ANALYSIS OF LETTER STRINGS IN WORD NONWORD CLASSIFICATION

By

GARY BRUCE FORBACH

Bachelor of Arts

University of Denver

Denver, Colorado

1969

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
MASTER OF SCIENCE
July, 1972

ANALYSIS OF LETTER STRINGS IN WORD

NONWORD CLASSIFICATION

Thesis Approved:

Robert F Stanners
Thesis Adviser

Larry T. Brown

R Weber

D Durham
Dean of the Graduate College

ACKNOWLEDGEMENTS

Since the initial work on this thesis which began in the Fall of 1971, several people have been most instrumental in transforming the rough, early conceptions of the study into a workable research project, and finally to this manuscript. For the use of his research facilities and continuous guidance and valuable insights, a special "thank you" is extended to Dr. Robert F. Stanners, who served as thesis advisor.

The author also wishes to thank the other members of his committee, Dr. Robert J. Weber and Dr. Larry L. Brown. A special note of thanks should go to my wife, Paula, for her continuous support and endless hours of typing and proofreading through various drafts and the final manuscript.

Finally, this thesis is dedicated to the advancement of human understanding via the science of Psychology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

STATEMENT OF THE PROBLEM

Purpose of the Study

Psychologists have long been concerned with the problem of recognition. Directly or indirectly, this problem is at the root of most learning and memory phenomena in the sense that *recognition* results from some memory operation which links a present event with a past event in a way that produces cognizance of the relationship between the two events. The important question is how is a particular stimulus input stored in memory so that it can be *retrieved* (or *located*, *activated*, *excited*) by the stimulus input from another event so that the relationship between these two events is detected, thereby producing recognition? William James in his *Psychology* referred to an argument relevant to this question which was put forth by James Mill in a discussion of memory. Mill first elaborated on the mechanism of association, and then established that what has been labeled *recognition* in this discussion was due to the same mechanism as *association*. Then he said ". . . the machinery of association /recognition/, as we know, is nothing but the elementary law of habit in the nerve-centres" (James, 1920, p. 290). This "explanation" of recognition is similar to more contemporary views which assume *content-addressable* memory. In the *content-addressable* scheme, the stimulus input indexes directly the address of the memory representation. Similarly, Mill assumed that specific "nerve-centres" were excited by the particular input associated with each neural center. This idea has been the basis for much of psychological theory, but has failed to specify those features of a stimulus event which are important for the formation of the "associations."

Contemporary investigation in the general area of recognition includes models for human recognition of distinctive features of phonemes (Halle, 1964), recognition of human speech (Paul, House, and Stevens, 1964), and models for machine recognition of human cursive script (Lindgren, 1965). The problem of word recognition is one of this class of problems. However, study of word recognition has several potential advantages over the study of some other types of pattern recognition. Language is highly organized. Thus organizational aspects such as phonology and frequency characteristics of a language can provide useful information for the study of word recognition. Words are formed by using letters. We know certain facts about these arbitrary symbols used to represent letters. But knowledge of the physical characteristics of these forms provides insufficient information to explain why certain letters in specific combinations produce a psychological reality--a word. However, if factors such as letter frequency, letter frequency in specific positions in words, frequency of letter pairs or groups, and frequency of words are considered, it may be possible to use this information to get at some of the factors governing word recognition. Thus structural and organizational aspects of language provide a context in which to study word recognition that is not available for the study of other types of recognition. It is with consideration of these aspects of the problem that the present study was formulated.

This study stems indirectly from a recognition phenomenon which has been called the mantiness problem (Norman, 1969, p. 162-163). The reader of English encountering mantiness determines virtually immediately that it is not a member of his vocabulary. However, the item is constructed according to orthographic and phonemic rules which eliminates the possibility that violation of rules of construction is what serves to alert the reader. And it is unlikely that in such a short time he could make an exhaustive search of all the representations of words in his memory to discover that mantiness is not in his vocabulary. The purpose of this study was to try to ascertain some of the factors which allow the extremely rapid memory search.

Models and Relevant Research

The present study was directed at the specific problem of word recognition rather than at an attempt to integrate aspects of this problem into a more general model of memory. For this reason, the only models and research discussed will be those which bear directly on the problem of word recognition.

Consideration of possible models of word recognition yields two basic types, those which involve search or scanning of memory, and those which assume content-addressable storage. The difference between the two types of models representing the extreme of each can be illustrated by the following examples. The task in the pure search system is analogous to the task of a person who is handed a deck of playing cards that have been shuffled, and is told to find the Ace-of-Hearts. He must look through the cards individually until the proper card is found. On the other hand, the content-addressable storage system is epitomized by the electronic computer. The operator commands the machine to read out the contents of storage register 27859367, and instantly he receives the information. Of course, to get the proper information, he must have the correct address. There are also less extreme versions of each type of theory because each version has disadvantages.

The search model requires some kind of recoding of the external form, and subsequent search through memory until the item is found. Time requirements apparent in the mantiness problem indicate that the pure search model is inadequate. The other possibility is some sort of content-addressable organization of memory. This type of system is very fast and efficient requiring only time to abstract the "address" of the sought-after material from the stimulus input. Norman (1969) has noted that there are several possible variations of a content-addressable organization of memory which may involve no search process, or which may have some component(s) of search. Since the pure search model is inadequate alone, it is likely that a potentially more fruitful approach to the problem of word recognition will involve some variation of a content-addressable type model.

An example of a content-addressable type model of word recognition by Morton (1969) is based on a hypothetical memory unit called a logogen. According to Morton,

> The logogen is a device which accepts information from the sensory analysis mechanisms concerning the properties of linguistic stimuli and from context-producing mechanisms. When the logogen has accumulated more than a certain amount of information, a response . . . is made available. [Here, a response is a single word, and an available response is a word implicitly available for verbalization.] Each logogen is in effect defined by

> the information it can accept and by the response it makes available. Relevant information can be described as the set of attributes . . . semantic, visual, and acoustic sets . . . incoming information has only a numerical effect upon any logogen which merely counts the number of members of its defining set which occur, without regard to their origin. When the count rises above a threshold value, the corresponding response is made available (pp. 165-166).

The most important aspect of this model is the fact that input is accepted directly by the "storage location," the logogen. In this sense, the model is a direct-access, content-addressable model. But it is important to note the assumption that input from one word may raise the mean count in several logogens due to the fact that some words have common features. Another important assumption is that over a long time interval, the logogen counts decay to some minimum "mean" value analogous to a baseline value, unless new input is accepted by the logogen.

The important features of the model involve factors that change the threshold and/or the mean count of the logogen. Context (c) is assumed to raise the mean count of a logogen. This has the net effect of moving the total count of the logogen toward the critical value. Stimulus (s) is the set of attributes associated with a particular word. Input of these attributes raises the count of the logogens sensitive to the particular attributes. However, the stimulus effect is momentary. This assumption is required since many words have similar construction. Stimulus properties of these items would raise the count in several logogens simultaneously. If the effect of stimulus did not disappear rapidly (Morton assumes decay is complete in less than one second) the mean count of many logogens would remain high and inappropriate responses would result from input of attributes which by chance would cause the logogen to fire. Thus Morton assumes that the effect of stimulus is transient, and the effect of context is self-sustaining. The combined effect of stimulus and context is assumed to be momentarily additive such that the count in the appropriate logogen is raised by (c + s).

The effect of word frequency on a logogen is quite complex. This effect is on the threshold of a logogen, rather than mean count. Both the effect of word frequency (frequency of the type estimated by the Thorndike-Lorge 1944 count) and the momentary effect of repetition of a word are due to the same hypothesized property of a logogen. When a word is presented, the threshold of the appropriate logogen is assumed to be lowered drastically--thus the same word repeated immediately would require much less input to fire the logogen. However, within a matter of minutes after firing, threshold moves back up to a point just below the original value. The hypothesized reason that high frequency words (Thorndike-Lorge frequency) have lower thresholds relative to low frequency words results from the cumulative effect of each single presentation of a word--each time a word fires a logogen, the net effect (on threshold) is a minutely lowered threshold. Therefore more presentations of a word result in higher frequency of experience, and lower threshold according to the logogen model.

The logogen model takes into account the effect of context, specific stimuli, frequency, and repetition of the same item, and predicts their effect on the logogen. This enables predictions about the probability of a response given input conditions, and the model explains many empirical results.

Another model (representative of a class of models) is a type proposed by (among others) Katz and Fodor (1963) called the dictionary model. This model is a variation of the content-addressable type. The model proposes that memory is the functional equivalent of a dictionary with various types of information "punched" on data "cards." These cards can be checked for certain features. Thus if a search for mantiness were being carried out, all words not belonging to the set "words with nine letters" would be ignored. The intersection of various other subsets of words would further reduce the search until the possible sets had been scanned, and mantiness was determined to be in the set "not in the vocabulary." Many aspects of this model are attractive. For example, it seems adequate to explain the "tip of the tongue" (TOT) phenomenon investigated by Brown and McNeill (1966). They found that when an individual is in the TOT state, certain "features" of the sought-after item are often available to the subject. These features include initial and/or terminal phoneme, stress pattern and/or number of syllables, and the general meaning of the word. The subject was instructed to report information of this type when in the TOT state. Brown and McNeill then used this information to fabricate a set of words, all of which had several or all of the features reported by the subject. These words were then presented to the individual, and the experimenters found that in this situation, the correct word can often be chosen. This leads to the hypothesis that the organization of memory for words might be based on "features" of the item. A possible approach to the problem of word recognition might be via a technique which is the approximate opposite of the TOT technique--present items with precisely controlled features, and see how long it takes for recognition. Since "features" can be specified and quantified in terms of frequency, etc., they can be manipulated for the purpose of determining which features are important for word recognition. For development of a model of word recognition, this approach seems potentially the most useful.

Another model, this one also involving search of a subset which has been "addressed" by some feature(s) of a word is based on the results of Rubenstein, Garfield and Millikan (1970). They found that in classification of items as words or not, response latencies were shorter for words than for nonsense items, shorter for high frequency words than for low frequency words, and shorter for homographs (words with more than one meaning) than for nonhomographs. Their proposed model involves four processes: (1) Quantization, the division of the word into segments for recoding; (2) marking, an operation using the output of quantization to mark a subset of the internal lexicon as the potential search area; (3) comparison of the outputs of quantization with each of the items in the marked subset; (4) selection of the item which S decides is a match of the input item. An important aspect of this model is the parallel function of quantization and the other processes after the initial output which starts the search; quantization continues providing new outputs for marking until the search is completed. Also, highest frequency items are assumed to be marked first. If a match is not found after search of these items, lower frequency items are subsequently marked until a match is found.

The interpretation of their results by Rubenstein et al. (1970) in light of the model is straightforward. The fact that responses to words are faster than those to nonsense is assumed to be a result of the nonsense items requiring an exhaustive search of the marked subset. The search for a word (when found) would always terminate prior to checking the complete subset.

If it is assumed that high frequency words are marked first and/or searched first, the differences attributable to frequency can be explained. The fact that homographs are recognized faster than nonhomographs is based on the assumption that with more lexical entries (homographs), the probability of finding a match in a given period of time is greater than for nonhomographs, which have only one lexical entry. A subsequent study by Rubenstein, Lewis, and Rubenstein (1971a) confirmed a predicted difference between homographs based on the relative frequencies of the meanings. They felt that if a word had two meanings, e.g., fork, "implement with prongs," and "division into branches," with relative frequencies of 95% and 5% respectively, the rare meaning might not be represented in the internal lexicon. Contrast this to a homograph, e.g., bulb, "electric light," and "part of a plant," where relative frequencies are 63% and 37%. The latter case would have two representations in memory--one for each meaning. Thus, homographic words with equiprobable meanings should have shorter latencies than the others with nonequiprobable meanings, which was confirmed by the results.

In addition to frequency and homography as language variables which have been used to study word recognition, there have been attempts to use semantic aspects of words, such as the abstract/ concrete dimension. A study by Winnick and Kressel (1965) compared abstract and concrete words (high and low frequency) in a tachistoscopic recognition task. The overall mean of abstract and concrete words was identical. However, a frequency effect was demonstrated.

The threshold of the low frequency items in both the abstract and concrete categories was significantly greater than the high frequency items. If semantic content of words is involved in the recognition process, there should be effects associated with semantic aspects of words, in this case, their degree of abstractness. The results of Winnick and Kressel (1965) rule out an explanation based only on semantic content in favor of some sort of word frequency hypothesis. And, although tachistoscopic recognition is a different task than word-nonword classification, it is reasonable to expect that some aspects of the two tasks are similar or overlapping, and consideration of word frequency is appropriate for a word recognition model.

The studies cited thusfar include the results that (1) as word frequency increases, classification latency decreases; (2) more meanings of a word produce a shorter classification latency (more representations of an item facilitate faster retrieval); (3) word frequency is more important than abstract/concreteness for tachistoscopic recognition; and (4) words require less time for classification than nonwords. Considered together, these results argue for at least two general characteristics of word recognition. First, there is a search of a "subset" of memory, where the "subset" contains only part of all the items in memory. Second, the order of search of these items is somehow related to some sort of frequency associated with each word. The means of establishing the boundary of the subset is unspecified, except by an implication of the Brown and McNeill (1966) study, that is, some kind of feature extraction. One possibility for indexing the correct subset involves the encoding of segmental information about a word. Support for this idea is provided by a study by Horowitz, White, and Atwood (1968). In their study, the S was presented a list of words to memorize for later recall. The words were then divided into segments and one of these segments was presented as an aid to recall of each item. They found that retrieval of an item occurred more often when the S was supplied with the first or last segment of a word rather than when the medial segment was provided as a cue for recall. This leads to the conclusion that there is important information in these segments of words for identification or retrieval of a word. Although the task was not classification of words and nonwords, the fact the segmental information produced a differential effect implies that segmental information may be used to index a word in memory.

Previous discussion has shown that the logogen model is compatible with the empirical results in most respects. However, this model provides no mechanism for the correct classification of nonwords. If no logogen fires, it could be assumed that the input was not a word. However, the model provides no mechanism which differentiates between the case A, when input is from a stimulus which is a word but is merely insufficient to fire an logogen, and the case B, when input is from a nonword and has by definition insufficient input attributes to fire a logogen. The logogen model can produce only responses to words and is able to account for neither correct classification of nonwords, nor for the empirical results that words are classified faster than legal nonwords.

Contrast the logogen model with a system in which a subset of memory may be indexed by partial information from the presented item. The search could be carried out on the subset which would contain the item if it were, in fact, stored. When the search of this subset is finished without success, it is apparent that the item in question is not stored. There would thus be a mechanism for producing negative responses, and the time requirements would match empirical results for words and nonwords. Such a system has characteristics of both a search model and a content-addressable model. Integration of properties of both models into a new content-addressable type model can be illustrated by example. A file clerk looking for a folder labeled "recognition" can quickly reject as irrelevant all file cabinets except the one labeled "R." With the search narrowed, she proceeds to search through a specific subset of folders until the correct one is found. Unsuccessful exhaustive search of the file indicates that particular folder is not in the file. Given the results discussed above, this type of model, such as the one proposed by Rubenstein et al. (1970) seems to be the most promising.

Immediate Antecedents of the Study

A study by Stanners, Forbach, and Headley (1971) investigated the effect of letter frequency in the mantiness problem. They used trigrams constructed by varying the frequency of the initial and terminal consonant. The reference for construction of the trigrams was a set of norms formulated by Venezky (1962). A somewhat more detailed description of the norms may be found in Stanners (1970). For the Stanners et al. (1971) study, the relevant data were the frequencies of letters as initial, and as terminal consonants. These frequency tabulations also took into account pronounciation of letters in different contexts. They found that within the categories consonant-vowel-consonant (CVC) word and CVC nonword, classification latency was a function of the frequency of the initial and terminal letters of the item. For consonant-consonant-consonant (CCC) nonwords, there was no effect of letter frequency. These effects seemed to be due to component frequency (the individual letter frequency). But a *post hoc* checking of the items in the Mayzner and Tresselt (1965) norms showed that frequency of the item as a unit *covaries* with component (letter) frequency. The reason for this is that trigrams (word and nonword) occur in English as complete *units*, because of their occurrence as words or parts of words. Thus "cat" is experienced as a word alone, but also as part of "category," "scat," "concatenate," "catatonic," etc. Likewise, even nonwords, for example "ter," "med," "sed," have been experienced as a unit by the reader of English. Thus for CVCs, the covariance of the frequency of components, and frequency as a unit cannot be untangled, and the results of Stanners el al. (1971) cannot be unambiguously interpreted as due to either component frequency or unit frequency. The need for determining whether the effects were due to component frequency or unit frequency was felt to be important since components (in this case, letters) *could* be the basis for indexing the memory subset to be searched. Frequency effects unambiguously attributed to component frequency would support this notion. The present study was done to provide information on which to base a choice between component frequency and unit frequency. Examination of the Mayzner and Tresselt (1965) norms revealed that five-letter nonwords of the form CCVCC virtually do not occur as complete units in English, and can be assigned a *unit* frequency of zero. Thus the use of five-letter items removes the confounding of the covariation of component frequency with unit frequency since the only frequency which can be attributed to CCVCC nonwords is frequency of components, that is, frequency of the initial and terminal pair of consonants.

The goal of this research was to increase understanding of what information in a letter string a reader of English used to restrict and perform a search for the representation of the word in memory. The study by Stanners et al. (1971) suggested the possibility that the S "analyzes" (breaks down into components) an item to organize and to carry out the search. If this is true, and the S does, in fact, analyze the item in question, the unit of analysis might correspond to the aspect of a word which indexes memory. Is it the consonant cluster that is this unit of analysis? Or is the item treated as a unit? The fact that no frequency effects were found for CCCs in their study led Stanners et al. to consider the possibility that the item was treated as a unit. However, frequency effects attributable to initial and terminal consonant clusters of CCCCCs or CCVCCs would provide strong support for the notion that the item is analyzed, and search is based on the components of an item.

Compared across categories, CCVCC word (WORDs), CCVCC nonword (CCVCCs), and CCCCCs, the items in the present study were constructed with identical consonant clusters. Therefore, the only difference between categories is the medial letter. Three analogous categories of trigrams produced results showing that CVC nonwords require the longest classification latencies, CCCs the shortest latencies, and CVC words intermediate latencies in the Stanners et al. (1971) study. Thus it was expected that the latencies of the three categories of five-letter materials used in the present study would be patterned similarly to these results. If component frequency is associated with the "analysis" of an item for search, frequency effects within categories would be expected. The occurrence of frequency effects within the category CCCCC would provide especially strong evidence that an analysis of the item takes place.

Included in the study were some filler items chosen to compare the effects on latency of frequency of words as units when component frequency was held constant. Rubenstein et al. (1971a) found that low frequency words require a longer classification latency than high frequency words. However, Rubenstein et al. (1971a) did not consider the possible differential effects of component frequency. To make a meaningful comparison of words based on unit frequency, component frequency should be held constant. If differences were found in the filler items chosen on this basis, they could be attributed solely to word frequency.

CHAPTER II

METHODOLOGY

Subjects

There were 23 volunteers from Introductory Psychology classes who served as subjects (Ss). Each S received a specified number of bonus points to be added to his final grade average.

To avoid the possibility of very widely deviant scores which could distort the data, minimum performance standards were set. The S's performance on the practice trials and first section of the experiment was compared to these criteria. If the S performed so that a total of 20% of his scores on the practice trials were considered errors, he was dismissed. Incorrect classification of items, or latencies above 2,000 milliseconds were scored as errors. Also, the first one third of the experimental trials were monitored closely for errors, and a total of 20% error scores resulted in the performance being classed as substandard. If performance was substandard, the S was thanked for his cooperation and dismissed after he was given credit for participation.

Three of the total number of Ss did not meet the minimum requirements and were dismissed. The results are therefore based on 20 Ss. Of these, eight were males, and twelve were females.

Apparatus

The central part of the apparatus was an eight channel Lafayette timer (Bank Timer 1431A) which controlled the timing sequence and the other equipment. Stimulus materials were presented by a Kodak Carousal projector with a five inch f3.5 lens which was equipped with a Lafayette I-24 solenoid operated shutter (power supply--Lafayette Tachistoscope VS1-E). Timing of latencies was done by a Hunter Klockounter which measured latencies to the nearest millisecond. The timing equipment was in a room adjoining the subject's room, the dividing wall fitted with a one-way mirror of the dimensions 50 x 70 cm., so that S could be observed while doing the task.

The subject's room was approximately 2 x 3 m., and was painted black to minimize ambient light reflection. The S was seated at a small table at a distance of approximately 50 cm. from a 18 x 13.5 cm. Plexiglas screen onto which the materials were backprojected to a height of approximately 1.8 cm. S was given a small thumb switch which initiated each trial when he was ready. The word-nonword responses were given via a lightly sprung toggle-type switch (normally open) in a circuit with a small latching relay which controlled the recording of the latencies. The switches (for right- or left-handed Ss) were mounted into the table top in such a way that Ss forearm and elbow rested comfortably on the table.

Materials

Three categories of materials were constructed, consonant-consonant-vowel-consonant-consonant words (WORDs), consonant-consonant-vowel-consonant-consonant nonwords (CCVCCs), and consonant-consonant-

consonant-consonant-consonant nonwords (CCCCCs). The basic reference for the materials was the Venezky (1962) norms. For the purposes of the present study, the data of interest were the frequencies of initial and terminal pairs of consonants, or consonant cluster (CC), based on their occurrence as specific sound units in words. In the norms, a given CC might have several frequency tabulations associated with it due to several pronounciations, e.g., "GH" pronounced "E," "G," or "K." An item was selected as being a high frequency CC only if it had a single pronounciation of relatively high frequency. A low frequency CC was one that had no pronounciations of relatively high frequency. An interesting fact about the English language was noted while constructing these items--a given CC may have multiple pronounciations, but only one will have a relatively high frequency, and the other pronounciations of the same CC will have very low frequency of occurrence, i.e., less than five in the Venezky sample.

A sample of high frequency initial CCs was chosen which ranged from 61-468 in the Venezky norms with a mean of 144.7, and median of 122.5; the corresponding range, mean, and median for the low frequency initial CCs were 1-39, 8.8, and 2.0. For the terminal CCs, the range of the high frequency sample was 31-729 with a mean of 159.4 and median 104.0. The range, mean, and median for the sample of low frequency terminal CCs were 1-18, 6.2, and 5.0, respectively.

The WORDs were organized into sets of four items defined by the combination of high (H) and low (L) frequency of initial and terminal CC. An example of a set so defined in order HH, HL, LH, LL is: THING, THUMB, WRONG, PSALM. Associated with each set of words was a set of CCVCCs. They were constructed by deleting the vowel in the WORD counterpart, and replacing it with another vowel that did not form a word. Thus for the example set above, the CCVCCs are: THENG, THAMB, WRENG, PSOLM.

The CCCCCs were constructed in a similar manner. The vowel in the WORD counterpart was replaced by a consonant. Several constraints were placed on the construction of the CCCCCs: (1) Only the consonants B, F, K, L, M, N, P, R, S, T, and Z were used in the medial position, since the other consonants were judged as possible to mistake for vowels; (2) no consonant was used twice until each had been used once; (3) a given consonant was not used if it preceded or followed in the alphabet the consonant immediately lateral in the CCCCC (thus TRSTH would not be allowed since S follows R in the alphabet and/or precedes T); (4) a given consonant was not used if it would cause the occurrence of a pair of the same letters in a medial position (thus neither TRRST nor TRSST were allowed; (5) a consonant which would form a recognizable acronyn was not used.

A total of 15 items were generated for each of the four sets of the main categories (WORDs, CCVCCs, and CCCCCs), with the exception of the LL WORDs. By definition the CCs used are low frequency--so low that only two LL items could be generated. Since there were no WORD exemplars for 13 of the CCVCCs and CCCCCs, these items were generated by random pairing of the low initial and low terminal CCs, subject to the constraints.

A total of 73 filler items were used to balance the number of words and nonwords at 120 each. All filler items were of the general form CCVCC, and were either HH or HL items. Within the filler words, two general subsets of items were formed for a secondary analysis:

(1) A set of 44 HH items (FILLER), half of which had low, and half high frequency in the Thorndike-Lorge (1944) count; (2) a set of 28 HL plural noun items (PLURAL), half of which had low, and half high frequency in the Thorndike-Lorge count. All the materials were typed in upper case with an IBM Executive typewriter, reproduced onto transparencies by the diazochrome method, and mounted in 35mm slide holders.

Experimental Design

The variables, frequency of initial CC, frequency of terminal CC, and category (WORD, CCVCC, and CCCCC) were manipulated completely within subjects. Direction of switch movement was balanced between subjects. Stimulus materials were randomly ordered for presentation to each S.

Procedure

S was seated in the subject room at the table and listened to a set of recorded instructions (see Appendix A), while the experimenter (E) pointed out the necessary switches. When ready, S initiated the trial by a handswitch held in his nonpreferred hand. A trial began with a buzzer which signalled the S to attend to the screen and to hold between thumb and forefinger of his preferred hand a lightly sprung toggle-type switch. Following the buzzer by two seconds, an item was projected on the screen. The S was instructed to indicate by the direction of the switch movement whether the item as a complete unit was or was not in his vocabulary. A small sign next to the switch specified the direction of the movement which was held constant for a given S. The Klockounter started with the presentation of the item, and stopped with S's response. The instructions discouraged responding positively to idiosyncratic nicknames, and stressed both speed and accuracy. The item remained on the screen until the S made his response. Except for a minimum of three seconds programmed by the apparatus, S controlled the intertrial-interval. This was to insure full concentration when he initiated the next trial, since he could pause and rest at the end of any trial if he so desired.

The experimental trials were preceded by 40 practice trials with material similar to the experimental materials. Each S attended one experimental session, which with the practice trials and instructions lasted 40-50 minutes.

Scoring of Data

An individual score in the data analysis was the antilogarithm of the mean of the log latencies for a given S in a given subcondition. The purpose of the transformation was to reduce the effect of one or a few highly deviant latencies on a score. Only "correct" (in agreement with the classification system of words and nonwords) responses were used in the transformed scores. No score was based on less than 10 of 15 possible latencies, and over 95% were based on 12 or more latencies. Thus after the transformations, the data for each S was a single score for each subcondition HH, HL, LH, and LL in each category of material, with the exception due to lack of LL WORDs. Subsequent analyses used these transformed scores.

The data were also scored by item to provide an estimate of the average latency of a particular item. This was done by recording the latency of each S for each item. This provided a maximum of 20 latencies, less if any Ss had made errors. These scores were then transformed in the same manner as were the category subcondition latencies. Output of this transformation produced an estimate of required processing time for each of the study items. (These items and latencies are given in Appendix B.)

CHAPTER III

RESULTS

Analysis of Nonwords

The main analysis was a three-factor analysis of variance on the data from the nonwords. The factors were initial consonant cluster (I), terminal consonant cluster (T), and category (C), either CCVCC or CCCCC. A summary of the analysis of variance is presented in Table I. The main effects were all highly significant ($p < .001$). The two-factor interactions I x C, T x C, and I x T were also significant ($p < .001$) as was the T x I x C interaction ($p < .025$).

A two-factor analysis of variance was employed to assess the effects within the categories CCVCC and CCCCC. For both the CCVCCs and CCCCCs, the main effects I and T, and the I x T interaction were significant well beyond the .001 level. Summaries of these analyses are presented in Table II.

Within the categories CCVCC and CCCCC, the means of each sub-condition were tested for differences with a Neuman-Keuls test. For CCCCCs, LL was significantly different ($p < .01$) from HH, HL, and LH, which were not statistically different from each other. For CCVCCs, LL was significantly different from each of the other three means ($p < .01$). HH was statistically different ($p < .01$) from HL, but not different from LH. HL and LH were not statistically different. A summary of all the Newman-Keuls tests performed is presented in

TABLE I

AOV OF LOG TRANSFORMED MEAN CLASSIFICATION LATENCY FOR CCVCCS AND CCCCCS

Source	df	SS	MS	F[1]
Total	159	4177245.000		
Subjects (S)	19	2013326.000	105964.500	
Terminal CC (T)	1	251064.000	251064.000	64.466***
ST	19	73996.375	3894.546	
Initial CC (I)	1	123543.188	123543.188	34.966***
SI	19	67131.500	3533.237	
Category (C)	1	1045228.750	1045228.750	82.367***
SC	19	241109.063	12689.949	
TI	1	54169.598	54169.598	24.521***
STI	19	41972.957	2209.103	
TC	1	69472.188	69472.188	17.083***
STC	19	77269.938	4066.839	
IC	1	44823.000	44823.000	34.001***
SIC	19	25047.098	1318.268	
TIC	1	11971.598	11971.598	6.127*
STIC	19	37125.750	1953.987	

[1]Individual error terms (indented) were used in each F-ratio. Significance levels for all tables are represented by the following: *** = $p < .001$; ** = $p < .01$; * = $p < .05$.

TABLE II

AOV OF LOG TRANSFORMED MEAN CLASSIFICATION LATENCY WITHIN NONWORD CATEGORIES

Source	df	SS	MS	F
CCVCCs				
Total	79	2433861.000		
Subjects (S)	19	1645671.000	86614.250	
Terminal CC (T)	1	292336.188	292336.188	40.685***
ST	19	136520.500	7185.289	
Initial CC (I)	1	158598.000	158598.000	40.837***
SI	19	73790.875	3883.730	
TI	1	58536.199	58536.199	16.258***
STI	19	68410.500	3600.552	
CCCCCs				
Total	79	698162.813		
Subjects (S)	19	608766.500		
Terminal CC (T)	1	28200.047	28200.047	36.334***
ST	19	14746.418	776.127	
Initial CC (I)	1	9768.199	9768.199	10.093**
SI	19	18388.297	967.805	
TI	1	7605.000	7605.000	13.519**
STI	19	10688.500	562.552	

Table III. The subcondition means for the analysis of variance are presented graphically in Figure 1. Actual values of these means are in Appendix C.

Analysis of Words

The three subconditions available for WORDs were HH, HL, and LH. One-way analysis of variance was performed on the WORD data. (Each subcondition was treated as a group rather than a product of two frequency manipulations.) The main effect was significant ($p < .001$). To compare the three means, a Neuman-Keuls test was used. The test indicated HH was statistically different from HL and LH, which were not statistically different. These three means and their relationship to the nonwords are shown in Figure 1.

The filler items, FILLER and PLURAL, were analyzed separately. Each set was comprised of half high and half low Thorndike-Lorge frequency items. One-way analysis of variance showed frequency to be a significant effect for both FILLER and PLURAL items ($p < .001$). Means of these items are presented with the means of the subconditions of WORDs in Figure 2.

Comparison of Words and Nonwords

Since the analysis of words and nonwords were done by separate analysis of variance, it was deemed appropriate to make some comparisons between categories. The relevant comparisons are HH (CCCCC)--HH (WORD)--HH (CCVCC), HL (CCCCC)--HL (WORD)--HL (CCVCC), LH (CCCCC)--LH (WORD)--LH (CCVCC), and LL (CCCCC)--LL (CCVCC). These comparisons were tested by the Neuman-Keuls test. With one exception, all means

TABLE III

A SUMMARY OF ALL NEUMAN-KEULS TESTS FOR DIFFERENCES AMONG MEANS

Test Category or Subcondition	Ordered Means				Comparisons
	1	2	3	4	
LL	CCCCC (586)	CCVCC (656)	-----	-----	2-1**
HH	WORDs (641)	CCCCC (646)	CCVCC (866)	-----	3-1** 3-2** 2-1 N.S.
HL	CCCCC (628)	WORDs (723)	CCVCC (799)	-----	3-1** 3-2** 2-1**
LH	CCCCC (643)	WORDs (737)	CCVCC (831)	-----	3-1** 3-2** 2-1**
WORDs	HH (641)	HL (723)	LH (737)	--	3-1* 3-2 N.S. 2-1*
CCVCC	LL (656)	HL (799)	LH (831)	HH (866)	4-1** 4-2** 4-3 N.S. 3-1** 3-2 N.S. 2-1**
CCCCC	LL (586)	HL (628)	LH (643)	HH (646)	4-1** 4-2 N.S. 4-3 N.S. 3-1** 3-2 N.S. 2-1**

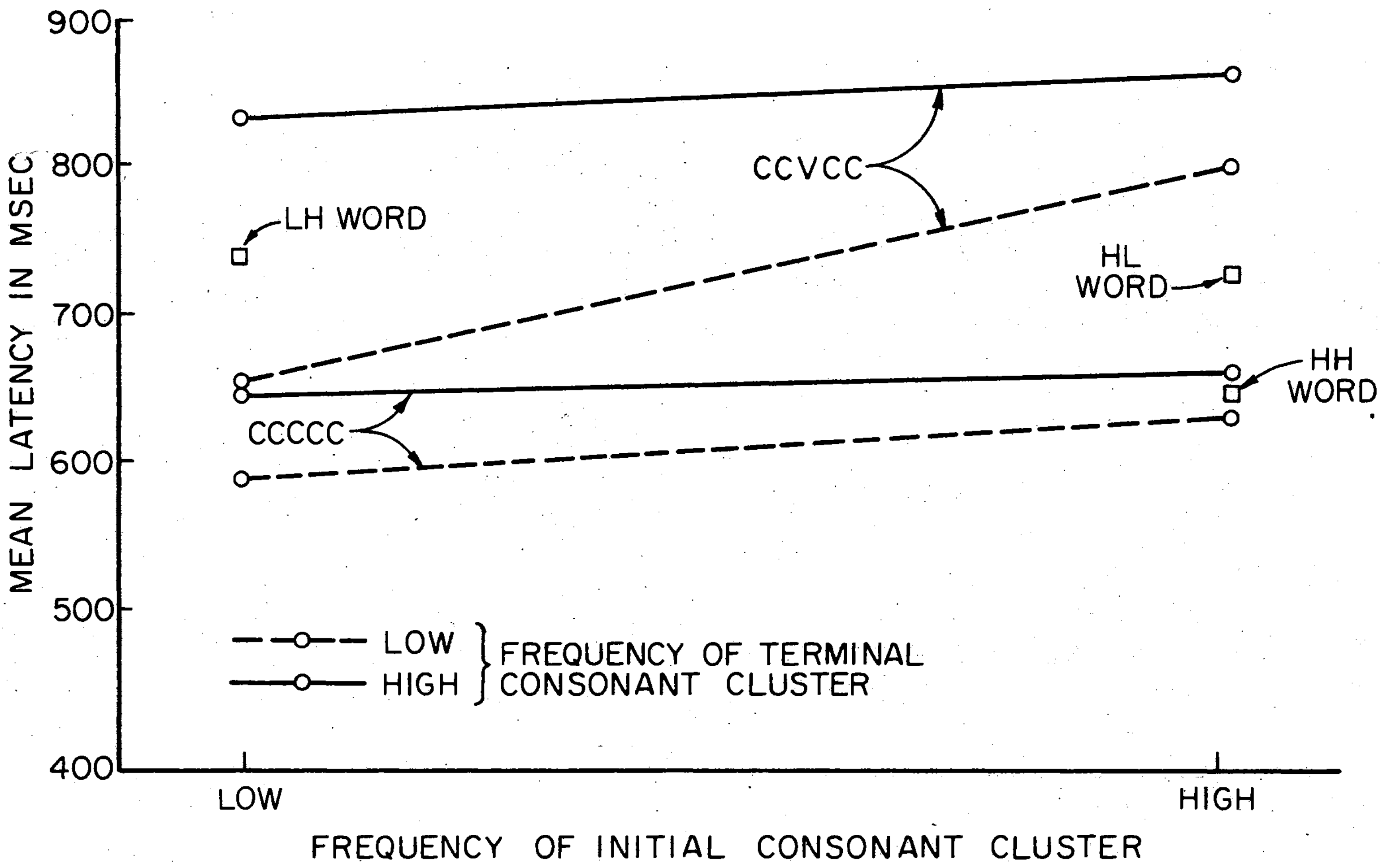

Figure 1. Subcondition mean classification latency for three categories of experimental material. Each value is a log transformed score based on twenty Ss.

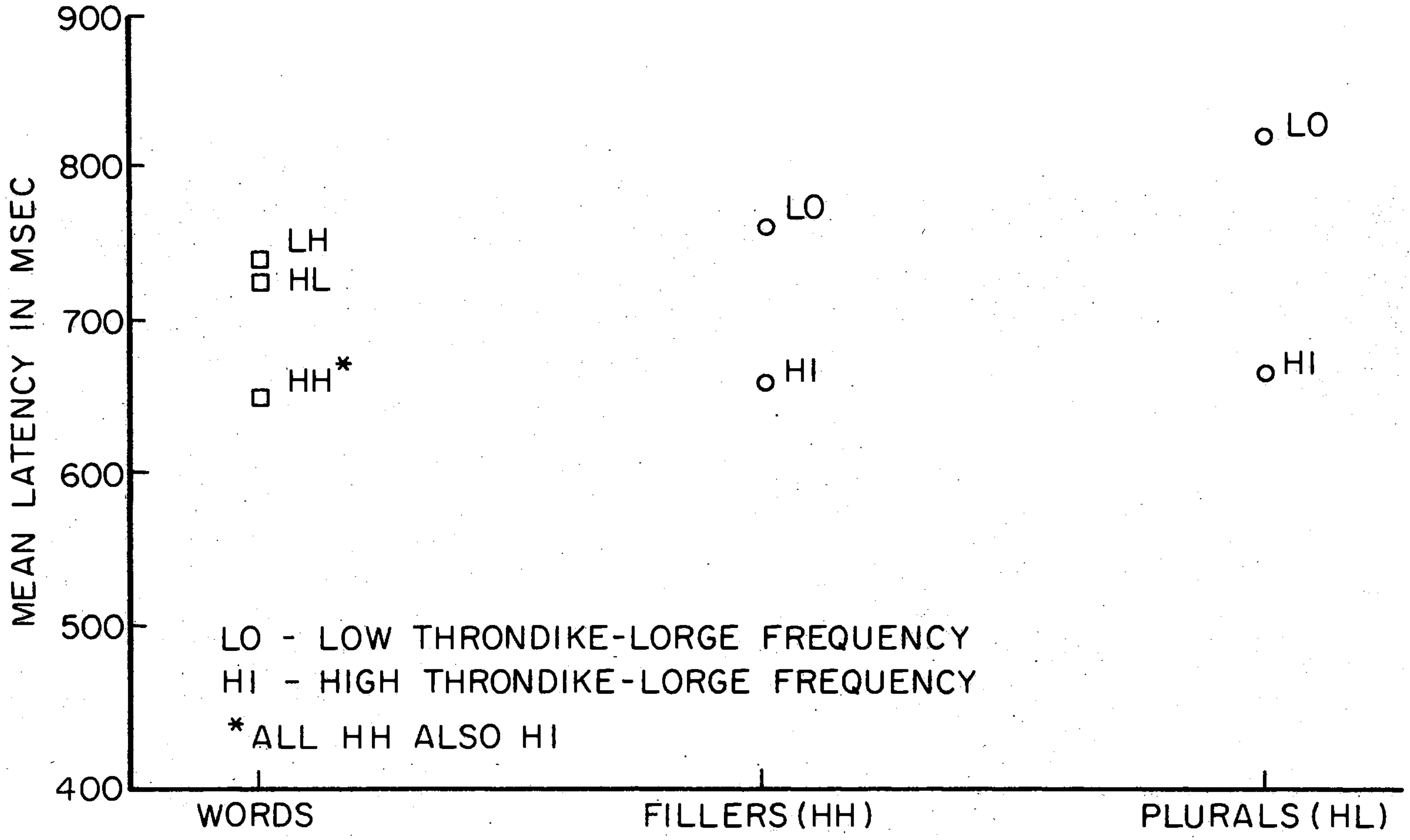

Figure 2. Comparison of means of filler words with WORDs. This demonstrates the effect of word frequency on classification latency when component frequency is held constant.

in the comparisons made were statistically different beyond the .01 level. The exception was the mean for HH (WORD) and HH (CCCCC). A summary of these comparisons is presented in Table III.

Error Data

Responses classed as errors were CCVCCs or CCCCCs which S had accepted as in his vocabulary (saying essentially, that these items were words), or WORDs which he had rejected. The error rate for WORDs was 9.7%, for CCVCCs was 6%, and for CCCCCs was 0.4%. [Generally for WORDs, more errors occurred on items with low frequency CCs; and for CCVCCs, more errors occurred on items with high frequency CCs.] Distribution of errors in each category and subcondition is presented in Table IV. Mean latency of errors in each category and subcondition is presented in Table V.

TABLE IV

TOTAL ERRORS IN EACH CATEGORY AND SUBCONDITION

Category	Subcondition			
	HH	HL	LH	LL
WORDs	5	39	35	12
CCVCCs	31	15	28	5
CCCCCs	1	1	3	0

TABLE V

MEAN LOG TRANSFORMED LATENCY OF ERRORS
IN EACH CATEGORY AND SUBCONDITION

Category	Subcondition			
	HH	HL	LH	LL
WORDs	678	854	781	632
CCVCCs	752	712	654	970
CCCCCs	622	592	550	*

* No errors

CHAPTER IV

DISCUSSION

The design and method of this study parallels the Stanners et al. (1971) study, but the two studies used different types of materials, so a comparison of results may be informative. Figure 3 presents the results of Stanners et al. (1971). The similarities of the two studies include the equivalence of pattern of the category latencies, and a simlilar pattern within categories. The main difference is the effect of frequency manipulation on the CCCCCs. Also, there was an enhanced interaction effect within the CCVCCs.

The results of this study strongly support the hypothesis that the item is "analyzed" because they clearly show a component frequency effect. Since the items used in the study have comparable unit frequency--essentially zero--the only source of variation in frequency is CC frequency. The effect of frequency variation for CCCCCs and CCVCCs requires that the <u>S</u> used this information sometime during the course of classification of the item.

A general model can be proposed to account for most of the results. The Stanners et al. (1971) paper proposed that the first stage of identification is an editing process which evaluates the phonological lawfulness of an item. It was assumed that a CCC would be rejected as unlawful before any search takes place. Thus the latencies for CCCs should be shorter than for items requiring search.

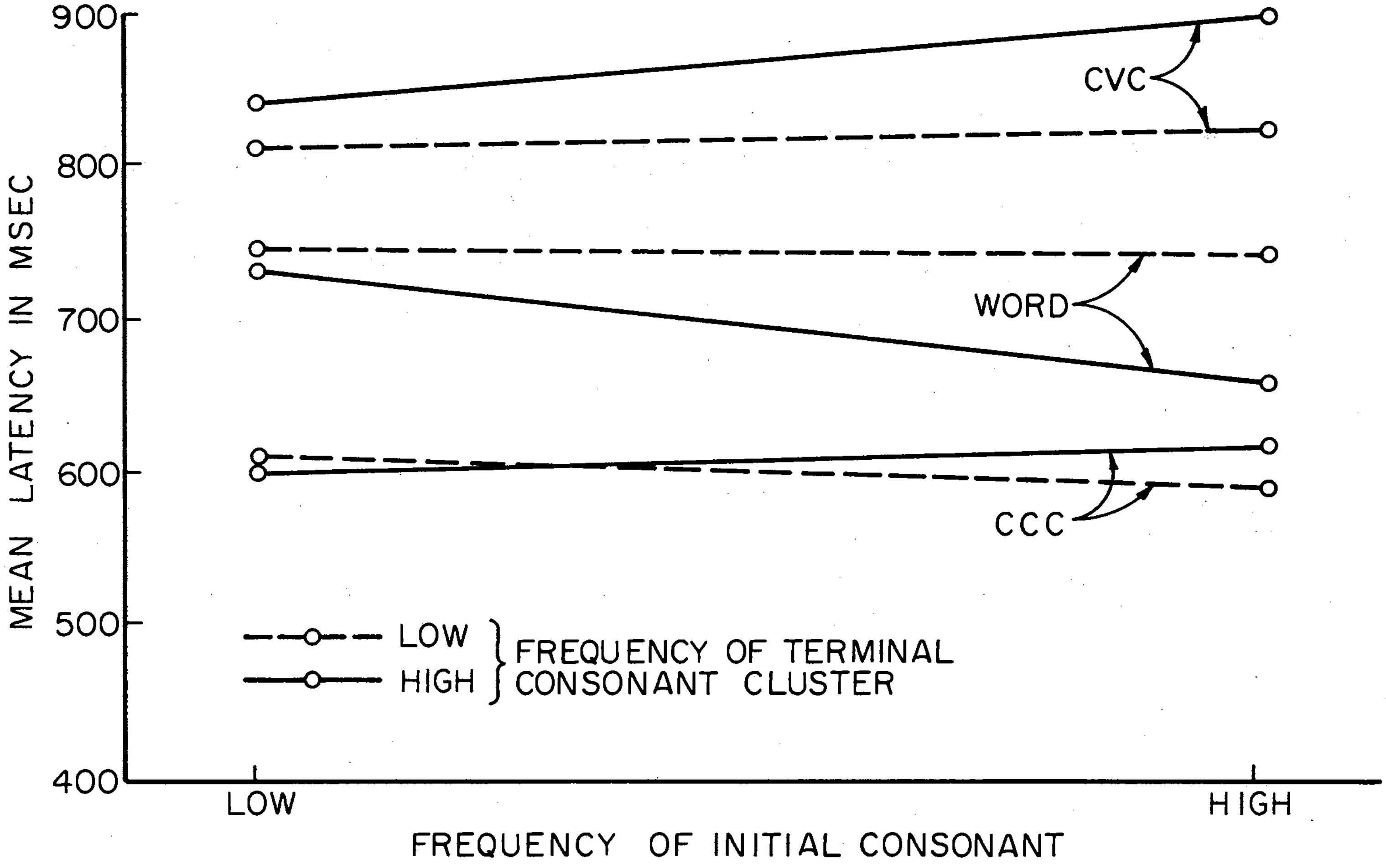

Figure 3. Subcondition mean classification latency for three categories of experimental material from Stanners et al. (1971). Each value is a log transformed score based on twenty Ss.

The results showed that latencies for CCCs were shorter than the other items. More important, however, there were no effects of component (letter) frequency on CCCs. This was interpreted as an indication that CCCs were rejected before the frequency effects, which are apparent for the Words and CVCs, could effect their latency. The fact that the error rate for CVCs and Words was approximately 80 times higher than that for CCCs also indicated a different type of process.

Some of the same arguments can be made for the results of the present study. First, the latencies were consistently lower for CCCCCs than for WORDs or CCVCCs. Second, the error rate for CCVCCs and WORDs was about 20 times higher than that for CCCCCs. However, the present study found that frequency does affect the latencies of CCCCCs. These effects would suggest a process which makes use of information from the CCs before the item is encoded as a unit for search. A process similar to that called "marking" (functional circumscription of the search area) by Rubenstein et al. (1970) can account for the observed results. To the extent that the words used in the Venezky (1962) norms (drawn not from running test, but sampled from the Thorndike Century Senior Dictionary) frequency count have representations in memory based on the frequency of occurrence, a high frequency CC should require that a larger subset of the internal representations of words should be designated for search. Thus both marking time and search time should be positively correlated with frequency. According to the present conception, the relationship among the means of the CCCCCs would reflect the marking time differences. The longest latencies among the CCCCCs would indicate the item requiring more extensive marking, while the shortest latencies would indicate the subsets requiring less marking. The relationship among the means of CCVCCs would reflect the marking time plus coding and search time. Since the construction of the CCVCCs and CCCCCs are identical except for the medial letter, and we have assumed that marking for both types of items is based on the same CCs, any difference in latency between the two types of items should be due to operations after marking. CCVCCs require marking, coding as a unit for search, and exhaustive search of a marked subset. CCCCCs require marking, and attempted coding for search. Thus a prediction can be made that CCVCC subcondition means should have longer latencies than comparable CCCCC subcondition means. As expected, in each case, the CCVCC mean is well above the corresponding CCCCC mean.

The explanation of differences in the means of the nonword items rests in part on the assumption that the marking operation uses information from the initial and terminal CC to determine the size of the search subset. Partial support for this assumption is provided by the results of Rubenstein, Lewis, and Rubenstein (1971b). They found that orthographically and phonemically illegal nonsense items produced different classification latencies when the pronounciability of the items was manipulated by changing the final pair of letters. The fact that the final CC of the item produced the effects lends credence to the interpretation of the results of the present study with respect to the information used in the marking operation.

The second part of the explanation of differences in the nonword means rests on the treatment of the items after marking is complete. The results of Rubenstein et al. (1971b) also showed that orthographically and phonemically legal nonsense words had a longer classification latency than either of two illegal types of nonsense words. These results and the results of the present study suggest that the item is phonologically encoded as a unit for search after marking. The attempted encoding of an illegal item results in immediate rejection (no search), while the encoding of a legal non-word leads to exhaustive search of a marked subset. This proposed explanation accounts for the substantially lower latencies of the CCCCCs, and also accounts for the results of Rubenstein et al. (1971b).

When the results for WORDs are examined, an apparent contradiction arises--the effect of component frequency is approximately opposite the effect for nonwords. There is no structural difference in the two types of items, and WORDs and CCVCCs are constructed from the same CCs. Whatever difference there is between the two sets results only from their status as words or nonwords. Why does frequency information seem to effect WORDs in an inverse manner from the effect on nonwords? A possible answer is based on the fact that words occur as units, and thus have a unit frequency in addition to component frequency which determines marking time. This unit frequency is approximated (and operationally defined) by the frequency tabulations in the Thorndike-Lorge (1944) count. A finding in support of the hypothesis that unit frequency effects classification latency is the correlation between mean latency (over 20 Ss) for each WORD and its unit frequency of -.65, $p < .001$. Another finding suggests that unit frequency effects latency of search. Among the FILLERs were a set of 13 pairs of words which had identical initial and terminal CCs but which were different in Thorndike-Lorge frequency, e.g., GLOSS and GLASS. The high-frequency group of these FILLERs had an average latency of 649 milliseconds while the mean of the low-frequency group is 765 milliseconds; $\underline{t}$ (12) = 4.20, $\underline{p} < .005$. This implies that the search through memory for the match to a study WORD is ordered, i.e., the most used items are toward the beginning of a subset, and are always searched first. Also, as Figure 2 indicates, the latencies for FILLERs (including all 46), and PLURALs lend further support to the notion that the search is ordered, since when component frequency is held constant (HH, and HL, respectively), change in unit frequency from high- to low-frequency results in marked classification latency differences.

This evidence for a nonrandom search within a marked subset suggests a testable implication--a set of WORDs with very high Thorndike-Lorge frequency should have a mean very close to the corresponding set of CCCCCs. Such a comparison is available because the HH WORDs all have a Thorndike-Lorge frequency of AA. Underlying the prediction is the assumption that for both HH WORDs and HH CCCCCs, the majority of classifcation time will be used for marking (both require the same amount since they use the same CCs) and search time for WORDs will be minimum since their high frequency places them at the beginning of the search subset. As indicated by Figure 1, the means are only five milliseconds different.

Comparison of the WORDs and CCVCCs indicates that search for WORDs is nonexhaustive, since latencies for WORDs are below those of CCVCCs. The error data supports the notion of nonexhaustive search for WORDs. The average latency for errors on the CCVCCs is less than that for correct CCVCC responses, and close to the mean for WORDs; CCVCC errors--718 milliseconds, correct WORDs--700 milliseconds. Presumably, S has (mistakenly) found a match before completing his search of the subset. The average latency of the errors for WORDs is about 100 milliseconds greater than that for correct responses and approximately the same as for correct responses on CCVCCs; WORD errors--799 milliseconds, correct CCVCCs--788 milliseconds. In this case a match is not found and the subset of representations is searched exhaustively.

It could be argued that the large effects resulting from the manipulation of the initial and terminal CC are in fact attributable to covariation of CC frequency with individual consonant frequency. To test this possibility ten pairs of CCVCCs were selected which had the same first letter and the same frequency of terminal CC, but which differed in frequency of initial CC. For example, the pair GRESS--GNUSH have the same first letter and the same terminal CC frequency, but are constructed HH and LH respectively. Average latency for each CCVCC (cf. Appendix B) was used for calculation of mean difference in each of the ten pairs chosen. If individual first letters had an effect on latency, the mean difference of pairs with identical first letters should be smaller than the comparable comparison for all CCVCCs. For pairs, difference was 112 milliseconds, and for all CCVCCs it was 89 milliseconds. Thus the effect of initial CC frequency seems unchanged when the first letter is held constant. An analogous analysis assessed the effect of the last letter by a comparison of seven pairs of CCVCCs with identical last letters and equal initial CC frequency when terminal CC frequency was varied. Pair difference was 109 milliseconds, while comparable difference for all CCVCCs was 120 milliseconds. Thus the effect of final CC is unchanged when the final letter is held constant also. The support for the argument that pair frequency is in fact the determiner of latency differences is based on only 17 paired latency differences; however, each individual score is based on the responses of 20 Ss. This should make the comparisons fairly stable, and make the conclusions based on the comparisons reasonably valid.

CHAPTER V

SUMMARY

Evidence is presented for a word recognition model with four distinct operations. The first stage involves abstraction of information from the components of an item. Next, this information is used to _mark_ a subset of items for search. Third, the item is encoded as a unit for search (with an editor evaluating phonological lawfulness). And finally, the marked set is searched in an ordered fashion with order determined by amount of use of each item as reflected by Thorndike-Lorge frequency count. Classification of a CCVCC (nonword) requires exhaustive search of the marked set, while search for a word requires a self-terminating search, producing shorter latencies for words than for nonwords.

A SELECTED BIBLIOGRAPHY

Brown, R., and McNeill, D. The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 325-337.

Halle, Morris. "On the Bases of Phonology" in Jerry A. Fodor and Jerrold J. Katz (eds.), The Structure of Language: Readings in the Philosophy of Language, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964, pp. 324-333.

Horowitz, L. M., White, M. A., and Atwood, D. W. Word fragments as aid to recall: The organization of a word. Journal of Experimental Psychology, 1968, 76, 219-226.

James, William. Psychology. New York: Henry Holt and Co., 1920.

Katz, J. J., and Fodor, J. A. The structure of a semantic theory. Language, 1963, 39, 170-210.

Lindgren, N. Machine recognition of human language. Part III--cursive script recognition. I.E.E.E. Spectrum, 1965, 2, 105-116.

Mayzner, M. S., and Tresselt, M. E. Tables of trigram frequency counts for various word-length and letter-position combinations. Psychonomic Monograph Supplements, 1965, 1, 33-78.

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

Norman, D. Memory and Attention. New York: Wiley, 1969.

Paul, A. P., House, A. S., and Stevens, K. N. Automatic reduction of vowel spectra: an analysis-by-synthesis method and its evaluation. Journal of the Acoustic Society of America, 1964, 36, 304-308.

Rubenstein, H. R., Garfield, L., and Millikan, Ja. A. Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 487-494.

Rubenstein, H., Lewis, S. S., and Rubenstein, M. A. Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 57-62. (a)

Rubenstein, H., Lewis, S. S., and Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657. (b)

Stanners, R. F. Language frequency correlates of related pronounciability. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 373-378.

Stanners, R. F., Forbach, G. B., and Headley, D. B. Decision and search processes in word-nonword classification. Journal of Experimental Psychology, 1971, 90, 45-50.

Thorndike, E. L., and Lorge, I. The Teacher's Word Book of 30,000 Words. Bureau of Publications Teachers College, Columbia University: New York, 1944.

Venezky, R. L. A computer program for deriving spelling-to-sound correlations. Unpublished M.A. Thesis, Cornell University, 1962.

Winnick, W. A., and Kressel, K. Tachistoscopic recognition thresholds, paired associate learning, and free recall as a function of abstractness-concreteness and word frequency. Journal of Experimental Psychology, 1965, 70, 163-168.

APPENDIX A

INSTRUCTIONS TO SUBJECTS

The following instructions were tape recorded and played to all _S_s before beginning the experiment.

This is an experiment concerned with simple judgements about verbal materials. It is not an intelligence test of any kind and should not be interpreted as such. Also, there is no electric shock nor any other unpleasant stimulus involved. Although the task may seem to be a very simple one, our research indicates that it can provide important information concerning language behavior. If for any reason during the course of the experiment you feel that you cannot fully cooperate, please let the experimenter know.

A five-letter item will be presented on the screen in front of you [_E_ indicates]. Your job is to decide, as quickly as possible, whether the item is or is not part of your vocabulary. If you decide the item is in your vocabulary, move the switch in the direction indicated on the card [_E_ indicates]. If the item is not part of your vocabulary, move the switch in the opposite direction. Make your judgement on the basis of whether the item is a complete unit in your vocabulary without adding anything to it. On this basis, the item S-P-A-R-C would _not_ be a member of most people's vocabulary, even though it is similar to and may remind you of the word S-P-A-R-K. In the same way the item S-L-A-N-D would not be in most people's vocabulary even though it is similar to and sounds like S-L-A-N-T. First or last names should also _not_ be treated as independent units. Examples of names which you might recognize, but should not be treated as independent units are C-H-U-C-K and S-M-I-T-H.

Slang terms _may_ be treated as independent units. If they are members of your vocabulary, you should indicate this with your response. Examples of fairly common slang terms are S-W-E-L-L and C-H-U-M-P. If items such as these are part of your vocabulary, then you should respond appropriately.

A complete trial sequence will proceed like this: You should hold the thumb button in your non-preferred hand [_E_ indicates]. Also, you should hold the switch between the thumb and forefinger of your preferred hand [_E_ indicates_7. When the experimenter is in the next room, and ready to start, a buzzer will sound indicating that you may begin. Start each trial by pressing the thumb button. About one second after you press it, the item will appear on the screen. As quickly as possible, decide whether the item is in your vocabulary or not, and move the switch in the appropriate direction. Both speed and accuracy are important. After your response, move the switch back to the middle position. Make sure that when you press the thumb button you are paying very close attention to the screen and that you are holding the switch properly. After your response, the white light will come on for a short rest interval. You may not activate the next trial until the white light goes off. After the white light goes off, you may start another trial when you wish, making sure you are paying very close attention to the screen before you press the thumb button.

Are there any questions?

APPENDIX B

INDIVIDUAL STUDY ITEMS AND THEIR LOG TRANSFORMED LATENCIES ALPHABETIZED WITHIN CATEGORY

<u>WORDs</u>

BRING	656	SHELF	668
BRONX	879	STALK	723
CHALK	628	STAND	636
CHECK	611	STAPH	755
CHILD	671	STICK	710
CLIMB	620	STILL	592
CROSS	723	THING	630
DWELL	653	THUGS	743
DWELT	688	THUMB	605
FRANC	780	TRUST	625
FRONT	614	TRUTH	602
GNASH	886	TWANG	842
GRAPH	738	TWIST	649
GRASS	640	TWIXT	695
GRITS	718	WHELP	823
KNACK	758	WRACK	833
KNOCK	644	WRATH	745
PLANT	625	WRECK	622
PLUMB	647	WRING	874
PRESS	616	WRIST	718
PROPS	841	WRONG	676
PSALM	826	WROTH	761
SCALP	670	WRUNG	801
SHALL	683		

FILLERs and PLURALs

BLANK	628
BLAST	636
BLEST	915
BLINK	712
BRAND	620
BRASH	676
BRASS	641
BRATS	681
BRIGS	775
BRUNT	747
CHANT	737
CHESS	628
CHEST	610
CHITS	847
CHUMS	789
CLICK	626
CLOCK	630
CLOGS	675
CLOTS	677
CRACK	646
CREPT	701
CROCK	872
CROPS	729
CRYPT	935
DRAMS	842
DREGS	847
DRESS	628
DRILL	641
DROLL	751
DROPS	623
DROSS	786
FLAGS	606
FLATS	636
FLICK	638
FLOCK	626
FROGS	625
GLASS	700
GLOSS	742
PLANS	654
PLUMS	671
PRAMS	881
PRINT	596
PROMS	796
PRONG	837
SCANT	718
SCENT	772
SHAGS	736
SHELL	612
SHILL	750
SHIMS	922
SHIPS	666
SHOPS	656
SHUNT	790
STASH	823
STEMS	660
STEPS	651
STING	679
STINT	811
SWAMP	679
SWELL	605
SWILL	713
TRACT	721
TRAMP	614
TRAMS	834
TRAPS	626
TRASH	588
TREKS	886
TRESS	701
TRIPS	689
TRUMP	794
TRUSS	772
WHICH	752
WHIPS	642

CCVCCs

BRENG	805	PSELM	729
BRUNX	780	RHOLP	693
CHACK	897	SCULP	839
CHELK	763	SHILF	744
CHOLD	761	SHULL	951
CLEMB	704	STECK	770
CRUSS	995	STIND	839
CZABT	615	STOLK	807
DWALT	802	STOLL	877
DWILL	836	STOPH	794
DWOLN	653	THAMB	780
FRENT	832	THENG	783
FRINC	637	THIGS	849
GHAMN	740	TMEKS	698
GNOMS	640	TRATH	913
GNUSH	719	TRIST	896
GRATS	951	TSENC	619
GRESS	850	TWEST	775
GROPH	771	TWEXT	695
KHINN	652	TWING	844
KNECK	866	TZULB	604
KNUCK	903	WHOLP	944
KRADZ	596	WRANG	788
LLAMT	634	WRENG	799
MNETZ	569	WRETH	860
PLEMB	715	WRICK	829
PLINT	773	WROCK	898
PNALC	622	WROST	867
PRASS	928	WRUTH	815
PRUPS	808	WRYNG	621

CCCCCs

BRKNG	641	PSNLM	638
BRKNX	661	RHKLP	591
CHLCK	711	SCTLP	588
CHMLD	601	SHMLL	679
CHPLK	638	SHRLF	653
CLFMB	598	STFCK	624
CRNSS	639	STNLK	694
CZBST	575	STRND	641
DWFLT	698	STZLL	602
DWRLL	733	STZPH	568
DWZLN	609	THKMB	619
FRBNT	654	THPGS	627
FRSNC	584	THSNG	623
GHTMN	559	TMBKS	550
GNPSH	587	TRFTH	683
GNSMS	577	TRLST	630
GRSPH	613	TSFNC	570
GRTSS	615	TWLNG	654
GRZTS	622	TWNST	647
KHPNN	565	TWPXT	543
KNLCK	645	TZMLB	574
KNMCK	593	WHSLP	674
KRLDZ	595	WRBNG	618
LLMNT	627	WRBTH	600
MNKTZ	568	WRKNG	620
PLPNT	614	WRKNG	666
PLRMB	626	WRMST	635
PNRLC	577	WRSCK	637
PRBSS	607	WRTCK	660
PRMPS	586	WRZTH	576

APPENDIX C

MEAN LATENCIES FOR CATEGORIES AND SUBCONDITIONS FOR STUDY ITEMS

	Subcondition			
	HH	HL	LH	LL
WORDs	641	723	737	**
CCVCCs	866	799	831	656
CCCCCs	646	628	643	586

**Too few English words for estimate

	Thorndike-Lorge Frequency	
	High	Low
FILLERs (HH)	649	763
PLURALs (HL)	662	810

VITA

Gary Bruce Forbach

Candidate for the Degree of

Master of Science

Thesis: ANALYSIS OF LETTER STRINGS IN WORD NONWORD CLASSIFICATION

Major Field: Psychology

Biographical:

Personal Data: Born in Casa Grande, Arizona, September 8, 1947, the son of Mr. and Mrs. A. L. Forbach.

Education: Graduated from Eastwood High School, El Paso, Texas, in May, 1965; received Bachelor of Arts degree from the University of Denver in 1969 with a major in Psychology.

Professional Experience: Served as Graduate Teaching Assistant, College of Arts and Sciences, Oklahoma State University, 1969-1970; served as Graduate Research Assistant 1970-1971; was National Science Foundation Summer Research Assistant, 1971; also, Graduate Research Assistant, August 1971 through July 1972; member Southwestern Psychological Association.