on vowels is that the feature nasal is especially important in judging similarity, since the nasal vowels are the problem in /græmz/ and /kræmp/ and nasal consonants in /membɚ/ and /wʌndɚ/.

The Nelson and Nelson Experiment

As previously mentioned, Nelson and Nelson (1970) provide data relevant to the model. They had 96 subjects rate 96 pairs of three-letter words. The words were typed on response booklets. Twelve of the pairs had no phonemes in common (PPD scores of 1); 36 had one phoneme in common, twelve pairs having a phoneme in common in each of the three different positions (PPD scores of .67 regardless of position); 36 had two phonemes in common, twelve with the first two positions in common, twelve with the last two positions in common, and twelve with the first and last position in common (all get PPD scores of .33); and twelve pairs were identical (PPD scores of 0). The correlation between the mean similarity rating of these pairs and the PPD is –.94 and is shown in Fig. 5.[2] If the twelve

[2] Professor Douglas L. Nelson graciously provided these data.

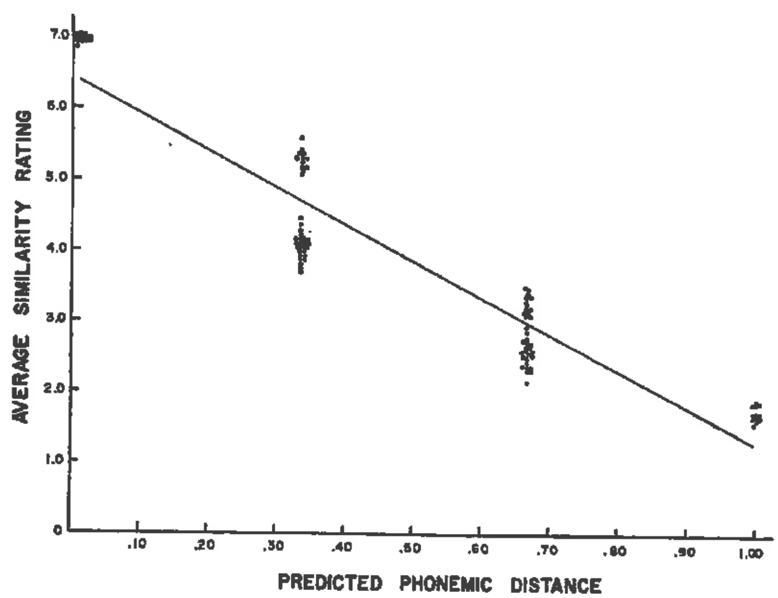

Fig. 5. Average similarity rating of 96 pairs of three letter words from Nelson and Nelson (1970) as a function of predicted phonemic distance.

pairs which were identical and had similarities of 7.00 or 6.98 are excluded, the new correlation is still a high −.90. Nelson and Nelson also provide statistically significant evidence for a rhyme effect since the twelve pairs which had the last two phonemes in common were rated more similar than those with the first two or first and last in common. In Fig. 5 the highest cluster of points at a PPD of .33 represents the rhyming pairs. They also found that words with only one phoneme in common were given a significantly higher rating if this was the first phoneme, that is, a first letter effect. In spite of these phoneme position effects, the correlation is quite high and accounts for most of the variance.

Nelson and Nelson also had subjects rate 15 word pairs which varied in both acoustic and spelling similarity. These 15 words, their phonemic code, their similarity ratings, and their PPD scores are shown in Table 6. The correlation between the PPD scores and the similarity ratings is −.70. This is low and we attribute this to the effect of the words' spelling, an interpretation Nelson and Nelson also make. To check on this, an additional experiment was run.

EXPERIMENT 5

Method

The 15 word pairs shown in Table 6 were tape recorded with each pair occurring in the order AB and BA. Four orders of presentation were used, four subjects hearing each order. The voice was the male speaker from Experiments 1–4. The rating scale from those experiments was also used.

Results

The new mean ratings are shown in Table 6 and the correlation between them and the PPD is −.77, a small improvement from −.70, but still low.

The ratings of Experiment 5 divide into two groups: high similarity scores if the vowels are the same and low if the vowels are differ-

TABLE 6

WORD PAIRS VARYING IN ACOUSTIC AND SPELLING SIMILARITY WITH THEIR PHONEMIC CODE, PREDICTED PHONEMIC DISTANCE, PREDICTED PHONEMIC CLUSTER DISTANCE, SIMILARITY RATING FROM NELSON AND NELSON AND SIMILARITY RATING FROM EXPERIMENT 5

Acoustic/spelling similarity[a]	Word pair	Phonemic code	Predicted phonemic distance	Predicted phonemic cluster distance	Average similarity rating: Nelson & Nelson (N = 96)	Average similarity rating: Experiment 5 (N = 16)
Low/high	1 yea–pea	yē–pī	1.00	1.00	3.04	0.29
	2 low–cow	lō–kau	1.00	1.00	2.91	0.62
	3 air–fir	er–fər	0.67	0.67	2.81	0.64
	4 wan–van	wān–væn	0.67	0.67	4.69	0.57
	5 out–hut	out–hʌt	0.67	0.67	2.38	0.65
High/high	6 sea–pea	sī–pī	0.50	0.50	5.45	2.89
	7 now–cow	nau–kau	0.50	0.50	5.43	3.05
	8 sir–fir	sər–fər	0.33	0.33	5.38	3.01
	9 tan–van	tæn–væn	0.33	0.33	5.42	3.00
	10 mut–hut	mʌt–hʌt	0.33	0.33	5.43	3.07
High/low	11 inc–arc	ink–ārk	0.67	0.75	2.41	0.75
	12 two–flu	tū–flū	0.67	0.50	5.00	2.85
	13 pea–ski	pī–skī	0.67	0.50	4.92	2.66
	14 rhy–why	rai–wai	0.50	0.50	5.38	2.96
	15 hoe–sew	hō–sō	0.50	0.50	4.72	2.87

[a] From Nelson and Nelson (1970).

ent. This may be another example of a rhyme effect but the low correlation of −.77 can also be attributed to the failure of the model to give vowels more weight than consonants. Evidence for the greater importance of vowels in a similar task was found by Wickelgren (1965a). A revision of the model which corrects, in part, for this weakness is described below.

DISCUSSION

Phoneme Similarity

On the whole, the model predicted well for such a simplified representation, but it is the factors which it fails to include that invite comment. The most obvious possible weakness is the absence of any measure of phoneme similarity. Besides the evidence cited, previous research has amply demonstrated, at least in perceptual and short-term memory tasks, that not all phonemes are equally confusable. This implies that what is needed is a phoneme matrix in which the entries represent the different distances between phonemes. One approach to such a matrix would be to derive it from a theoretical description such as the distinctive features of Chomsky and Halle (1968); from the work of Wickelgren (1965b, 1966); or from systems based on articulatory characteristics, for example, Jones (1966). All of these systems have an important common problem—how to weight the different features—since they do not all contribute equally. A different approach would be to empirically scale the rated differences between such segments without regard to a theoretical system. Regardless, a phoneme distance matrix combined with a rationale for the distance between a phoneme and silence would allow a revised predicted phonemic distance based on the minimum distance between aligned phonemes.

Revision 1: Phoneme Similarity Model

A rough attempt was made to evaluate the possible improvement in prediction that would result if a measure of phoneme similarity was taken into account. We used the Chomsky–Halle distinctive feature system (1968, Table 1, p. 176) and defined two segments as similar if they differed by only one feature; otherwise, they were defined as before, that is, as different. Two similar phonemes were given a distance of 0.5. These new PPD values were frequently identical to the original scores, but for those word pairs having phonemes which are supposedly very similar the distance was less. This revised PPD measure was applied to the words of Experiments 1–4 and to our surprise the new measure either did not improve the correlation or lowered it! The difficulty was not hard to find. As an example of the problem, in the Chomsky–Halle system /i/ and /ī/ differ by only one feature, tense, and are scored similar, but /i/ and /ʌ/ differ by two, high and back, and remain with distance scores of 1. But, from the ratings, /i/ and /ī/ are quite different while most of the vowels in the middle of the vowel quadrilateral are quite similar. That the Chomsky–Halle feature distance has only a very moderate correlation with perceptual and memory data can be seen quickly by comparing distances derived from their system with the perceptual confusion data of Miller and Nicely (1955, Tables I–IV) and the intrusion data of Wickelgren (1966). One interpretation is that this occurs because the Chomsky–Halle features vary greatly in how much they should be weighted for perceptual or psychological significance, for example, manner features being more important than place (Cole, Haber & Sales, 1968). This argument is consistent with the problem of /i/ and /ī/ versus /i/ and /ʌ/ since tense is a manner feature and high and back are not, and with the position made above that nasal's are especially important.

Deriving distance measures from other systems, for example, Wickelgren (1965b, 1966) and Jones (1966), proved unsatisfactory primarily because the conventional phonetic analysis of the vowels by two dimensions does not allow adequate discrimination among them. Perhaps an empirically derived rated phoneme matrix would give better results. Instead, our interpretation is that relatively little of the variance in the rating of complete words is due to factors existing at a lower or more molecular level than the phoneme. The phoneme certainly accounts for a large amount of the variance but it is probable that most of the remaining variance is at a still higher level. Possible higher level factors are: (1) stress, both the within syllable stress sometimes called "prominence," which usually means vowels get extra weight, and the common between syllable stress found in multisyllable words; (2) the contribution of higher order units, such as clusters or syllables; and (3) serial position effects.

Revision 2: Phonemic Cluster Model

A revision of the model that did lead to significant improvement was based on a new and somewhat higher structural unit, the phonemic cluster. A consonant cluster was defined as a single consonant or two or more adjacent consonants within a syllable. A vowel cluster is simply a vowel. A phonemic cluster is a phoneme representation of a consonant or vowel cluster. Thus, a phonemic cluster representation is /pl.æ.nt/, with dots designating cluster boundries. The other three standard words are unchanged by this new representation. The distance between two phonemic clusters is the proportion of phonemes which do not match after alignment. For example,

/st/
/s*/

has a distance of .50;

/st/
/pl/

a distance of 1;

/st*/
/*ts/

a distance of .67, and so on. The alignment rule for the phonemic cluster model is to minimize the phonemic cluster distance between the words being compared. Examples of this new alignment and distance are

/pl.æ.nt/
/gr.æ.nt/ $= (1 + 0 + 0)/3 = .33$;
/spl.i.*t/
/*pl.æ.nt/ $= (\frac{1}{3} + 1 + \frac{1}{2})/3 = .61$.

This predicted phonemic cluster distance, PPCD, is closely related to the old measure. In Experiments 3 and 4 all phonemic clusters are also single phonemes and no changes result. In Experiment 1 the words /end/, /imp/, and /its/ have a cluster code different from their phonemic code but their distance scores happen to remain unchanged. The only changes occur in Experiments 2 and 5. The PPCD scores for the words of experiment 2 are shown in Table 7. The PPCD scores for Experiment 5 are shown in Table 6. This revision makes all single consonants or clusters equal to each other and equal to a vowel. The major effect is to give vowels a proportionally greater weight in those words having clusters containing more than one consonant. The revised model's correlation for Experiment 2 is now −.90, a gain in the variance accounted for from 65 to 81%, a substantial improvement. In Experiment 5, the correlation is improved from −.77 to −.86. Since the correlations in the other experiments are unchanged, the phonemic cluster model appears to be superior to the simpler phoneme model. These improvements are, in addition, support for the argument that when the stimuli are complete words their code is primarily at a phoneme or higher level.

Serial Position, Rhyme and Alliteration

The Nelson and Nelson (1970) data make it clear that the first phoneme and the last two are relatively more important. This, plus the modest evidence from Experiments 1–4, suggests a serial position curve with a strong recency effect and a moderate primacy effect as in the typical free recall serial position curve. An explanation of the common poetic devices of

TABLE 7

REVISED MODEL'S PREDICTED PHONEMIC CLUSTER DISTANCE FOR THE WORDS OF EXPERIMENT 2[a]

English spelling	Phonemic cluster coding	Phonemic cluster distance	Obtained average similarity	English spelling	Phonemic cluster coding	Phonemic cluster distance	Obtained average similarity
1. Screech	scr.ī.č	1.00	0.11	14. Blond	bl.ɔ.nd	.67	0.90
2. Bricks	br.i.ks	1.00	0.15	15. Pierced	p.ī.rst	.72	0.98
3. Crowd	cr.au.d	1.00	0.18	16. Stand	st.æ.nd	.50	1.51
4. Waves	w.ē.vz	1.00	0.18	17. Plots	pl.ɔ.ts	.72	1.79
5. Harmed	h.ā.rmd	1.00	0.42	18. Split	spl.i.t	.61	0.80
6. Limps	l.i.mps	0.83	0.32	19. Prank	pr.æ.nk	.50	1.85
7. Drink	dr.i.nk	0.83	0.28	20. Grant	gr.æ.nt	.33	3.43
8. Smart	sm.ā.rt	0.83	0.75	21. Plast	pl.æ.st	.33	2.27
9. Blamed	bl.ē.md	0.83	1.01	22. Plint	pl.i.nt	.33	2.28
10. Parks	p.ā.rks	0.83	1.27	23. Plans	pl.æ.nz	.17	2.84
11. Grams	gr.æ.mz	0.67	1.79	24. Prant	pr.æ.nt	.17	3.39
12. Cramp	kr.æ.mp	0.67	2.37	25. Slant	sl.æ.nt	.17	3.49
13. Flushed	fl.ʌ.ʃt	0.67	0.29				

[a] See Table 3 for original coding and PPD.

alliteration and rhyme is that they take advantage of this curve of relative importance within a word. Deriving rhyme and alliteration effects from a serial position argument avoids postulating them as new and independent effects. These effects should be common to all subjects although this does not imply that every culture should use them in their poetry. Evidence for a serial position effect in long-term memory for words is presented in a study by Horowitz, White, and Atwood (1968) who show that the first and last phonemes of a word are especially good cues for word retrieval as compared to middle position phonemes.

REFERENCES

CHOMSKY, N., & HALLE, M. *The sound pattern of English.* New York: Harper and Row, 1968.

COLE, R. A., HABER, R. N., & SALES, B. D. Mechanisms of aural encoding. I. Distinctive features for consonants. *Perception and Psychophysics,* 1968, **3**, 281–284.

COLE, R. A., SALES, B. D., & HABER, R. N. Mechanisms of aural encoding: II. The role of distinctive features in articulation and rehearsal. *Perception and Psychophysics,* 1969, **6**, 343–348.

GLANZER, M., KOPPENAAL, L., & NELSON, R. Effects of relations between words on short-term memory storage and long-term storage. *Journal of Verbal Learning and Verbal Behavior,* 1972, **11**, 403–416.

GREENBERG, J. H., & JENKINS, J. J. Studies in the psychological correlates of the sound system of American English. *Word,* 1964, **20**, 157–177.

HOROWITZ, L. M., WHITE, M. A., & ATWOOD, D. W. Word fragments as aids to recall: The organization of a word. *Journal of Experimental Psychology* 1968, **76**, 219–226.

HUBEL, D. H., & WIESEL, T. N. Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology,* 1965, **28**, 229–289.

JONES, D. *The pronounciation of English.* Cambridge: Cambridge University Press, 1966.

MCFARLAND, J. H. Some evidence bearing on operations of "analysis" and "integration" in visual form perception by humans. In: W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form.* Cambridge, Massachusetts: M.I.T. Press, 1967.

MASSARO, D. W. Perceptual images, processing time, and perceptual units in auditory perception. *Psychological Review,* 1972, **79**, 124–125.

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

NELSON, D. L., & NELSON, L. D. Rated acoustic (articulatory) similarity for word pairs varying in number and ordinal position of common letters. *Psychonomic Science*, 1970, **19**, 81–82.

SALES, D. B., COLE, R. A., & HABER, R. N. Mechanisms of aural encoding: V. Environmental effects of consonants on vowel encoding. *Perception and Psychophysics*, 1969, **6**, 361–365.

SAVIN, H. B., & BEVER, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 293–302.

STUDDERT-KENNEDY, M. The Perception of Speech. In: T. A. Sebeok (Ed.), *Current Trends in Linguistics*, Vol. XII. The Hague: Mouton, in press, 1973.

VITZ, P. C. & TODD, T. C. A model of the perception of simple geometric figures. *Psychological Review*, 1971, **78**, 207–228.

WARREN, R. M. Perceptual restoration of missing sound. *Science*, 1970, **167**, 392–393.

WICKELGREN, W. Acoustic similarity in intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965a, **70**, 102–108.

WICKELGREN, W. A. Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, 1965b, **38**, 583–588.

WICKELGREN, W. A. Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, 1966, **39**, 388–398.

(Received December 18, 1972)