DOCUMENT RESUME

ED 108 130 CS 001 896

AUTHOR	Travers, Jeffrey R.
TITLE	Formal Models of Word Recognition. Final Report.
INSTITUTION	Swarthmore Coll., Pa.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C. Office of Research Grants.
BUREAU NO	BR3-0454
PUB DATE	2 Apr 75
GRANT	NE-G-00-3-0032
NOTE	160p.
EDRS PRICE	MF-$0.76 HC-$8.24 PLUS POSTAGE
DESCRIPTORS	Decoding (Reading); Models; Reading; *Reading Comprehension; *Reading Processes; *Reading Research; Reading Skills; Verbal Stimuli; *Word Recognition

ABSTRACT

Existing mathematical models of word recognition are reviewed and a new theory is proposed in this research. The new theory integrates earlier proposals within a single framework, sacrificing none of the predictive power of the earlier proposals, but offering a gain in theoretical economy. The theory holds that word recognition is accomplished by filtering visual feature information from the printed word through a hierarchy of letter, letter-cluster, and word detectors. The detectors are Bayesian decision devices which estimate the likelihood of the presence of their target configurations by combining information from lower detectors with a priori knowledge about the structure of words in English. In addition, several empirical studies on issues related to the theory were conducted. Two of these studies demonstrated that skilled readers draw visual information from all the letters in a word at once, rather than from one letter at a time; and that statistical co-occurance of letter sequences affects the perceptibility of those sequences, independent of their pronounceability. A third study, on whether covert pronunciation of words is necessary to apprehend their meaning, proved inconclusive. (Author/WR)

ED108130

Final Report

Project No. 3-0454
Grant No. NE-G-00-3-0032

Formal Models of Word Recognition

Jeffrey R. Travers
Department of Psychology
Swarthmore College
Swarthmore, Pennsylvania 19081

April 2, 1975

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
NATIONAL INSTITUTE OF EDUCATION
OFFICE OF RESEARCH GRANTS

S 001 896

Final Report

Project No. 3-0454
Grant No. NE-G-00-3-0032

Formal Models of Word Recognition

Jeffrey R. Travers
Department of Psychology
Swarthmore College
Swarthmore, Pennsylvania 19081

April 2, 1975

The research reported herein was performed pursuant to a grant with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U.S. DEPARTMENT OF

HEALTH, EDUCATION, AND WELFARE

NATIONAL INSTITUTE OF EDUCATION

OFFICE OF RESEARCH GRANTS

Contents

Abstract

Existing mathematical models of word recognition are reviewed and a new theory proposed. The new theory integrates earlier proposals within a single framework, sacrificing none of the predictive power of the earlier proposals, but offering a gain in theoretical economy. The theory holds that word recognition is accomplished by filtering visual feature information from the printed word through a hierarchy of letter, letter-cluster and word "detectors." The detectors are Bayesian decision devices which "estimate" the likelihood of the presence of their target configurations by combining information from lower detectors with *a priori* knowledge about the structure of words in English. The theory accounts for such phenomena as the ease with which words and wordlike nonwords can be read (relative to random letter strings), the effects of word and letter-cluster frequency on recognition, and the effects of reader expectations based on prior syntactic and semantic context.

In addition, several empirical studies on issues related to the theory were conducted. These demonstrated (1) that skilled readers draw visual information from all the letters in a word at once, rather than from one letter at a time; and (2) that sheer statistical co-occurrence of letter sequences affects the perceptibility of those sequences, independent of their pronounceability. A third study, on the question of whether covert pronunciation of words is necessary to apprehend their meaning, proved inconclusive.

The results of the theoretical and empirical studies imply that skilled readers process words as perceptual wholes.

Introduction

Word recognition is of practical interest because it is a central process in reading; moreover, because it involves fundamental perceptual and cognitive skills, it is of broad theoretical interest for psychology as well. The principal purpose of the research described in this report was to develop a mathematical model of the information processing which underlies the skilled reader's ability to recognize words. The value of such a model lies not in the mathematics per se but in the fact that formalization requires the theorist to be precise and complete, thus either forcing him to understand the phenomenon in depth or revealing his ignorance of crucial aspects of it. Results of the modeling effort are described in detail in Appendix A of this report. The appendix, a paper entitled "Formal Models of Word Recognition" attempts to integrate existing mathematical treatments within a more comprehensive framework that carries the predictive power of all the previous models together. The body of the report briefly chronicles the efforts which produced the "Models" paper and summarizes the paper's contents.

The proposal for this research (Travers, 1973a) outlined the presuppositions of the modeling effort and the specific problems with which the effort would deal. To recapitulate briefly some of the key points:

(1) It was assumed that a complete model of word recognition must be integrated with a subordinate model of letter perception and a superordinate model of language comprehension. Letter perception is a special case of visual pattern recognition, a process which has received extensive formal theoretical treatment. Following Neisser's (1967) review of the literature, it was proposed that letter recognition is accomplished by a hierarchical feature-extraction system like that modeled in Selfridge's (1959) computer simulation. In contrast to the situation with letter perception, where a reasonably adequate prior model provides us with theoretical building blocks, language comprehension remains an unsolved problem, and one that lies far outside the scope of word recognition per se. Therefore no attempt could be made to borrow or construct a comprehension model. At the same time, it was clear that any useful model of word recognition must give some account of the effects of syntactic and semantic context. Resolving this dilemma was to be one of the tasks of the modeling effort.

(2) Perhaps the central fact to emerge from nearly a century of empirical work on word recognition is the fact that letters within words can be reported more accurately than letters within random strings of letters. This phenomenon, dubbed the "word apprehension effect" (WAE) by Neisser (1967)

must be explained in terms of some sort of integrative mechanism which combines individual letter percepts into wholistic representations of words. (Such processes might occur in perception, memory, response organization, or any combination of these three loci.) Constructing such an integrative mechanism was to be the primary task of the model. However, the model was not to be ad hoc or limited to the WAE alone; it was to be sufficiently comprehensive and flexible to explain a wide range of results in the area, e.g., those concerning frequency effects, subject expectations, and such other phenomena as might emerge from a review of the experimental literature.

(3) In line with the author's previous research (1970, 1973b, 1974) it was assumed that the integrative mechanism would operate "in parallel", i.e., that visual feature information is extracted from all letter positions within a word simultaneously. A "contingent parallel" model structure was proposed--i.e., one in which feature analyzers are integrated into letter, letter-cluster and word analyzers, with economy in feature extraction introduced at these higher levels due to redundancies in the language. (That is, the model proposes that words can be reported more accurately than letter strings because less feature information is needed to identify a letter in a word.) The mathematical details of the various analyzers remained to be worked out during the modeling effort; however, it was suggested that existing theoretical structures, e.g., those of statistical decision theory or signal-detection theory, might be adapted to describe the operation of the hierarchy of detectors.

A secondary aspect of the funded research was execution of several new experiments on word recognition, dealing with issues relevant to the model but not directly treated in the "Models" paper. These experiments, two successful, one unsuccessful, are described in Appendices B, C and D of this report. Again, the body of the report contains only a brief summary of the empirical work conducted. Three empirical questions were considered:

(1) Could the author's earlier empirical work on parallel and serial processing (1970, 1973b, 1974) be extended to visual stimuli which resemble normal print, and would the earlier findings be confirmed when more stringent experimental controls were introduced? That is, would the assumption that feature information from multiple letter locations is processed simultaneously stand up to new tests?

(2) Many investigators (e.g., Miller, Bruner and Postman, 1954; Gibson, Pick, Osser and Hannond, 1962; Baron and Thurston, 1973) have shown that "wordlike" nonwords exhibit some of the perceptual, mnemonic or response advantages

shown by words. What structural features of "wordlike" nonwords cause them to be so accurately reported--and what would an answer to this question tell us about word perception itself?

(3) Many people, even skilled readers, "pronounce" words silently as they read; indeed, reading is often defined or described as translation of visual signals to internal speech. But is covert auditory recoding really necessary in extracting meaning from visual symbols?

Method

A. Theory Construction

The primary, or theoretical, effort of the project had two components--first, an extensive review of existing theories and relevant experimental findings, and second, construction of the theory itself.

The literature review phase of the project proved to be a more demanding and revealing task than had been anticipated. The task was demanding in that the body of potentially relevant data was simply too vast to be reviewed exhaustively, particularly when sources in the educational literature were added to those in experimental psychology itself. Fortunately, as the theory developed, it began to provide selectivity principles by which many otherwise important findings could be set aside. To cite some examples: (1) The literature on differential effectiveness of "whole-word" vs "phonic" teaching techniques (Chall, 1970) was ignored on the grounds that (a) processes involved in learning may differ from processes used by the skilled reader, and (b) the effectiveness of teaching techniques depends on many factors, such as motivation, curriculum design, etc., which lie outside the information-processing strategies under consideration in the theory. (2) Experimental findings bearing on such tasks as visual search through a letter list, search through a letter list in short-term memory, word-nonword discrimination, etc. were ignored, on the grounds that these tasks are unlike reading and may introduce task-specific cognitive strategies which replace or obscure those used in reading. (3) Eye-movement studies of reading were ignored, simply because they yield no information about the processes which take place <u>within</u> a single visual fixation. (Most words can be recognized with a single fixation.) Ultimately, the theoretical effort focused on full- and partial-report tachistoscopic tasks, which attempt to elucidate the processes which occur during a single fixation. In particular, the ingenious task devised by Gerald Reicher (1969), which controls most memory and response factors, throwing into sharp relief the perceptual processes in word recognition, received a great deal of attention.

The literature review task was revealing in that it unearthed several new theoretical papers, some of them published in recent months, some still unpublished, which anticipated many of the ideas outlined in the grant proposal (Travers, 1973a). The work of Estes (1974, 1975) and of Rumelhart and Siple (1974), in particular, contains many of the key ideas which the model was to develop. However, close examination of the paper just cited, as well as two other recently proposed formal models (Smith and Spoehr, 1974; Morton, 1969) showed that important theoretical work remained to be done. Though each of the four models has considerable predictive power with respect to some set of word-perception findings, they appear to focus on somewhat separate empirical domains. It became clear that a general model which integrated the four would represent a considerable advance in theoretical simplicity, with no loss in predictive power. Certain formal similarities among the models emerged under close scrutiny, and it was possible to construct an integrated model without major distortions of any of the four. Appendix A describes both the distinctive predictions of the models and their formal similarities, ending with the proposed integration.

B. Empirical Studies

1. Parallel vs. serial processing. This issue was addressed by a technique previously developed by the author (Travers, 1970, 1973b, 1974). In this technique, subjects are forced to process letters within words one at a time, by means of serial display of letters with a backward mask following each letter. Such displays markedly impair word recognition, suggesting that parallel, rather than serial processing, is the preferred strategy for the skilled reader. As noted above, the new research attempted to confirm and extend earlier findings in this regard, using new visual displays and improved experimental controls. The previous work had been done using light-on-dark uppercase letters displayed by a computer-controlled oscilloscope. The new displays were black-on-white lowercase typed letters displayed via a stroboscopic tachistocope. Obviously, the new displays are far more like ordinary print than the old; should differences in performance be obtained, the new results would clearly be the more relevant to ordinary reading. Also, one of the earlier studies (Travers, 1974) lacked a crucial experimental control and therefore did not give clear evidence on the question of whether simultaneous availability of feature information actually enhances word perception. The relevant control was included in the new study.

2. Structural properties of nonwords. "Wordlike" nonwords have many properties--pronouncability, orthographic

regularity, statistical resemblance to English letter sequences, etc.--any of which might account for their ease of recognition relative to random letter strings. Prevailing opinion attributes this effect to pronounceability and/or orthographic regularity, rather than to statistical factors. However, closely controlled studies, which vary pronounceability and statistical "Englishness" orthogonally, have not been performed. Using a new measure of statistical Englishness, strings high and low in Englishness, and also either high or low in pronounceability, were constructed. These were presented to subjects in a tachistoscopic report task, in an effort to determine whether either (or both) of the two factors exert an effect independent of the other.

3. Semantics and phonology. Chomsky (1970) has argued that many of the "irregularities" of English spelling in fact permit the written language to represent underlying meaning relations among words more accurately than would an orthography more faithful to phonetics. (For example, in the word-pair "courage-courageous"), the letter sequence courage has different sound values, but clearly represents the underlying kinship of meaning.) A reaction-time experiment was conducted in order to determine whether semantic relations are easier to detect when variations in sound pattern like that exemplified by "courage-courageous" are not involved. For example, would subjects be quicker to detect the semantic kinship between "outrage-outrageous", which involves no shift in vowel sound, than in "courage-courageous", which involves such a shift? If so, the RT result would constitute evidence that semantic judgments are affected by phonological factors, possibly because a phonological recoding stage intervenes between visual processing of a word and apprehension of its meaning.

Methodological details of the studies sketched in 1-3 above are given in Appendices B-D, respectively.

Results

A. The Theory

The hierarchical feature based system outlined in the grant proposal (Travers, 1973a) readily incorporated the proposals of the four mathematical models mentioned earlier: (1) Estes (1974, 1975) describes a hierarchy of feature, letter, cluster and word analyzers virtually identical to the one proposed, so that problems of integration obviously do not arise in the case of his theory. However, Estes has formalized only a small portion of his model, and in particular has not given a formal account of how the hierarchy of feature, letter, cluster and word-detectors interact;

therefore further mathematization was clearly necessary. (2) Rumelhart and Siple (1974) propose a similar model but one with a less elaborate hierarchical structure; however they provide an explicit, Bayesian decision rule to describe the operation of their detectors. Therefore the Rumelhart-Siple mathematics was borrowed and applied to the richer Estes structure. (3) Morton (1969) proposes a model with only one level of detectors--word detectors, or "logogens" as he calls them. However, he attributes to his logogens a formal decision principle quite like that of Rumelhart and Siple. Therefore, the Morton model could be seen as a special case of the hybrid Rumelhart-Siple-Estes model. (4) Smith and Spoehr (1975) propose a model with an elaborate parsing rule for segmenting printed words into syllable-like units. Though the parsing rule, conceived as a set of real-time psychological processes, could not be incorporated into the hybrid model, the units themselves could be incorporated, by the simple expedient of setting up cluster analyzers whose target clusters were those prescribed by the Smith-Spoehr rule.

This elaborate effort to collapse the four models into one another was not an arbitrary theoretical exercise, but was motivated by a desire to create a single theory with the power to predict a wide range of human performance data. Each of the four models was constructed to explain a particular set of data, and each has proved successful in making accurate quantitative predictions. In particular: (1) Estes' model predicts the results of experiments using the Reicher paradigm and of a new variant introduced by Estes himself, including intricate predictions about the pattern of errors in the Estes procedure. (2) The Rumelhart-Siple model predicts the results of full-report tachistoscopic tasks, including many subtle results having to do with frequencies of words and letter clusters. (3) The Smith-Spoehr model predicts the results of experiments showing the effects of syllable structure on word recognition, and on perceptibility of nonwords which resemble English in varying degrees and ways. (4) The Morton model is of special interest because it predicts the effects of syntactic and semantic context on word recognition. As noted in the introduction, this is a special problem for theories of word recognition because we lack an adequate theory of language comprehension. However Morton sidesteps this problem by treating context effects in terms of the reader's expectations, operationalized as his ability to predict particular words in context. His model gives accurate quantitative predictions about the interaction of stimulus and context effects. In short, the comprehensive model which incorporates the four previous models gains the ability to predict a very broad range of results having to do with word frequency, reader expectations, syntactic and semantic context, and structural characteristics of words and nonwords. In this sense the model represents an advance in theoretical economy.

B. The Experiments

1. Parallel vs. serial processing. Although certain of the author's earlier findings (Travers, 1970, 1973b) proved to be confined to the rather unusual computer displays used in those experiments, the essential outcomes were replicated using displays more like those or ordinary reading. In particular, it was shown that subjects can recognize words much more easily when all letters are available simultaneously than when letters become available one at a time--even when the display time for a word shown as a whole is equal to that for each letter shown sequentially. This is strong evidence that skilled readers process words as complex perceptual gestalts, and not as sequences of letters; moreover the finding articulates perfectly with the hierarchical model, which proposes that visual features from multiple letter locations are simultaneously filtered through a network of detection devices.

2. Pronounceability vs. statistical Englishness. Strings of letters which exhibited high statistical transition probabilities among letters, but which could not be easily pronounced (e.g., SPHST) and strings with low probability but high pronounceability (e.g., UMFIK) were both recognized more easily than strings low in both statistical Englishness and pronounceability. However, when a very conservative statistical test was applied, only the transition-probability effect proved significant. Clearly, this result leaves open the question of whether pronounceability exerts a perceptual/mnemonic effect independent of cluster frequency; however it demonstrates unequivocally that cluster frequency has an effect independent of pronounceability. Again, the result articulates with the model, which assumes that cluster detectors are established through long-term perceptual learning, and performs truly "visual" functions in word recognition.

3. Phonetic recoding and semantic relations. No interaction was found between the phonetic relations between pairs of words and the speed with which subjects could judge their semantic relatedness. (Pairs with phonological shifts, like "courage-courageous" were judged as rapidly as pairs without such shifts, like "outrage-outrageous.") A variety of potentially interfering factors, such as word length and frequency, were uncontrolled in this pilot experiment; however close examination of the data revealed no systematic relation between these variables and the phonological structure of the word pairs; hence there seemed little promise that a more careful experiment would produce an effect of phonology on semantic judgments. Clearly, such negative findings do not permit strong conclusions; however the null result is at least consistent with the belief that

skilled readers can apprehend meaning without recourse to phonological coding.

Conclusions

Both the theoretical and empirical work described in this report suggest that skilled readers, through repeated exposure to English words, build up complex perceptual representations of letters, words and of frequent letter configurations. These representations, best characterized as lists of visual features, enable skilled readers to construct complex percepts on the basis of limited visual input. This is why a word or wordlike nonword can be read at a glance, while a string of unrelated letters requires close attention. While learn[illegible] read, the child may need to go through a laborious [illegible] of letter-by-letter, or cluster-by-cluster phonetic recoding--and adults may do so when confronted with unfamiliar words. However, most of the words encountered by the skilled reader are like familiar faces--complex sets of visual features that can be apprehended simultaneously, rather than through successive focusing. In this limited sense, the process of learning to read does not end when the child has mastered English phonics (spelling-to-sound correspondence rules); exercise of his new recoding skills leads him (unconsciously) to undergo a process of perceptual learning which changes reading from a tedious process to an efficient and comfortable one. (Of course, nonperceptual skills also can enhance the efficiency of reading--e.g., the ability to guess and predict words from context, which as we have seen reduces the amount of perceptual input necessary to identify words correctly.) While it would be absurd to claim that these broad conclusions are forced on us by the theory and data reported here, they are surely suggested by the present report and by a wide range of previous data as well--and their practical importance makes them worthy of further investigation.

Bibliography

Baron, J. and Thurston, I. An analysis of the word-superiority effect. Cognitive Psychology, 1973, 4, 207-228.

Chall, J. Learning to read: the great debate. New York: McGraw-Hill, 1967.

Chomsky, N. Phonology and reading. In H. Levin and J. P. W[illegible]s (Eds.) Basic studies on reading. New York: B[illegible]ooks, 1970.

Estes, W. K. On the locus of inferential and perceptual processes in letter identification. Unpublished paper, 1974.

Estes, W. K. Memory, perception and decision in letter identification. In R. L. Solso (Ed.), Information processing and cognition: the Loyola Symposium. Washington, D.C.: Erlbaum Associates, 1975.

*
Miller, G. A., Bruner, J. and Postman, L. Familiarity or letter sequences and tachistoscopic identification. Journal of General Psychology, 1954, 50, 129-139.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 274-280.

Rumelhart, D. E. and Siple, P. Process of recognizing tachistoscopically-presented words. Psychological Review, 1974, 81, 99-118.

Selfridge, O. G. Pandemonium: a paradigm for learning. In The Mechanization of thought processes. London: H. M. Stationery Office, 1959.

Smith, E. E. and Spoehr, K. T. The perception of printed English: a theoretical perspective. In B. H. Kantowitz (Ed.) Human information processing: tutorials in performance and cognition. Published by Lawrence Erlbaum Associates; distributed by Halsted Press division of John Wiley and Sons, New York, 1974.

Travers, J. R. Sequential and parallel processes in reading. Unpublished doctoral dissertation, Harvard University, 1970.

Travers, J. R. Development of a formal model of word recognition. Research proposal submitted to the National Institute of Education, March 1, 1973. (Project later funded under grant number NE-G-00-3-0032) (a)

Travers, J. R. The effects of forced serial processing on identification of words and random letter strings. Cognitive Psychology, 1973, 5, 109-137. (b)

Travers, J. R. Word recognition with forced serial processing: effects of segment size and temporal order variation. Perception and Psychophysics, 1974, 16, 35-42.

*Gibson, E. J., Pick, A. D., Osser, H. and Hammond, M. The role of grapheme-phoneme correspondences in the perception of words. American Journal of Psychology, 1962, 75, 554-570.

FORMAL MODELS OF WORD RECOGNITION[1]

Jeffrey R. Travers

Swarthmore College

What is a theory of word recognition for? The question is intentionally ambiguous. On one hand, it is a question about motivation: Why do we wish to construct a theory of word recognition? On the other hand, it is a question about goals and conditions of adequacy: What are the data for which the theory must account, and how can a satisfactory account be characterized?

With respect to motivation, it seems obvious that there are compelling practical and theoretical reasons to attack the problem: Word recognition is presumably an important component of reading; if we understood the skill better, perhaps we could learn to teach it more effectively to children and illiterate adults. At the same time, reading is a complex ability which taps the most basic processes of perception, cognition and language comprehension; if we make significant advances in understanding any aspect of reading, we must necessarily penetrate more deeply the nature of human information processing. The study of word recognition in particular promises to unlock some basic issues having to do with the perceptual integration of elements in complex patterns.

Unfortunately, this tidy statement of motivation sidesteps a host of thorny questions: Can word recognition be studied meaningfully, apart from reading as a whole? If not, the tasks we impose on subjects in the laboratory are robbed of their immediate practical interest; whether or not the tasks, and the fragmentary theories we construct to explain our subjects' behavior, retain more general scientific value then depends on whether the tasks reflect fundamental cognitive skills. But how do we discriminate fundamental skills from transitory, task-created strategies? Questions like these should not be resolved on the basis of preconceptions and cannot be resolved on the basis of existing evidence. Nevertheless it is important to raise such questions, to keep our general aims in mind as we review and evaluate specific theories.

With respect to the goals of theory, and the constraints which theory must meet, theorist and reviewer alike are faced with a major problem of selectivity. Since the late 19th century psychologists have accumulated a great deal of information relevant to the recognition of words. In the experimental literature there are countless studies on recognition of isolated letters, strings of unrelated letters, structured nonword strings, isolated words and words in syntactic and/or semantic context. Subjects' tasks have included full report (naming), precued and postcued forced-choice recognition, precued and postcued yes-no recognition, lexical decision (word/nonword discrimination), search for a target in a list and apprehension of semantic content. Dependent measures have included accuracies, reaction times, duration and brightness thresholds. This research has produced a number of reasonably reliable empirical generalizations about the effects of such variables as word frequency, orthographic regularity of letter strings, pronounceability, statistical resemblance to English, and experimental "set." In the educational literature there exists an equally large array of studies on such issues as the effectiveness of phonics vs. whole-word teaching techniques (Chall, 1967), skill and strategy differences between good and poor readers (e.g., Sticht, Beck, Hauke, Kleiman and James, 1974), "speed reading" (see Berger,1970, for an annotated bibliography) and other topics of practical interest.

It is unrealistic to expect a single model of word recognition to account for more than a small fraction of the available information bearing directly and indirectly on that skill. At best, we can hope for a model which accounts in detail for some central core of the data, and which gives us a nonarbitrary basis for excluding other data, i.e., for invoking factors external to the theory which interact with factors specified in the theory to account for performance in situations other than those on which the theory is based. This hope, of course, requires the theorist to select in advance, on more or less intuitive grounds, the "core" of data which he will try to explain. The value of his theory will then depend as much on his choice of data as on the adequacy of his theory in explaining the data chosen.

Not surprisingly, the growth of our theoretical understanding has not kept pace with the accumulation of facts. Today, some ninety years after the first studies of tachistoscopic word recognition, we are still unable to provide a precise and complete account of the process by which the skilled reader converts the information in light reflected from the printed page to an internal representation of a word's identity or its meaning. To be sure, there have been many attempts to conceptualize the process, but only recently have there been detailed accounts susceptible to quantitative formulation and testing. The chief purpose of this paper is to review several recently proposed formal models of word recognition. Though none of these is without faults, and though none accounts for all aspects of word recognition, they exhibit a remarkable degree of convergence and collectively suggest that we are now close to a basic understanding of part of the process.

The remainder of this introduction is devoted to (a) a brief sketch of some important facts for which existing formal theories of word recognition attempt to account, and for which any complete theory must account, and (b) a brief discussion of informal and quasi-formal "theories" which have previously been advanced to account for the facts. The main body of the paper discusses in detail four formal proposals published in the last half-dozen years, each of which focuses on some distinctive aspect of word recognition, and each of which has been shown to generate accurate quantitative predictions in its chosen domain. The final section of the paper attempts to integrate the models, stressing their common points rather than their distinctness, as well as suggesting their collective limitations.

Some Facts about Word Recognition

Most of the data for which existing theories of word recognition attempt to account fall under the headings of "lexical" and "structural" effects, in the useful terminology of Manelis (1974). Lexical effects relate to the word as a unit; chief among these are the effects of semantic and syntactic context, and the effects of word frequency. It has been clearly established that report accuracies are higher and/or brightness or duration thresholds lower, for words which fit into some prior context known to the subject than for words which do not fit (e.g., Tulving and Gold, 1963; Tulving, Mandler and Baumal , 1964). Similarly, it has been shown that high-frequency words are more easily reported than low-frequency words (e.g., Howes and Solomon, 1951). Several of the theories to be reviewed, especially that of Morton (1969) give detailed accounts of the frequency effect. No contemporary theory could possibly give a full account of semantic/syntactic effects, for to do so would presuppose an adequate psycholinguistic theory of the way in which sentences are parsed and analyzed for meaning. Such a theory is not currently available, and its development lies far outside the scope of word recognition <u>per se</u>. It is not surprising, therefore, that existing theories treat sentential context as a kind of extraneous variable, though the theories do attempt to show how context-based <u>expectations</u> can affect perceptual recognition.

Structural effects, in Manelis' terminology, are those which relate to letter sequences within words. Letter strings which obey English structural rules are more easily perceived than those which do not. In particular, words are more perceptible than nonword strings of the same length, as has been known at least since the work of Cattell (1885) and Erdmann and Dodge (1898). The ability of subjects to report more letters from word than nonword stimuli has been dubbed the "word apprehension effect" by Neisser (1967). (Neisser's term, abbreviated WAE, will be used throughout this paper.) Nonword strings which resemble English have also been shown to produce higher tachistoscopic report accuracy; however it has not yet been possible to specify the dimension(s) of resemblance clearly. Gibson, Pick, Osser and Hammond (1962) showed that pronounceable nonwords (e.g., GLURCK) are more perceptible than unpronounceable nonwords formed from the same letters (e.g., CKURGL). However, other data (e.g., Gibson, Shurcliff and Yonas, 1970) suggest that orthographic regularity, the presence of English spelling patterns, rather than pronounceability _per se_ accounts for the effect. Miller, Bruner and Postman (1954) found that strings which approximate English in terms of transition-probabilities among letters also produce higher levels of report accuracy. Clearly, pronounceability, orthographic regularity and statistical Englishness are intercorrelated variables; which, if any is "the" crucial structural property for word recognition is not known for sure. At present, the weight of published opinion is with orthographic regularity (but for another opinion, see Travers, 1975.)

The lexical and structural effects so far mentioned can be explained by a simple theory variously termed "fragment theory" (Neisser, 1967), "sophisticated guessing" or "response bias." According to the theory, visual feature extraction depends only on visual variables such as brightness, contrast, exposure duration, etc. Available feature information does not differ between words and nonwords, wordlike and unwordlike nonwords, high and low frequency words, words consistent with context and words inconsistent with context. When a subject extracts too little information to identify a stimulus uniquely, he guesses. His guesses conform to his previous experience with the language--i.e., he guesses words rather than nonwords, wordlike rather than nonwordlike letter strings, etc. His guesses will coincide with actual stimuli more often when those stimuli are themselves words, wordlike nonwords, etc. Hence report "accuracy" will be higher for such strings.

Fragment theory accounts for virtually all the data available until the late 1960's. However, beginning with the work of Gerald Reicher (1969) a plethora of new studies appeared, challenging that straightforward explanation. Reicher presented subjects with common four-letter words (e.g., WORD), scrambled letter strings (e.g., ORWD) or single letters (e.g., D) for brief periods (around 50 milliseconds) and followed each display with a backward mask. Simultaneous with the mask, subjects were presented with a forced choice between two letters, one of which had appeared in an indicated position in the stimulus. In the case of word stimuli, both letters of the choice pair completed common English words. (Thus a subject might be shown WORD and then asked whether D or K had occurred in the last position.) This procedure minimizes effects of memory and the advantage of guesses based on knowledge of English words; nevertheless, letters within words were chosen correctly more often than letters within scrambled strings--and even better than letters presented alone. Thus Reicher's experiment seemed to show that every letter within a word is perceived more accurately than any one letter in isolation. This effect has been termed the "word-letter phenomenon," "word-superiority effect" or "Reicher-Wheeler effect" (after Reicher and Daniel Wheeler, who in 1970 followed Reicher's study with a complex experiment which ruled out many possible artifacts.) The WSE (an abbreviation for "word superiority effect" to be used throughout the present paper) provoked a new burst of theorizing which has not yet subsided.

Though variations in procedure can cause disappearance of reversal of the WSE (Bjork and Estes, 1973; Johnston and McClelland, 1973; Massaro, 1973; Mezrich, 1973; Thompson and Massaro, 1973; Estes, Bjork and Skaar, 1974), several successful replications have been reported (e.g., Smith, 1969; Smith and Haviland, 1972; Manelis, 1974; Spoehr and Smith, 1975.) Some of these have incorporated refinements and extensions of the Reicher-Wheeler data which both specify the phenomenon more precisely and constrain possible explanations. For example: Spoehr and Smith (1975) and

Baron and Thurston (1973) have shown that the "word" superiority effect obtains for wordlike, pronounceable nonwords as well as for words themselves, although Manelis (1974) has shown that the advantage for words is greater than for wordlike nonwords. Smith and Haviland (1972) have shown that sequential and distributional redundancy is not sufficient to produce the effect; even after hundreds of trials of training, subjects showed no perceptual advantage for letters embedded in redundant but unpronounceable strings. Several authors (Bjork and Estes, 1973; Estes, Bjork and Skaar, 1974; Massaro, 1973, Thompson and Massaro, 1973) have shown that the effect is reversed, i.e., single letters are reported more accurately than letters in context, when subjects are knowingly tested on the same pair of letters on repeated trials. Finally, some of the most revealing new data come from a modification of the Reicher-Wheeler procedure introduced by Estes (1974); the Estes data will be discussed in detail in connection with his model.

With the exception of Morton's (1969) logogen model, which predates work on the WSE, the contemporary theories of word recognition to be discussed below all offer explicit or implicit explanations for the outcome of the Reicher-Wheeler procedure and its variants. As will become obvious, explaining the WSE automatically explains the WAE and associated findings on structural effects in full report tasks. In addition, most of the theories to be discussed offer at least potential explanations for frequency and context effects, and of an additional group of effects which crosscuts Manelis' lexical/structural distinction, namely the effects of set and expectation. (For example, demonstrations by Aderman and Smith, 1971, that the WSE occurs only when the subject expects at least some stimuli to be words, and by Manelis, 1974, that the effect is enhanced in blocked designs, when the subject can reliably expect words on particular trials and nonwords on others.) It is perhaps fortunate that the appearance of the WSE, a new challenge to theory, coincided with a general movement in cognitive psychology toward complex and precise formal theorizing.

Models: Formal, Informal and Quasi-Formal

The earliest theories of word recognition were wholly informal, verbal and analogical. For example, one finds general claims--quite likely correct, as far as they go--that words are recognized "as wholes" or as "gestalts," rather than as strings of isolated letters. Such theories, obviously, do not lend themselves to precise formulation and testing, unless a large number of assumptions, often inessential and occasionally alien to the general conception at issue, are added. Psychologists have long recognized the pitfalls of such theorizing and have generally proposed theories of greater rigor and explicitness. The "fragment theory" discussed above is one such example. Fragment theory is not, in its general form, capable of generating quantitative predictions; however, it could readily be converted into a formal theory with some further specification. Theories of this type will be termed "quasi-formal" here. Of course, the models of central interest are those which have been given precise form and have been tested quantitatively against some body of data. The term "formal" will be reserved for models of this type.

The four formal models on which this paper focuses all grow out of the "information-processing" approach to cognitive psychology which has developed over the past decade or two. The models borrow freely from successful attempts at formal theory in other areas of information processing; for example, concepts from signal-detection theory and from mechanical pattern recognition and other areas of artificial intelligence are appropriated and modified as required. The paper attempts to show that this approach has brought us to the brink of a solution for a range of problems in the area of word recognition. However the paper should not be construed as arguing that only the information-processing approach lends itself to quantification and successful prediction. To cite some counterexamples: the old information theory lent itself to quantification and provided a useful tool for studying some aspects of word recognition. And fragment theory, rooted as it is in a general S-R approach, could easily be formalized as many other areas of learning theory have been. Conversely, many recent attempts to conceptualize word recognition in information-processing terms do not lend themselves to formalization; many information-processing "models" are really just conceptualization of component processes, without clear specification of how these processes operate or interrelate. Frequently, such theories are presented in "flow-diagram" form, but they are far from being programmable on a computer. This is not to deny the usefulness of such conceptual clarification; it is to deny their status as formal models. Examples, selected with no pejorative intent, include the conceptualization of Mackworth (1971), Gough (1972), and any of dozens of "models" discussed in the useful collection edited by Davis (1971).

A final point should be made concerning the four formal models before launching into a discussion of their details. All, with the possible exception of Morton's (1964) "logogen" model, presuppose that identification of individual letters is accomplished by means of a feature-analysis scheme (Neisser, 1967). That is, all assume that letters are represented internally by a list of properties, rather than by a physical analogue or template. All, again with the exception of the logogen model, deal primarily with the question of how letter-analysis is integrated into a larger word-analysis mechanism. Morton's model has different aims; therefore it is of special interest to show that it is compatible with the feature-analysis and letter-integration mechanisms proposed by the other models.

Four Models of Word Recognition

Morton's Logogen Model

Morton's logogen model is proposed in various publications (Morton, 1964a, 1964b, 1964c, 1968; Morton and Broadbent, 1967) but is explicated most completely in a paper in the Psychological Review five years ago (Morton, 1969). The general purpose of the model is to account for the interaction of various types of information which contribute to word recognition--visual or auditory information from the stimulus word itself, and "semantic" information from prior context. (Though Morton does not say so explicitly, it is clear that his model could incorporate syntactic context information as well.) More specifically, the model makes successful predictions about (1) the effects of word frequency on recognition accuracy; (2) the effects of limiting the number of alternatives in a recognition accuracy task; (3) the effects of repeated presentations of the same stimuli; and (4) the effects on recognition accuracy of predictability of the particular stimulus from prior semantic and/or syntactic context. Morton explicitly denies that the effects of context can be entirely explained by "response bias" or "guessing"; he holds that a genuinely perceptual interaction takes place, or, more precisely, that the perception vs. response distinction loses its meaning within the framework of his model.

Morton postulates a system of "logogens," one such entity for each word in the reader's vocabulary. The logogen is essentially a counting device. The count in a given logogen is increased when visual and/or auditory stimulus information, and/or semantic information from context, make the occurrence of a particular word likely. Morton does not specify the nature of the stimulus information, but it does no violence to his model to represent the information as a (visual or auditory) feature list. Thus, for example, one can readily imagine a set of extracted visual features which would simultaneously increase the logogen counts for "cat," "cut," and "cot." Similarly, one can imagine that the logogen count for "cat" is increased by prior context such as "the mouse was chased by the _____." A key feature of logogens is that their counts are increased regardless of the source of input information; thus, to pursue the above example, the logogen for "cat" will simply add the count increase due to visual feature input together with that due to context. When the count in a logogen exceeds some threshold, a response corresponding to an articulatory program for uttering the relevant word becomes available. Thus, in the above example, the combination of context and stimulus information would almost certainly make available the verbal response "cat." (Given appropriate context and/or stimulus information, several word responses might become available simultaneously.) Potential responses (articulatory programs) are stored in an "output buffer" from whence they may be executed as overt responses, or recirculated to the logogen system through covert rehearsal. (The exit to the output buffer from the logogen system is a single channel; hence, in the case where several responses become available, only one can be formed into an articulatory program, stored in the buffer and rehearsed or executed overtly. Morton states that "the first such response to become available will have precedence," but he does not pursue in detail an explanation of response completion within the framework of his model.)

Morton suggests that logogens behave in a manner analogous to the "detectors" of signal detection theory. (Green and Swets, 1966) The operation of the logogen is illustrated in Figure 1. In the absence of stimulus and context, logogens have some "normal" activity level arbitrarily designed as zero activity (Figure 1a). In ordinary reading and in certain word recognition tasks, the continuous interaction of context with the logogen system produces some additional excitation in each logogen, despite the absence of stimulus information. That excitation (the logogen count) has a probability distribution as illustrated in Figure 1b. The presence of relevant stimulus information, without context, also shifts the distribution upward (Figure 1c). The magnitude of the shift corresponds to d' in signal detection theory. The effects of context and stimulus information add to produce an even greater upward shift in logogen activity (Figure 1d). When the combined effects of context and/or stimulus exceed a fixed threshold (t in Figure 1), analogous to the "criterion" of signal detection theory, the relevant response becomes available.

Insert Figure 1 about here

Logogen counts are assumed to decay rapidly, within one second or so. However, it is also assumed that, once a response has become available, the threshold for that response is lowered to a new level γ , ($\gamma < t$). The threshold then returns slowly to a level $\underline{L}$, close to but less than the original threshold, ($\underline{\gamma} < \underline{L} < \underline{t}$; $\underline{t} - \underline{L}$ small) over a period of time which is very long relative to the period required for the count to decay. Thus, effects of stimulus repetition (except for very rapid repetitions) are marked by a threshold shift, and not maintenance of the logogen count. Similarly, word frequency, which in effect is equivalent to stimulus repetition in ordinary reading, exerts its effect by threshold shifts, with frequent words having lower thresholds than medium or low-frequency words (Figure 1).

Though Green and Birdsall (1958) have applied signal detection theory directly to auditory word recognition data, Morton opts for a somewhat different mathematization of his model. He assigns to logogens the properties of Luce's (1959) response strength model, which he calls a logarithmic transform of the signal detection model. In particular, he proposes that the probability of a response's becoming available is given by the ratio of the response strength for that item to the total of response strengths for all possible responses. Further, he proposes that increments in response strength due to stimulus and context may be multiplied, rather than added as shown in Figure 1.

In situations like typical tachistoscopic experiments, where stimulus information is present but context is absent, Morton arbitrarily assigns a value of unity to the average of response strengths for all logogens. The value for any particular logogen, $\underline{\gamma}$, fluctuates around this average, with $\underline{\gamma}$ presumably highest for logogens representing the target word and other words which share visual features with the target. For most applications

Morton finds it convenient to make a stronger simplifying assumption that the correct logogen has a response strength of $\underline{\alpha}$, while all other logogens have strengths of exactly unity. Thus the sum of response strengths for all logogens is $\alpha + (N - 1)$, where N = the total number of logogens (words) in the reader's vocabulary. Then, following the "ratio rule" above, P_s, the probability of a correct response based on the stimulus alone is given by

$$P_s = \frac{\alpha}{\alpha + N - 1} \tag{1}$$

Logogen counts (response strengths) also vary on the basis of context alone, independent of stimulus information. For example, subjects can often guess missing words accurately, given sentence contexts, suggesting that contexts can occasionally raise logogen counts above threshold, even in the absence of stimulus information. More generally, in ordinary reading and in certain tachistoscopic experiments, context "primes" the reader to "see" certain words and not others (Tulving and Gold, 1963; Tulving, Mandler and Baumal, 1964). The effects of context are represented in the model by the variable $\underline{V}_i$, which represents the response strength of each logogen based on context alone. The sum of response strengths for all logogens, $\underline{T}$, is given simply by $T = \sum_{i=1}^{N} V_i$; and the probability of selecting a correct response on the basis of context alone, P_c, is given by

$$P_c = \frac{V_i}{T} \tag{2}$$

When both stimulus and context information are present, the response strength (count) for each logogen is the product of strengths due to stimulus and context taken independently. Thus, for the correct logogen, designated by the subscript *i*, response strength is given by αV_i, and for all incorrect logogens, designated by the subscript *j* ($j \neq i$), response strengths are equal to $(1)(V_j)$. Note that the sum of response strengths for all incorrect logogens will equal $T - V_i$. Then the sum of response strengths for all logogens will equal

$$\alpha V_i + (T - V_i) = T + (\alpha - 1) V_i$$

The probability of a correct response based on both stimulus and context, P_{sc}, is then given by

$$P_{sc} = \frac{\alpha V_i}{T + (\alpha - 1) V_i} \qquad (3)$$

By simple algebraic manipulations of equations 1-3 above, Morton is able to demonstrate the predictive power of his model with respect to details of performance in several published experiments. For example, the model predicts that in a stimulus-only experiment, for a given signal-noise ratio, we should expect linear functions relating the log of a certain ratio based on performance data to the number of stimulus alternatives. In particular,

$$\log\left(\frac{P_n}{1 - P_n}\right) = \log\alpha - \log(N-1),$$

where P_n = probability of selecting the correct response *n* from

N alternatives, and α, as indicated above, is a fixed value. If the theory were false, i.e., if α were not fixed or if P_n were not the designated function of N, data relating $\log\left(\frac{P_n}{1 - P_n}\right)$ to $\log(N - 1)$ should depart appreciably from linearity. Using data from Miller, Heise and Lichten (1951), Morton shows that the function is indeed linear for a range of signal-to-noise ratios and for N's varying from 2 to 1000. To cite a second example, the model predicts that, in cases where stimulus and context interact,

$$\log\left(\frac{P_{sc}}{1 - P_{sc}}\right) = \log\left(\frac{P_s}{1 - P_s}\right) + \log\left(\frac{P_c}{1 - P_c}\right) + \log(N - 1)$$

The equation suggests that $\log\left(\frac{P_{sc}}{1 - P_{sc}}\right)$ should vary linearly with $\log\left(\frac{P_s}{1 - P_s}\right)$ for given context and fixed N, and that the resulting line should have a slope of unity. Morton shows this to be true of data from Tulving, Mandler and Baumal (1964), in which recognition accuracies with 0, 2, 4 and 8 words of prior context were assessed. Other examples falling into the four classes of data mentioned in the introduction to this section could be cited, but such citation should be unnecessary to demonstrate the predictive power of the logogen model.

Morton's model is designed to describe the interaction of stimulus and context information. However, it should be clear that the model sidesteps the thorny issue of how context exerts its effects on the logogens. As indicated in the introduction, to achieve such an explanation would require a detailed theory of syntactic and semantic processing of natural language. In the absence of such a theory, Morton adopts a pragmatic course: he uses the predictability of a word in a given context as an index of the degree to which prior syntactic and semantic analyses activate particular logogens. Morton's approach is necessarily limited by the present state of psycholinguistic knowledge. However, his is the only existing formal model of word recognition which attempts to take account of context at all. Smith and Spoehr (1974) point out that other writers on the subject of context effects, particularly those who focus on tasks which approximate normal reading (e.g., Levin and Kaplan, 1970), postulate analytic units larger than the single word. Smith and Spoehr suggest that Morton's model may be incompatible with units larger than words. However, an alternative view is that Morton's model describes effects on perceptual processing of single words due to postperceptual processing of larger context units. While Morton's model offers no account of how larger units are processed, it does not preclude and perhaps presupposes such processing.

As noted earlier, the logogen model focuses on what Manelis (1974) calls "lexical," rather than "structural" effects; that is, the model incorporates variables bearing on the word as a whole (e.g., frequency, predictability from context) rather than on letter sequences within the word (e.g., transition probabilities, orthographic regularity). In contrast, the rest of the models considered in this report focus on structural effects and on data from experiments in which words or nonword letter strings are displayed without context. It is thus serendipitous, or perhaps revealing of some deep regularity, that the formal structure of the logogen model resembles the structures of several of the structure-oriented models, in particular those of F. Smith (1970), Rumelhart and Siple (1974), and Estes (1974, 1975). In all cases a fixed detection device analogous to the logogen is postulated (in contrast, for example, to possible models which might propose that words are somehow "synthesized" anew with each new presentation.) In all cases, feature information extracted from the current stimulus is combined with prior information reflecting the likelihood of a particular word or letter sequence (e.g., information about the frequency of a word or letter sequence.) Finally, in all cases the combinatorial rule is multiplicative, as has already been shown for the logogen model. Though these resemblances are relatively superficial and do not in themselves point to underlying agreement among the models, they do raise that tantalizing possibility, which is explored in the final section on integration of the models.

The Smith-Spoehr Model

Smith and Spoehr (1974; see also Spoehr and Smith, 1973) propose a two-stage model of word perception, incorporating both a stage of visual feature extraction and a stage of interpretation, in which the extracted information is assigned to some stored category (e.g., letters, syllables, or words). Their model is presented in the context of a lengthy review of other theories of word perception. In the course of that review they reject theories which explain the WAE and WSE in terms of differences at the first, or feature extraction stage. Further, they reject theories like those of F. Smith (1971) and Rumelhart and Siple (1972), which operate at the interpretation stage but which are based on feature redundancy within words and/or letter-clusters. Feature-redundancy models, particularly in the form proposed by Estes (1974, 1975) will be defended later in this report; the Smith-Spoehr arguments against such models will be considered at that point. Here, however, the report focuses on the Smith-Spoehr proposals concerning what they call the "translation" process.

Smith and Spoehr subdivide their interpretation stage into three component processes: (1) matching, in which extracted feature information is compared to stored lists of features demarcating letter categories; (2) decision, in which the best letter match is selected, and (3) translation, in which the visual category is translated into an acoustic or phonological equivalent. The authors choose to call all of these processes "perceptual"; even "translation" is seen as an intrinsic part of visual perception, and not a postperceptual recoding or mnemonic process. As will be seen later, this somewhat counterintuitive usage appears to be required in order for their theory to account for certain results on the role of syllables in word perception (Spoehr and Smith. 1973) as well as for the WAE, the WSE and related effects.

Smith and Spoehr assume that the reader first goes through the decision and matching processes to determine the identities of separate letters within a word. This categorical information is then preserved in what the authors call a "sensory" store (although more common usage presumes that information in sensory storage is precategorical--a fact which will be stressed in the critique of the model below). While the letter categories are preserved in the sensory store, a parsing process is applied, subdividing the string of letters into syllable-like units called "vocalic center groups" or VCGs (after Hansen and Rodgers, 1965). It is the explicit nature of this parsing process which qualifies the Smith-Spoehr model as "formal" in the sense defined earlier. The parsing rules are shown in Table 1.

After the parsing rules have been applied, the reader maps each VCG into an acoustic (or perhaps articulatory) unit corresponding to a syllable in oral speech. Two facts are important here: (1) The acoustic products of translation are not individual letter names, unless the reader has been presented with a highly unwordlike string which cannot be parsed into VCGs; (2) translation does not take place on a single-letter-to-single-phoneme basis. (Indeed, this cannot occur, since the phonemic value of a letter is determined by context. VCGs are intended to represent the minimum units of printed text for which sound values can be specified.)

In the cited paper (1974) and elsewhere, Smith and Spoehr have marshalled several lines of evidence in support of their model: (1) Spoehr and Smith (1975) show that nonwords which form VCGs (e.g., BLOST) are more easily perceived than comparable strings which lack the crucial vowel on which the parsing rules depend (e.g., BLST); moreover the difficulty of perceiving non-VCG strings can be predicted by the number of transformations required to convert these strings into VCGs, using explicit rules proposed by MacKay (1972). (2) Spoehr and Smith (1973) show that words containing multiple VCGs are harder to perceive than words containing single VCGs. They also show that perceptual accuracy scores for successive letters are more highly correlated when both letters are part of a VCG than when they are drawn from both sides of a VCG boundary. (3) Spoehr (1973) has obtained reasonably close matches to data from a range of experiments, using a computer simulation model which incorporates the VCG parsing rules shown in Table 3, as well as a number of other assumptions widely shared among modelers of the word recognition process. (4) Finally, the model accounts for most existing findings on perception of words and structured nonwords, in terms of the number of VCGs, or number of transformations required to create VCGs, that the various stimuli entail.

In sum, the Smith-Spoehr model accounts for a range of existing findings and has shown considerable heuristic value in generating new and interesting data. Yet the model embodies logical difficulties which have forced the authors to assumptions which, at the very least, violate common usage of terms: the model assumes that processing of words, wordlike nonwords and random letter strings does not differ up to the point of letter identification; it is only at the stage of parsing and acoustic coding that differences between structured and unstructured strings emerge. If the authors were interested in explaining the results of full report tasks, which necessarily incorporate coding and memory effects, the relatively late appearance of word-nonword differences in their processing model would be reasonable. But the authors in fact wish to explain genuinely perceptual effects, such as those assumed to be observed in immediate two-choice recognition experiments, where memory and response factors are minimized. The authors are therefore forced to assume that "translation" is a perceptual process, which intervenes even before choice in such tasks. Further, since letter identification is assumed to be equal for structured and unstructured strings, they must assume that information available after letter identification is unavailable at the point of choice even though that point may follow immediately after stimulus display (cf. Reicher, 1969; Wheeler, 1970). This necessary assumption is incorporated in the proposal that (categorical) letter identities are maintained in a "sensory" store. This store in turn is assigned the properties usually attributed to iconic memory (Neisser, 1967) or the visual information store (Sperling, 1963) --i.e., rapid exponential decay, presumably leaving the reader with little or no useful information after a few hundred milliseconds--thus forcing him, in a choice task, to rely on "translated" information, which shows the effects of string structure. However, the sensory store, as usually interpreted (e.g., by Neisser and Sperling) contains only <u>pre</u>categorical information--letter features rather than letter identities. Of course, Smith and Spoehr are free to redefine the sensory store; however, since the "precategorical" conception of the sensory store has proved useful in structuring so much of the available information on tachistoscopic recognition (cf. Neisser, 1967) it ought not be lightly abandoned.

Smith and Spoehr are not free to use the conventional "precategorical" conception of the sensory store, however. Their parsing algorithm requires that letter-identities be available before letter strings can be appropriately segmented. (More precisely, their model requires that letters be tagged as vowels or consonants before segmentation takes place. They consider the possibility that letters can be so tagged on the basis of crude feature information, prior to letter identification; however they reject this alternative proposal because the only relevant data on the subject (Posner, 1970) seems to show that letters cannot be tagged as vowels or consonants until their identities are known.) Thus the authors are forced to assume, on the one hand, that letter identities are available very early in perceptual processing but, on the other, that identities are not directly available at the response stage, even when responses are cued immediately after stimulus presentation.

In a later section of this report, an attempt will be made to show that the important insights of the Smith-Spoehr model can be preserved, within the framework of a model that also preserves more usual conceptions of sensory storage and letter identification. That model builds on the proposals of Rumelhart and Siple (1973) and Estes (1974, 1975), which are examined below.

Feature-Redundancy Models

The models of F. Smith (1971), Rumelhart and Siple (1974) and Estes (1974, 1975) are similar in certain essential respects. In particular, all are "feature redundancy models" in the terminology of Smith and Spoehr (1974). That is, all three models assume

(a) that individual letters are recognized by a feature-extraction scheme of the general type described earlier; (b) that <u>groups</u> of letters, perhaps whole words, can also be stored as feature lists against which input information from novel stimuli can be matched directly; (c) that, due to distributional and sequential redundancies in printed English, a word, spelling pattern, syllable, VCG or whatever common letter cluster, can be uniquely matched using a smaller list of features than a random letter string of the same length. (These assumptions will gain clarity as particular models are explained.)

F. Smith's model may be taken as a simple prototype of this class. Smith proposes that readers develop a set of discriminating features for whole words, just as they develop a set of features for letters in the early stages of learning to read. Thus words become perceptual units, perhaps akin to the ideograms of Chinese and Japanese. Because the permissible sequences of letters in English are constrained by orthographic or phonological rules, the feature list for a word can in theory be shorter than would be required if letters within words were identified separately. Smith proposes that this featural redundancy accounts for the efficient perceptual processing of words. Smith's model will not be discussed further since it is unformalized and untested against detailed data; therefore it does not meet the criteria outlined earlier. Its essential ideas are given precise form in the work of Estes and of Rumelhart and Siple. However one calculation of Smith's is worth keeping in mind, since it conveys the general power of feature redundancy models: If English were nonredundant, there would be about 12 million five-letter words (26 letters in each of five positions; $26^5 = 11,881,376$). In fact, there are perhaps 20,000 five-letter words. If the perceptual feature tests of English were (a) binary and (b) maximally efficient, we would need 4-5 such tests to discriminate 26 letters ($2^4 = 16$; $2^5 = 32$). We would need about 23-24 feature tests to discriminate among the 12 million five-letter alternatives in a nonredundant language ($2^{24} = 16,777,216$) but only 14-15 tests to discriminate among the 20,000 alternatives that actually exist ($2^{14} = 16,384$). Thus the redundancies of English would allow us to discriminate five-letter English words with 58% as many binary feature tests as are needed to discriminate random letter strings of the same length. Clearly, this gross calculation tells us nothing about the perceptual mechanisms involved--the models outlined in the following section are designed to accomplish that--but it does give us some idea of the potential saving in processing capacity when readers deal with printed stimuli which obey known structural rules.

The Rumelhart-Siple Model

Like F. Smith, Rumelhart and Siple (1973) assume that a letter, syllable or word can be represented in long-term memory as a list of features, with less features necessary to uniquely define a syllable or word than a random letter string of the same length. Their model for word perception is a special case of a more general "multi-component" model for tachistoscopic perception proposed by Rumelhart (1970).

A "component" is a line segment in a display; components have fixed length, orientation and retinal location. Functional "features" are line segments, composed of one or more components, such that the presence of a single component guarentees the presence of the entire feature segment, given the set of alternatives from which the display is drawn. (Thus, for example, in some simple uppercase typefaces, the presence of any segment of a medial upright line guarantees the presence of at least a medial line extending from the bottom to the midpoint of the letter, and, by extension, the presence of an I, T, Y and perhaps J, depending on the exact shape of J in the particular typeface.) Thus components are defined purely in terms of stimulus geometry, while features depend on both geometry and on the particular set of stimuli to be recognized. The probability of extracting a particular feature, f_i, present in a stimulus display depends on (a) the number of components in the feature, i.e., its length; (b) the signal-to-noise ratio in the display; (c) the duration of stimulus exposure; and (d) the duration of retention of the stimulus in iconic memory. For fixed experimental conditions,

$$P(f_i \mid f_i \in S) = 1 - \alpha^{l_i} \qquad (4)$$

where the term on the left should be read "the probability of extracting feature f_i, given its presence in the stimulus, S"; l_i is the length of the feature segment, and α is a parameter embodying values of b - d above, which are fixed for a particular experiment.

The probability that a reader will report a particular stimulus (letter, syllable or word) depends on the set of features which he extracts, together with his *a priori* expectations about what will be presented. When the subject (correctly) expects an

English syllable or word, he can identify the stimulus with less feature information than he would require for a random string of letters, because his expectations conform to actual stimulus properties. Formally, Rumelhart and Siple define a "candidate set," C(F), or list of possible stimuli (letters, words or syllables, consistent with F, the extracted set of features. A particular response, r_i, is a member of C(F) if

(a) the set of extracted features, F, is consistent with s_i, the stimulus corresponding to r_i, and

(b) the total number of features in s_i does not exceed the number of extracted features by too wide a margin (i.e., by more than some arbitrary criterion.)

Given these constraints, a particular response, r_i, is selected by the following rule:

(1) If the candidate set is empty ($C(F) = \emptyset$), response r_i is selected according to the a priori probability of s_i's occurrence ($P(r_i) = P(s_i)$).

(2) If the candidate set is not empty, but r_i is not in the set, r_i is not selected ($P(r_i) = 0$).

(3) If the candidate set is not empty, and r_i is in the set (along with other potential responses), r_i is selected according to a Bayesian decision rule:

$$P(r_i) = \frac{P(F \mid s_i)\, P(s_i)}{\sum_j P(F \mid s_j)\, P(s_j)} \qquad (5)$$

where: $P(r_i)$ = probability of selecting response r_i

$P(s_i)$ = *a priori* expectation that stimulus s_i will be presented

$P(s_j)$ = *a priori* expectations for each of *j* alternative stimuli in the candidate set, including s_i

$P(F | s_i)$ = probability of extracting feature set F, given stimulus s_i

$P(F | s_j)$ = probabilities of extracting feature set F, given each of the *j* alternatives in the candidate set.

Finally, Rumelhart and Siple assume that subjects determine their *a priori* expectations of particular stimuli, $P(s_j)$, on the basis of (a) their expectations that stimuli will be letters, syllables or words and (b) their implicit knowledge of objective frequencies of particular stimuli within each of the three stimulus categories. Formally,

$$P(s_i) = f_w(s_i)P(\text{WORD}) + f_s(s_i)P(\text{SYL}) + f_l(s_i)P(\text{LETTER}) \qquad (6)$$

where

P(WORD), P(SYL), P(LETTER) = subject's *a priori* expectation that stimulus will be a word, syllable or letter. (These probabilities are assumed to be "sets," constant for a particular experiment.)

and

$f_w(s_i)$, $f_s(s_i)$, $f_l(s_i)$ are the subjective probabilities of particular stimuli, given the general categories word, syllable and letter. These subjective probabilities are assumed to mirror objective frequencies in the reader's experience with the language.

Rumelhart and Siple test their model against data from a tachistoscopic recognition experiment in which five subjects reported 726 three-letter strings, 510 words and 216 nonwords, with syllables falling in both categories. The stimuli were all the three-letter strings tabulated in the Kučera-Francis (1967) count of one million words of printed English. Letters were presented in a simplified type font which allowed convenient analysis into features as defined above. Signal-noise ratios were fixed at a level that allowed 50% correct recognition of single letters. Parameters required for application of the model were estimated by a variety of procedures too complex to report here, and the recognition data were then simulated by computer with generally excellent fits. In particular, the model successfully predicts the following aspects of the human data:

(a) The frequency class of error responses, which tends to fall in the same frequency class as actual stimuli.

(b) The distribution of correct responses across classes of frequency, letter predictability and letter confusability.

(c) The fact that words with intermediate frequency of occurrence are reported most accurately when those words contained improbable letter sequences. As Smith and Spoehr point out, the model can also predict the WSE, since, in general, the number of feature tests per letter position is smaller for letters within words than for isolated letters.

In short, the Rumelhart-Siple model is exemplary in its explicitness and has predicted human data with high fidelity, including in its predictive scope several nonobvious effects. Smith and Spoehr, however, raise several objections to the model: First, they point out that some of the predicted effects are obtained only with full-report experimental procedures, and not with forced-choice paradigms. This fact suggests that the effects are postperceptual, and Smith and Spoehr point out that the model's powerful decision procedure may in fact capture response processes rather than perceptual processes. Second, Smith and Spoehr question whether human memory could plausibly contain feature lists for all the words that can be recognized on sight, particularly when our ability to deal with variations in type font is taken into account. Finally, Smith and Spoehr contend that the model cannot predict differences in perceptibility among different kinds of nonwords. In their words "... either a feature set of the input exists in memory, making perceptual performance quite good, or no such set exists and performance is quite poor. In other words, if feature redundancy is incorporated only at the level of a word, then there is no room in the model for gradations between nonwords."

All three objections of Smith and Spoehr have some force, though it is possible to muster counterobjections. For example, gradations among nonwords are to some degree handled by the fact that the model incorporates detection devices at a syllable level, as well as letter and word level. In fact, if "syllables" were redefined as VCGs, many predictions of the Rumelhart-Siple model would coincide with those of the Smith-Spoehr model considered earlier. This point is considered in more detail, along with the "response-bias" and "memory-load" objections in the final section of this report. First, however a model proposed by Estes (1974, 1975) is considered. The Estes model shares some of the basic architecture of the Rumelhart-Siple model, but it incorporates several refinements which, in the view of the present writer, effectively counter the major objections to all of the other models discussed.

The Estes Model

Estes (1974, 1975) has proposed a model similar in design to a quasi-formal model proposed independently by Travers (1970, 1973), and also bearing a noticeable resemblance to Selfridge's "Pandemonium," a computer program which recognizes letters and nonsense syllables (Selfridge, 1959;Selfridge and Neisser, 1960). The Estes model postulates a hierarchy of "control elements," which may be conceived as memory structures, perceptual filters, detectors or "demons" in the "Pandemonium" sense. "Control elements" are devices which signal the presence of particular configurations in the stimulus, i.e., particular letter features, letters, letter clusters or words. The control element for each stimulus configuration may be activated by two kinds of input: (1) stimulus information gleaned from "lower" detectors, which may include other control elements, and (2) expectations based on prior context, experimental set, etc. The two types of information are combined multiplicatively to yield an output which corresponds to each control element's "estimate" of the probability that its target configuration (feature, letter, cluster or word) is present in the external stimulus. This "estimate" may then be used as input to a "higher" detector.

The hierarchical organization of control elements represents the reader's enduring knowledge of orthography and morphology, i.e., of the features comprising letters, the letters comprising orthographically regular clusters (e.g., syllables, spelling, patterns, VCG's), the clusters comprising words. Stimulus information is filtered upward from the retina through the feature, letter, letter-group and word control elements, until a level is reached at which no match to current information is found (i.e., all control elements at that level signal very low probabilities that their target configurations are present in the stimulus.) Responses are based on the highest level of matching achieved. Thus, for example, a random letter sequence would be lik ly to generate several matches at the level of letter control elements, but no match at the level of orthographically regular clusters. Responses would then be based on individual letter identities, in typical situations taking the form of covert or overt naming of the recognized letters. If the stimulus were a nonsense syllable instead of a random letter sequence, it would be likely to excite one or more control elements at the letter cluster level. Responses could then be based on phonetic recodings of whole clusters, rather than individual letters.

Figure 2 illustrates the design of the system. The figure is a hybrid based on both Estes (1974, 1975) and Travers (1970, 1973). By tracing the path of stimulus information from the word "WORD" upward through the control element system, we can illustrate the general operation of the model more clearly than has been done so far, as well as introducing certain details which add to the model's predictive power.

Insert Figure 2 about here

Light reflected from the printed page or tachistoscope screen casts a pattern on the retina, exciting receptor cells by differential amounts, depending on whether dark areas (the stimulus configuration) or light areas (the background) happen to hit particular receptor cells. This pattern of excitation is then effectively converted to a list of features (lines, angles, curvilinearity vs. angularity, elongation, etc.) Additional information, not precisely accurate, about the location of features in the external display is also extracted. (As will be seen, inaccuracy of position information plays a key role in explaining various experimental effects.) This characterization of feature control elements is obviously reminiscent of the single-cell analyzers described by Hubel and Wiesel (1963). However, in the absence of evidence on the anatomical or physiological basis for detection of complex letter features in humans, Estes avoids speculation about the neural mechanisms underlying his feature control elements. Our ignorance regarding the process of translation from the retinal to the feature list representation of stimulus information is indicated by dotted line segments in Figure 2.

For expository purposes, the feature extraction and letter detection processes are exemplified in a somewhat unrealistic manner in Figure 2. It is assumed that sufficient information is drawn from the W, the O and the R to distinguish each letter uniquely from all other letters of the alphabet. For example, the two oblique line segments and one angle (circled on the "retinal" represenation of the W) excite their associated feature control elements. In the particular typeface for which the system is "set," these three features uniquely specify the letter W -- a fact represented by the transmission of the output of the three feature detectors to W and to no other letter. In the case of the D, however, it is assumed that the extracted features (circled on the "retinal" representation) are sufficient to limit possible letter candidates to D and B, but not to distinguish the two candidates from each other.

Information from the letter control elements is now passed upward to the cluster control elements. Again, to simplify the exposition, only the final consonant cluster is considered. The two candidates for the final cluster are RD and RB. (The system "knows" that the cluster is in final position because it has detected the blank space to the right of the final letter. The blank is symbolized by # in the Figure.) Since both RD and RB are permissible final consonant clusters in English, the output of the cluster control elements is passed to the word level, where only one sequence consistent with all available input information exists--the word "WORD," which is therefore correctly recognized.

How does the model explain the WAE and the WSE? The WAE can be explained in a straightforward manner, by reference to the example just given: Knowledge of word structure allows the subject to eliminate incorrect responses which are consistent with extracted visual features but which are not consistent with English words. Thus WORD is selected over WORB in the example, though both sequences are consistent with available feature information. Clearly, this explanation is a variant of "fragment theory" or "sophisticated guessing," although it represents the guessing process as more intimately connected with visual processing than other variants of the same explanation. The subject does not mull over available features and consciously select a word consistent with those features; there is no clear separation of visual and verbal processes as Neisser (1967) suggests. Rather, extracted feature information makes contact with memory structures which carry some "verbal" information (e.g., about permissible letter sequences in the language). Visual and verbal processes are in a sense continuous; their interaction is rapid and presumably unconscious. (All of the latter interpretations go beyond Estes' explicit statements, but they seem consistent with the tone of his comments and the time parameters of the experiments cited in support of the model.)

Explanation of the WSE is less straightforward. It could be explained within the model by an extension of the sophisticated guessing argument: If we add the assumption that feature information decays rapidly from iconic memory, the subject must then retain letters either as abstracted visual codes or as acoustic representations of letter names. In either case, the subject does not have available at the moment of choice in a Reicher-Wheeler procedure all of the feature information from which he derived the letter code. Consider the example once again: The level of feature information extracted from the final letter position is insufficient to distinguish D from B. If the subject has erroneously coded the input as B, he cannot recall the particular features which led him to that code. Now faced with a choice between D and K, he must guess which is more likely to have provoked his perception of B, and this guess entails some probability of error. If, however, D is presented in a word context, the subject will perceive "WORD" rather than "WORB" for reasons already outlined. Because he is less likely to code the final letter erroneously in the word context, he is more likely to have a correct letter identity available at the moment of choice.

Despite the fact that this explanation is consistent with his model, Estes rejects it on the basis of data from a novel experimental procedure which he introduced (Estes, 1974). In this procedure, single letters, words, or four-letter nonwords are displayed briefly and followed by a mask, as in the Reicher-Wheeler procedure. However, instead of presenting the subject with a forced choice, Estes simply indicates the position of the letter to be reported with an arrow. Like the Reicher-Wheeler method, the Estes procedure yields higher report accuracy for letters in word contexts than for letters in isolation or in nonword contexts. However, unlike forced-choice, the procedure permits revealing error analyses. In particular, the explanation advanced above leads to the prediction that erroneously identified letters should be those that, together with context letters, form English words. Thus, K should be a frequently chosen erroneous response when WORD is shown and the last position is selected for report. To test this prediction, Estes included in his study a large number of trials for which the same two letters, R and L, were the only alternatives which completed words in context with other letters. The prediction did not hold: the key letters were erroneously selected only at chance levels. Subjects did not appear to be using context to restrict response alternatives.

How, then, could the WSE be explained in terms of the model? The data revealed two important features of erroneous responses: (1) The advantage for letters in word context over letters in isolation was due largely to errors of omission for isolated letters; (2) the advantage for letters in word context over letters in nonword contexts was due largely to errors of transposition for the nonword stimuli. These facts led Estes to conjecture that inaccuracies of position information at the feature or letter levels cause the WSE:

On single-letter trials feature information may be too degraded to allow a correct identification not just because the subject fails to extract enough features from the target location, but also because he may attempt to extract features from a nontarget location (i.e., from a flanking mask character in Estes' study, or from a blank space in the Reicher-Wheeler procedure.) In either case, poor feature information forces the subject to guess (as he must in the Reicher-Wheeler procedure) or to omit any response (as he often does in the Estes procedure.) On nonword trials, when input from the target is insufficient to allow clear identification and/or localization, input from adjacent letters may lead to correct identification of those letters. Because of positional uncertainty, the adjacent letters may be reported in place of the target, producing transposition errors in the Estes procedure and guesses in the Reicher-Wheeler procedure. Finally, on word trials where target input is degraded, information from adjacent letters together with knowledge of word structure may allow the subject to base his response solely on feature information from the target location, thus increasing the likelihood that he will generate a correct response.

Estes' explanation for the advantage of words over nonwords, and for the prevalence of transposition errors in response to nonwords, seems eminently reasonable. However, his explanation for the superiority of words over single letters is not so clearly plausible and requires closer examination. The latter explanation clearly presupposes that <u>some</u> information from the target location be available--else why would WORD be generated more often than WORK? Yet this information cannot be sufficient to allow the subject to select D over K in isolation with the same probability. Estes' explanation for the word-single letter discrepancy is that the subject cannot localize feature information as accurately when it comes from an isolated letter as he can when it comes from a letter in a word context. But why does the subject need to localize feature information for isolated letters? Why does he not use whatever feature information is available, even if it seems to come from the wrong location? The answer may lie in the fact that, in Estes' procedure, single letters are always flanked by masking characters which are roughly letter-like (number symbols--#--or dollar signs--$). Thus the subject will expect feature input at all locations and might attempt to assign a letter interpretation to feature information actually drawn from a mask character. However, it is not so plausible that location errors explain the word-letter discrepancy in the case of the Reicher-Wheeler procedure, where flanking masks were not used for single-letter stimuli, nor in the case of blocked designs in which the subject knows on every trial whether a single letter or word will be presented. This issue will be discussed in more detail below; first, however, more formal aspects of the model will be treated.

Estes has developed mathematical applications of his model to two types of experiment--two-choice detection procedures and the probe procedure described above. The latter application will be described here, because the probe procedure reveals more about the process of word recognition and the workings of the model:

The formalization requires four parameters: (1) α, the probability that any letter in the array is correctly identified, i.e., that its control element is activated; α is assumed to depend only on visual conditions and to be the same for single letters (SL), nonword (NW) and word (W) displays. (2) γ_i, the probability that the subject locates any particular letter correctly, whether or not he identifies it; γ is assumed to vary across SL, NW and W displays--hence the subscript, which denotes the three different values that γ can take. (3) C, the probability that the subject will guess the letter in the position which he perceives, correctly or incorrectly, as the probed position. (4) g, the probability that such a guess will be correct; Estes assumes that such guesses are random, i.e., that $g = .04$, for his probe study.

According to the model, a correct response occurs if (a) the target is identified and correctly localized, or (b) the target is correctly localized and a correct guess occurs as to its identify; (c) the target is neither localized nor identified, but a correct guess occurs. That is,

$$P(C) = \gamma_i \alpha + \gamma_i \left[(1-\alpha)C(.04)\right] + (1-\gamma_i)(1-\alpha)C(.04) \qquad (7)$$

An intrusion error will occur if the subject guesses (with or without correct localization of the target), and his guess is a letter not in the display. With four-letter displays, the probability of such a choice is 22/26. That is,

$$P(IE)_4 = (1-\alpha)C(.85) \qquad (8)$$

A transposition error occurs if (a) there is incorrect localization and the letter in the wrong location is correctly identified, or if (b) an incorrect guess occurs and the letter chosen is one of the nontarget letters in the display, the probability of such a choice being 3/26 for four-letter displays. That is,

$$P(TE)_4 = \alpha(1-\gamma_i) + (1-\alpha)C(.12) \qquad (9)$$

An omission error occurs if the target is not identified and the subject declines to guess:

$$P(O)_4 = (1-\alpha)(1-C) \qquad (10)$$

Equation (7) specifies the probability of a correct guess for both SL and four-letter displays. However, equations 8-10 predict errors for four-letter displays only. For SL displays there can be no transposition errors; all errors are intrusions or omissions. If the subject fails to identify the target and guesses incorrectly, he produces an intrusion with probability given by

$$P(IE)_1 = (1-\alpha)C(.96) \qquad (11)$$

(.96, of course, is the probability of an incorrect random guess, or 25/26.)

In order to make the probabilities of correct and incorrect sequences sum to unity, it is necessary to add the term $\alpha(1-\delta_i)$ from equation (9) to equation (10), yielding, for omission errors:

$$P(O)_1 = \alpha(1-\delta_i) + (1-\alpha)(1-C) \qquad (12)$$

The first term of equation (12) has a natural psychological interpretation when it appears in equation (9); there it represents the probability of correctly identifying a letter from a nontarget position in a multiletter array, and of reporting that letter in place of the target, producing a transposition error. However, the interpretation of the term for single-letter arrays is a little less straightforward. It appears to represent the probability of misplacing the probe, and of correctly identifying the contents of the apparently probed position--i.e., of perceiving a mask character as the probed item, and therefore of omitting any response.

To apply equations 7-12 to the data from his probe experiment, Estes first estimates $\underline{\alpha}$, $\underline{C}$ and $\underline{\gamma}_i$ from the observed proportions of omissions and correct responses for the W and NW condition. He then uses the obtained parameter values to predict the proportions of intrusions and transpositions for the same stimuli, and of omissions for the SL stimuli. The obtained predictions fit the data well, as shown in Table 2. (Estes goes on to demonstrate equally good fits to additional data from other variants on the probe experiment; these variants will not be described in detail here, though one is mentioned below.)

Whenever a mathematical model with several parameters is fit to a set of data, a question can be raised as to whether the behavior of those parameters is sufficiently constrained by the data to reveal anything interesting about underlying processes. It is often possible to fit data equally well with several alternative choices of parameter values, making it difficult to draw any conclusions about the entities said to be described by those parameters. Thus it is interesting and important that Estes has explored alternative ways of assigning parameter values and has found that all alternatives produce qualitative errors of prediction, as well as quantitative deviations from the rather accurate predictions summarized in Table 2. In particular, he has explored the behavior of the model when $\underline{\gamma}$, the location parameter is fixed instead of variable across the SL, NW and W conditions and one of the other parameters--$\underline{\alpha}$ or *C*-- is allowed to vary. Varying $\underline{\alpha}$, *C*, or both, produces errors of prediction, suggesting that $\underline{\gamma}$ is indeed the parameter affected by linguistic context. The psychological implication is that differences in report accuracy across the SL, NW and W conditions is not due to differences in probability of identifying individual letters (α) or of guessing (C) but to differences in the accuracy with which letters are localized (γ_i).

It is clear from the foregoing discussion that only a limited portion of Estes model has thus far been formalized. No attempt has been made to show formally how the hierarchical structure of the control elements comes into play, nor how expectancies interact with feature information. In the concluding section of this paper, some speculations will be offered about how the formalization could be extended to account for the effects of such variables as set and word frequency. However, even the limited formalization has proved to have some predictive power, and it supports one of Estes' key substantive contentions, namely that differential localization accuracy accounts for perceptual differences among SL, NW and W stimuli.

That substantive contention creates a dilemma which deserves some discussion, however. A recognition advantage is obtained for words over single letters even when the latter are presented in isolation, without flanking mask characters (Reicher, 1969; Wheeler, 1970). This occurs despite the fact that lateral interference almost certainly inhibits recognition of letters within words. (Presumably, a desire to control lateral masking effects motivated Estes' use of flanking mask characters for his SL displays.) Although localization of input from isolated letters may be inaccurate, it seems unlikely that such inaccuracy should lead to omission errors. When only a few visual features are available from anywhere in the display, one would expect subjects to base their responses on those features even if they seem to come from the wrong location; one would not expect subjects to ignore such feature information and attempt to assign letter interpretations to the contents of nontarget locations when those locations are blank. Thus one horn of our dilemma lies in the fact that localization errors do not seem to provide a satisfactory explanation for the WSE under all experimental conditions.

We can avoid being impaled on this horn by appealing to the alternative explanation of the WSE proposed earlier--that word context restricts the set of letter guesses which a subject will generate on the basis of partial feature information. But then we are prodded by the other horn: the choice-restriction, or feature-redundancy, explanation implies that intrusion errors for word stimuli in the probe experiment should, more often than chance, be letters which complete words when taken in the context of the remaining letters of the display. Unfortunately, as Estes shows in his analysis of trials for which only the letters L and R complete English words, subjects do not substitute the letters L and R for one another to any great extent. (For word stimuli, only 2% of all responses, and 6% of all errors, were L-R intrusions. Moreover, L-R intrusions were almost as frequent for single-letter displays as for word displays, and they were more frequent for nonword displays--1% and 4% of all responses, respectively.)

We can extricate ourselves from this dilemma by the simple expedient of arguing that both the choice restriction and localization mechanisms operate, but that the Estes and Reicher-Wheeler procedures create differential probabilities of their operation and/or observation. We have already made the case that localization errors should be rare for the SL conditions of the Reicher-Wheeler procedure and that they should be more common in the Estes procedure. It remains for us to make the case that Estes' procedure inhibits use of the choice-restriction strategy and/or makes it difficult to observe the effects of the strategy.

Unfortunately, I have been unable to discover any reason why choice-restriction should be inhibited by Estes' procedure to a greater degree than by Reicher-Wheeler. In fact, both procedures probably inhibit use of the strategy, since both require the subject to respond at the level of letters rather than words, and both present him with randomly interspersed W, NW and SL trials. The type of response required, and the random stimulus sequence, may prevent subjects from "setting" themselves for words, thus inhibiting their use of knowledge of word structure. Perhaps inappropriate set accounts for the small size of the WSE--typically about 10%.

There is, however, a reason why Estes' procedure might not yield clear evidence for the L-R intrusions predicted by a choice-restriction or feature-redundancy theory. The reason is that the letters L and R are visually quite distinct; hence, on a large majority of trials the subject is likely to extract feature information sufficient to distinguish L and R and therefore sufficient to prevent L-R intrusion errors, though not necessarily sufficient to allow unambiguous identification of the relevant letter. (For example, one can easily imagine a subject who is not sure, on a particular trial, whether he has seen a R, B, or P, but is quite sure that he has not seen an L.) "True" L-R intrusions will occur only when feature information is insufficient to distinguish L from R and when context is accurately perceived, so that choices can be limited to L and R. It is probably rare that a subject sees the target letter so poorly as to be unable to distinguish L from R, yet sees the context letters so well as to identify all of them accurately. If the expected proportion of true L-R intrusions for words is very small, perhaps we should not be surprised that we cannot detect an excess of L-R intrusions over the proportion expected on the basis of random guessing.

Alas, the foregoing argument is speculative and the troublesome data cannot be wished away. However the argument entails at least two predictions, one testable against currently available data and one requiring new information. The first prediction is that the proportion of L-R intrusions should rise if the subject can be given accurate and undeniable evidence about the context letters. Estes accomplished exactly this through a position-probe experiment in which the context letters appeared at the time of the probe and remained in view while the subject decided on his response. (The target letter was presented briefly beforehand and was masked during the presence of the context letters. The position of the mask indicated the position of the letter to be reported.) In this condition, L-R intrusions rose from 2% to 7% of all responses to word stimuli, and from 6% to 18% of all errors. The increase, while not especially dramatic, confirms the first prediction.

Estes' interpretation of the data, of course, is that context can be used to restrict choices <u>only</u> when it is available for an extended period. He views the new procedure as <u>creating</u>, not merely enhancing, the opportunity for feature-redundancy mechanism to operate. The second prediction, if confirmed, would counter this interpretation. The second prediction is that use of a blocked rather than randomized design with the standard probe procedure would also increase the proportion of L-R errors. This prediction is based on the assumption that a blocked design will allow the subject to "set" himself for words, increasing the likelihood that he will restrict his reports to letters which complete words. Since context would be available only during the display of the target, an increase in L-R intrusions would show that subjects can use context to restrict choices, even when the context is displayed very briefly.

An Overview of the Models

Despite their obvious differences of emphasis, the four models reveal significant areas of agreement and potential agreement. The chief purpose of this section is to show some of the ways in which the models can be integrated.

Estes' model provides a general framework within which the integration may be achieved. His hierarchy of "control elements" is a structure which readily incorporates the hypothetical detection devices of the other theories. His word control elements correspond directly to Morton's logogens, and to the word detectors of Rumelhart and Siple. His feature, letter and cluster detectors correspond to similar detectors postulated by Rumelhart and Siple. Also, his cluster detectors could easily be designed to detect VCG's, the units postulated by Smith and Spoehr.

If Estes' model provides, as it were, the static architecture of an integrated model, does it also provide the rules to put those static parts in motion? It does, but only in a weak sense: Estes proposes that each control element obeys a multiplicative rule--i.e., that its probability of activation depends on a multiplicative combination of prior expectations with new information from the stimulus. Estes does not formalize and test this proposal, however. Rumelhart and Siple, on the other hand, offer a specific multiplicative principle (equation 5), which in spirit and in letter could become the operative rule for Estes' system. Morton also proposes a multiplicative principle for the activation of logogens, which could be incorporated within Estes' framework.

What is crucial here is that Morton's rule translates into the Rumelhart-Siple rule rather directly: In both cases, the probability of correct detection of a target i is given by a rule of the general form:

$$\text{P(i's detector firing)} = \frac{\begin{matrix}\text{(activation level of} \\ \text{i's detector due} \\ \text{to stimulus information)}\end{matrix} \times \begin{matrix}\text{(activation level of} \\ \text{i's detector due} \\ \text{to prior expectations)}\end{matrix}}{\begin{matrix}\text{(sum of multiplied activation levels of all} \\ \text{detectors)}\end{matrix}}$$

In Morton's model, the relevant activation levels are unanalyzed parameters ($\mathcal{A}$ and V_i) which are measured and manipulated in various ways to yield the successful predictions described earlier. In the Rumelhart-Siple model, the activation levels are further analyzed as reflections of the subject's knowledge about configurations of visual features, letter, syllables and words. That is,

$$\text{P}\begin{pmatrix}\text{activation of }\underline{\text{i's}}\text{ detector} \\ \text{due to stimulus information}\end{pmatrix} = P(F|S_i)$$

where $P(F|S_i)$ means "the probability of extracting feature set F, given stimulus i. The higher this probability, the more likely that the i-detecter will fire in the presence of feature set F. The activation level of i's detector due to prior expectations is given by

$$P(S_i)=f_w(S_i)P(WORD) + f_{syl}(S_i)P(SYL) + f_l(S_i)P(LETTER) \quad (6)$$

where all terms of the form P(UNIT) represent the subject's expectation that a given type of unit--word, syllable or letter--will be shown, and terms of the form $f_{unit}(Si)$ represent the relative frequency of stimulus i within the relevant class of units.

There are, of course, important differences between the Morton and Rumelhart-Siple formulations. In Morton's model the target i is always a word. In the Rumelhart-Siple model it may be a word, syllable or letter. In Morton's model, expectation may be based on prior syntactic and semantic content; in the Rumelhart-Siple model, expectations are based purely on frequencies. However these differences do not constitute an unbridgeable gap. The two models were designed to account for different data--Morton's for the interaction of stimulus and context, Rumelhart-Siple's for the identification of stimuli in isolation; therefore differences of emphasis are to be expected. But the two models could be combined without great damage to either: Nothing in Morton's model prevents it from being extended to subword structures by incorporation in a framework like that proposed by Estes. And Rumelhart-Siple's equation (6) could easily be generalized by substituting e_w, e_s, e_l for f_w, f_s, f_l where e represents subjective expectation rather than objective frequency. The terms e and f would become identical for the special case of stimuli presented without context. With context, expectations would deviate markedly from relative frequencies. (As stated several times, we have no way of predicting e, but we can measure it, as Morton has, by assessing predictability of a given stimulus in a given context.)

The Rumelhart-Siple model assumes that subjects select a level of response (word, syllable or letter), assign to each possible target an expected frequency within the chosen class, and weigh the set of expected frequencies together with feature information from the input in order to determine the response most likely to be accurate. The various operations could be performed separately and sequentially in the order just indicated, and Rumelhart and Siple's simulation program may well operate in this manner. However, sequential, ordered execution of the operations does not appear to be essential to the psychological content of the model; what is essential is the claim that humans weigh the various sources of evidence in the manner indicated. There is no reason why a model like Estes' cannot achieve this sort of weighing, though it does so without performing the operations in quite the way suggested above:

Long-run frequency information can be built into Estes' model by adjusting the thresholds of control elements such that elements whose targets are common configurations are triggered relatively easily, i.e., on the basis of relatively impoverished visual input.

Such an adjustment is identical to the threshold shift proposed by Morton to account for frequency effects, and it is analogous to Rumelhart and Siple's use of frequency information to adjust P(Si).

Expectations can be built into Estes' model by allowing context, broadly defined, to increase the activation of relevant control units. Thus, a subject who is led, by experimental instructions, or by experience with a prior sequence of stimuli, to expect to be shown words can increase the activation level for words as a group. Similarly, a subject who is given an incomplete sentence can increase the activation levels for words which fit the prior syntactic or semantic context. In both cases, increased activation in the control elements will allow them to fire with relatively little visual input. Essentially the same thing could be accomplished by allowing thresholds, rather than activation levels, to vary with context. The former alternative, varying activation with context, is the course proposed by Morton early in the text of his 1969 article; however, in his formal treatment, both frequency effects and context effects are treated in terms of the parameter V, suggesting either that the alternatives are formally identical, or that expectations due to context, like expectations due to frequency, are best treated in terms of threshold shifts. As we have seen, Rumelhart and Siple do not treat expectations due to prior semantic or syntactic context; in effect, however, they do treat expectations due to experimental instructions, subject "set" or experience with prior experimental trials. Such expectations are built into the parameters P(WORD), P(SYL) and P(LETTER), which are multiplied with frequency information to yield P(Si). Thus, they too treat frequency and expectations alike, suggesting again that both effects operate on detector firing thresholds, and, more generally, that all three models potentially treat both classes of effects in compatible ways.

The Smith-Spoehr model has been ignored in most of the previous discussion, because it requires special treatment. Smith and Spoehr propose a two-stage model, incorporating a stage of visual feature extraction followed by a stage of translation into acoustically codeable units. Clearly, these two stages fit readily into the integrated model sketched so far: The model postulates a level of feature extraction, followed by a filtering of feature information through higher, more abstract control elements (letter, cluster and word detectors). The feature control elements carry out Smith and Spoehr's stage 1; the higher units carry out their translation stage. Moreover, the first substage of their translation stage corresponds to establishment of letter identities--a process identical to that carried out by the letter control elements, which are directly above the feature elements in our postulated hierarchy.

A major difficulty arises, however, when their next substages are considered. According to Smith and Spoehr, letter identities are uniquely determined at a decision substage, and letter sequences are then segmented into VCG's by the rules shown in Table 1. The

VCG's are then coded acoustically. According to the filter-hierarchy model, letters may not be uniquely identified by the letter control elements. More significantly, there are no translation operations in the filter hierarchy which correspond to VCG parsing rules. On their face, these discrepancies seem to preclude any mutual compatibility between the Smith-Spoehr model and the model being proposed here.

Fortunately, the apparent incompatibility is not beyond resolution. I suggest that we are faced here with a confusion between linguistic competence and psychological performance. The VCG parsing rules proposed by Hansen and Rodgers (1965) do not, I suggest, correspond to real-time operations in word recognition as Smith and Spoehr propose. Rather, the rules capture the reader's intuitions about how written words should be segmented into units which approximate syllables in oral speech. These linguistic intuitions do play a role in perceptual performance; namely, they define memory units or perceptual filters (control elements) against which feature input is compared--they map into the cluster detectors which lie between letter and word control elements. When such units exist in an input string of letters, the subject can take advantage of within-unit redundancy, just as he takes advantage of within-word redundancy in the WORD example detailed earlier, accounting for the results of Spoehr and Smith (1973, 1975) as well as other results on perception of wordlike nonwords. Just as early attempts to incorporate other linguistic ideas (transformational grammar) directly into psychological models (viz. the psycholinguistic research of the early 1960's) proved to be too simple and gave way to less direct incorporations, so the valuable concept of the VCG may find its psychological representation in a manner somewhat less straightforward than that proposed by Smith and Spoehr. If this speculation is correct, the apparent incompatibility of the Smith-Spoehr model and the filter hierarchy model largely disappears.

At this point it is well to consider the objections which Smith and Spoehr raise regarding feature-redundancy models as a class, since we have just argued that the Smith-Spoehr model itself is compatible with a feature-redundancy formulation. First, Smith and Spoehr object to the particular feature-redundancy models of F. Smith (1971) and of Rumelhart and Siple (1974) on the grounds that they do not explain perceptual effects obtained with wordlike nonwords, since they propose that feature lists are assigned to letters and words only. (In fact, both models make some provision for nonwords; Smith and Spoehr criticize F. Smith's attempt in this regard, and ignore that of Rumelhart and Siple.) In any case, the objection clearly does not apply to the filter hierarchy model proposed here, which explicitly accounts for "wordlike nonword" effects by means of the cluster detectors. Smith and Spoehr anticipate this way of extending feature redundancy models, and object that it presupposes an ability to parse words into clusters before letters are identified. But it does not:

Letter detectors are simply connected to detectors for clusters of which those letters are part. If detectors for one or more letters in a cluster are activated, the cluster detector is activated to some degree. If enough component letters are activated strongly enough, and/or if the cluster is common enough so that its threshold is low, the cluster detector will fire. The operation of cluster detectors does not depend on prior segmentation of letter groups; segmentation is accomplished by the operation of the cluster detectors. Finally, Smith and Spoehr object that feature redundancy models require skilled readers to possess separate feature lists for every word of the language in every possible typeface--a highly implausible demand on memory. But this is where the hierarchical structure of Estes' model plays a crucial role. The network of relations among letters, clusters and words is fixed and independent of feature input. The feature lists which map into the 26 letters must be redefined whenever a new typeface or handwriting style is encountered, but redundancy rules above the letter level continue to operate without change. In fact, I suspect that it is the continued operation of these higher-level redundancy rules that enables us to cope so easily with new writing styles.

One objection raised by Smith and Spoehr concerning feature-redundancy models cannot be countered, though it is not clear that it should be countered. Smith and Spoehr point out, specifically with respect to the Rumelhart-Siple model, that it does not distinguish between perceptual and postperceptual processes of memory and decision. Since Rumelhart and Siple focus on data from full report experiments, Smith and Spoehr suggest that the predictive power of their model actually derives from its ability to describe response selection and organization, rather than perception *per se*. However, as long as we do not think of response selection as a slow, conscious, verbal process, it can be argued that memory and response are inseparable from perception. If we define perception broadly, to incorporate all processes by which information from a stimulus makes contact with paradigmatic representations stored in long term memory, the decision component of the Rumelhart-Siple model can be interpreted as a description of the way in which the nervous system routes incoming information to the appropriate permanent representation. In this regard, Rumelhart and Siple's decision procedure is exactly like the translation processes (including letter decisions) proposed by Smith and Spoehr.

In summary, the filter-hierarchy model proposed here, an integration of proposals by Estes (1974, 1975), Smith and Spoehr (1974), Rumelhart and Siple (1974), Morton (1969) and Travers (1970, 1973) appears to cope with an impressive range of phenomena in the field of word recognition. Though the integrated model has not itself been formalized, we may regard each of the four formalizations reviewed here as a special case of the model's quantitative predictive power. Anything that can be predicted by the mathematical models reviewed here can be predicted by the integrated model; the various empirical tests discussed above may be claimed as support for the integrated model.

The model also has, it is hoped, some potential heuristic virtues which range beyond issues discussed here. One is the fact that it portrays word recognition in a manner which can readily be extended to other forms of pattern recognition. That is, by focusing on a special case of pattern recognition--one in which perceptual elements and their interrelations are relatively well-defined--we may have unearthed principles which can be applied to recognition of complex objects more generally. (Whether this is so, or whether other forms of pattern recognition differ fundamentally from word recognition, precisely because other patterns are not based on well-defined elements, is a matter for future research.) Closely related is the possibility that some aspects of the model may admit neurophysiological study in the not-impossibly-distant future. This claim must be advanced with utmost diffidence; all that can be said is that the model has a certain neurological plausibility, in that one can readily imagine neural circuits which accomplish some of what detectors are said to accomplish. Finally, and more immediately, the model suggests some directions for developmental research: If skilled adult readers possess the postulated hierarchies of control elements, how do children acquire them? Can we discover an important dimension in the acquisition of reading skill, using the model as a guide? Can we relate acquisition of the postulated structures to particular methods of reading instruction, or to particular experiences which occur in the process of learning?

In closing, it is probably well to suggest some problems and limitations of the model. Perhaps most obvious is the fact that there are unresolved internal issues, chief among them Estes' emphasis on location information and his rejection of the redundancy explanation for the WSE. This issue must be resolved before harmony can truthfully be claimed among the theories reviewed here. A second issue already raised is the fact that semantic and syntactic context effects at present must be treated as wholly extraneous to the model; such effects alter parameters in the model but cannot themselves be explained. Perhaps this is a virtue; perhaps a qualitatively different explanatory system is required for such effects--but we cannot be sure that fuller exploration of context effects will not force upon us a reconceptualization of word recognition itself. Finally, there is a wide range of general problems in pattern perception which the model thus far sidesteps altogether. To cite just a few: The model implicitly assumes that letters within words are always presented in more-or-less normal orientation and spatial distribution. Yet Kolers, Eden and Boyer (1964) have shown that skilled readers show remarkable adaptability to drastically rotated texts. Conversely, Mewhort (1966) has shown that increasing the angular separation of letters in a word reduces the perceptual advantage of words over nonwords. The model must be extended to account explicitly for these somewhat paradoxical facts. Also, the model appears to assume that extraction of features is a spatially parallel process, a contention which I support (Travers, 1970, 1973b) but which others have disputed (e.g., Gough, 1972). This is another perceptual issue which must be resolved if the model is to be extended to other areas of pattern recognition.

References

Aderman, D. and Smith, E. E. Expectancy as a determinant of functional units in perceptual recognition. <u>Cognitive Psychology</u>, 1971, <u>2</u>, 117-129.

Baron, J. and Thurston, I. An analysis of the word-superiority effect. <u>Cognitive Psychology</u>, 1973, <u>4</u>, 207-228.

Berger, A. <u>Speed reading: An annotated bibliography</u>. Newark, Del: International Reading Association, 1970.

Bjork, E. L. and Estes, W. K. Letter identification in relation to linguistic context and masking conditions. <u>Memory and Cognition</u>, 1973, <u>1</u>, 217-223.

Cattell, J. McK. Über die Zeit der Erkennung and Bekennung von Schriftzeichen, Bilden and Farben. <u>Philosophical Studies</u>, 1885, <u>2</u>, 635-650.

Chall, J. <u>Learning to read: The great debate</u>. New York: McGraw-Hill, 1967.

Davis, F. B. (Ed.) <u>The literature of research in reading with emphasis on models</u>. Final report, Contract No. OEC-0-70-4790 (508), Project No. 0-9030, Office of Education, National Center for Educational Research and Development, 1971. (Document distributed by the Graduate School of Education, Rutgers University, New Brunswick, New Jersey.)

Erdmann, B. and Dodge, R. <u>Psychologische Untersuchungen über das Lesen</u>. Halle: M. Niemeyer, 1898.

Estes, W. K. On the locus of inferential and perceptual processes in letter identification. Unpublished paper, 1974.

Estes, W. K. Memory, perception and decision in letter identification. In R. L. Solso (Ed.), Information Processing and Cognition: The Loyola Symposium. Washington, D.C.: Erlbaum Associates, 1975.

Estes, W. K., Bjork, E. L. and Skaar, E. Detection of single letters and letters in words with changing versus unchanging mask characters. Bulletin of the Psychonomic Society, 1974, 3, 201-203.

Gibson, E. J., Pick, A. D., Osser, H. and Hammond, M. The role of grapheme-phoneme correspondences in the perception of words. American Journal of Psychology, 1962, 75, 554-570.

Gibson, E. J., Shurcliff, A. and Yonas, A. Utilization of spelling patterns in deaf and hearing subjects. In H. Levin and J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970.

Gough, P. B. One second of reading. In J. F. Kavanaugh and I. A. Mattingly, (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Green, D. M. & Birdsall, T. G. The effect of vocabulary size on articulation score. University of Michigan: Electronic Defense Group, Technical Memorandum No. 81 and Technical Note AFCRC-TR-57-58, 1958. (Reprinted in J. A. Swets, Ed., Signal detection and recognition by human observers. New York: Wiley, 1964.)

Green, D. M. and Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.

Hansen, D. and Rodgers, T. S. An exploration of psycholinguistic units in initial reading. In *Proceedings of the symposium on the psycholinguistic nature of the reading process*. Detroit: Wayne State University, 1965.

Howes, D. H. and Solomon, R. L. Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 1951, *41*, 401-410.

Hubel, D. H. and Wiesel, T. N. The visual cortex of the brain. *Scientific American*, 1963, *209*, 54-62.

Johnston, J. C. and McClelland, J. L. Visual factors in word perception. *Perception and Psychophysics*, 1973, *14*, 365-370.

Kolers, P. A., Eden, M. and Boyer, A. Reading as a perceptual skill. *MIT research laboratory of electronics, quarterly progress report*, 1964, *74*, 214-217.

Krueger, L. E. Search time in a redundant visual display. *Journal of Experimental Psychology*, 1970, *83*, 391-399. (a)

Krueger, L. E. Visual comparison in a redundant display. *Cognitive Psychology*, 1970, *1*, 341-357. (b)

*Levin, H. and Kaplan, E. L. Grammatical structure and reading. In H. Levin and J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Luce, D. *Individual choice behavior*. New York: Wiley, 1959.

MacKay, D. C. The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 1972, *3*, 210-227.

*Kučera, H. & Francis, W. N. *Computational analysis of present-day American English*. Providence: Brown University Press, 1967.

Mackworth, J. Some models of the reading process: learners and skilled readers. In F. B. Davis (Ed.) The literature of research on reading with emphasis on models. Final report, Contract No. OEC-0-70-4790(508), National Center for Educational Research and Development, 1971.

Manelis, L. The effect of meaningfulness in tachistoscopic word perception. Perception and Psychophysics, 1974, 16, 182-192.

Massaro, D. W. Perception of letters, words and nonwords. Journal of Experimental Psychology, 1973, 100, 349-353.

Mewhort, D. Sequential redundancy and letter spacing as determinants of tachistoscopic recognition. Canadian Journal of Psychology, 1966, 20, 435-444.

Mezrich, J. J. The word superiority effect in brief visual displays: elimination by vocalization. Perception and Psychophysics, 1973, 13, 45-48.

Miller, G. A., Bruner, J. and Postman, L. Familiarity of letter sequences and tachistoscopic identification. Journal of General Psychology, 1954, 50, 129-139.

Miller, G. A., Heise, G. A. and Lichten, W. The intelligibility of speech as a function of the context of the test materials. Journal of Experimental Psychology, 1951, 41, 329-335.

Morton, J. The effects of context on the visual duration threshold for words. British Journal of Psychology, 1964, 55, 165-180. (a)

Morton, J. The effects of context upon speed of reading, eye-movements and the eye-voice span. Quarterly Journal of Experimental Psychology, 1964, 16, 340-354. (b)

Morton, J. A preliminary functional model for language behavior. *International Audiology*, 1964, *3*, 216-225. (Reprinted in R. C. Oldfield and J. C. Marshall, Eds., *Language*. London: Penguin Books, 1968.) (c)

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, *76*, 165-178.

Morton, J. and Broadbent, D. E. Passive versus active recognition models, or Is your homunculus really necessary? In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form*. Cambridge, Mass.: MIT Press, 1967.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Posner, M. I. On the relationship between letter names and superordinate categories. *Quarterly Journal of Experimental Psychology*, 1970, *22*, 279-287.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, *81*, 274-280.

Rumelhart, D. A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, *7*, 191-218.

Rumelhart, D. E. and Siple, P. Process of recognizing tachistoscopically-presented words. *Psychological Review*, 1974, *81*, 99-118.

Selfridge, O. G. Pandemonium: a paradigm for learning. In *The mechanization of thought processes*. London: H. M. Stationery Office, 1959.

Selfridge, O. G. and Neisser, U. Pattern recognition by machine. Scientific American, 1960, 203 (Aug), 60-68.

Smith, E. E. and Haviland, S. E. Why words are perceived more accurately than nonwords: inference vs. unitization. Journal of Experimental Psychology, 1972, 92, 59-64.

Smith, E. E. and Spoehr, K. T. The perception of printed English: a theoretical perspective. In B. H. Kantowitz (ed.), Human information processing: tutorials in pe :ormance and cognition. Published by Lawrence Erlbaum/Associate , distributed by Halsted Press division of John Wiley and Sons, New York, 1974.

Smith, F. The use of featural dependencies across letters in the visual identification of words. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 215-218.

Smith, F. Understanding reading: a psycholinguistic analysis of reading and learning to read. New York: Holt, Rinehart and Winston, 1971.

Sperling, G. A model for visual memory tasks. Human Factors, 1963, 5, 19-31.

Spoehr, K. T. Linguistic processes in the perception of letter strings. Unpublished Doctoral Dissertation, Stanford University, 1973.

Spoehr, K. T. and Smith, E. E. The role of syllables in perceptual processing. Cognitive Psychology, 1973, 5, 71-89.

Spoehr, K. T. and Smith, E. E. The role of orthographic and phonotactic rules in perceiving letter patterns. Journal of Experimental Psychology: Human Perception and Performance, 1975, in press.

Sticht, T. G., Beck, L. J., Hauke, R. N., Kleiman, G. M. and James, J. H. Auding and reading: a developmental model. Human Relations Research Organization Technical Report, 1974.

Thompson, M. C. and Massaro, D. W. Visual information redundancy in reading. Journal of Experimental Psychology, 1973, 98, 49-54.

Travers, J. R. Sequential and parallel processes in reading. Unpublished doctoral dissertation, Harvard University, 1970.

Travers, J. R. Development of a formal model of word recognition. Research proposal submitted to the National Institute of Education, March 1, 1973. (Project later funded under grant number NE-G-00-3-0032.)

Travers, J. R. Effects of pronounceability and statistical "Englishness" on recognition of tachistoscopically-presented letter strings. Final report, NIE Grant NE-G-00-3-0032, Appendix C. March, 1975.

Tulving, E. and Gold, C. Stimulus information and contextual information as determinants of tachistoscopic recognition of words. Journal of Experimental Psychology, 1963, 66, 319-327.

Tulving, E., Mandler, G. and Baumal, R. Interaction of two sources of information in tachistoscopic word recognition. Canadian Journal of Psychology, 1964, 18, 62-71.

Wheeler, D. Processes in word recognition. Cognitive Psychology, 1970, 1, 59-85.

NOTES

1. Preparation of this paper was supported by a grant from the National Institute of Education, number NE-G-00-3-0032. Requests for copies should be addressed to Jeffrey Travers, Department of Psychology, Swarthmore College, Swarthmore, Pennsylvania, 19081.

Table 1

Vocalic Center Group Parsing Process

(After Smith & Spoehr, 1974)

1. Mark Positions of Vowels
2. Unitize Initial Consonants(s) with Initial Vowel and Final Consonant(s) with Final Vowel
3. Parse Intermediate Consonant(s) According to Following:
    a. . . . VCV . . . → . . . V + CV . . . .
    b. . . . VCCV . . → . . . VC + CV . . .
    c. . . . VCCCV . . → . . . VC + CCV . . .
4. If Previous Rules Yield an Inappropriate Result, Reparse Intermediate Consonant(s) According to the Following:
    a. . . . VCV . . . → . . . VC + V . . . .
    b. . . . VCCV . . → . . . V + CCV . . .
    c. . . . VCCCV . . → . . . V + CCCV . . .

Table 2

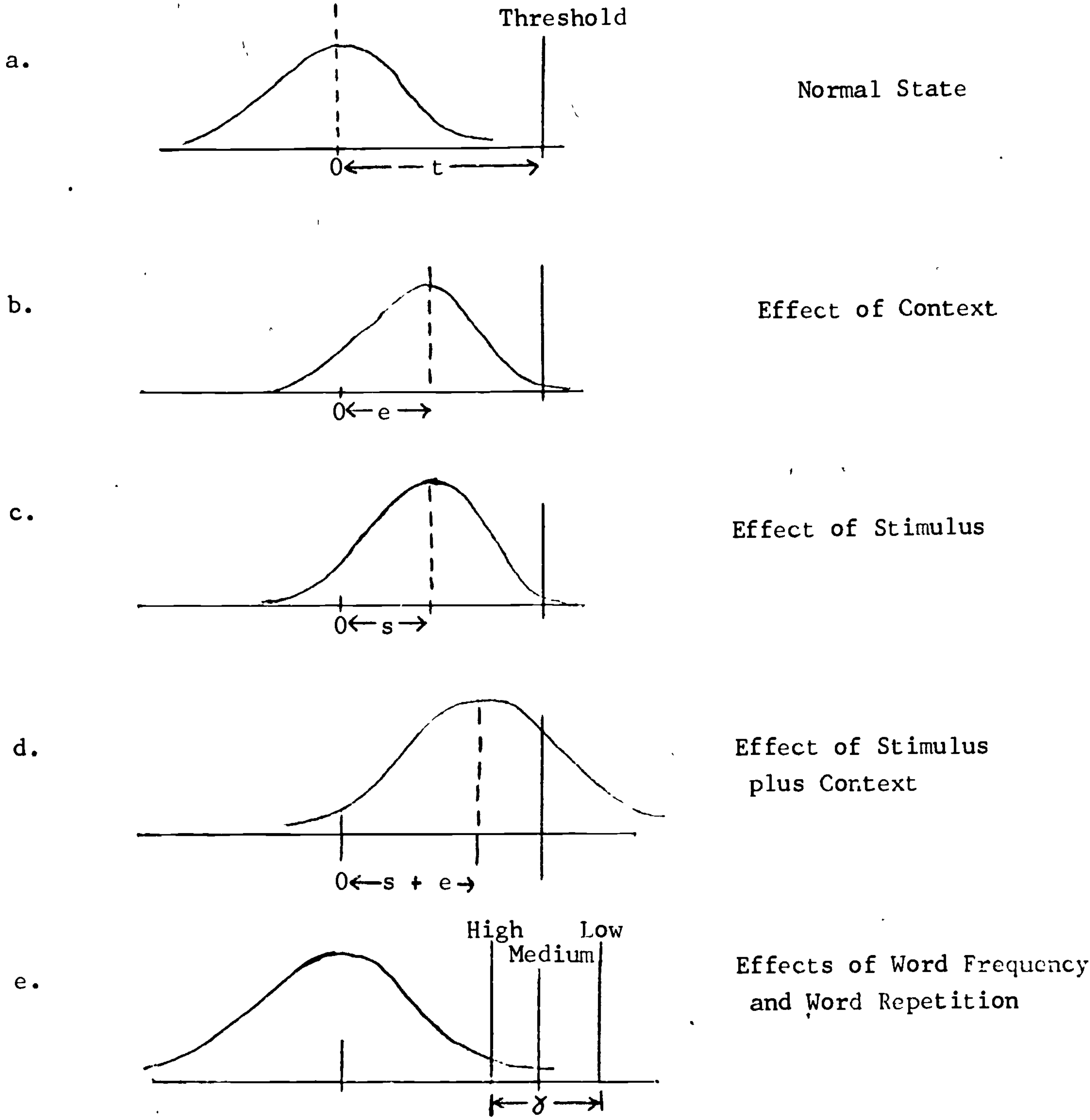
Predicted and Observed Percentages of
Error Types in Probe Experiment
(From Estes, 1974)

	CONTEXT					
	Single Letter		Word		Nonword	
RESPONSE TYPE	Observed	Theoretical	Observed	Theoretical	Observed	Theoretical
Correct	59	59	69	68	63	63
Incorrect						
Intrusions	24	25	21	22	23	22
Transpositions	-	-	5	6	10	11
Omissions	17	16	5	4	4	4

Figure 1

Operation of a Logogen

(After Morton, 1969, Figure 2)

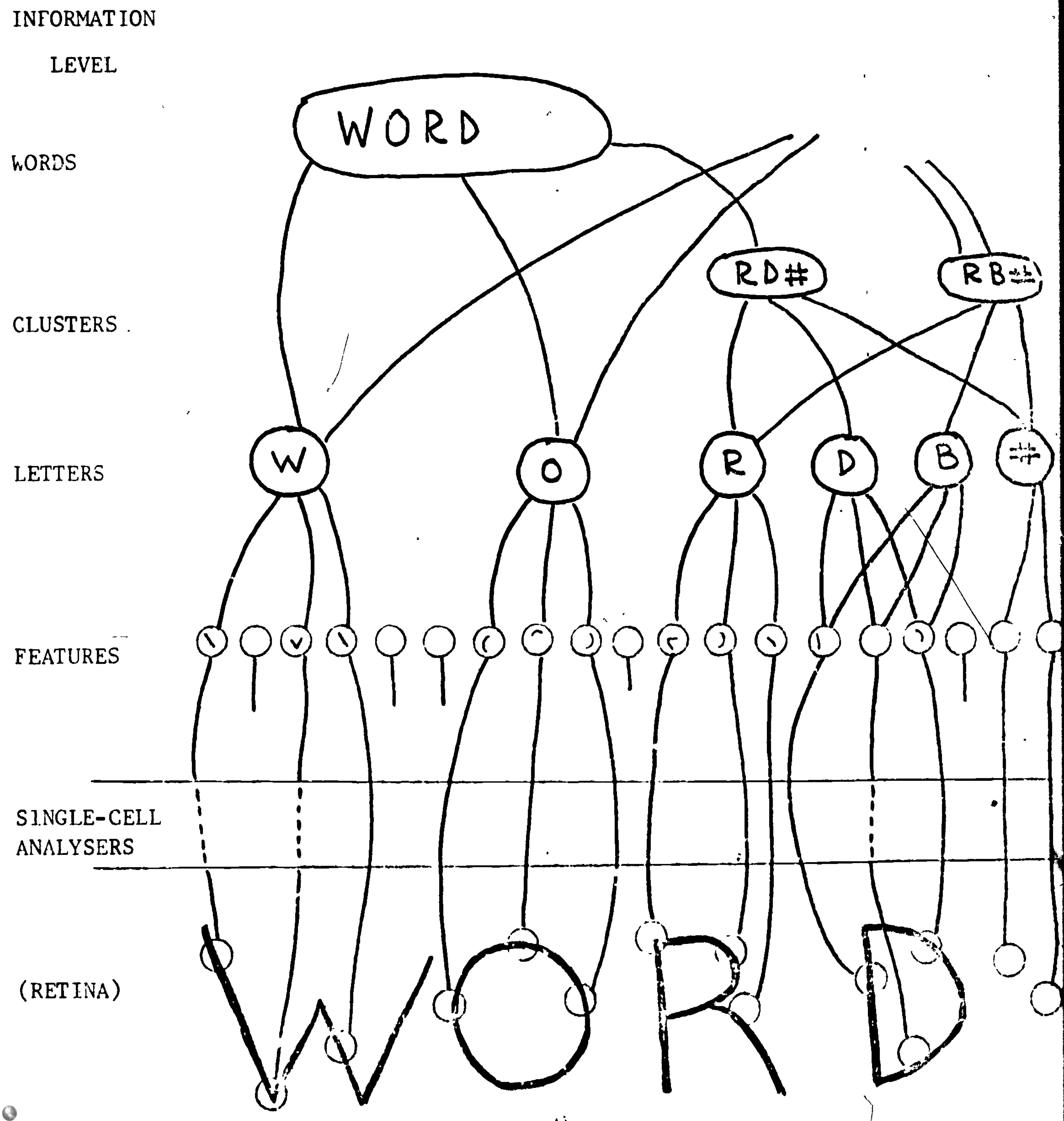


Horizontal axis represents level of activity in a logogen. Curves correspond to probability distributions of activation. When activation level exceeds threshold, a response becomes available.

Figure 2

Operation of a Control Element System

(After Estes, 1974, 1975; Travers, 1970)



FORCED SERIAL PROCESSING OF WORDS AND LETTER STRINGS: A RE-EXAMINATION[1]

Jeffrey R. Travers

Swarthmore College

Short Title: Forced Serial Processing

Send correspondence to: Jeffrey R. Travers
Department of Psychology
Swarthmore College
Swarthmore, Pennsylvania 19081

Abstract

The experiments of Travers (1973, 1974) on "forced serial processing" of words and nonword letter strings were repeated in a single study using new display characteristics and instructions to subjects. Most of the earlier findings were replicated but some were not. Words and nonword strings, three or seven letters long, were displayed serially (i.e., one letter at a time) or simultaneously, with and without backward masking. Recognition of words, and of individual letters within words, was markedly impaired in the masked serial condition relative to the unmasked serial, unmasked simultaneous and masked simultaneous conditions. Analogous effects for seven-letter nonwords were smaller or nonexistent, but three-letter nonwords produced relatively "wordlike" data. Implications of the results for the issue of serial vs. parallel processing in word recognition are discussed.

FORCED SERIAL PROCESSING OF WORDS AND LETTER STRINGS: A RE-EXAMINATION[1]

Jeffrey R. Travers

Swarthmore College

Travers (1973, 1974) used a technique of "forced serial processing" to demonstrate that, when recognizing words, skilled readers extract visual feature information from several letter positions at once and code the extracted information in chunked or unitary form. Serial processing was forced by displaying words one letter at a time, with letters in normal adjacent spatial positions and in temporal order corresponding to their left-right sequence within the word. Each letter was followed immediately by a mask, in order to prevent retention of letters in iconic memory. Such display conditions produced poor recognition at rapid exposure durations (e.g., 50 msec. per letter) which do not allow subjects enough time to code individual letters verbally; at slower rates (e.g., 200 msec. per letter), which allow a substantial amount of coding, recognition was much superior.

In both of the earlier papers, performance under forced serial processing was contrasted with performance under conditions designed to allow parallel processing.[2] In the 1973 paper, the contrast condition was one of serial, adjacent display without masking, designed to allow retention of serially-input letters in iconic memory. This condition produced uniform high levels of word recognition (about 85%) across all exposure durations from 50 to 200 msec per letter. In the later paper (1974), three contrast conditions were used--unmasked serial display, simultaneous display with masking, and simultaneous display without masking. The two unmasked contrast conditions produced near-perfect word recognition (over 95%). Accuracy in the masked simultaneous condition, though lower than in the unmasked simultaneous condition, was much better than for the masked serial condition (84% vs. 33%). The latter finding seemed particularly dramatic in view of the fact that letter exposure durations were kept constant at 48 msec. Thus, total display time for an N-letter word presented serially was 48 x N msec, while display time for the same word under simultaneous presentation was only 48 msec. The advantage of simultaneous display appeared despite the presumably countervailing effect of total display time.

As indicated above, the author interpreted these results as evidence that skilled readers normally code information from several (or all) letter positions within a word at once. Masked serial displays were assumed to interfere with this "parallel processing" by erasing or degrading the traces of letters in iconic memory before other letters were available for feature extraction--forcing the reader to an unnatural and inefficient letter-by-letter coding strategy. To rule out a competing explanation, namely that the difficulty of reading masked serial displays is due solely to the effects of the mask on perceptibility of individual letters, a nonword control was run in the 1973 study. It was assumed that random letter strings permit relatively little parallel encoding of the type suggested above; therefore report accuracies for random strings were expected primarily to reflect differences in item perceptibility across masked and unmasked display conditions. When random strings were shown in serial formats identical to those just described for words, there was no significant difference in report accuracy between masked and unmasked displays, whether the data were scored for whole strings or individual letters correct. This strong result appeared somewhat counterintuitive at first; however it could be understood in light of the fact that all exposure durations used in the experiment were well in excess of durations normally required for identification of individual letters with masking and with dark pre-exposure fields (Sperling, 1963). Given this fact, it seemed reasonable that report differences between masking conditions might be due solely to a postperceptual coding process, and not to differences in visual feature extraction. Since this conclusion will be called into question below, it is important to note that the result was stronger than required by the parallel processing hypothesis; a greater effect of masking with words than with random strings would have been sufficient to establish the greater utility of a parallel processing strategy for words.

There are reasons for doubting the generality of Travers' finding that masking does not affect report accuracy for serially-displayed random letter strings. One reason is that the mask employed in the 1973 study, a crosshatched number symbol, has been found to be relatively ineffective (cf. Travers, 1974 ; Estes, Bjork and Skaar, 1974). To this objection a counterobjection may of course be offered: The mask <u>did</u> prove effective for words, and the word-nonword difference was the datum of primary interest.

However this counterobjection loses force when a subtle error in the word-nonword comparison is pointed out: The large word-nonword difference was obtained for words <u>scored</u> <u>as</u> <u>wholes</u>. No difference was obtained for random strings scored as wholes, but such a difference might have been obscured by floor effects, since few random strings were recognized in their entirety under any display conditions. While no significant difference was obtained for letters within random strings--data for which floor effects did not apply--it is possible that the true effect for individual letters was merely very small and failed, by chance, to reach significance. Small differences in the probability of recognizing individual letters might aggregate to produce a large difference at the level of whole words. This rather speculative objection is reinforced by the fact that <u>S</u>'s were encouraged to guess freely on the word displays, and to report whole words whenever possible. Small differences in the quality of visual information available in the masked and unmasked conditions may have been magnified by the guessing strategy. Since guessing could not be of much help in the random strings, the word-nonword difference may have been exaggerated by the method chosen. In the replication study reported below, both problems with the 1973 study were avoided insofar as possible: A highly effective mask was used, and subjects in both word and nonword conditions were instructed not to guess, but to report only letters they were sure they saw.

In addition to the methodological problems just mentioned, two (as yet unpublished) contradictory empirical findings have come to the author's attention since publication of the earlier papers:

(1) In extensive studies involving serial displays of words and nonword letter strings without masking, Haber (personal communication) has repeatedly obtained U-shaped curves of accuracy vs. processing time, in contrast to the flat curve for words and the upward-sloping curve for nonwords obtained by Travers (1973). Haber varies processing time by manipulating not letter exposure duration but inter-stimulus interval (ISI). He uses high-contrast stimuli with durations on the order of microseconds, and varies ISI from zero (i.e., simultaneous display of all letters) up to several hundred msec. Previous research (e.g., Haber and Nathanson, 1969) gives ample reason to believe that such displays should be perceptually equivalent to displays in which stimulus on-time is manipulated directly, as in Travers (1973). However, Haber consistently finds that report accuracy for unmasked serial displays is worst for ISI's in the neighborhood of 100 msec, and improves with ISI's of greater or lesser duration.

Haber explains his U-shaped functions in terms of two countervailing processes: ISI's below 100 msec facilitate retention of several letters at once in iconic memory, while ISI's above 100 msec permit increasing amounts of letter-by-letter naming. This plausible explanation raises the possibility that differences in the curves for words obtained by Haber and Travers may be due to differences in visual persistence produced by the different displays used by the two authors. Travers used luminescent green characters on the dark gray face of a computer-controlled oscilloscope. Displays with dark pre- and post-exposure fields can produce visual persistence up to several seconds (Sperling, 1963). Therefore, Travers' word displays may have elicited ceiling performance--for the particular subject group and character set--at all of the exposure durations he studied. This hypothesis is tested indirectly in the experiment reported below, which replicates the relevant portion of Travers' 1973 study, but uses black-on-white tachistoscopic displays, which may be expected to produce icon durations on the order of hundreds of msec.

The reasons for the discrepancy between the nonword results of Haber and Travers is not at all clear. However, the empirical contradiction is important for Travers' (1973) argument, which hinges on the minimal differences in report accuracy induced by masking for rapid serial displays of nonwords. Haber finds an upswing in accuracy for unmasked serial displays of nonwords at rapid rates (i.e., faster than 100 msec per letter); Travers found a downswing, paralleling the downswing for masked serial displays. If Haber's finding proves to be the more general one, relatively large differences in report accuracy between masked and unmasked rapid serial displays might be the rule for nonwords as well as words. And if this is the case, it becomes necessary to demonstrate that the difference for words is significantly larger, if the parallel processing hypothesis is to be maintained. The present study uses display characteristics different from those of Haber and radically different from those of Travers (1973). Thus the study investigates the replicability of both sets of results across variations in display parameters, and, more important, permits a retest of the parallel processing hypothesis in the event that significant masking differences are obtained with nonwords.

(2) Arabie (personal communication) has done a series of studies analogous to those of Travers (1974) in which report of letters in serial masked displays is contrasted to report of letters in simultaneous masked displays. Arabie finds marked superiority of recognition in the simultaneous condition, even when the stimuli are nonword,

nonpronounceable trigrams. Travers (1974) obtained a similar effect (of much greater magnitude) for words of length 4-8 letters. He attributed the superiority of the simultaneous condition to parallel processing of letters within words. However, Travers did not run a nonword control. Arabie's results may be due to either or both of two factors which Travers did not consider: (a) Short nonwords, such as the 3-letter strings used by Arabie, may be processed in parallel; (b) More important, the serial-simultaneous difference may be due to general perceptual effects independent of coding processes, string structure and the subject's knowledge thereof. To select the most obvious example of such an effect, serial masked displays entail lateral masking of a given letter by the interference pattern intended as a backward mask for the preceding letter. (See Travers, 1974 , Figure 1, for clarification.) Unless such effects are assessed through a nonword control, the data of Travers (in press) provide at best ambiguous support for the parallel processing hypothesis. In order to address these issues, the experiment reported below replicates the study of Travers (1974) with two crucial differences: (1) Very short (3-letter) words and nonwords are included as stimuli, in order to determine whether such strings elicit results different from those produced by longer strings; (2) Both words and nonwords are shown in serial and simultaneous conditions, thus allowing a partial separation of general perceptual effects from coding effects specific to words.

The general outlines of the replication study emerge from the foregoing discussion of problems with the earlier work. Following Haber, words and unpronounceable, "un-wordlike" nonwords were shown to subjects with varying intervals between onsets of individual letters. ISI's were zero (simultaneous display), 50 msec, 100 msec or 200 msec. Stimulus strings were either three or seven letters in length and were either masked or unmasked. The design permitted simultaneous comparison of the effects on recognition of words and nonwords of forced serial processing (in the masked serial condition) with three conditions designed to allow varying degrees of parallel processing (serial, unmasked displays; simultaneous unmasked displays; simultaneous, masked displays.) Further, it allowed separate comparison for very short and relatively long strings. As indicated above, instructions to subjects were as uniform as possible across word and nonword displays. Furthermore, stimuli were presented in ordinary lower-case type on a stroboscopic tachistoscope, permitting a test of the generality of the earlier findings across rather different display parameters.

Method

Display Apparatus and Materials

Stimuli were displayed on a stroboscopic tachistoscope designed by Douglas Lawrence. The apparatus consists of an aluminum frame which is drawn upward past a horizontal slit at a fixed rate--in the present experiment, 1/6", or a line of IBM type, per 50 msec. Stimuli are typed on ordinary 8½" X 11" sheets which are fixed to the frame. A high-intensity strob light illuminates the sheet from behind for a period of a few microseconds, timed to coincide with the centering of a line of type in the slit. The subject views the typed stimulus from the front of the slit. By typing successive letters of a word or nonword string on successive lines of the sheet, it is possible to display letters serially. ISI is varied by skipping varying numbers of lines between letters. Further details on the construction of the apparatus and the visual characteristics of its displays are available in Lawrence and Sasaki (1970).

Stimulus strings were typed in lower case on 8½" X 11" sheets (Gray's Harbor Bond, No. 16) using an IBM Selectric typewriter equipped with a carbon ribbon and a Courier 72 ball. Strings designated for simultaneous (zero ISI) display were typed on a single line. Strings designated for serial display at 50 msec per letter were typed with letters on successive lines. Strings designated for display at 100 or 200 msec per letter were typed with one or three line skipped between successive letters. A pair of parentheses--()--

was typed ten lines (500 msec) above the first letter of each display. The parentheses served as a warning signal and bracketed the space in which the 3- or 7-letter string was to appear. Strings subtended a vertical visual angle of approximately 0°24'. Three-letter strings subtended a horizontal angle of approximately 1°10', and 7-letter strings an angle of 2°45'.

The mask was a capital "X" superimposed on a capital "O" (⊗). Pilot work showed it to be highly effective. In serial displays, the mask for a given letter appeared simultaneously with the following letter. In simultaneous displays a row of masks, one for each letter position, appeared one line (50 msec) after the stimulus string.

Design

A repeated-measures design was used, in which eight subjects each viewed a total of 640 stimuli, 20 in each of 32 experimental conditions. The 32 conditions were defined by the intersection of the four independent variables described in the introduction: There were two stimulus classes (words and nonwords), two masking conditions (masked and unmasked), two stimulus lengths (three and seven letters) and four ISIs (0, 50, 100 and 200 msec.)

The 32 experimental conditions were presented as blocks of 20 items. Presentation order of the blocks was counterbalanced as far as possible, given the constraints imposed by the number of subjects (eight). Half of the Ss saw words first, and half saw nonwords first. Within each of these two groups, half saw masked items first, and half unmasked items. Within each of the groups defined by joint orderings of stimulus types and masking conditions, half (i.e., one subject) saw 3-letter items first and half 7-letter items. Each subject saw half of the stimuli in ascending order of ISI and half in descending order; however ISI order obviously could not be varied within the cells defined by joint orderings of stimulus type, masking condition and length, since such cells contained only a single subject. The 20 stimuli within each block were shown in a different random order for each S.

Stimulus Strings

Stimulus words all had frequencies in printed English greater than ten and less than 250, according to the Kučera-Francis (1967) count. Words were selected as follows: All the 3-letter words falling in the specified frequency range were listed; technical terms, contractions and proper names were excluded (except for proper names that doubled as common words, e.g., rob, rod, sue, guy). This list was only a little longer than the 160 words required for the experiment. Seven-letter words were then picked by finding the 7-letter word closest to each 3-letter word on the Kučera-Francis list. In most cases, this procedure produced exact matching of frequencies between 3- and 7-letter items. Perfect matching was not possible at the upper range of frequencies, however. Matched pairs were then distributed across the eight display conditions of the experiment (two masking crossed by four ISI conditions) so as to equalize frequency distributions as exactly as possible. This required discarding some high-frequency items which could not be matched across display conditions, or for which the 3- and 7-letter matches were not sufficiently close. The procedure yielded a very close matching of frequency distributions across masking conditions, ISIs, and word lengths. Means for all 16 cells fell in the range 60.1 to 60.8 occurrence per million.

Nonword stimuli were created from the population of letters appearing in the word stimuli by arranging the 20 words assigned to each cell of the design in columns and going down the columns, selecting each vertical sequence of three or seven letters to appear as a (horizontally displayed) nonword string under the same visual conditions. The only constraints on this process were (1) that no string appeared (intuitively) to be pronounceable or "wordlike"; and (2) that no string was used more than once in the entire experiment. Internal rearrangement of strings prevented violation of these constraints.

Subjects

Ss were eight Stanford University undergraduates, three men and five women. All were native speakers of English. None reported uncorrected defects of vision. All were paid volunteers.

Procedure

Ss were run in four or five sessions of approximately two hours' duration. At the beginning of the first session, Ss were given a minimum of 32 practice trials, one or two on displays for each block of the experiment, in order to familiarize them with the apparatus and the general characteristics of the displays. Ss were also given five additional practice trials preceding each of the 32 experimental blocks, in order to allow them to form appropriate strategies for dealing with the forthcoming display type.

Ss were told that the purpose of the experiment was to determine the effects of various displays upon the readability of words and letters. They were instructed to identify stimuli aloud, as rapidly as possible. In the case of word stimuli, Ss were told to name the whole word if they felt they saw all of its letters, and to name individual letters otherwise. In cases where they deduced the identity of words while in the process of reporting individual letters, they were asked to supply the deduced word, but these "afterthoughts" were not scored as correct word identifications. Only words and letters reported as "actually seen" are taken into account in the analyses below.[3] Data were recorded by one experimenter, while a second changed stimulus sheets in the tachistoscope. Ss initiated each trial by pressing a button which caused the moving frame and strob timer to begin operation.

Results

The percentage of letters correctly identified by each S for each of the 32 experimental conditions is shown in Table 1. Although absolute levels of performance vary widely across Ss, the pattern of results is fairly consistent. (Data for the 32 conditions, averaged across Ss, are represented graphically in Figure 1.) Table 1 also shows (a) data averaged across Ss on percentages of words and nonwords correctly reported as wholes, i.e., with all letters reported in proper order; and (b) relevant comparison data from Travers (in press) and from the dissertation on which the 1973 report was based (Travers, 1970).

Insert Table 1 and Figure 1 about here.

Two types of statistical analysis were applied to the data: (1) A six-way analysis of variance was performed, using string type, masking condition, string length and ISI as fixed independent variables, subjects and stimulus items as random independent variables, and proportion of letters correctly identified as the dependent variable. Data were first subjected to an arcsin transformation, as recommended by Winer (1971, pp. 399-400).[4] Significance was tested by means of quasi-F ratios, which take account of error variance due to both items and subjects (Clark, 1973; Winer, 1971, pp. 375-378).[5] Selected results of this analysis are shown in Table 2. The quasi-F ratios convey the reliability of the various effects. However, only a few of the ratios directly test relevant hypotheses; these will be discussed where appropriate. (2) Since the most instructive contrasts are buried in multi-way interactions, several planned comparisons were also performed and are also discussed below.

Insert Table 2 about here.

Word data

In most theoretically relevant respects, the word data replicate the findings of Travers (1973, in press), although discrepancies may also be noted:

(1) The masked serial conditions (ISI=50, 100 and 200 msec) exert a marked detrimental effect on report of letters within words; the size of this effect diminishes as ISI increases, i.e., as the time available for coding individual letters grows. (Note the significant main effect for masking and the significant interaction of ISI and masking. These effects of course incorporate nonwords as well as words. The significant string type x masking x ISI interaction shows that the patterns for words and nonwords are different, as discussed in a later section on the nonword data.)

(2) Letter recognition is near-perfect for conditions which allow parallel processing, i.e., unmasked simultaneous displays, unmasked serial displays at rapid rates (ISI=50 msec) and masked simultaneous displays.

(3) There is a weak tendency for unmasked serial displays to produce U-shaped curves, with poorest recognition at 100 msec per letter, in line with Haber's results.

(4) The effects of the mask and of ISI are much larger for the whole-word data than for the individual letter data, in line with the methodological point raised in the introduction. Comparison of the whole-word data with analogous data from earlier studies using computer displays suggests that quantitative results for the unmasked displays are roughly similar. For masked displays, however, display parameters affect performance markedly: (a) the mask in the present study depresses performance on serial displays far more than the crosshatch used in the 1973 paper, and somewhat more than the

improved mask used in the more recent study (in press.) However, (b) masked *simultaneous* displays on the tachistoscope appear to be more readable than comparable displays on the oscilloscope.

(5) Performance is better for 3-letter words than for 7-letter words, particularly in the masked conditions. Travers (1973) also obtained significant length effects, especially for masked displays, but the differences were considerably smaller than those in the present study, presumably because 3-letter words were not studied.

Nonword Data

The nonword conditions strengthen the conclusion of Travers (in press) but weaken somewhat the conclusions of Travers (1973):

(1) Whereas both masked and unmasked simultaneous presentations produce virtually perfect recognition of 3- and 7-letter words, no such effects are observed for random strings. There is a facilitating effect on recognition of 3-letter strings for simultaneous masked presentation versus serial masked presentation at 50 msec, confirming Arabie's findings. However, this facilitation is less than that observed for words. (A *t*-test on the difference of differences was performed using the arcsin transformation of the data in order to compensate for ceiling effects in the 3-letter case. The resulting *t* was 1.96, *df*=7, $p < .05$.) In the case of 7-letter strings, serial presentation is actually better than simultaneous presentation (*t* for the difference of differences, performed on raw scores, = 6.7, *df*=7, $p < .005$). As suggested earlier, in connection with Arabie's results, the 3-letter data may reflect parallel encoding of short nonwords, or they may reflect perceptual impairment in the serial

case, due to lateral masking or some other form of interference. However, since the effect for words remains significantly larger than the effect for nonwords, the hypothesis of greater parallel encoding of words remains viable. The data on 7-letter strings may also be explained in terms of post-perceptual coding: Simultaneous (masked) display of 7-letter random strings, whatever its perceptual advantages, allows little time for [illegible] individual letters. Serial display, whatever its perceptual [illegible] advantages, allows greater total coding time--hence the slightly better performance in the serial case. Presumably the difference between the results for 3- and 7-letter strings reflects the different relative importance of coding vs. perceptual factors for long and short strings. With words, coding is unitary; when S sees a whole word at once, whether three or seven letters long, he names it without difficulty. His performance is damaged by masked serial presentation, which forces him to abandon his natural strategy of parallel encoding.

(2) Unlike the serial-simultaneous contrast, the serial-masked vs. serial-unmasked contrast of Travers (1973) becomes less clearcut when visual display conditions are altered. Overall, masking exerts a greater detrimental effect on words than on nonwords, as required by the parallel encoding hypothesis. This fact is evidenced by the marginally significant masking condition x string type interaction in Table 2. However, as is readily apparent in Figure 1, the effect is due largely to 7-letter strings, a fact also shown by the highly significant length x string type x masking interaction. The effects of the mask are greater for 7-letter words than 7-letter nonwords at

rapid display rates of 50 and 100 msec per letter. (t for the difference of differences at 50 msec = 4.3, df=7, $p < .005$; t at 100 msec = 5.05, df=7, $p < .005$) In contrast, 3-letter nonwords produce relatively "wordlike" data; in fact, the effects of masking on words are [illegible] than the effects of nonwords at the 50 msec display rate. This outcome may be due in part to ceiling effects with 3-letter words, and in part to parallel processing of 3-letter nonwords, as suggested by other aspects of the data.

The discrepant results for 3-letter strings do not constitute a direct empirical disconfirmation of Travers' (1973) data, since 3-letter strings were not examined in that study. However, one data point from the present study does directly contradict an earlier finding. Seven-letter words displayed without masking at 50 msec per letter were reported more accurately than 7-letter words displayed with masking at 50 msec. (A post-hoc t test yields a significance level of .005 for the masked-unmasked comparison.) In the earlier study, masked and unmasked words of all lengths from 4-8 letters were identified equally poorly at 50 msec; moreover, recognition at 50 msec was worse than at 100 msec for both the masked and unmasked cases. In the present data, however, the curve for unmasked nonwords turns up as the display rate goes from 100 to 50 msec. The curve is noticeably bowed, with a floor at 100 msec, as in Haber's data. The explanation for the contradiction between the present and the 1973 data is not readily apparent. One possibility is that prolonged visual persistence in the 1973 displays produced some type of lateral masking effect for serial displays at 50 msec. (There was, however, no evidence of such an effect for words, possibly because of ceiling effects.)

The existence of a significant masking effect for 7-letter strings at the 50 msec display rate brings up an issue raised in the introduction: Can this effect, interpreted as an estimate of the degree to which masking interferes with visual feature extraction, explain the large difference in report accuracy for whole words under masked and unmasked display conditions? A simple probability analysis suggests that it cannot. The probability of identifying a letter within a 7-letter nonword without masking at ISI=50 is .729; with masking the probability drops to .584. Taking these values as estimates of fixed probabilities of letter-recognition under masked and unmasked conditions, we may calculate the probability of getting 0, 1, 2....7 letters correct by a simple binomial:

$$P(C) = (C^7)\ p^C\ (1 - p)^{7 - C}$$

where:

P(C) = probability of getting exactly C letters correct

p = fixed probability of getting any one letter correct (p = .729 or .584)

The results of such a calculation are shown in Table 3. We do not know exactly how many letters must be independently identified in order to identify a whole word correctly. Despite instructions to subjects to report only "seen" letters, we must assume that a considerable amount of guessing occurs. Fortunately, our uncertainty on this point does not matter for present purposes. For any reasonable value of C, the difference between predicted whole-word accuracy levels for masked and unmasked presentations is substantially less than the actual difference. For example, if we assume that S can identify a whole word given that he has independently identified four or more letters, we would predict (by summing the relevant probabilities) that he should recognize whole words with probability .908 in the unmasked case, which is fairly close to the observed value. However, the same assumption applied to the masked strings yields a whole-word probability of .679, substantially higher than the observed value. The lesson is clear: In order to recognize a whole word, more letters must be identified under masked conditions than under unmasked conditions.

Discussion

Despite some deviations from earlier findings, the data on balance add support to the hypothesis that visual feature information from different letter positions within a word is encoded in parallel, rather than going through a preliminary process of serial letter-by-letter coding before the whole-word code is retrieved. The new data suggest, however, that very short nonwords, even "unwordlike" nonwords, may also be coded in parallel, i.e., that verbal codes for up to three unrelated letters may be retrieved simultaneously (though of course three letters cannot be rehearsed simultaneously.) Perhaps the latter result should not surprise us; there is no reason to believe that English orthography has evolved so as to produce a perfect match between the information content of a single letter and the simultaneous retrieval capacity of the human information-processing system. Likewise there is no reason to believe that our inability to execute three "verbal rehearsal routines" at once must reflect an inability of the system to use visual input to "call" several such routines at once.

The data also make an obvious but often overlooked methodological point--that general conclusions about "information processing" should be based on a reasonably broad sample of input conditions, and not on a single experiment. Travers (1973) reached the premature conclusion that forced serial processing (i.e., masked serial display) does not impair report of nonword letter strings. The present results indicate that masking <u>can</u> affect report of serially-presented nonwords. The greater effect of masking upon words than upon nonwords, a crucial datum for the parallel processing hypothesis advanced by the author, holds only for strings longer than three letters, and is considerably less dramatic in the present study than in the 1973 study.

In contrast to the masked serial vs. unmasked serial comparison technique, the method used in the later study (1974) seems fairly robust across display conditions. There are dramatic differences in report accuracy for masked serial vs. masked simultaneous displays of words. This effect is weak for 3-letter nonwords and is actually reversed for 7-letter nonwords. Thus, whatever the general perceptual advantages of simultaneous display, there appears to be a special facilitating effect for words. At present, the best explanation for this effect appears to lie in the utility of parallel processing for stimuli which map into unitary verbal codes.

It is valuable to have a reliable technique for demonstrating parallel processing, and one that produces such large effects.

Presumably it will be of interest to learn whther parallel processing is useful for wordlike nonwords of various kinds, in order to construct a model of word recognition that takes account of subword structure. The larger the basic effect, the more likely it is that the effect will differ measurably for nonwords with relatively subtle structural differences. The author attempted a study of this type, using strings of varying "orders of approximation to English" in the 1973 paradigm. The results were ambiguous, perhaps because the technique does not produce sufficiently large or reliable effects. The masked serial/masked simultaneous contrast appears to offer hope for successful investigations along these lines.

References

Clark, H. The language-as-fixed effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Estes, W. K., Bjork, E. L., & Skaar, E. Detection of single letters and letters in words with changing versus unchanging mask characters. Bulletin of the Psychonomic Society, 1974, 3, 201-203.

Haber, R. N. & Nathanson, L. S. Processing of sequentially presented letters. Perception and Psychophysics, 1969, 5, 359-361.

Kučera, H. & Francis, W. N. Computational analysis of present-day American English. Providence, R.I.: Brown University Press, 1967.

Lawrence, D. H. & Sasaki, E. H. An N-channel stroboscopic tachistoscope for typewritten stimuli. Behavior Research Methods and Instrumentation, 1970, 2, 63-65.

Sperling, G. A model for visual memory tasks. Human Factors, 1963, 5, 19-31.

Travers, J. R. Sequential and parallel processes in reading. Unpublished doctoral dissertation, Harvard University, 1970.

Travers, J. R. The effects of forced serial processing on indentification of words and random letter strings. Cognitive Psychology, 1973, 5, 109-137.

Travers, J. R. Word recognition with forced serial processing: Effects of segment size and temporal order variation. Perception and Psychophysics, 1974, 16, 35-42.

Winer, B. J. _Statistical principles in experimental design_. (2nd ed.) New York: McGraw Hill, 1971.

Notes

1. The research reported in this paper was supported by grant number NE-G-00-3-0032 from the National Institute of Education. The author wishes to thank Keith Harris and Lisa Friedman, who assisted in conducting this experiment. Special thanks are due to Professor Douglas Lawrence, who provided the display apparatus.

2. For reasons of clarity, many conditions of the two earlier experiments are omitted from the present discussion.

3. *S*s usually reported words as such, rather than reporting individual letters within words, except in cases where they saw too few letters to identify the words. Often, however, they followed their whole-word reports with the information that they had only "seen" certain of the letters; in such cases, only the "seen" letters are scored in the data. The instruction to report only "seen" letters seems to have been taken seriously by at least some *S*s, though it clearly cannot be claimed that the instruction eliminated guessing entirely.

4. The transformation used was: $\Phi = 2 \arcsin \sqrt{X}$, where Φ = the transformed score and X = the proportion of letters correct, out of three or seven, on a given trial. However, a value of .999 was substituted whenever the actual X was 1.0. The correction for ceiling effects suggested by Winer was not used, because it depends on the number of observations underlying each proportion, i.e., on the number of letters in each string. This "correction" has the effect of wiping out most length effects in the ANOVA.

5. Quasi-F ratios are all of the form:

$$F' = \frac{MS_E + MS_{I \times S}}{MS_{E \times S} + MS_I}$$

where:

MS_E = mean square for the main effect or interaction of interest

MS_I = mean square for the (nested) item effect

$MS_{I \times S}$ = mean square for the subject-item interaction

$MS_{E \times S}$ = mean square for the interaction between subjects and the effect of interest

Degrees of freedom are calculated from formula given in Winer (1971) and Clark (1973).

Table 1

Percentage of Letters Correctly Identified by Subject, String Type, Masking Condition, String Length and ISI

WORDS

	Unmasked								Masked							
	3-letter				7-letter				3-letter				7-letter			
ISI=	0	50	100	200	0	50	100	200	0	50	100	200	0	50	100	20
S1	96.7	100.0	100.0	98.3	97.1	99.3	97.1	95.7	100.0	78.3	95.0	98.3	99.3	77.1	91.4	86
S2	100.0	100.0	100.0	100.0	100.0	98.6	95.7	98.6	100.0	83.3	100.0	100.0	100.0	53.6	81.4	97
S3	100.0	100.0	98.3	100.0	100.0	97.1	97.9	97.9	98.3	96.7	100.0	100.0	95.7	88.6	87.9	92
S4	96.7	100.0	98.3	100.0	100.0	100.0	97.9	100.0	98.3	91.7	100.0	100.0	100.0	70.0	92.9	99
S5	98.3	93.3	93.3	93.3	100.0	99.3	89.3	100.0	96.7	93.3	90.0	95.0	97.8	86.4	86.4	99
S6	98.3	98.3	100.0	100.0	100.0	96.4	81.4	87.9	98.3	58.3	86.7	100.0	94.3	56.4	75.7	87
S7	100.0	100.0	96.7	98.3	100.0	97.1	98.6	99.3	100.0	80.0	86.7	96.7	97.1	67.3	86.4	98
S8	100.0	100.0	96.7	96.7	100.0	99.3	84.3	89.3	100.0	98.3	98.3	100.0	100.0	65.7	80.7	87
ean	98.8	99.0	97.9	98.3	99.6	98.4	92.8	96.1	99.0	85.0	94.6	98.8	98.0	70.6	85.4	93
ean of hole ords	96.9	96.9	95.6	96.3	99.4	93.1	79.4	83.1	96.9	68.1	88.1	95.6	93.1	33.8	50.6	74
o[illegible]rable data rom Travers (1970)						91	81	86						51	71	86
omparable data rom Travers (1974)					98	96							82	44		

NONWORDS

	Unmasked								Masked							
	3-letter				7-letter				3-letter				7-letter			
ISI=	0	50	100	200	0	50	100	200	0	50	100	200	0	50	100	20
S1	100.0	98.3	98.3	100.0	88.6	76.4	75.7	76.4	81.7	78.3	81.7	95.0	70.7	66.4	71.4	93
S2	98.3	100.0	98.3	96.7	81.4	77.9	70.7	82.9	88.3	80.0	91.7	98.3	62.1	58.6	62.9	75
S3	100.0	96.7	96.7	100.0	88.6	82.9	80.0	86.4	96.7	83.3	91.7	96.7	65.7	70.0	77.9	82
S4	93.3	96.7	95.0	96.7	72.9	68.6	67.9	85.0	85.0	83.3	86.7	100.0	49.3	59.3	66.4	77
S5	96.7	90.0	93.3	95.0	77.1	69.3	72.1	80.0	66.7	58.3	63.3	85.0	57.1	70.0	76.4	76
S6	93.3	90.0	93.3	98.3	77.9	82.9	69.3	77.1	73.3	55.0	80.0	93.3	51.4	57.1	71.4	77
S7	86.7	81.7	73.3	91.7	68.6	60.0	60.0	70.0	75.0	63.3	83.3	88.3	40.7	40.7	55.0	71
S8	100.0	100.0	96.7	100.0	68.6	65.7	58.6	80.0	91.7	80.0	86.7	100.0	57.9	45.0	58.6	85
ean	96.0	94.2	93.1	97.3	77.9	72.9	69.4	79.7	82.3	72.7	83.1	94.6	56.9	58.4	67.4	77
M[illegible] of whole strings	89.3	83.1	79.4	91.9	8.1	1.3	0.0	3.1	53.1	27.5	41.3	83.8	0.0	0.0	0.0	
Co[illegible]e data fro[illegible]avers (1973)						0	0	0					0	0	0	

Table 2

Analysis of Variance

(See Notes 4 & 5)

Source	Quasi-F Value	df	p
String Length	73.2	1, 9	$<.001$
String Type	95.2	1, 9	$<.001$
Masking Condition	72.9	1, 12	$<.001$
ISI	25.0	3, 50	$<.001$
String Length x String Type	84.0	1, 14	$<.001$
String Length x Masking Condition	.369	2, 12	N.S.
String Length x ISI	1.06	4, 48	N.S.
String Type x Masking Condition	3.39	1, 16	$.05 < p < .1$
String Type x ISI	6.63	3, 52	$<.001$
Masking Condition x ISI	17.1	3, 49	$<.001$
Length x Type x Masking	9.92	1, 15	$<.01$
Length x Type x ISI	3.42	3, 94	$<.05$
Length x Masking x ISI	.67	8, 110	N.S.
Type x Masking x ISI	9.36	3, 99	$<.001$
Length x Type x Masking x ISI	5.66	3, 147	$<.005$

Table 3

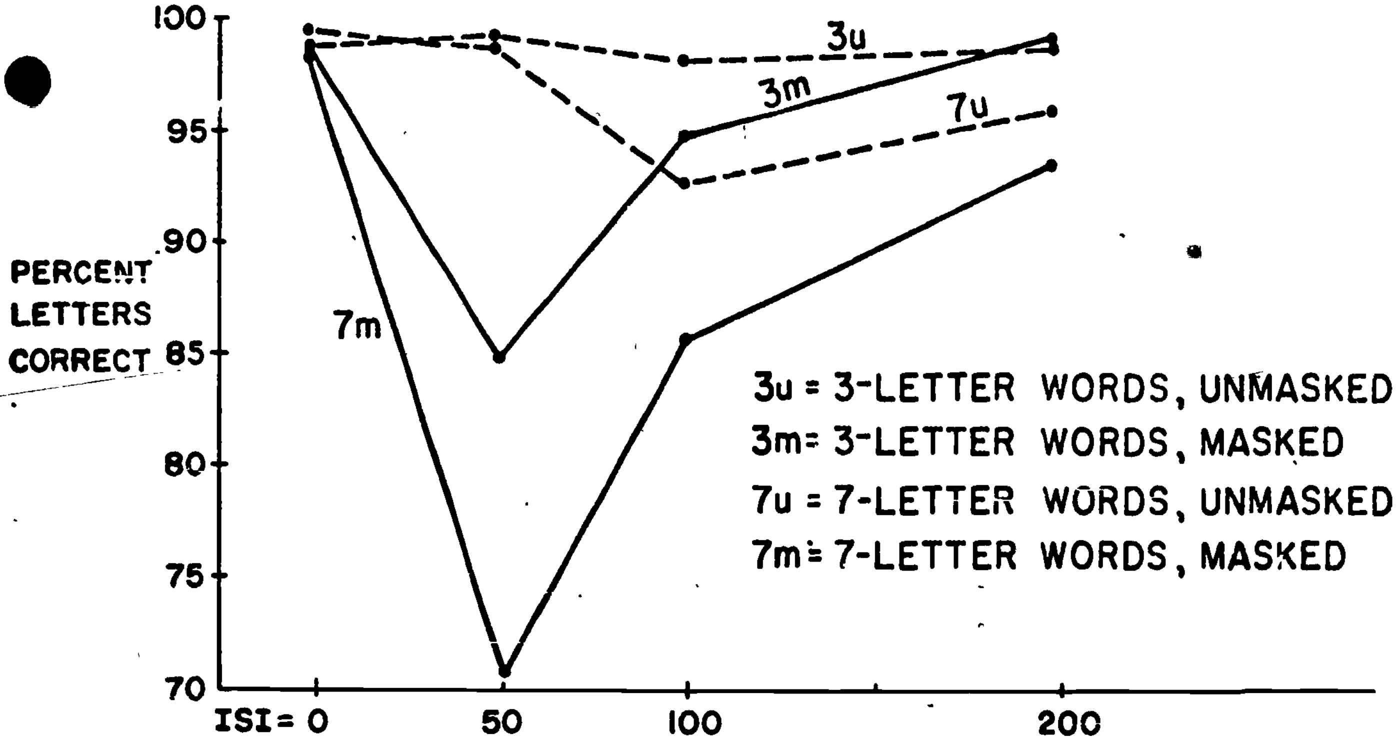
Probability Analysis for 7-Letter Strings at ISI = 50

Unmasked	Masked
p = .729	p = .584
p(0) = .0001	p(0) = .0022
p(1) = .0015	p(1) = .0213
p(2) = .0167	p(2) = .0895
p(3) = .0732	p(3) = .2085
p(4) = .1967	p(4) = .2931
p(5) = .3174	p(5) = .2468
p(6) = .2847	p(6) = .1156
p(7) = .1094	p(7) = .0232
Actual p(word) = .931	Actual p(word) = .338

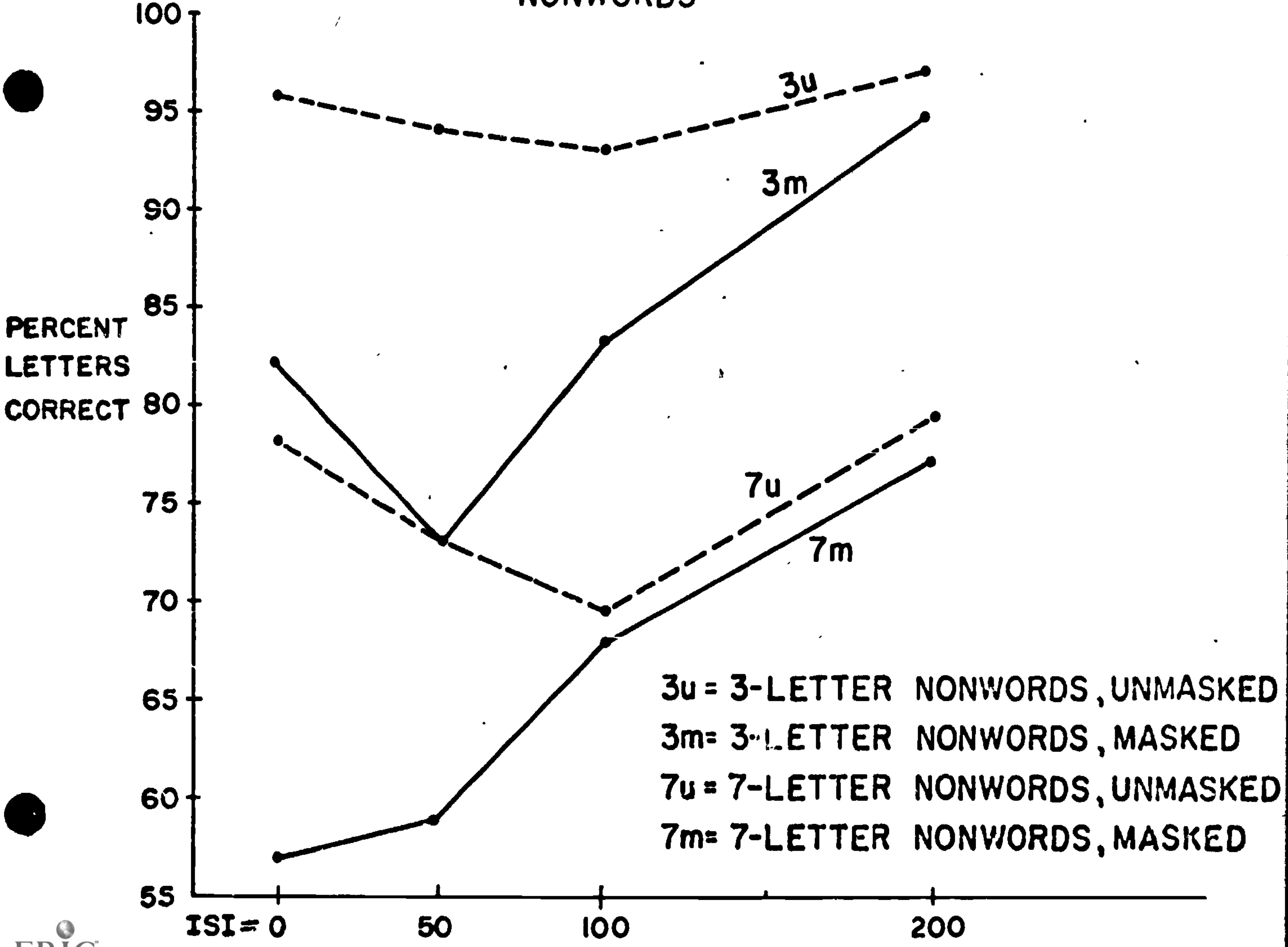
Figure Captions

Figure 1. Percent letters correct as a function of stimulus string type, stimulus string length, masking condition and inter-stimulus (i.e., inter-letter) time interval (averaged across eight subjects.)

WORDS



NONWORDS



EFFECTS OF PRONOUNCEABILITY AND STATISTICAL "ENGLISHNESS" ON IDENTIFICATION OF TACHISTOSCOPICALLY-DISPLAYED LETTER STRINGS[1]

Jeffrey R. Travers
Swarthmore College
and Donald C. Olivier

Short Title: Pronounceability and "Englishness"

Send correspondence to: Jeffrey R. Travers
Department of Psychology
Swarthmore College
Swarthmore, Pennsylvania 19081

Abstract

The statistical "Englishness" of letter strings, as assessed by a measure based on letter-cluster frequencies, exerts a significant effect on report accuracy, independent of pronounceability, despite previous suggestions to the contrary. This claim is supported by an experiment on tachistoscopic recognition of a set of nonword strings for which rated pronounceability and "Englishness" vary orthogonally. Implications for a theory of word recognition are discussed.

EFFECTS OF PRONOUNCEABILITY AND STATISTICAL "ENGLISHNESS" ON IDENTIFICATION OF TACHISTOSCOPICALLY-DISPLAYED LETTER STRINGS

Jeffrey R. Travers
Swarthmore College
and Donald C. Olivier

An impressive array of studies demonstrates that "wordlike" nonwords of various kinds exhibit some of the perceptual and/or response characteristics of words (e.g., Miller, Bruner and Postman, 1954; Postman and Rosenzweig, 1956; Gibson, Pick, Osser and Hammond, 1962; Baron and Thurston, 1973; McClelland and Johnston, 1974 ; Spoehr and Smith, 1975). The study of wordlike nonwords is of interest because it bears promise of revealing an important aspect of the word-perception mechanism, in particular, of showing what kind of knowledge about morphological and orthographic structure the skilled reader uses in recognizing words.

Perhaps the simplest hypothesis about the skilled reader's knowledge is that he knows which letter clusters occur frequently in his (printed) language. However, several studies have found little or no relationship between cluster frequencies within words or nonwords and perceptibility of those strings as assessed by a variety of measures (Postman and Conger, 1954; Gibson, 1964; Gibson, Shurcliff and Yonas, 1970; McClelland and Johnston, 1974; Spoehr and Smith, 1975). Recent papers on the subject have generally advanced nonstatistical conceptions of the psychologically relevant aspects of word structure, such as grapheme-phoneme correspondence (Gibson et al, 1962), orthographic regularity (Gibson et al, 1970; McClelland and Johnston, 1974) or syllabic organization (Spoehr and Smith, 1975).

It is important not to read into the aforementioned papers more than the data actually permit. In most cases, the authors demonstrated that nonstatistical structural features exerted perceptual and/or response effects even when average bigram or trigram frequencies were controlled. Their conclusions regarding nonstatistical aspects of structure are not called into question here, although it will be argued that average bigram or trigram frequency is but one of many possible frequency-based measures of structure; other statistical measures might prove more powerful, and the demonstration of effects independent of such measures might prove more difficult. What the cited papers do not show, but might be misread as showing, is that statistical aspects of structure exert no independent effect on letter-string recognition. The principal purpose of the present paper is to demonstrate the existence of such an effect.

The demonstration involves three steps: (1) A brief review of the cited papers, pointing out why the question of frequency-based structural effects is still open; (2) A critique of standard frequency-based measures of statistical "Englishness," together with the introduction of a new measure which avoids some of the inadequacies of previous ones; (3) Description of an experiment in which "Englishness" is shown to exert an effect on free report of letter strings, independent of their pronounceability.

Previous Work on Cluster Frequency Effects

The early work of Postman and Conger (1954) is often cited as proving the ineffectiveness of cluster frequencies in word perception. Postman and Conger showed that free report accuracy for trigrams is unrelated to frequency of occurrence of those trigrams in printed English. However, their measure of frequency did not take position into account. Their stimulus list included items such as CTI, which is very frequent in printed English, but always appears in a medial position (in words such as ACTION, FRICTION, etc.) Had they used a position-dependent frequency count (or simply treated "space" as a letter, and taken account of the frequency of such "trigrams" as space-C-T) their results might have been different.

Spoehr and Smith (1975) showed that perceptibility of letters within nonword strings could be predicted by the amount of recoding necessary to convert those strings into pronounceable sequences of syllables. Average bigram frequencies for the various strings were unrelated to letter perceptibility. However, Spoehr and Smith also used a position-independent measure of cluster frequency. Moreover, their method of concatenating separate bigram frequencies--averaging--is subject to criticism developed in a later section.

McClelland and Johnston (1974) contrasted the effects of cluster frequency with those of "orthographic regularity," operationally defined as pronounceability. Cluster frequency was assessed by summing bigram frequencies across their 4-letter stimulus strings. Bigram frequencies were counts of the number of word types in which a given bigram occurred in a given position in a crossword puzzle dictionary. Cluster frequency was found to exert minimal impact on letter perceptibility, as assessed by both forced choice and free report. McClelland and Johnston's measure is suspect on two grounds: First, token-based, rather than type-based counts presumably reflect the reader's visual experience with letter clusters. Type-based counts, especially counts using rare words such as those in crossword dictionaries, are likely to overestimate the frequencies of certain clusters, as the recent work of Landauer & Stree (1973) suggests. Second, McClelland and Johnston concatenated frequencies by simple summing, a technique criticized below.

Perhaps the most extensive examination of the relative importance of cluster frequencies and alternative conceptions of structure has been conducted by Eleanor Gibson and her colleagues (Gibson et al, 1962; Gibson, 1964; Gibson et al, 1970). Gibson et al (1962) showed that pronounceable nonwords (e.g., GLURCK) were reported more accurately than unpronounceable nonwords formed by reversal of initial and final consonant clusters of the pronounceable set (e.g., CKURGL). Anisfeld (1964) pointed out that summed bigram and trigram frequencies were higher for pronounceable than matched unpronounceable items for almost every pair; thus he raised the possibility that cluster-frequency rather than pronounceability might explain Gibson's results. Gibson (1964) replied that summed bigram and trigram frequencies were confounded with string length, itself a strong predictor of report accuracy. She showed that average bigram and trigram frequencies were uncorrelated with report accuracy, while pronounceability had a correlation of .65.

Gibson originally interpreted the correlation between pronounceability and report accuracy as showing that letter clusters which map consistently into sounds become perceptual chunks. Later she amended this interpretation when she found "pronounceability" effects in perceptual reports of deaf subjects (Gibson et al, 1970). In the latter paper she concluded that sheer orthographic regularity could lead to perceptual chunking without the direct mediation of sound. However, Gibson continued to reject the notion that sequential letter dependencies might be a basis for perceptual chunking. In the 1970 study, sequential dependencies were rejected after a stepwise regression showed that they explained at most about 1% of the variance in report accuracy, after length and pronounceability were taken into account. (Gibson examined both position-dependent and raw bigram and trigram frequencies.)

Gibson's analysis shows that orthographic regularity (again, operationally definable as pronounceability) exerts an effect independent of average bigram and trigram frequency. However, pronounceability and frequency are correlated in her stimulus set. (For pronounceability and average position-dependent bigram frequency, Gibson reports an r of .63.) It is possible that an independent frequency effect was obscured in her data, since her stimulus set did not include many relatively unpronounceable items with relatively high-frequency clusters, or pronounceable items with low frequency clusters, which would have lowered the pronounceability-frequency correlation and--perhaps--have increased the proportion of variance explained by frequency independent of pronounceability. In addition, Gibson, like the other investigators cited, used averaging as a means for concatenating cluster frequencies across strings. This method is criticized in the following section.

Measures of Statistical "Englishness": A Critique and a Proposal

With one exception, the studies cited above used summed or averaged bigram or trigram frequencies as overall measures of the statistical "Englishness" of nonword strings. (The exception is the study of Postman and Conger, which used raw frequencies for trigram stimuli.) None of the authors gave an explicit rationale for choosing this measure, presumably because it bears an obvious intuitive relation to "Englishness," conceived in terms of cluster frequencies. However, it is likely that the authors hoped to rule out the general class of frequency-based conceptions of psychologically relevant structure, and not merely to rule out conceptions tied to the specific measure chosen. It is therefore relevant to examine some of the shortcomings of the measure, and to explore other measures equally consistent with a frequency-based notion of "Englishness." Two points may be made in this connection.

(1) Frequency-based concepts of word structure do not in general require that cluster frequencies be summed or averaged; other combinatorial principles, e.g., taking continued products or geometric means, are equally consistent with such concepts. Summing or averaging can in fact produce intuitively misleading estimates of the central tendency of transition probabilities within a string, especially where very high frequency clusters are

involved. For example, GLURCK and THXZQP have about the same average bigram frequency (according to the Underwood-Schulz, 1960, combined count) because of the presence of the high-frequency bigram "TH" in the latter string. A continued product would assign THXZQP a statistical "Englishness" rating of zero, because no other bigram in the string occurs in the language. The point is not that multiplying is "right" and summing "wrong," merely that a case can be made for either within a frequency framework.

(2) Raw frequencies of occurrence may not be as relevant psychologically as certain conditional probabilities or relative frequencies. It may not matter how often a reader has seen a particular cluster, if other, similar clusters are equally frequent. Partial visual information could trigger perception or report of any of the similar clusters with roughly equal likelihood; therefore, despite their high frequencies, members of the set might not appear to show perceptual advantages. For example, the trigram "THI" is more frequent than the trigram "QUE", according to the Underwood-Schulz (1960) count. However, "QUE" is the most frequent trigram beginning with QU, while "THE" is almost ten times more frequent than "THI." Thus "QUE" might be reported with high "accuracy" under visual conditions that permitted only partial feature information to be extracted, while "THI", presented under identical conditions, might show many "THE" intrusion errors. A measure which reflects the frequency of a trigram relative to other, similar trigrams, e.g., $\frac{F(THI)}{F(TH)}$, the frequency of THI divided by the frequency of all trigrams beginning with TH, might predict perceptibility/reportability better than the raw frequency measure. Again, the point is not

that raw frequencies are wrong and relative frequencies right, but that both are plausible ways of linking cluster frequencies to performance and must be evaluated empirically.

We propose the following measure of associative strength, or statistical "Englishness," which uses relative rather than absolute frequencies and a multiplicative rather than additive combinatorial rule:

Let the "Englishness" (E) of an n-letter string $\#L_1L_2\ldots L_k\ldots L_n\#$ (where # denotes "space" and L_i denotes the letter in the ith position in the string) be defined as the probability that the string will be generated by the rule:

$$E = P(\#L_1L_2\ldots L_k\ldots L_n\#) =$$
$$P(\#) \,.\, P(L_1|\#) \,.\, P(L_2|\#L_1) \,.\, P(L_3|L_1L_2)\ldots P(L_k|L_{k-2}L_{k-1})\ldots \quad ($$
$$P(\#|L_{n-1}L_n)$$

where each conditional probability $P(L_k|L_{k-2}L_{k-1})$ is interpreted as the probability that letter L_k follows letters L_{k-2} and L_{k-1} in printed English. This rule can be rationalized in at least two ways. First, it can be seen as a Markov approximation to English orthographic rules. It is formally analogous to the "Shannon guessing game" technique for generating statistical approximations to English (Shannon, 1951; Miller et al., 1954), but it is a way of assessing the "Englishness" of existing strings, rather than producing new ones. Second, a rationalization with less theoretical loading can be given in terms of a general linear model for predicting English letter strings.[2]

The probability expression in formula (1) may be converted into a usable measure by means of the following simplifications and transformations: (1) The conditional probabilities may be estimated by frequencies of the relevant trigrams and bigrams, i.e.,

$$P(L_k | L_{k-2}L_{k-1}) \overset{\wedge}{=} \frac{F(L_{k-2}L_{k-1}L_k)}{F(L_{k-2}L_{k-1})}$$

where "$\overset{\wedge}{=}$" means "is estimated by" and capital F denotes relative bigram and trigram frequencies. (2) When the estimation is performed, certain terms may be cancelled or ignored:

$$P(\#L_1L_2 \ldots L_k \ldots L_n) \overset{\wedge}{=} \frac{F(\#)}{F(All)} \cdot \frac{F(\#L_1)}{F(\#)} \cdot \frac{F(\#L_1L_2)}{F(\#L_1)} \ldots$$

The terms $F(\#)$ and $F(\#L_1)$ may be cancelled, and the term $\frac{1}{F(All)}$ ignored, since it appears in all strings and thus contributes nothing to measuring their relative "Englishness." (3) Since the continued product of formula (1) will yield very small values for E, it is convenient to take the negative logarithm of the estimator product, yielding positive values for E, generally between 1 and 20, depending on the length of the string. Note that large values of the negative log denote low "Englishness." These operations leave us with the following formula for estimating the relative "Englishness" of a string:

$$E \overset{\wedge}{=} -\sum \log F(\#L_1L_2) + \log \frac{F(L_1L_2L_3)}{F(L_1L_2)} + \ldots \log \frac{F(L_{k-2}L_{k-1}L_k)}{F(L_{k-2}L_{k-1})} + \ldots \log \frac{F(L_{n-1}L_n\#)}{F(L_{n-1}L_n)}$$

(4) Finally, the confounding of the measure with string length may be circumvented by the simple expedient of calculating "Englishness-per-letter" by dividing E by the number of letters in the string.

The second author has tabulated the frequencies of all letters, bigrams and trigrams appearing in the Kučera-Francis (1967) count of one million words of printed English. (Note that the counts are partially position-sensitive, since "space" is treated as a character and n-grams incorporating "space" are included in the count.) The table also includes logarithms needed to calculate the E-measure, as defined above; for each trigram the value of $\log \frac{F(L_1L_2L_3)}{F(L_1L_2)}$ is given. In addition, the second author has prepared a computer program for calculating the "Englishness" measure for any input string. Investigations of the empirical properties of the measure are now underway; in particular, data from some of the experiments cited above are being reanalyzed to determine whether the measure predicts performance as well or better than summed or averaged n-gram frequencies. One preliminary finding may be reported here, in order to show that the measure is at least equivalent in predictive power to previous measures:

Report accuracy data from Gibson <u>et al</u>.(1962) were regressed on various combinations of stimulus string length, pronounceability rating and "Englishness." The results are shown in Table 1. The multiple $\underline{R}^2$s shown in the table are quite similar to those obtained by Gibson <u>et al</u>. (1970), using average position-dependent bigram frequency as a measure of statistical "Englishness", and using different data. The table suggests that pronounceability is again the dominant variable, even with respect to the new "Englishness" measure, and that the two variables are correlated both with each other and with length, though "Englishness" is more confounded with length than is pronounceability. "Englishness" contributes 9% to

Insert Table 1 about here.

the explained variance when length is controlled and pronounceability ignored, while pronounceability contributes 13% with length controlled and "Englishness" ignored. "Englishness" contributes 1% when both length and pronounceability are taken into account. The authors naturally hoped, and failed, to show that the new "Englishness" measure predicts more powerfully than summed or averaged bigram or trigram frequency. However, for purposes of the present paper it suffices to show that the new measure is equivalent to older ones, for one main interest is to show experimentally that statistical "Englishness", as defined by the measure, can contribute to tachistoscopic report accuracy, independently of pronounceability.

An Experiment on Pronounceability and Statistical "Englishness"

Since pronounceability and statistical "Englishness" in the cluster-frequency sense are correlated in most stimulus sets, it is difficult to separate their contributions to perceptibility/reportability. For example, Gibson's data (and our reanalysis thereof) suggest that pronounceability is the dominant variable with respect to predicting report accuracy; yet neither predictor explains much variance independent of the other. Moreover, as indicated earlier, a stimulus array that included items of high "Englishness" but low pronounceability, and/or high pronounceability but low "Englishness," might have given a different picture of the relative strengths of the two variables.

In the present experiment, four sets of stimuli were constructed: One set ($H_E L_P$) high in "Englishness," as defined in the previous section, but low in rated pronounceability (e.g., SPHLB); one

(L_EH_P) low in "Englishness" but high in pronounceability (e.g., UMFIK), one (H_EH_P) high on both measures (e.g., PALEB) and one (L_EL_P) low on both measures (e.g., UDRSL). Stimulus sets were designed so that the distributions of pronounceability were closely matched between the L_EH_P and H_EH_P sets, as well as between the L_EL_P and H_EL_P sets. Similarly, distributions of "Englishness" were closely matched between the L_EH_P and L_EL_P sets, and between the H_EL_P and H_EH_P sets. Thus pronounceability and "Englishness" varied orthogonally in the total stimulus set.

It should be noted that the ranges of variability of both pronounceability and "Englishness" were severely restricted by the requirements of the design: The H_EH_P and L_EH_P strings could be no higher in rated pronounceability than allowed by the low "Englishness" of the L_EH_P set; similarly, the H_EH_P and H_EL_P sets could be no higher in "Englishness" than allowed by the low pronounceability of the H_EL_P set. Analogous restrictions held at the low end of the pronounceability and "Englishness" scales. Significant effects of "Englishness," pronounceability or both would thus indicate high sensitivity of tachistoscopic report to one or both of these variables.

Since it was desirable to assess the effects of the two variables independent of string length, and unnecessary to duplicate Gibson's demonstration of the powerful effects of length, all stimuli were five letters long. Twenty strings of each of the four types were shown to subjects tachistoscopically. The dependent variable was free report of letters in the displayed strings--a measure clearly reflecting both perceptibility and response factors. It was felt that separation of the effects of "Englishness" on the two types of factors could wait until after a general effect had been demonstrated.

Method

Stimulus Strings

Olivier's computer count of frequencies and log-relative-frequencies was used to construct 140 strings, 35 in each of the four stimulus categories. "Englishness" values were calculated directly, by summing the relevant logarithms. Pronounceability was initially judged intuitively. The 140 strings were then presented to 12 subjects (Stanford University undergraduates) who rated them on a 7-point scale of pronounceability. (A rating of one corresponded to "unpronounceable" and a rating of seven to "easily pronounceable.") Mean ratings and "Englishness" values were then used to select the four sets of 20 stimuli used in the experiment, with pronounceability and "Englishness" distributions closely matched where appropriate. An additional ten stimuli in each category were saved for use as practice strings, and five in each category were discarded.

Table 2 lists the experimental stimuli, together with mean "Englishness" scores and pronounceability ratings for each. The

---

Insert Table 2 about here.

---

table shows that E-scores and P-ratings were controlled closely across the cells of the design. The number of syllables was also approximately equal for the two pronounceable sets. The proportion of vowels and consonants was fairly well controlled across the two levels of "Englishness," though not across the two levels of pronounceability. This did not seem a critical failing, since a confounding of letter-types and pronounceability could at most render effects of pronounceability somewhat ambiguous. It could not affect the outcome for "Englishness," the variable of prime interest here. The low-Pronounceability stimuli clearly show the effects of abbreviations, contractions, Roman numerals and foreign words. This is not an undesirable feature for purposes of testing the effects of visual familiarity.

Display Apparatus and Materials

Strings were typed on 6" x 9" white cards, in large upper-case letters, using an IBM Selectric typewriter with a carbon ribbon and an "Orator" ball. Cards were displayed in an Iconix model 6137 3-field tachistoscope, controlled by a model 6010 Preset Controller and a model 6255 Timebase and Counter. Stimuli subtended a visual angle of approximately 50' vertically and 4° 12' horizontally. Illumination was approximately 21.3 mL. The pre-exposure field was a large dot, displayed at the location of the center of the string for 500 msec. and followed by a 1000 msec blank The postexposure field was a masking pattern consisting of three overlapping lines of five number symbols (#), displayed for 1000 msec.

Subjects

Nine subjects were run in the experiment. All were Stanford

University undergraduates and paid volunteers. All were native speakers of English, and none reported uncorrected defects of vision.

Procedure

The 80 stimulus strings were arranged in random order, subject to the constraint that equal numbers of strings from each of the four experimental categories be included in each block of 20 trials. The 80 strings were shown to all Ss in the same order.

Prior to the experiment proper, Ss were given 40 practice trials with non-experimental items of each of the four types. During practice, exposure durations were adjusted to a level which produced correct reports of approximately three letters out of the five presented on each trial. Durations thus obtained were used for the first block of 20 experimental trials. If an S's performance drifted noticeably above or below three letters per trial, exposure duration was adjusted by 10 msec in a compensating downward or upward direction for the next block of trials. If necessary, this procedure was repeated for succeeding blocks of 20 trials. (Such adjustments could have no systematic effect on performance across experimental conditions, since items from the four conditions were distributed equally across blocks of trials.) Exposure durations varied from 45 msec to 200 msec across Ss and blocks of trials.

Results

Mean numbers of letters reported for each of the four experimental conditions are shown in Table 3. The effects of

---

Insert Table 3 about here.

---

pronounceability and "Englishness" are small (about 9% for pronounceability, 12% for "Englishness") but in the expected direction. Their statistical reliability was tested by the conservative analysis-of-variance procedure recommended by Clark (1973). Pronounceability and "Englishness" were treated as fixed independent variables, subjects and items as random independent variables. Significance was tested by Quasi-F ratios (Winer, 1971, pp. 375-378) which incorporate both subject and item variance in their error terms. The main effect of "Englishness" proved highly reliable ($F'' = 9.90$; $df=1,51$; $p < .005$). The main effect of pronounceability was at best marginally significant ($F'' = 3.70$; $df=1,27$; $.05 < p < .1$). The pronounceability-"Englishness" interaction was nonsignificant ($F'' = 2.64$; $df=1,47$; $p > .1$).

Discussion

The small size of the effect of statistical "Englishness" is not surprising, in view of the method by which stimulus strings were constructed and selected. As indicated above, the variation in "Englishness" permitted by the design was severely limited, restricting the size of the effects we could expect to observe. The restriction, of course, was necessary in order to permit orthogonal variation of pronounceability and "Englishness," which normally show a high correlation. Given the restriction, it is noteworthy that "Englishness" nevertheless exerted a measureable effect, one that met a rather rigorous statistical test.

It is also noteworthy that the effects of "Englishness" proved more than equal to those of pronounceability when the two variables were forced to operate independently. Of course, variations in pronounceability were also restricted by the requirements of the design. Moreover, we have no idea whether the ranges of variation in pronounceability and "Englishness" were in any sense commensurate, since the scales were constructed independently. Therefore we are in no position to claim that previously observed effects of pronounceability are artifacts of statistical "Englishness," and no such claim is intended. We do claim, however, that the effects of "Englishness" in a purely statistical sense may previously have been underestimated.

With respect to theories of word recognition, this claim may be significant in either or both of two quite distinct ways: (1) It may imply perceptual learning, unmediated by the auditory/articulatory mapping characteristics of letter strings--a possibility already entertained by Gibson et al. (1970). (2) It may imply that apparently different structural conceptions are highly correlated with "Englishness" and with each other; hence they may prove more difficult to separate empirically than has previously been thought. A good measure of statistical "Englishness" ought to reflect the contribution of whatever structural factors operate in word perception; thus it is not actually an alternative to other conceptions, except insofar as it captures purely associative perceptual learning that other conceptions ignore.

Finally, it should be stressed that various conceptions of the skilled reader's knowledge of word structure are consistent with many alternative hypotheses about how that knowledge is put to use--and this general point applies to the statistical conception explored here. For example, identification of one or two letters from a high-frequency cluster, or identification of a set of letter fragments consistent with such a cluster, might predispose the reader/subject to respond with the names of all letters in the cluster (a response-bias explanation of the relative ease with which words and wordlike nonwords are reported in tachistoscope experiments). Alternatively, frequent, familiar clusters might function as perceptual units; thus the probability of "visual synthesis" (Neisser, 1967) of an entire cluster might be high, given the extraction of partial feature information consistent with the cluster (and, in general, with other, lower-frequency units also). Clear specification of both knowledge and process are needed for development of an adequate theory of word-perception.

References

Anisfeld, M. A comment on "The role of grapheme-phoneme correspondence in the perception of words." *American Journal of Psychology*, 1964, *77*, 320-321.

Baron, J. & Thurston, I. An analysis of the word-superiority effect. *Cognitive Psychology*, 1973, *4*, 207-228.

Clark, H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335-359.

Gibson, E. J. On the perception of words. *American Journal of Psychology*, 1964, *77*, 667-669.

Gibson, E. J., Pick, A., Osser, H. & Hammond, M. The role of grapheme-phoneme correspondence in the perception of words. *American Journal of Psychology*, 1962, *75*, 554-570.

Gibson, E. J., Shurcliff, A. & Yonas, A. Utilization of spelling patterns by deaf and hearing subjects. In H. Levin and J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Kučera, H. & Francis, W. N. *Computational analysis of present-day American English*. Providence: Brown University Press, 1967.

Landauer, T. K. & Streeter, L. A. Structural differences between common and rare words: failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 119-131.

McClelland, J. & Johnston, J. What makes words so easy to see? Unpublished manuscript, 1974.

Miller, G. A., Bruner, J. S. & Postman, L. Familiarity of letter sequences and tachistoscopic identification. Journal of General Psychology, 1954, 50, 129-139.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Shannon, C. Prediction and entropy of printed English. Bell System Technical Journal, 1951, 30, 50-64.

Spoehr, K. T. & Smith, E. E. The role of orthographic and phonotactic rules in perceiving letter patterns. Journal of Experimental Psychology: Human Perception and Performance, 1975 (in press.)

Underwood, B. J. & Schulz, R. W. Meaningfulness and verbal learning. Chicago: Lippincott, 1960.

Winer, B. J. Statistical principles in experimental design, 2nd ed. New York: McGraw-Hill, 1971.

Notes

1. The research reported in this paper was supported in part by a grant to the first author from the National Institute of Education (number NE-G-00-3-0032). Requests for reprints should be addressed to Jeffrey R. Travers, Department of Psychology, Swarthmore College, Swarthmore, Pennsylvania, 19081. Requests for further information on the cluster frequency counts and the program for computing the "Englishness" measure should be addressed to Donald C. Olivier, 20 Fairfield Street, Cambridge, Massachusetts, 02138. The authors wish to thank Nancy Adams who ran subjects and conducted much of the data analysis.

2. A paper presenting and rationalizing the measure in detail, as well as applying it to data from many of the experiments cited in the text, is now in preparation.

3. The reader should not be misled into thinking that the "left-right" structure of the measure implies left-right serial processing of letters in word recognition. The measure is in fact neutral with respect to the issue of serial vs. parallel processing, and is consistent with a variety of different recognition models, as the discussion section endeavors to show. The skeptic on this point may be interested to note that the measure yields identical "Englishness" scores whether organized left-right or right-left. That is,

$$E = -\log\left[F(\#L_1L_2) \times \frac{F(L_1L_2L_3)}{F(L_1L_2)} \cdots \frac{F(L_{n-2}L_{n-1}L_n)}{F(L_{n-2}L_{n-1})} \times \frac{F(L_{n-1}L_n\#)}{F(L_{n-1}L_n)} \right].$$

$$= -\log \left[F(L_{n-1}L_n\#) \times \frac{F(L_{n-2}L_{n-1}L_n)}{F(L_{n-1}L_n)} \cdots \frac{F(L_1L_2L_3)}{F(L_2L_3)} \times \frac{F(\#L_1L_2)}{F(L_1L_2)} \right]$$

4. The summed log measure, uncontrolled for length, appears in this regression. Since some of the trigrams in the unpronounceable strings have zero frequencies in English, it was necessary to insert a very low "penalty value" in such cases, in order to avoid infinite logarithms. The penalty value is discussed in a paper by Oliver and Travers, now in preparation (see Note 2).

Table 1

Stepwise Regression of Report Accuracy on
Length, Pronounceability and "Englishness"
(Data from Gibson *et al.*, 1962)[3]

Predictor Variables	Multiple R^2
Length	.69
Length + Pronounceability	.82
Length + Pronounceability + "Englishness"	.83
Length + "Englishness"	.78
Pronounceability (Length uncontrolled)	.42
"Englishness" (Length uncontrolled)	.57
Pronounceability + "Englishness" (Length uncontrolled)	.63

Table 2

Stimulus Strings

High "Englishness" High Pronounceability			High "Englishness" Low Pronounceability		
String	"E" Score	"P" Rating	String	"E" Score	"P" Rating
DYSTE	7.004	4.5	CHNST	7.010	2.4
EFUET	7.568	4.8	STSTE	8.170	2.0
XYGES	7.576	4.8	SMSTS	8.238	2.0
THRYS	8.261	5.5	MRSHM	8.466	2.5
ODISE	8.404	6.4	SHMST	8.596	2.3
AMEAP	8.612	5.8	SPHST	8.604	2.5
EBETE	9.336	6.8	XYDNT	8.725	2.4
XYMON	9.490	4.6	MRSTR	9.177	2.1
ATAUL	9.490	6.1	KRZFE	9.203	2.1
PALEB	9.610	6.4	SPHLB	9.314	1.6
ALPOE	9.790	6.7	XYGNS	9.335	2.3
AMBAE	9.871	5.9	PHLBS	9.592	2.3
XEDIT	9.875	4.0	THSTH	9.611	1.8
ZWESH	10.055	5.6	KHMST	9.727	2.4
KRUKA	10.824	6.5	ZWKST	9.957	2.1
ZAKIT	10.935	6.5	EAUEE	10.015	2.8
DASOS	11.493	6.3	XYTTS	10.859	2.0
OMSOF	11.514	6.5	AEAUE	10.935	2.4
IVOMS	11.622	6.0	DRSTR	11.339	3.0
SUHAB	11.955	6.5	OUIEO	11.397	3.0
Mean	9.666	5.81	Mean	9.414	2.30

2 1-syllable
18 2 or more syllables

55 consonants
45 vowels

80 consonants
20 vowels

Table 2 (Continued)

Stimulus Strings

Low "Englishness"
High Pronounceability

String	"E" Score	"P" Rating
OSBIM	13.860	6.6
UBNAK	14.267	6.1
OMSBI	14.743	6.0
OMLOK	14.897	6.8
OMSUZ	15.019	5.6
UMLOX	15.124	6.1
IPRUX	15.194	5.3
LYDOV	15.288	6.0
OOVOP	15.298	5.8
NYDOB	15.401	5.5
IKAKK	15.501	5.5
IKLUF	15.593	6.0
UMFIK	15.633	6.3
UBRYM	15.694	7.0
TYMSU	15.874	5.1
IPRUK	15.887	5.7
OOGMU	16.281	5.8
UCOKK	16.462	5.0
TYBIV	16.590	5.5
OSBIV	16.654	6.3
Mean	15.463	5.90

20 2-syllables

58 consonants
42 vowels

Low "Englishness"
Low Pronounceability

String	"E" Score	"P" Rating
IEUXI	13.514	2.9
UGNPH	14.265	2.0
LTBLD	14.386	1.6
XIGPD	14.655	1.6
IEWNP	14.698	2.5
GHNNH	14.704	1.6
EWNRL	14.882	2.5
ILFTF	14.964	2.0
OAIAU	15.347	3.0
XACSF	15.409	2.0
BLDBR	15.423	2.0
GLDYM	15.452	2.8
XJISQ	16.060	1.6
PTHNU	16.182	2.5
YRNKH	16.191	2.0
GMSKN	16.389	1.9
UDRSM	16.460	3.5
GMSBR	16.791	2.1
IUATU	18.026	2.8
EOEUI	19.118	2.8
Mean	15.648	2.29

69 consonants
31 vowels

Table 3

Mean Number of Letters Reported as a Function of "Englishness" and Pronounceability

		Pronounceability			
		Low	High		
Statistical "Englishness"	Low	2.81	3.29	3.05	Difference: +.39
	High	3.40	3.48	3.44	
		3.10	3.39		
		Difference: +.29			

PHONOLOGICAL ALTERNATION AND SEMANTIC RELATEDNESS JUDGMENTS

(A PILOT STUDY)

Jeffrey R. Travers

Swarthmore College

Many writers have pointed out the "irregularity" of English spelling--the lack of a one-to-one relation between letters and sounds--and the difficulties which this irregularity creates for the child learning to read. Some (e.g., Makita, 1968) have suggested that the existence of simple grapheme-phoneme mapping in other languages accounts for the low rates of reading disability observed among children learning to read those languages. Others (e.g., Fries, 1963; Veneszky, 1967; Berdiansky, Cronnell & Koehler, 1969) have argued that orthographic-phonetic regularities do exist in English, but not at the level of letters; letter clusters, appearing in specified environments within the word, do tend to have relatively stable sound values. (Consider, for example, the variable sound values of "i" in ice, it, machine, first, action, etc., vs. the relative stability of "tion" in action, friction, suction, diction, etc.) However, as Smith (1971) points out, attempts to specify even a fraction of the rules linking orthography and phonetics in English have led to very long and complex lists. It is not clear that such rules can, even in principle, account for all of the orthographic-phonetic connections of English. Moreover, even if they could they would surely pose a major learning problem for the beginning reader.

A radically different view of the relation between English orthography and phonetics has been taken by Chomsky (1970). Chomsky argues that English orthography does not and should not map directly into sounds. He points to pairs of words, like "courage-courageous" in which identical letter sequences ("courage-") have different pronunciations and yet convey the same underlying meaning. If orthography were faithful to sound, the members of pairs which show such "phonological alternations" would have to be spelled differently; thus the orthography would fail to represent the--more important--fact that "courage" and "courageous" incorporate the same underlying lexical entry. Since the rules governing phonological alternations are known (intuitively) to the adult speaker, an orthography which represents the underlying lexical entry will allow him to generate appropriate pronunciations if necessary. At the same time, such an orthography will exhibit semantically-relevant equivalences and differences directly, whereas a phonetic orthography would fail to do so. Chomsky goes so far as to argue that "conventional English orthography ... appears to be a near-optimal system for representing the spoken language." (Chomsky, 1970, p. 4)

Klima (1972) has pointed out that Chomsky's enthusiasm for English orthography is justified only if one makes certain assumptions about what the adult speaker/reader knows. Without pursuing Klima's complex argument in detail, the present paper attempts to explore some of the ways in which psychological assumptions about the reading process and the adult reader's knowledge might interact with the features of English orthography to which Chomsky has directed attention.

There are two simple conceptions of reading, both of which may be true for some people some of the time, but which make opposed predictions about the effects of phonological alternations on recovery of semantic information in reading.

(1) The skilled reader may typically recode many or most words from a visual into an auditory or articulatory internal representation before recovering meaning. That is, he may subvocalize printed words, and "understand" his internally-generated speech rather than "understanding" words and sentences directly from their representations in visual memory. If he does, it might well take him longer to decide that phonologically dissimilar pairs like "potent-impotent" are closely related in meaning than phonologically similar pairs, like "patient-impatient." The assumption here is that when internally generated phonetic sequences are somewhat dissimilar, the reader must scrutinize them more closely in order to decide that they bear a similar meaning. By the same argument, it should take the reader relatively long to determine that phonetically similar sequences are unrelated in meaning, as in pairs such as "peach-impeach."

(2) The skilled reader may make little use of subvocalization in retrieving semantic information. Semantic analysis may be based (in some unknown way) on "visual" representations of printed words. In this case, we might expect phonological alternations to make no difference in the time required to decide whether word pairs are semantically related.

In the pilot experiment reported here, subjects were asked to decide whether pairs of words bore a close semantic relation.

They were pretrained to understand that the meaning relations were very close; "related" pairs represented antonym relations (e.g., patient-impatient), part-of-speech shifts based on the same lexical entry (pursue-pursuit), or, in a few cases, tense changes for verbs (hid-hide) or number changes for nouns (knife-knives). Word pairs were presented in a single-field tachistoscope. Displays were terminated when Ss hit one of two response keys, signalling either "yes" (i.e., that a close semantic relation existed between the members of the pair) or "no" (i.e., that the semantic relation was nonexistent or very remote.)

The dependent variable of interest was reaction time necessary to make correct responses. Concept (1) above predicts that mean RT for "yes" responses should be greater for pairs with phonological shifts than without, and that RT for "no" responses should be greater for pairs without shifts than for pairs with shifts. Concept (2) predicts similar distributions of RT's for "shift" and "no-shift" pairs.

Method

Four sets of stimulus pairs were constructed--one with phonological alternations and close semantic relations (e.g., courage-courageous), one with no alternations and close relations (e.g., possible-impossible), one without alternation but also without semantic relations (e.g., peach-impeach) and one with apparent alternations but with no semantic relations (e.g., leg-legal). All pairs were visually similar, with most letters of the shorter member of the pair incorporated in the longer member.

The author's intuitive judgments of semantic relatedness were checked against ratings on a seven-point scale, provided by four subjects. Twenty-five pairs of each type were constructed.

Stimulus pairs were typed in lower case letters on white 3" x 5" cards, approximately centered. They were presented in a small one-field tachistoscope, illuminated by an ordinary incandescent bulb. Viewing distances were approximately those of ordinary reading. The two response buttons were placed side-by-side and operated by the S's preferred hand. Ss hit the left button for "yes" (or semantic similarity) and the right button for "no" (or semantic dissimilarity). The 100 stimuli were presented in fixed random order, after 10-20 practice trials to familiarize subjects with the apparatus and task. Seven Ss, all Stanford University undergraduates, were run.

Results

Mean RT's for the four conditions of the experiment are shown in Table 1. Overall, it took Ss an average of 69 msec

---

Insert Table 1 about here.

---

longer to make negative judgments than positive judgments, consistent with the usual finding in RT work that "no's" take longer than "yeses". The presence of a phonological shift slowed RT by 14 msec, suggesting that shifts may increase processing time. However, the interaction predicted by concept (1) above did not materialize. RTs were faster for semantically related pairs when a phonological shift was present. RTs were slower for semantically unrelated pairs when a shift was present. Thus the interaction was opposite to that predicted.

Discussion

The data gave little support to either of the conceptions advanced above. Phonological shifts did appear to affect RT, contrary to concept (2), but did so in a manner contrary to concept (1).

There were many uncontrolled factors in the experiment. For example, the lengths of pairs were not equated across conditions, nor were word frequencies. However, neither of these factors seemed to predict the pattern of results. Also, there were occasional extreme values of RT, but the pattern did not become clearer when extremely slow RTs (presumably caused by lapses in attention) were deleted.

The data did not seem to merit elaborate statistical analysis, and none was performed. However, it is likely that the small (14 msec) difference produced by phonological alternation is unreliable, given the small n and variability of the scores. It may well be that concept (2) above is the more accurate--and therefore that Chomsky is right to imply that phonological factors do not intervene between the visual stimulus and the recovery of meaning by skilled readers. However, such a conclusion would be entirely premature on the basis of the present pilot work. It might well be the case that concept (1) is not in error, but that subsidiary assumptions linking the concept to predictions about reaction time are. Further exploration of this issue, an important one for understanding and teaching reading, clearly requires far more careful theoretical analysis and more sensitive empirical techniques.

References

Berdiansky, B., Cronnell, B. & Koehler, J. Spelling-sound relations and primary form-class descriptions for speech-comprehension vocabularies of 6-9 year olds. Southwest Regional Laboratory for Educational Research and Development, Technical Report No. 15, 1969.

Chomsky, N. Phonology and reading. In H. Levin & J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970.

Fries, C. C. Linguistics and reading. New York: Holt, Rinehart & Winston, 1963.

Klima, E. S. How alphabets might reflect language. In J. F. Kavanaugh & I. A. Mattingly (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Makita, K. The rarity of reading disability in Japanese children. American Journal of Orthopsychiatry, 1968, 38, 599-614.

Smith, F. Understanding reading. New York: Holt, Rinehart & Winston, 1971.

Veneszky, R. L. English orthography: Its graphical structure and its relation to sound. Reading Research Quarterly, 1967, 2, 75-106.

Table 1

Mean Reaction Times for Judging Semantic Relatedness

(RTs in seconds; N = 7 subjects)

		Semantic Relation?		
		Yes	No	
Phonological Alternation?	Yes	1.070	1.178	1.124
	No	1.095	1.125	1.100
		1.083	1.152	