1-1-1986

Modularity and auditory word recognition.

Cynthia M. Connine
University of Massachusetts Amherst

MODULARITY AND AUDITORY WORD RECOGNITION

A Dissertation Presented

by

CYNTHIA MAY CONNINE

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 1986

Psychology

MODULARITY AND AUDITORY WORD RECOGNITION

A Dissertation Presented

by

Cynthia May Connine

Approved as to style and content by:

Charles E. Clifton, Jr., Chairperson of Committee

James Chumbley, Member

Nelson Cowan, Member

Lyn Frazier, Member

Seymour Berger, Chairperson

Department of Psychology

Supported in part by:

National Institute of Health Grant HD 18708

and

National Institute of Mental Health Grant MH 16745

ACKNOWLEDGEMENT

I would like to express my appreciation and thanks to the members of my committee, Chuck Clifton, Lyn Frazier, Jim Chumbley and Nelson Cowan for their time and energy throughout this project.

I thank Jim for disagreeing in such a congenial manner and for his expertise in deriving the quadratic function equations and writing the Basic program, Lyn for her unique ability of looking for reasons and theories, not just evidence, and Nelson for cheerfully providing his long distance comments.

Special thanks go to Chuck for providing his expert guidance and advice while at the same time allowing me to make my own mistakes. I am especially grateful to Chuck for his almost neverending patience.

I also want to express my gratitude to Joanne Miller at Northeastern University who generously provided the materials used in Experiment 2 and was the source of the initial spark of my interest in speech perception.

My parents have been unceasingly supportive and loving throughout my education. I thank them for this, and their encouragement but mostly for being so very proud of my accomplishments.

I thank Sue Lima who should know what a rare and precious person she is, Luchy Toyofuku for the fine hospitality of "the dissertation hotel" and who is never a failure as a hostess and Pam Hardiman who manages somehow in the midst of it all to get everything done.

Last but not least, I thank Albrecht in true Teutonic style.

ABSTRACT

Modularity and Auditory Word Recognition

February, 1986

Cynthia May Connine, B.S., Northeastern University

M.S., Ph.D., University of Massachusetts

Directed by: Professor Charles Clifton, Jr.

Three experiments investigated the role of lexical (Experiment 1) and sentence context (Experiments 2 and 3) information in speech perception. In Experiment 1, subjects were presented a list of stimuli (words and non-words) taken from two voicing speech continua and asked to identify the initial phoneme. In one continuum, the voiced endpoint formed a word and the voiceless endpoint formed a non-word. A second continuum ranged from a voiced non-word to a voiceless word. Subjects labeled ambiguous stimuli in the mid-range of the continua as the phoneme that formed a word. The reaction time functions showed a clear shift in the peak reaction time. Word responses were faster than non-words at the category boundary while no difference was found at the continua endpoints. These results are

consistent with an interactive account of lexical status effects on phoneme identification. At the boundary, word responses show an advantage due to feedback from the lexically stored phonological representation. Unambiguous acoustic information allows a response prior lexical feedback.

In Experiments 2 and 3, stimuli from voicing continua were presented in biased sentence contexts. Sentence contexts were pragmatically biased toward the word formed at the voiced endpoint or biased toward the voiceless endpoint. Subjects identified stimuli in the mid-range of the continua as the pragmatically appropriate word. In Experiment 2, the reaction time functions showed an unreliable tendency for a shift in peak reaction time. Responses that formed a word consistent with the sentence context were facilitated at the boundary and endpoints. Since responses to stimuli at the endpoints can be made based on acoustic information alone, it is argued that the endpoint effects reflect post-perceptual processes. This component accounts for the consistency effects at the category boundary. In Experiment 3, voiced responses were faster than voiceless responses at the boundary regardless of sentence bias while a consistency effect was found at the endpoints.

The experiments suggest that analysis of speech is directly influenced by prestored phonological form while sentence context effects are post-perceptual. Discussion of these findings includes the role of syllable and word (morpheme and idioms) structure. A model incorporating these elements is sketched.

TABLE OF CONTENTS

LIST OF TABLES

LIST of FIGURES

CHAPTER I

INTRODUCTION

Comprehension of spoken language requires a listener to interpret a complex, highly encoded auditory signal. The seeming ease with which this task is performed places language behavior as one of the most accomplished abilities of humans. Language processing is assumed to involve the computation of representations organized in terms of traditional lines of linguistic inquiry, that is phonology, syntax and formal semantics. The manner in which these representations develop, interact and are made available to non-linguistic knowledge (world knowledge) has been a primary focus of concern in psycholinguistics (Fodor, Bever & Garrett, 1974).

Fodor (1983) has recently put forth a theory of mind that espouses the notion of modularity as a general characteristic of mental architecture, in particular of language. In this work, Fodor develops a characterization of modular subsystems as operating quickly, efficiently and employing only a subset of the information relevant for a processing decision. Further, the internal computations of a modular subsystem can not be directly influenced by knowledge contained in another subsystem. In the weak form of modular theory described by Fodor, sources of information within the linguistic system are not constrained in terms of the scheduling of information use within a linguistic domain or the availability of information between linguistic domains. Modularity exists in this formulation of the theory only in that world knowledge does not directly intervene in linguistic processing decisions.

It is important to note that Fodor does not specifically argue against modularity within the linguistic system. Further structuring of information within the language processing system is not incompatible with the weak modularity hypothesis and in fact a stronger view of modularity in language processing has been proposed for some specific linguistic domains. These proposals assume that language processing is modular with respect to world knowledge but further assume that subsystems exist. In a series of experiments based on eye movements, Frazier and her co-workers (Frazier, 1979; Rayner, Carlson & Frazier, 1983; Frazier & Rayner, 1982; see also Frazier, Clifton & Randall, 1983) have argued for a modular syntactic processor. These experiments have provided evidence that a syntactic representation of a sentence is initially constructed based on syntactic strategies. The syntactic processing system is modular in that syntactic strategies are applied to initially structure a string without reference to semantic input.

Other recent work (Tanenhaus, Leiman & Seidenberg, 1979) has systematically investigated the modularity hypothesis with respect to access of lexical meaning (see also Cairns & Kammerman, 1975). These researchers used a cross modal priming technique first developed to investigate associative priming effects (Swinney, Onifer, Prather & Hirschowitz, 1979) and later utilized to investigate lexical ambiguity (see also Swinney, 1979; 1982; Seidenberg, Tanenhaus, Leiman, & Bienkowski; 1982). In one experiment, listeners were presented noun-verb ambiguities in syntactically unambiguous contexts (e.g. John began to tire). A visual probe was presented for naming at 0 msec or 200 msec after offset of the ambiguous word. Naming responses for probes related to the appropriate and inappropriate senses of the ambiguous word were both facilitated at 0 msec delay. However, facilitation for only the appropriate probe was found at the 200 msec delay. Comparable results were found for contexts in which one sense of an ambiguous word was more compatible with plausibility relations in the sentence (e.g. You should have played the spade). These studies present a strong case for a modular lexicon (see Yates, 1978; Hoagaboam & Perfetti, 1975; Simpson, 1981 for investigations that have implicated the importance of meaning frequency).

The modularity hypothesis will be explored here specifically with reference to speech perception and auditory word recognition. The hypothesized existence of a speech perception module amounts to the claim that a constrained class of linguistic knowledge is consulted to interpret the speech signal. The kind of information considered relevant to the operation of the speech perception module will be formulated in terms of general operating characteristics. Identifying general properties of speech processing will be well served by focusing on the kinds of information that are used in developing an initial hypothesis (or hypotheses) concerning the acoustic input.

One general operating characteristic developed in this thesis is the lexical storage hypothesis. The lexical storage hypothesis postulates that the phonological information stored with a lexical entry is used to develop hypotheses during word recognition. The lexical storage hypothesis is outlined in Chapter II and tested in Experiment 1.

The lexical storage hypothesis implies that information outside the domain of speech processing may be used to choose among hypotheses or used to integrate a specific hypothesis with a sentence and/or discourse context (see Norris (1982) for a specific modular language processing system with this general characteristic). Prestored phonological information contrasts with pragmatic, semantic and syntactic information in that the latter is used to choose among lexical hypotheses. Sentential context is used to choose an appropriate lexical item from among activated entries but does not contribute to the developement of activated lexical entries. The lexical storage hypothesis violates a strictly modular position in that lexical level information can directly influence perceptual processing. However, a great deal of constraint on the nature of the interaction is provided since the lexical storage hypothesis rules out feedback from higher levels (pragmatics, semantics and syntax).

The hypothesis that higher level information is used to choose among potential words but is not influential in the development of word hypotheses is developed in Chapter III and tested in Experiments 2 and 3.

Activation of multiple lexical hypotheses may be a general principle of auditory word recognition. Information in the speech signal relevant to the sounds that make up a word becomes available over time. The temporal nature of auditory language suggests that multiple activation of lexical entries may result. This is also an interesting hypothesis in that it may be a dimension along which modules within language processing differ. Frazier & Rayner (in preparation) have suggested that if the representation computed during the course of language comprehension is not already stored as such in memory (e.g. syntactic structure) then only a single analysis is considered at any one time. If, however, the computed representation is one that is also prestored in memory (e.g. the phonological properties of a lexical item stored with the entry) multiple hypotheses may be considered simultaneously.

Chapter I is organized into four major sections. The first section outlines two major theories of word recognition in psycholinguistics (hierarchical and interactive theories). Next, the evidence in support of an interactive use of sentence context in auditory word recognition is reviewed. The major hypothesis developed in this section is an alternative interpretation of these findings within a hierarchical theory in which the effects of sentence context are attributed to post-perceptual mechanisms. In a third section, evidence supporting post-perceptual effects in visual word recognition research are presented. A final section provides an overview of models proposed in speech perception. The considerations outlined in Chapter I provide a framework within which the lexical storage hypothesis and its implications is developed.

Models of Word Recognition

Interactive models. One specific model of auditory word recognition, the cohort model, is particularly relevant to the modularity issue. The cohort model of auditory word recognition represents one of a class of models, interactive models (cf. Marslen-Wilson, 1983 for a recent discussion). Interactive models may be characterized by the general principle that all sources of knowledge (world and linguistic knowledge) contribute in concert with acoustic analysis during comprehension. In the cohort model, the acoustic information at the beginning of a word is used to establish a pool of items in the lexicon with which it is consistent. Analysis of additional portions of the utterance, in conjunction with constraints imposed by syntactic and semantic expectations based on previous context, removes inconsistent lexical items from consideration (cf. Marslen-Wilson & Tyler, 1980). Word recognition occurs when a single item remains in the cohort.

Morton (1969) presents another interactive model, the logogen model. All sources of contextual information contribute to the activation of a logogen associated with each word. Lexical access occurs when a given logogen reaches a threshold level of activation.

Hierarchical models. In contrast to the cohort model is a language processing system outlined by Forster (1978, 1979). This model represents a class of models, hierarchical models, in which the processing system is highly structured and consists of a set of linguistic subprocessors (lexical, syntactic and semantic) each of which has its own domain of computations. The flow of information from lexical to syntactic to semantic subprocessors is strictly linear. The representation completed by each subprocessor is available to the next processor in the architecture only when a completed representation has been computed. Each processor has access to the lexicon which is conceptualized as a language oriented data structure. In addition to the linguistic system, there exists a general problem solver (GPS). The GPS receives the output of each of the subprocessors but cannot influence their operation.

The lexical processor is one of the most developed aspects of the model. In contrast to the cohort model in which lexical items are directly activated and accessed, Forsters lexical access model involves a search ordered in terms of word frequency. Further, lexical access proceeds only on the basis of sensory information since the architecture of the lexical processing system denies the use of contextual influences.

The model contains three peripheral access files organized according to the stimulus features necessary for production, auditory perception and visual perception. Words are grouped in bins in each access file according to similarity in features. The first stage of lexical access involves extracting the relevant feature information from the stimulus in some unspecified fashion. The features are used to locate the bin containing the item. Each bin is organized by word frequency. Once an entry is located in a peripheral access file, a pointer associated with each entry in the access file is used to access an entry in a master lexicon containing all information about a word.

Contextual Factors and Auditory Word Recognition

Psychologists have been interested in the role of context in auditory word recognition since at least the turn of this century (see Cole & Rudnicky, 1983, for a review of the work of William Bagley, 1874-1946) and numerous studies have reported that world knowledge can effect auditory word recognition. In this section evidence arguing for interactive accounts of contextual contributions during auditory word recognition will be reviewed. The review will be organized by the particular task employed with the intent of providing an alternative class of explanations for these phenomona in terms of post-perceptual decision factors.

Intelligibility. Some of the earliest contemporary work focussed on the intelligibility of words presented in noise. Miller, Heise & Lichten (1951) found that identification of words was more accurate in sentence contexts than in isolation. Other studies (Pollack & Pickett, 1963; Pickett & Pollack, 1963,1964) have examined intelligibility of words excised from fluent speech. These studies showed that identification accuracy increased when an item was presented in the context of one word relative to accuracy for isolated words.

Shadowing. More recent research interpreted in support of an interactive class of models is based on shadowing tasks (Marslen-Wilson, 1975). In these experiments, subjects listen to sentences in which mispronunciations have deliberately been introduced. Subjects must repeat aloud, 'shadow', the speech as quickly as it is heard. Marslen-Wilson & Welsh (1978) presented listeners with sentences in which mispronunciations occurred in a highly constrained context (e.g. "He wanted to smoke a cigarette"...; the underlined item received the mispronunciation) or in a low constraint condition (e.g. It was his misfortune...). It was found that subjects tended not to shadow a mispronunciation in the high constraint sentences but instead fluently restored the mispronounced item to the appropriate form. Further, an analysis of shadowing response latencies for words occurring after the restored mispronunciation revealed no disruptions whereas performance after shadowed mispronunciations showed an increase in response latency. Marslen-Wilson & Welsh argued that the constraints imposed by the context in the high constraint condition decreased the amount of acoustic input necessary for unique identification of a lexical item. Thus, subjects were able to identify the lexical item prior to analysis of all the acoustic input necessary to uniquely identify the item.

Mispronunciation detection. An extensive series of experiments has been conducted employing a mispronunciation detection technique developed by Cole (1973). A target word is embedded in a sentence context and produced with a deliberate mispronunciation. Typically, the mispronunciation consists of a change in a feature (or features) of a consonant (e.g. place of articulation: make is mispronounced as nake). It is assumed that mispronunciation detection reflects how quickly the intended word can be recognized and this process will occur more quickly if the context is highly constraining.

In one experiment, Cole & Jakimik (1978; see also Cole & Jakimik, 1979) manipulated the transitional probability of the target words. Transitional probability of a word is the probability of a word's occurrence given a preceding string of words. Faster mispronunciation detections were found for high than low transitional probability items (see also Jakimik, 1979). The mispronunciation effect was largest if only a single feature was altered in the critical item. A smaller effect on reaction time was seen for mispronunuciations in which more than one feature was altered.

The mispronunciation effect was replicated in experiments in which contextual constraint was determined by the immediately preceding word (e.g. mink coat, where coat is the target), implied instruments of action (e.g. pound the nail, nail is the target) and thematic organization supplied by a picture or a title to a spoken passage. The effects of predictability in this task has also been replicated for spoken language in children (Cole & Perfetti, 1980).

Gating. Other experiments have used a gating paradigm to investigate auditory word recognition. Here, a subject is presented with successively larger fragments of a word until the entire word is heard. On each trial, the subject must indicate the identity of the lexical item based on the information available. Grosjean (1980) presented gated words in sentence contexts with high or low degrees of contextual constraint. He found that the presence of greater contextual constraint decreased the amount of acoustic information necessary to identify a word correctly. Listeners required only 37% of a gated word heard in a high constraint context compared to 60% in a low constraint context in order to correctly identify the item.

Grosjean notes that the estimate of the amount of information necessary for a correct response in the gating paradigm is similar to that found in the speech shadowing paradigm (Marslen-Wilson & Welsh, 1978). These findings were replicated (Cotton & Grosjean, 1984) when a variation of the gating paradigm was used in which subjects were presented only one gate duration. This design eliminated successive repetitions of an item.

Other experiments conducted in Dutch (Tyler & Wessels, 1983) have argued that contextual constraints used in narrowing the pool of possible lexical items in this task are primarily semantic. Syntactic constraints of the type used (the word preceding the target word constrained it to an infinitive verb form or a small number of adjectives) did not reliably decrease the amount of acoustic information necessary to correctly identify a word. Finally, strong contextual constraints reduce the pool of items elicited from the first fragment presentation (Grosjean, 1980; Tyler, 1984). Grosjean suggests that this indicates pre-selection of likely lexical items, based on the context, prior to any acoustic input. Context also increases the rate with which the pool of word candidates are eliminated from subjects' responses (Tyler, 1984). Presumably, context is used to further narrow the subset of items consistent with the auditory information.

<u>Guessing and decision factors</u>. The data from intelligibility studies, shadowing, mispronunciation detection and the gating paradigm have been interpreted as support for a theory of auditory word recognition in which linguistic and world knowledge are presumed to contribute directly to the analysis of speech. It is not clear, however, that the evidence reflects perceptual processing and lexical access. The nature of the intelligibility and gating paradigms

suggests that world knowledge and contextual constraints can be used to guess a word's identity, given sufficient time for this information to be consulted.

During shadowing and mispronunciation detection, it is quite possible that world knowledge and linguistic constraints available from the context can facilitate decision processes in making a response. For example, mispronunciation detection may reflect the influence of context on decisions concerning the identity of the intended word, which is at a point subsequent to analysis of the acoustic input.

In the shadowing task, unlike listening, a decision must be made immediately for an overt response (articulation of just received input). This pressure may encourage subjects to make a commitment (a "best guess") based on partial acoustic information in conjunction with an assessment as to whether the "best guess" under consideration is consistent with the context. Such a decision process that operates on an analyzed input may, by its nature, be facilitated by greater contextual constraint (i.e. less information from the context will have to be consulted in high constraint sentences in order to determine the appropriateness of the lexical hypothesis).

Decision Factors in Visual Word Recognition

The notion that context facilitates decision processes in a particular task, and not perceptual processing, is currently a matter of considerable debate in visual word recognition research. Many studies have used the lexical decision task to demonstrate facilitated perceptual processing of words in predictive sentence contexts (Fischler & Bloom, 1979; Schuberth & Eimas, 1977). For example, Fischler & Bloom (1979) required subjects to perform a lexical decision to a target word that appeared as the last word in predictable, semantically appropriate or semantically anomalous sentences. It was found that lexical decision responses were facilitated in the predictable contexts.

Also representative of the work using the lexical decision task is a series of studies by Stanovich & West (1978, 1979). Stanovich & West formulate contextual effects on word recognition within a theoretical framework established by Posner & Snyder (1975). Briefly, the Posner & Snyder theory proposes two processes that influence recognition. A fast-acting, automatic spreading activation process spreads activation from an encoded stimulus memory location to semantically related items in memory. This process is excitatory for expected items and does not cause inhibition of incongruous items. A second conscious attention mechanism is slow acting and attention demanding. This process directs attention to the memory location of expected items. As a result, responses to incongruous items are inhibited since attention is focussed elsewhere in memory. Under unencumbered viewing conditions, lexical decision times to a word are facilitated in a congruous context but no inhibitory effects are seen in an incongruous context. When the target word is degraded, large inhibitory effects in incongruous contexts are found. Stanovich & West argue that under normal viewing conditions, recognition occurs so quickly, only the spreading activation mechanism is engaged and produces facilitated responses. If stimulus encoding is delayed, responses in the incongruous contexts are inhibited by the misdirection of the conscious attention mechanism.

However, other studies have argued that effects of this type can be attributed to post-perceptual decision processes (e.g. Forster, 1981; Cairns, Cowart & Jablon, 1981). A recent example of the role of decision factors has been reported by Seidenberg, Waters, Sanders, & Langer (1984). These researchers have presented evidence consistent with the view that many contextual effects may be attributed to decision factors associated with a particular task. In this series of experiments, a context word was presented and followed by a target word. In one condition, subjects were required to perform a lexical decision on the target word. In a second condition, a naming response was produced.

Two classes of priming conditions were used. Associated (e.g. dream-sleep) and semantic priming (boy-prince) reflect connections within the lexicon and are mediated by a spreading activation mechanism. Syntactic (e.g. whose-planet) and backwards priming (fly-fruit) effects are the result of post-lexical judgements concerning the relatedness of the items. The lexical decision task showed facilitation under associate and semantic priming conditions as well as in syntactic and backwards priming. Responses in the naming task were facilitated only in associate and semantic priming contexts. Seidenberg et al attributed these task specific patterns of facilitation to an additional post-lexical processing component in the lexical decision task. Immediate lexical encoding is influenced only by associative and semantic priming conditions that reflect the structural organization of the lexicon.

Decision Factors in Auditory Word Recognition

The visual word recognition literature just described suggests that it is crucial to consider post-perceptual decision factors in tasks used to show context effects on perception. One task used in the auditory domain that has been investigated with respect to contextual and post-perceptual factors is the phoneme restoration phenomonon first reported by Warren (1970) (see also Warren & Obusek, 1971). In these experiments, a portion of the acoustic signal is electronically spliced out (e.g. /s/ in legislature) and replaced with a non-linguistic sound. For example, the /s/ in legislature may be replaced with a 'cough'. Listeners will report hearing an intact word with (in this example) a 'cough' in the background.

Recently, Samuels (1981a, 1981b) has conducted a series of studies on phoneme restorations using signal detection methodology to separate perceptual effects (d') from post-perceptual bias (B). In these experiments, word targets were presented with a sound replaced by white noise or a sound presented intact plus white noise. Samuels (1981a) found that listeners tended to report the targets as intact, rather than replaced, when the item occured in a predictable context. If higher level syntactic and semantic information were directly influencing phonetic processing, discriminability (of intact vs. replaced items) for predicted words would decrease. The data indicate that discriminability for predicted words actually increased. While interpretive problems exist in this application of signal detection theory, these data suggest that context does not influence directly perceptual processing.

There are no available data on the role of decision factors in the other tasks considered earlier (intelligibility, shadowing, mispronunciation detection, and gating) used to investigate auditory language comprehension. The identification of decision processes as explanatory mechanisms for context effects in the auditory domain has obvious implications for the modularity hypothesis.

Models of Speech Perception

The hypotheses developed in the previous sections are relevant to the concerns of researchers in both speech perception and word recognition. The kinds of issues studied by word recognition theorists have been discussed above. This section will briefly review the major problems investigated by workers in speech perception and the kinds of theories proposed.

Theories of speech perception have been largely motivated by observations concerning the form of the acoustic event during production of spoken language. One of the obvious facts about speech is the variability evident in the acoustic waveform when a visual display of the speech signal is examined. Sources of variability include between-speaker and as well as within-speaker variability. In addition, the acoustic information for a phoneme can change dramatically depending upon the particular phonetic environment (e.g. a change in vowel environment can alter the shape of the formant transitions for a preceding consonant). It is equally striking that listeners are typically unaware of this tremendous variability. The ability to maintain perceptual constancy given intra- and inter-speaker variability and effects of phonetic context has been a dominant focus of theoretical and empirical work in speech perception.

Motor theory. One class of models in speech perception can be characterized by the notion that the listener imposes an organization on the speech signal. One early proposal, the motor theory of speech perception, attempts to explain speech perception by hypothesizing that the speech signal is decoded through mediation by production. A major aspect of motor theory is its assumption of specialized speech processing since unlike other modes of perception, speech perception requires reference to production (Liberman, 1982).

The seminal research exploring the motor theory involved a phenomonon called categorical perception (Liberman, Harris, Kinney & Lane, 1965). In a typical experiment on categorical perception, a synthetic speech continuum varying along a single acoustic parameter is presented to listeners for classification. Generally, the stimuli are divided by listeners into two (or more) categories. In order to determine the degree to which listeners can detect absolute physical differences among the stimuli an ABX task is employed. The ABX task requires a subject to indicate whether the third stimulus (stimulus X) of a triad of stimuli is identical to the first (stimulus A) or second stimulus (stimulus B) presented. If the stimulus continuum tested involves stop consonant categories, within category discrimination is poor and typically near chance, while between category discrimination is quite good. If, however, the ABX task is used to test discrimination of vowels that vary along an acoustic dimension, performance is equally good between and within categories.

The motor theory explains the facts of categorical perception by appeal to articulatory organization (cf. Repp, 1983). If production of a phonetic contrast is characterized by discrete articulatory gestures, as is the case for stop consonant contrasts of voicing, for example, perception will be categorical. Conversely, the continuous articulatory gestures that describe vowel production are mirrored in the continuous perception of vowels.

It should be noted that the motor theory explanation for the labeling and discrimination difference between vowels and consonants has been challenged. Some researchers claim the phenomonon is due in part to two modes of perception with different properties (Pisoni, 1971; Pisoni & Tash, 1974; see also Fujisaki & Kawashima, cited in Repp, 1983). One mode is strictly categorical and represents phonetic classification and an associated verbal short term memory. The other mode is continuous and involves processes common to all auditory perception. The observed differences in vowel and consonant recognition reflect differences in the strength of the representation in auditory memory.

One important assumption of motor theory as originally stated in its strongest version (Liberman, 1957) is that the variability seen in the speech signal is attenuated in production. This would provide a simpler relationship between perceived phoneme and acoustic attribute. However, electromyographic research has indicated that the motor movements in speech production are a source of considerable variability (Harris, Bastian & Liberman, 1961; MacNeilage, 1970). More recent accounts of motor theory have placed the mediation of articulation at the level of neuromotor commands (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967).

While the notion of a role for articulation in speech perception is intuitively appealing, the theory has been criticized for its lack of precision in stating exactly how the underlying articulatory gestures are extracted by a listener (cf. Jusczyk, 1984). Other research has challenged motor theory by demonstrating categorical perception for non-speech sounds and in chinchillas (see Repp, 1983 for an extensive review). These findings have been used to argue that categorical perception is not unique to speech or to humans and thus speech perception is not a specialized processing system.

More recent research has provided a new line of evidence for a specialized speech mode. This research employs ambiguous sinewave stimuli. In these synthetic stimuli, the formant frequency structure of a speech sound is replaced by sinewaves. The result is a stimulus that is perceived by some listeners as the original speech sound and by others as a mixture of tones and noise. Depending on the perceptual experience of the listener, the stimuli are identified and discriminated to a greater or lesser extent as speech (e.g. Best, Morrongiello & Robson, 1981).

Template models. Other approaches in speech perception have attempted to analyze the acoustic signal for invariant cues to phonemes. This class of models assumes invariance is a property of the speech signal and the problem is to determine the most appropriate way to analyze the signal to discover the relevant patterns. In a series of studies, Stevens & Blumstein (1978; Blumstein & Stevens, 1979, 1980) have focussed the search for invariant properties on integrated cues, each of which alone may not show invariance but will exhibit a characteristic pattern together.

This approach has lead to some success particularly for place of articulation distinctions in stop consonants. Blumstein & Stevens have identified a different pattern of spectral energy at the onset of the burst associated with syllable initial labial, bilabial and alveolar stops.

A similar model has been recently developed by Kewley-Port (1980, 1983). One important aspect in which Kewley-Port's approach differs from Blumstein & Stevens is its emphasis on spectral change. Kewley-Port argues that the onset spectra identified by Blumstein & Stevens is insufficient to characterize speech patterns. She has incorporated the notion of how spectral patterns change over time as critical for stop consonant identification.

Massaro (1980, 1981) has argued for the necessity of a perceptual mechanism with a continuous rather than the binary output tacitly assumed by most models. In the fuzzy logical model Massaro has developed, features are specified by the degree to which a presented sound has that relevant acoustic feature or property. Multiple features are extracted for a given sound, represented probabilistically and then combined via logical integration rules. A discrete percept is obtained when the integrated feature values are compared against the relative degree of fit to a prototype.

One characteristic these classes of models share is that the speech signal is ultimately internally represented as a series of discrete segments and features. This is not a universal characteristic of models of speech. For example, Klatt (1979) has developed a model with no intermediate representation between analysis of the waveform and lexical access. The speech signal is simply analyzed into spectral chunks of 10 msecs. The obtained spectral sequences are used to trace an optimal path in a prestored network of spectral chunks.

There are, however, a number of lines of evidence suggesting speech is represented segmentally. These have recently been reviewed by Pisoni (in press). The perceptual and production evidence that is discussed generally involves errors occuring in phoneme or feature size units suggesting listeners have access to segmental representations. For example, Garnes & Bond (1980) have catalogued a corpus of misperceptions of fluent speech and found a large number of phoneme and feature errors. The phoneme production errors found in production error corpuses have been used to argue for the phoneme unit as crucial in speech planning (e.g. Shattuck-Hufnagel & Klatt, 1979).

Summary

Two classes of language processing models are outlined (hierarchical and interactive). It is suggested that evidence originally interpreted as support for interactive models of auditory word recognition may be accounted for by hierarchical models. A brief review of speech perception outlines the two major theoretical positions. While some progress has been made, the evidence is equivocal. However, the issues that are addressed overalap with issues in auditory word recognition in attempting to determine how the relevant information is extracted from the auditory signal.

CHAPTER II

SPEECH PERCEPTION AND AUDITORY WORD RECOGNITION

This chapter will focus on a class of information, phonological knowledge stored with a particular lexical item, that has generally been ignored in speech perception research. The lexical storage hypothesis will be formulated as a general operating characteristic of the speech perception module. The relevant literature will be reviewed and Experiment 1 will be presented in support of this hypothesis.

Lexical storage hypothesis. The lexical storage hypothesis claims that phonological information stored with a lexical item will directly influence acoustic analysis of speech. If we assume that the acoustic information activates a set of lexical hypotheses, then the prestored information associated with a particular item will be readily available. The probability of occurrence of a particular phoneme is high given activation of a lexical entry containing that phoneme. Scheduling the use of prestored phonological information early in processing will facilitate recovery of further information from the acoustic signal. The phonological representation stored with a lexical entry can reinforce correct hypotheses and rule out incorrect hypotheses at the sound level. The lexical storage hypothesis also implies that structural aspects of the lexicon resulting in differential availability of classes of words will also operate in an interactive fashion. This will allow a characterization of the speech perception module as a fast, efficient system, one of the characteristics identified by Fodor (1983) as typical of modular systems.

The lexical storage hypothesis is supported by some recent research investigating the role of lexical knowledge in speech processing. The first kind of evidence to be considered is the demonstrated influence of lexical status on speech perception. Samuels (1981a, 1981b) presented a series of experiments in which the effects of lexical status on speech perception were examined using a signal detection analysis of the phoneme restoration effect. Samuels found more phoneme restorations for real words than for phonologically legal pseudo-words. Specifically, discriminability (d') of added & replaced versions of pseudowords was better than for words. This suggests that the phonological form of a lexical item is accessed and directly influences perceptual processing.

A similar influence of lexical status in speech perception was reported by Ganong (1980). In this experiment, listeners were presented with a synthesized voicing continuum that varied in voice-onset-time (VOT) of the initial phoneme. In one series (e.g. dust-tust) the voiced end point of the continuum formed a word. In a yoked series (e.g. test-dest) the voiceless stimulus was a word. It was found that listeners' identification functions (percentage voiced responses as a function of point along the continuum) differed depending upon which end of the continuum (voiced or voiceless) formed a word. Specifically, subjects tended to label ambiguous stimuli (typically, those stimuli in the mid-range of a continuum) so that the phonetic sequence resulted in a word (henceforth, the "lexical effect").

Since the lexical effect was found primarily for acoustically ambiguous stimuli, Ganong argued that a strict categorical perception account of the effect was not supported.[1] A categorical perception model asserts that the listener has available only a binary feature description of the stimulus. Thus, this model predicts a lexical effect at all points along a continuum since relative acoustic information is lost. If this were the case, lexical information would influence the phonemic decision throughout the range of acoustic values.

Experiment 1

The first experiment reported here (see also, Connine & Clifton, in preparation) is designed to replicate the lexical effect (see also, Miller, Dexter & Pickard, 1984; Fox, (1984)[2]) and use reaction time data to distinguish between hierarchical and interactive classes of explanations for the lexical effect. As previously described, the general principle defining interactive models conceptualizes several linguistic knowledge sources as acting in concert to contribute to mental linguistic representations. Hierarchical models provide a view of auditory word recognition in which information from prestored knowledge sources influences decision processes only after initial interpretation of the acoustic information. Both classes of models are consistent with the lexical effect but may be distinguished based on patterns of reaction time.

Predictions. In terms of the lexical effect, hierarchical models assume that information from the lexical domain is used only at a decision stage after the acoustic information is encoded. If the acoustic information relevant for a sound is ambiguous but consistent with two potential alternative sounds, lexical knowledge biases the choice toward the sound that forms a word with the surrounding context. Interpretation of the acoustic information proceeds without consulting the lexicon if the acoustic information is unambiguous.

This explanation accounts for the lexical effect since lexical status is influential only for stimuli in the mid-range of a continuum, typically perceived as a mixture of the end point sounds. Reaction time to ambiguous information will include a stage in which lexical status is consulted. The resulting delay in categorization of ambiguous information causes slower reaction times to stimuli in the middle of a continuum than to unambiguous end point stimuli. However, the peak of the reaction time function (the position along the continuum at which the slowest reaction times occur) should not depend on which end point forms a word.

In the present context, interactive models assume that lexical status alters directly the interpretation of the acoustic signal. One specific interactive mechanism assumes that a positive feedback loop exists from a lexical representation and a representation of the sound segments. The positive feedback can affect the rate of accrual of evidence, perhaps to an asymptotic value, for a particular phoneme. Information is continuously extracted from the acoustic signal, providing evidence for one or more alternative sound segments. However, if one of these segments combined with other surrounding hypothesized segments corresponds to a word, then the lexical entry may provide positive evidence in favor of that sound alternative. Lexical status is thus not viewed as providing input to a decision mechanism, but rather viewed as influencing the accrual of information for a particular sound alternative. The proposed mechanism is similar to that proposed by connectionist models for language comprehension and production (cf. Feldman & Ballard, 1982; see also Dell, 1985).

Within the connectionist model framework, positive feedback from a lexical entry contributes to the evidence for a response consistent with a word. This predicts that categorization of ambiguous acoustic information as a sound consistent with a word will occur more quickly while non-word responses to ambiguous acoustic information will be slow. The occurrence of non-word responses to ambiguous acoustic input may be explained within the context of positive lexical feedback by assuming variable availability of positive feedback from a lexical representation. On the trials in which lexical entry feedback is available relatively slowly, a sound forming a non-word alternative may reach the critical amount of acoustic evidence necessary for a response prior to a sound forming a word alternative.

Importantly, if the critical phoneme target is in word initial position, effects of lexical status on word and non-word responses should be seen primarily at the category boundary. For word initial targets, lexical feedback has an oppurtunity to accrue only if the acoustic information is insufficient for a categorization. If the acoustic information is unambiguous, the lexicon may obligatorily be consulted but a response can be made early on the basis of the acoustic information alone.

As described, an interactive account of the lexical effect predicts faster word responses than non-word responses in the region along a speech continuum that is perceptually ambiguous. The region of the category boundary provides the best means to test this hypothesis since word and non-word responses are approximately evenly distributed for the boundary stimulus. Further, a word advantage is not predicted for unambiguous acoustic information. If this is the case, any word advantage found at the category boundary cannot be attributed to a reaction time advantage at a late decision stage for word responses.

With the exception of the boundary stimulus, the distribution of word and non-word responses for stimuli in the mid-range region of the continuum will not permit a direct test of facilitated word responses to ambiguous stimuli. Stimuli toward the word end of a continuum from the boundary will include a disproportionate number of fast word responses and few slow non-word responses. If both word and non-word responses are considered, response times toward the word end of the continuum should decrease for two reasons. First, the percentage of fast word responses will increase and second, the acoustic information will be more consistent with a word categorization.

Consider the reaction times to all responses (word and non-word) in the region around the category boundary for a given continuum. For the stimuli toward the word endpoint and stimuli toward the non-word endpoint of a continuum, the following predictions emerge. At the category boundary, we expect the slowest overall reaction times since the acoustic information is ambiguous (by definition). The stimulus toward the word endpoint of a continuum from the boundary (but still within the boundary region of ambiguity) will be relatively more consistent with a word response. An initial higher activation of a categorization consistent with a lexical response will result. This will increase the likelihood of a word response and, those word responses that are made will be facilitated. The greater percentage of fast word responses to this stimulus will result in an overall faster reaction time (word and non-word) compared with the reaction time to the boundary stimulus.

Responses to stimuli toward the non-word end of a continuum (but still within the boundary region) will include relatively more non-word responses. These non-word responses will not necessarily be slow compared with word responses since the acoustic information is more consistent with a non-word response. The acoustic information increases the likelihood of a non-word categorization. A combination of acoustic information and fast word and (potentially) fast non-word responses will result in an overall faster reaction time compared with the reaction time to the boundary stimulus.

The net result of the factors contributing to the pattern of reaction times within the boundary region is that the peak of the reaction time function will depend on the location of the lexical anchor. Specifically, the peak of the reaction time function will be shifted away from the lexical anchor.

Method

Subjects. A total of 17 members of the University of Massachusetts community were pretested. Five subjects were rejected on the grounds that they did not consistently label the VOT endpoint stimuli. The remaining subjects participated in the full experiment which lasted approximately one hour and were paid $5.00 for their participation.

Materials. The Klatt (1980) software formant synthesizer at Sussex

University was used to prepare instances of the words TYPE and DICE. Natural pronunciations of each item were used as models. Eight variants of each item were prepared, varying in VOT from 10 to 45 msec. VOT is defined articulatorily as the time between the initial release of the consonant and onset of vocal fold vibration. VOT can be measured acoustically by the duration between the abrupt increase in energy typically found at the consonant release and the onset of periodic energy (vocal fold vibration). Voiceless sounds are produced with a longer VOT than voiced sounds.

The synthesized VOT section of each item was replaced with an equal duration of an attentuated natural t-burst, terminated at a zero-crossing. Otherwise, all parameters of each item, aside from the parameters that determined the final segments, were identical. The stimuli differed from each other before 200 msec only in that the first format fell more slowly from its early peak for DICE than for TYPE.

Tokens varying in VOT were also made of the words TIGRESS, DIGRESS, PREFER and BRIEFER. These were presented to subjects along with the crucial TYPE and DICE items, as described below, as part of a series of experiments designed to explore the effects of lexical stress on word recognition.

Procedure and apparatus. Subjects listened to tokens of the stimuli through headphones adjusted to a comfortable listening level. The subjects were pretested on another set of synthesized stimuli to insure consistent labeling of the synthesized speech. The pretest session lasted approximately 20 minutes and used the same procedures as the main experiment.

Each of the 8 TYPE-DYPE and the 8 DICE-TICE stimuli was presented ten times, in a random order, and randomly intermixed with ten occurrences each of the 32 TIGRESS, DIGRESS, PREFER and BRIEFER stimluli. A soft 200 Hz warning tone, 50 msec in length, preceded each stimulus by 100 msec.

Two audio tapes were prepared, which presented the items in opposite orders. Six subjects heard one tape, and six the other. Subjects received instructions to press a right-hand button if they heard a word or non-word beginning with a /p/ or /t/ and a left-hand button if they heard a word or non-word beginning with a /b/ or /d/.

Responses for each trial were recorded. The subjects were told that their reponse times were recorded and were instructed to attend to the beginning of each item and to make their responses quickly and intuitively. They were informed what stimuli they would hear and 10 stimuli from a random position in the pretest tape were played for each subject in order to familiarize them with synthesized speech. After every 50 trials, stimulus presentation was interrupted and a message was presented on a CRT screen instructing the subject to take a break. Presentation of the stimuli resumed when the subject pressed a response key on the keyboard.

Results and discussion

Identification functions. The percentages of /d/ responses made by each subject to each token were calculated[3]. The percentage of /d/ responses as a function of VOT (identification function) for each continuum is shown in Figure 1. As expected, more voiced responses were found for the DICE-TICE continuum relative to the DYPE-TYPE continuum.

In order to quantify this observation, a VOT boundary value corresponding to the 50% /d/ responses was calculated for each subject's identification function. This calculation consisted of fitting a linear regression equation to each identification function excluding the tails of the function (i.e. in the boundary region of the identification function). The obtained linear regression equation was used to calculate the VOT value at 50% /d/ responses. A relatively longer boundary value reflects an identification function with a greater proportion of voiced responses.

The boundary values associated with the DYPE-TYPE and DICE-TICE continua are 22.84 and 26.06 msec, respectively. A one way ANOVA (lexical anchor- voiced or voiceless) showed a reliable identification function shift of 3.21 msec ($F(1,11) = 18.29$, $p < .001$). These data clearly replicate the lexical effect found by Ganong (1980).

Reaction time functions. The reaction time values (collapsed across word and non-word responses) as a function of VOT are presented in Figure 2. In order to compare the reaction time data for

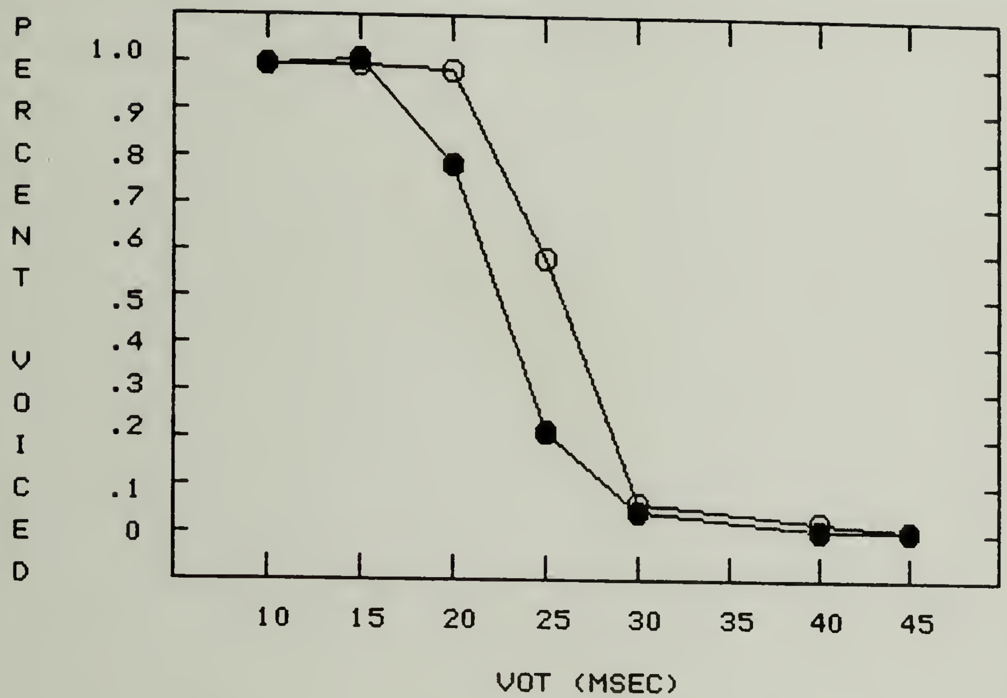


FIGURE 1. PERCENTAGE OF VOICED RESPONSES AS A FUNCTION OF STIMULUS AND LEXICAL ANCHOR. OPEN CIRCLES INDICATE DICE-TICE; FILLED CIRCLES INDICATE DYPE-TYPE. EXPERIMENT 1.

the two continua and determine the VOT value associated with the slowest reaction time, the data were analyzed in the following manner. Four reaction time points in the boundary region for each subject were determined for the 2 continua (two stimuli towards the voiced endpoint and 2 stimuli towards the voiceless endpoint from the boundary). The boundary value was based on the obtained value from the linear regression analysis. The best fitting quadratic function was calculated for the boundary reaction times[4]. The obtained quadratic function was used to calculate the VOT associated with the slowest reaction time (the predicted maximum of the function)[5].

A number of comments are in order concerning the quadratic function analysis. A quadratic function was chosen as the appropriate curve for the simple reason it appeared, on inspection, to be a reasonable fit of the reaction time data in the category boundary region. A subset of the reaction time data, those around the boundary value, were chosen in order to exclude spurious maxima, that is those maxima not in the boundary region [6]. This analysis resulted in 2 reaction time functions (8%) in which the computed function did not include the observed peak of the function. In these cases, the observed peak occurred an average of 3 stimuli from the boundary.

Finally, while a quadratic function appeared to adequately describe the reaction time functions in the boundary region, local fluctuations in the shape of an individual subject's function sometimes resulted in an equation that misrepresented its shape and

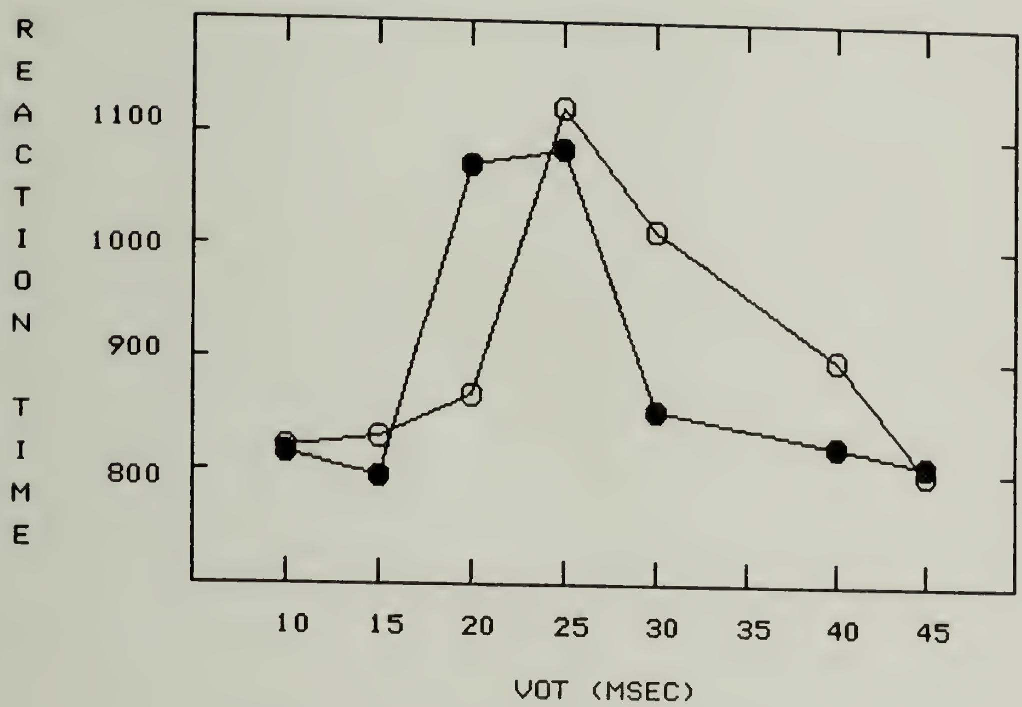


FIGURE 2. REACTION TIMES AS A FUNCTION OF STIMULUS AND LEXICAL ANCHOR. OPEN CIRCLES INDICATE DICE-TICE; FILLED CIRCLES INDICATE DYPE-TYPE. EXPERIMENT 1.

thus the predicted maximum. These cases include functions in which the fitted quadratic function described a u-shaped curve, as indicated by a negative _a_ coefficient.

In other cases, the quadratic function analysis yielded a reaction time peak value beyond the range of the stimulus VOT values used in the experiment (i.e. a peak VOT value greater than 45 msec). Here, the reaction times in the boundary region tended to be relatively linear (i.e. the observed peak occurred at the largest or smallest stimulus VOT value within the four stimuli fit by the quadratic function). The VOT corresponding to the maxima of these functions was conservatively estimated as the VOT at the observed peak within the range of 3 stimulus steps to the left or the right of the boundary. If the reaction times adjacent to the observed peak in the 3-stimulus range were within 100 msec of the observed peak, the maximum VOT was taken to be the average VOT value of these adjacent reaction times. Inspection of the individual subject reaction time functions indicates This is an accurate method to estimate the VOT value associated with the slowest reaction time.

The VOT associated with the peak reaction time for the voiced and voiceless lexical anchor are 27.7 and 24.3 msec VOT respectively. The 3.37 msec difference is reliable ($F(1,11) = 5.12$, $p < .03$) in a one way (lexical anchor, voiced and voiceless) ANOVA. The peak of the DICE-TICE continuum is shifted away from the lexical anchor and thus located at a relatively longer VOT than the peak of the

DYPE-TYPE continuum.

Word and non-word reaction times. The peaks of the reaction time functions are composed of word and non-word responses. In order to determine the relative contribution of the each response category to the peak reaction time, the stimulus at the category boundary for each lexical bias was partitioned into word and non-word reaction times. The following procedure was followed to determine the boundary reaction times. If the category boundary fell between two stimuli, the stimulus closest in VOT value to the boundary was taken as representative of the reaction time at the boundary. If the stimulus nearest the boundary did not receive approximately equal numbers of word and non-word responses, the boundary reaction time was taken as the average of the two reaction times straddling the boundary. This procedure was necessary to insure comparable percentages of word and non-word responses in the reaction time values. A disproportionate number of responses in one category may result in a biased sample of reaction times and misrepresent the mean value for a response category.

The reaction time values partitioned into the word and non-word components of the boundary stimuli are shown in Table 1. The values in parenthesis indicate the relative percentage of responses at the boundary for a given response category. These data show a 69 msec advantage for word (voiced) responses given a voiced lexical bias and a 79 msec advantage for word (voiceless) responses given the voiceless lexical bias. Thus, the voiced lexical anchor peak is composed of two components: relatively fast word (voiced) responses and slow non-word (voiceless) responses. The relationship between word and non-word response times holds for the voiceless lexical anchor peak. A two way ANOVA (word-nonword x lexical bias) shows a significant advantage for word responses (74 msec, $F(1,11) = 5.7$, $p < .03$). No other effects were significant.

The reaction time advantage for word responses is confined to the category boundary region. Reaction times to endpoint stimuli (see Table 2) show no evidence of faster word responses. A two way ANOVA (word-nonword x endpoint stimulus - voiced vs voiceless) confirmed this observation. No effects were reliable.

The absence of a word advantage for endpoint (unambiguous) stimuli is not due to a simple response bias advantage for word responses. Further, the word reaction time advantage cannot be explained by acoustic factors. The stimulus at the category boundary for the voiced bias continuum contains relatively more voiceless information than the boundary stimulus for the voiceless bias continuum. Acoustic factors alone predict an advantage for voiceless responses in the vicinity of the category boundary for the voiced bias continuum.

Summary. Lexical status influenced phoneme identification in addition to reaction times. The reaction time functions showed a clear peak shift. Word responses were faster than non word responses

		WORD BIAS	
		VOICED	VOICELESS
RESPONSE	WORD	1076 (52)	1095 (48)
	NON-WORD	1145 (48)	1174 (52)

TABLE 1. WORD AND NON-WORD RESPONSES AT THE CATEGORY BOUNDARY AS A FUNCTION OF WORD BIAS. THE NUMBERS IN PARENTHESES INDICATE THE PERCENTAGE OF RESPONSES FROM THE CATEGORY BOUNDARY FOR A GIVEN REACTION TIME. EXPERIMENT 1.

		WORD BIAS	
		VOICED	VOICELESS
ENDPOINT	VOICED	816	812
	VOICELESS	798	805

TABLE 2. ENDPOINT RESPONSES AS A FUNCTION OF WORD BIAS. EXPERIMENT 1.

at the category boundary. No advantage for word responses was found at the continuum endpoints. The lexical effect and the reaction time data are consistent with a characterization of auditory word recognition in which prestored phonological information operates very early during acoustic analysis. A natural account for the pattern of data is an interactive mechanism in which positive feedback from the lexical level contributes directly to the evidence for component sound segments.

CHAPTER III

EXPERIMENT 2

The approach developed here is to assume that interactive principles apply between acoustic analysis and linguistic knowledge but in a relatively constrained fashion. One specific constraint has been outlined in which prestored phonological knowledge will contribute directly to interpretation of the speech signal. According to this view, pragmatic, semantic and syntactic information contrast with phonological and lexical information in that the former are used to select among alternative word hypotheses, not to guide the development of such hypotheses.

A number of studies (Garnes & Bond, 1975; Spencer & Halwes, 1981; Miller, Green & Schermer, 1984) have reported that listeners tend to report ambiguous stimuli such that the resulting lexical item is consistent with the sentential context (henceforth, the "semantic effect"). For example, Garnes & Bond varied place of articulation of a word-initial consonant. Specifically, the starting value of the second formant transition was modified to create a series that ranged from BAIT to DATE to GATE. Each item in the continuum was presented in three different contexts which were biased toward one of the target words (Check the time and the...; Here's the fishing gear and the...; Paint the fence and the...). It was found that subjects labeled acoustically ambiguous target items to make them conform with the semantic bias of the sentence. Acoustically clear stimuli were not systematically influenced.

The semantic context effect was replicated by Miller, Green & Schermer (1984) who in addition found that the influence of semantic context depended on instruction set. Miller et al embedded a voicing series (BATH-PATH) in semantically biased carrier sentences (She ran some hot water for the...; She likes to jog along the...). One instruction set required the subjects to report the sentence context, as well as the identity of the target item. In another instruction set, listeners merely reported the identity of the target word (but still heard the entire sentence). Listeners reported ambiguous stimuli as consistent with the bias of the sentence only when the subjects were required to attend to the sentence context.

The primary concern of Miller et al was the contrast between the effect of instruction set on the semantic effect and the imperviousness of the effects of rate of speech (as speech rate increases, a relatively longer VOT is required to label a phoneme as voiced) to changes in instructions. However, the data suggest that the semantic effect may be accounted for by a post-perceptual decision mechanism. The goal of Experiment 2 was to distinguish between competing explanations of the semantic context effect based on a distinction between hierarchical and interactive classes of models using the reaction time peak shift paradigm.

Predictions. In one specific hierarchical model, sentential context is consulted to choose between two (or more) potential lexical items if the acoustic information is insufficient for a unique identification. Given ambiguous input, the time required for a response consistent with the context should be comparable to responses inconsistent with the context. A peak in reaction time to stimuli in the mid-range of the continuum will not depend on which word endpoint is semantically consistent with the context.

Consistent responses could be facilitated relative to inconsistent responses due to a late stage reaction time component. Specifically, consistent responses occur in a contextual environment in which the target word maybe integrated into a sentence context more readily. If this is the case, then response times may reflect this component and result in facilitation for consistent responses across the entire range of acoustic stimulus values. A reaction time peak shift would be found (due to relative precentages of consistent responses) but the peak shift will be accompanied by facilitation for consistent responses at the endpoints of the continuum.

In interactive models, sentence context facilitates the availability of a categorization consistent with a contextually appropriate word. One possible mechanism considered earlier included an increase in rate of accrual of information. In this model, a categorization that combines with the surrounding sounds to form a contextually appropriate word (consistent response) will be available earlier. The overall rate of accrual of evidence will increase for a consistent response relative to an inconsistent responses. An increase in the rate of accrual of evidence will result in faster consistent relative to inconsistent categorizations. At the category boundary, this predicts faster reaction times to consistent responses compared to inconsistent responses. Contextual effects on the rate of evidence accrual should be seen primarily for ambiguous stimuli since unambiguous acoustic information should provide the critical amount of evidence very quickly for a response.

If reaction times for all responses are plotted as a function of stimulus, we should find an inverted u-shape, with slowest times to stimuli in the middle of the continuum. The peak in the response time function should be shifted away from the consistent context endpoint. Response times to stimuli in the boundary region for a given sentence bias will reflect two components, a fast component due to consistent responses and a slower component to inconsistent responses. At the end points of a continuum, consistent responses should not be faster than inconsistent responses.

Method

Subjects. Twenty subjects from the University of Massachusetts participated in the experiment for course credit.

Materials

Sentence and word pairs. The materials used in this experiment are a subset of those from Miller, Green & Schermer(1984). Two lexical items were selected that differed in the voicing feature (bath and path) and a semantically biased carrier sentence form each item was constructed ("She needed hot water for the..." and "She likes to run along the...).

Normative data were collected from 24 subjects to corroborate the experimenters intuitions concerning the semantic bias of the carrier sentences. None of the subjects who served as raters participated in the auditory experiment. Subjects were asked to complete the carrier sentences with the most appropriate lexical item and then rate the appropriateness of the sentence/lexical item combinations on a scale from 1 to 10. A low number was assigned if "bath" was the most appropriate and a number was assigned if "path" was the most appropriate. The mean ratings were 1.33 for "the bath" bias sentence (s.d. = .92) and 9.88 for "the path" bias sentence (s.d. = .45).

A third semantically neutral sentence was constructed ("She was thinking about the...). Each sentence and lexical item combination was recorded by a male speaker. The neutral carrier sentence recordings were digitized and an wave form editing program was used to excise the phrases "the bath" and "the path". A voicing series was constructed from the excised phrases by manipulating voice onset time (VOT) of the initial stop consonant using a technique developed by Lisker (1978) (see also Ganong, 1980).

In order to create a voicing continuum, successive portions of periodic energy were excised from "bath", beginning at the end of the closure for /b/. The excised portions were replaced with equally long acoustic segments of aperiodic energy taken from "path", beginning at the end of the closure for /p/. The excised segments were cut at zero-crossings in order to avoid abrupt changes in amplitude in the hybrid stimuli. Abrupt changes in amplitude can be perceived as clicks or pops superimposed over the speech and are potentially disruptive. The duration of the excised acoustic segment was successively lengthened to create a voicing continuum that consisted of 16 stimuli. The length of each bath/path stimulus was 418 msec and the duration of "the" plus the closure (the period of silence prior to the release of the stop consonant) was 180 msec.

In order to minimize the number of conditions in the experiment, a subset of 10 of 'the bath/path' stimuli was selected by presenting each of the original 16 stimuli to listeners for identification. The 10 stimuli chosen for inclusion in the sentence context experiment were those stimuli in the region of the voicing boundary stimulus (50% voiced responses). 5 stimuli toward the voiced endpoint and 5 stimuli toward the voiceless endpoint were chosen. The VOT values of these stimuli were: 15, 19, 21, 26, 32, 35, 39, 45, 48 and 53 msec.

A production of each sentence context was selected. The 'bath' bias sentence frame was originally produced with the lexical item "path" and conversely, the 'path' bias sentence frame was originally produced with the lexical item 'bath'. Each sentence frame was digitized and the final phrase was excised ("the path" and "the bath", respectively). Next, each 'the bath/path' stimulus was combined with each of the two sentence frames to create a total of 20 stimuli (2 sentence contexts x 10 the bath/path stimuli).

Procedure. The stimuli were re-digitized and a warm-up audio and an experimental audio tape were prepared. The warm-up tape consisted of 6 randomizations of the 10 "the bath/path" stimuli without sentence context. The warm-up tape served to familiarize subjects with the stimuli. The experimental tape consisted of one instance of each of the 20 stimuli randomly presented in each block of trials and eight blocks were presented. Three seconds of silence was recorded between each trial and a rest break was given after every 20 trials. A saturated square wave click was placed at the beginning of the burst of each target word on the right channel of the audio tape, inaudible to the subject. During the experiment, the click activated a Schmitt trigger starting a timer. The timer was stopped when a subject made a response. The sentences were presented binaurally over headphones at a comfortable listening level in a quiet room on a Teac dual capstan drive X-10 tape recorder. The tape recorder was operated by a CompuPro microcomputer and subjects' responses and reaction times were stored on disc.

In the warmup and experimental phases, subjects were instructed to identify the initial sound of the target by pressing the appropriate key of a labeled response panel ('B' or 'P') quickly and intuitively. In the experimental phase, subjects were required to indicate whether the /b/ or /p/ response made on a trial formed a word that was sensible or anomalous with the sentence frame by circling 'sensible' or 'anomalous' on an answer sheet. This response was required in order to insure the entire sentence context was attended to on every trial.

Results and Discussion

Identification functions. The percentage of voiced (/b/) responses as a function of point along the stimulus continuum are presented in Figure 3 for each sentence context. The identification function is extremely orderly showing a high percentage of voiced choices at the short VOT end of the continuum and nearly 0% voiced choices at the long VOT end of the continuum. Responses to stimuli towards the middle of the VOT continuum were influenced by sentence context. As predicted, 'the bath' bias sentence context resulted in relatively more voiced responses than 'the path' bias sentence.

A category boundary value was computed for each sentence context identification function using the linear regression analysis described in the preliminary experiment. 'The bath' sentence context produced a boundary value shifted toward the voiceless end of the continuum (35.11 msec) relative to 'the path' sentence context (33.27 msec). The 1.84 msec difference was reliable in a one-way ANOVA

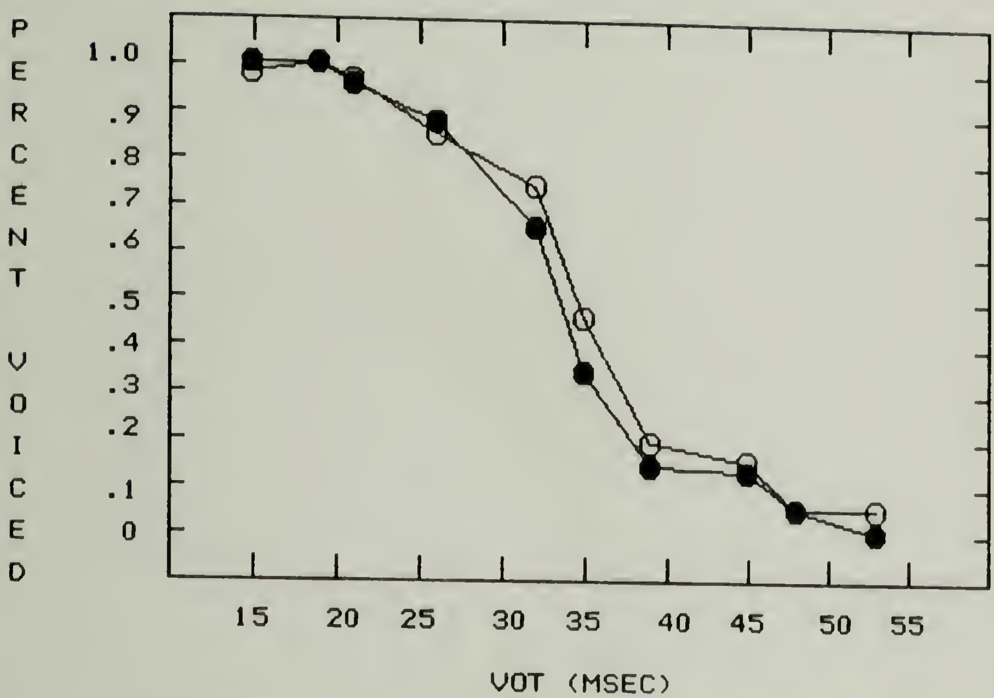


FIGURE 3. PERCENTAGE OF VOICED RESPONSES AS A FUNCTION OF STIMULUS AND SENTENCE CONTEXT. OPEN CIRCLES INDICATE 'BATH' BIAS; FILLED CIRCLES INDICATE 'PATH' BIAS. EXPERIMENT 2.

($F(1,19) = 4.4$, $p < .04$). In addition, an analysis of the isolated word identification functions from the warmup session revealed the isolated boundary was shifted toward the voiced endpoint relative to the voiced and voiceless sentence context (32.6 msec)[7]. The explanation for the shift of the isolation boundary value toward the voiced endpoint is not known.

Reaction time functions. The reaction time functions are presented in Figure 4. In order to determine the vot value associated with the slowest reaction time for the two continua, the quadratic function analysis described in the preliminary experiment was employed. The subset of the reaction time values around the category boundary value included the observed peak of the function in 80% of the functions. In the remaining 20% of the functions, the observed peak ranged from 4 to 8 steps from the boundary.

The VOT values found at the peak reaction time for the voiced bias and voiceless bias sentence contexts are 34.57 and 33.93 msec respectively. The .64 msec difference is in the direction predictd by an interactive account of the identification function shift, but it was highly unreliable ($F(1,19) < 1$).

Consistent and inconsistent reaction times. At the category boundary value found for the voiced bias sentence context, the reaction time values were partitioned into consistent and inconsistent responses. Similarly the voiceless sentence bias boundary reaction times were separated into consistent and

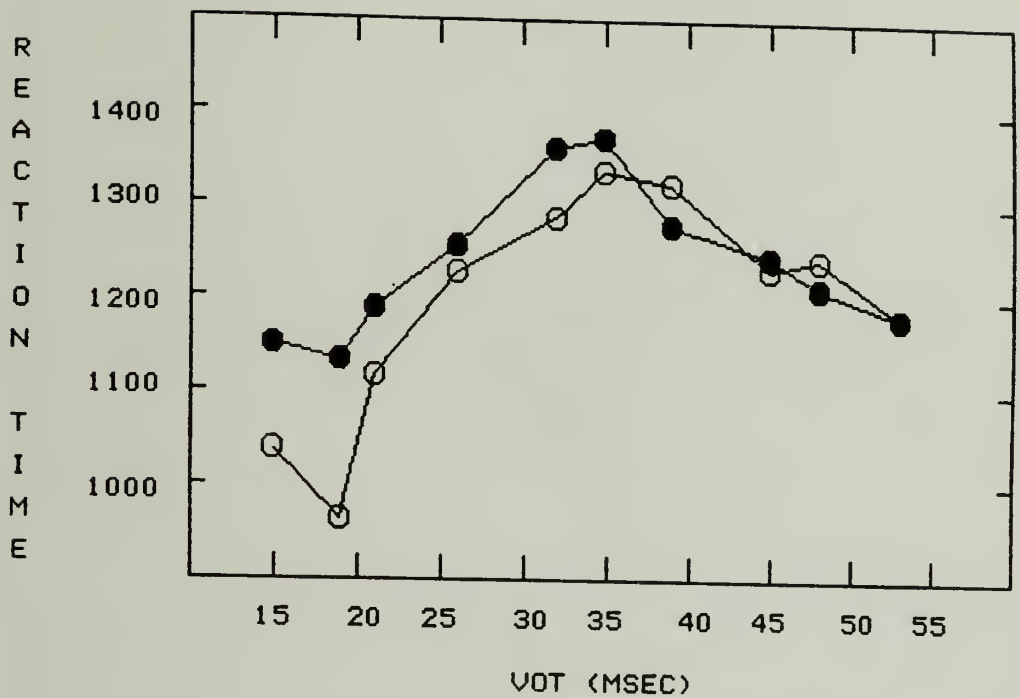


FIGURE 4. REACTION TIMES AS A FUNCTION OF STIMULUS AND SENTENCE CONTEXT. OPEN CIRCLES INDICATE 'BATH' BIAS; FILLED CIRCLES INDICATE 'PATH' BIAS. EXPERIMENT 2.

inconsistent response components. These data are shown in Table 3. The values in parentheses are the percentage of responses for the boundary stimulus on which the partitioned reaction time values are based.

The voiced sentence bias boundary response showed a 144 msec advantage for consistent (voiced) relative to inconsistent (voiceless) responses and the voiceless sentence bias boundary stimulus showed a 26 msec advantage for consistent (voiceless) relative to inconsistent (voiced) responses. The 85 msec advantage for consistent responses was not reliable ($F(1,19) = 3.3$, $p < .08$) in a two way ANOVA (sentence bias x response consistency - consistent vs. inconsistent). No other effects were significant.

Although the main effect of consistency did not reach an acceptable level of significance, the pattern of data are in accord with the lexical effect in that consistent responses show a reaction time advantage at the boundary. The advantage is quite asymmetrical, however, in that consistent responses in the voiced sentence bias show an advantage nearly seven times that seen for consistent responses in the voiceless sentence bias. The observed asymmetry is in contrast to the extremely symmetrical advantage for word responses seen in the lexical effect data (69 msec, voiced bias and 79 msec, voiceless bias).

An analysis of the reaction time responses at the endpoints of the continuum provides a possible explanation for the asymmetry seen at

		SENTENCE BIAS	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1290 (51)	1371 (46)
	INCONSISTENT	1434 (49)	1397 (54)

TABLE 3. CONSISTENT AND INCONSISTENT RESPONSES AT THE CATEGORY BOUNDARY AS A FUNCTION OF SENTENCE CONTEXT. THE NUMBERS IN PARENTHESES INDICATE THE PERCENTAGE OF RESPONSES FROM THE CATEGORY BOUNDARY FOR A GIVEN REACTION TIME. EXPERIMENT 2.

the boundary. The endpoint reaction times for each sentence bias are shown in Table 4. A 51 msec advantage is found for consistent responses. A two-way ANOVA (endpoint - voiced vs. voiceless X response consistency) did not show a reliable main effect of consistency ($F(1,19) = 1.8$, $p < .20$). The interaction (endpoint X consistency) was reliable ($F(1,19) = 4.7$, $p < .05$). A large advantage for consistent (voiced) responses is seen at the voiced endpoint (117 msec) and inconsistent (also voiced) responses show a slight advantage at the voiceless endpoint (17 msec).

This endpoint stimuli pattern of data suggests an overall advantage for voiced responses. It should be noted that the voiceless portion of the reaction time function appears to be slightly slower than the voiced portion of the function. The explanation for the observed asymmetry may lie in an artifact of the technique used to create the voicing series from natural speech. As described, the technique involved creating hybrid stimuli in which the major portion of the item was taken from a voiced exemplar. Although the voicing information is relevant for a voiceless token at the voiceless end of a continuum, the following vowel may contain information more appropriate for a voiced token.

Subjectively, the voiceless stimuli are perceived as 'good' voiceless tokens and are labeled as such, but the mismatch between voicing information and other information in the syllable may have contributed to overall slower reaction times. The observed advantage for voiced responses (81 msec) did not reach an acceptable level of significance in a two way ANOVA (endpoint X sentence bias) ($F(1,19) = 3.3$, $p < .10$). A main effect of bias (voiced vs. voiceless) indicated a reliable advantage for responses in the voiced bias context ($F(1,19) = 4.7$, $p < .04$). The reason for faster responses in the voiced bias context is unclear.

The large advantage for a voiced response in a voiced sentence bias at the endpoint (117 msec) is comparable to that seen at the category boundary (144 msec). Responses to endpoint stimuli can be made based on the acoustic information alone. This suggests that the advantage for consistent voiced responses at the boundary is due to a late stage process in which a faster response can be made if it is consistent with the sentence context.

Consistent responses in the voiceless bias sentence show a different pattern of data. At the boundary, an advantage is found for consistent (voiceless) responses (26 msec). At the endpoint, an advantage for inconsistent (voiced) responses is found (17 msec). The reason for this discrepancy is not known, although both voiceless sentence bias effects are quite small.

Consideration of the boundary and endpoint reaction time data suggests an explanation for the (non-significant) peak shift tendancy. The voiced bias boundary region consists of a relatively fast voiced response component. The stimulus toward the voiced endpoint from the boundary will have relatively more fast voiced

		ENDPOINT	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1031	1180
	INCONSISTENT	1148	1163

TABLE 4. ENDPOINT REACTION TIMES AS A FUNCTION OF SENTENCE CONTEXT. EXPERIMENT 2.

responses and thus an overall faster reaction time. The peak of the reaction time function will be slightly shifted away from the voiced endpoint.

The sentence context effect pattern of data contrasts with the lexical effect in a number of aspects. First, the sentence context effect showed an unreliable peak shift nearly half the size of the identification function shift (.63 msec vs. 1.84 msec, respectively). The lexical effect data showed a reaction time peak shift slightly larger than the identification function shift (3.78 msec vs. 3.2 msec, respectively).

Further analysis of sentence context consistent and inconsistent responses at the category boundary suggests that the unreliable tendency for a peak shift is primarily accounted for by an advantage for context consistent voiced responses. Because the voiced endpoint response advantage mirrors the voiced boundary response advantage in the voiced bias sentence context, the advantage for context consistent voiced responses at the boundary is attributed to a stage after stimulus encoding that is sensitive to sentence integration processes.

As described in Experiment 1 (lexical effect), no comparable word reaction time advantage is found at the voiced or voiceless endpoint. However, a word advantage was found for the boundary stimuli. Given the absence of a word advantage at the endpoints, the boundary effect cannot be explained by appeal to a simple overall advantage for word responses. The lexical effect pattern of data seems to reflect an interactive use of prestored phonological form during speech perception.

Summary. A speech sound from the mid-range of a voicing continuum was labeled such that it formed a word consistent with the sentence context. Reaction time data showed a small and unreliable shift in the reaction time functions. Further analysis of consistent and inconsistent responses at the category boundary and at the continuum endpoints suggests that the tendency for a peak shift may be accounted for by fast consistent responses in the voiced sentence bias. Since responses to stimuli at the endpoints can be made based on the acoustic information alone, it is argued that the endpoint effects reflect a late stage component of the reaction time. It is this late stage component that is claimed to account for the consistency effects at the boundary.

CHAPTER IV

EXPERIMENT 3

Experiment 2 replicated the sentence context effect on phoneme labeling and provided evidence for non-contextually driven lexical access. Before concluding that the speech perception module takes account of lexical information but not sentential context, a number of troublesome aspects of the data must be addressed. The size of the identification function shift found in Experiment 2 was reliable but rather small (1.84 msec). Certainly, the potential shift is small given that the number of perceptually ambiguous stimuli is a small subset of the continuum. One or two stimuli in the mid-range of the continuum are subjectively ambiguous. However, the size of the shift in the original Miller et al experiment from which the stimuli were taken, was 4.4 msec, more than twice the present effect. It may be that the reaction time peak shift found here was too unreliable to be detected give the size of the identification function shift.

A second problem with Experiment 2 is methodological. In order to obtain sufficient numbers of observations per stimulus along the continuum, the two sentence contexts and lexical items were presented repeatedly with little novelty. It may be that the number of repetitions produced subject strategies particular to the experimental situation. Such strategies may account for an asymmetry observed in the data in which only the voiced endpoint showed an advantage for consistent responses. Accordingly, a second experiment was conducted in which the number of sentence contexts and target word pairs were increased.

Method

Subjects. Twenty subjects from the University of Massachusetts participated in the experiment for a combination of course credit and money.

Materials. Seven target word pairs that differ in voicing value of the word-initial stop consonant were chosen as endpoint stimuli (DENT-TENT, DIME-TIME, DRIP-TRIP, GUARD-CARD, GOAL-COAL, GOAT-COAT, GOLD-COLD) and two context sentences were constructed for each word pair[8]. As in Experiment 2, one context sentence was pragmatically biased toward the voiceless member of the pair and the second context sentence was biased toward the voiced member. The target words occurred as the last word in each sentence. Sentence pairs were matched on number of syllables and prosody. The context sentence associated with the word pairs are listed in Table 5.

In order to corroborate the experimenters' intuitions concerning the pragmatic bias of the text sentences, the sentences were rated by 15 subjects (none of whom participated in the auditory experiment). The rating procedure was identical to that used in Experiment 2 (1 =

WORD PAIR	**SENTENCE CONTEXT**
DENT/TENT	SHE DRIVES THE CAR WITH SHE SAW THE SHOW IN
DIME/TIME	SHE FEARED THAT THE PAY PHONE WOULD NOT TAKE SHE HOPED THAT THE CHURCH BELLS WOULD TELL HER
DRIP/TRIP	SHE WAS CERTAIN HER HUSBAND COULD NOT FIX SHE INVITED HER COUSINS ALONG ON
GUARD/CARD	SHE SAID HELLO TO SHE PAID ALOT FOR
GOAT/COAT	SHE HURRIED TO FEED SHE WANTED TO WEAR
GOAL/COAL	SHE HAD INTERCEPTED THE SOCCER BALL JUST INCHES FROM SHE HAD BEEN COVERED WITH GRIMY DUST AFTER SHOVELING
GOLD/COLD	SHE TOLD THE CLIENT TO SELL ALL SHE CLOSED THE WINDOW TO KEEP OUT

TABLE 5. SENTENCE CONTEXTS AND WORD PAIRS. THE VOICED SENTENCE CONTEXT APPEARS FIRST FOR EACH WORD PAIR. EXPERIMENT 3.

very compatible; 6 = very incompatible). As expected, a voiceless word was rated as more compatible with a voiceless bias sentence context than a voiced bias sentence (1.00 and 5.66, respectively). Similarly, a voiced word was rated as very compatible in the voiced context compared with a voiceless word (1.06 and 5.86, respectively).

The rating values for individual sentence context/stimulus combination are shown in Table 6. The reliability of the observed differences was tested in a two-way ANOVA with word (voiced or voiceless) and sentence bias (voiced or voiceless) as factors. As expected, a significant interaction of word and sentence bias was found ($F(1,14) = 1061.2$, $p < .001$).

A voicing series was prepared for each word pair using the procedure in Experiment 2. A subset of ten stimuli for each voicing series was chosen based on pretesting. The VOT values of the stimuli included in the experiment are listed in Table 7. Table 8 lists the duration of each of the sentence contexts, the target words and 'the' plus the closure.

Procedure. Each of the ten stimuli in the 7 voicing series was combined with the context sentences for a given target word pair (2 context sentences x 10 stimuli per voicing series x 7 voicing series = 140 sentence context/stimulus series combinations. Five randomizations of the entire sentence context/stimulus series set were recorded on audio tape. Over the course of 3 days of testing,

	SENTENCE BIAS	
	VOICED	VOICELESS
DENT	1.4	5.6
TENT	6.0	1.3
DIME	1.06	6.0
TIME	5.9	1.06
DRIP	1.3	5.9
TRIP	5.8	1.13
GUARD	1.2	5.7
CARD	5.5	1.3
GOAT	1.2	6.0
COAT	6.0	1.03
GOAL	1.0	5.8
COAL	5.6	1.06
GOLD	1.13	5.9
COLD	5.9	1.0

TABLE 6. PLAUSIBILITY RATINGS FOR WORD/SENTENCE CONTEXT COMBINATIONS.
1 = VERY PLAUSIBLE, 6 = VERY IMPLAUSIBLE.
EXPERIMENT 3.

DENT/TENT	14,18,23,28,33,38,43,48,55,60
DIME/TIME	15,20,24,29,34,40,46,50,55,60
DRIP/TRIP	68,74,82,86,92,98,105,111,115,120
GUARD/CARD	30,35,42,47,55,62,68,73,80,86
GOAT/COAT	27,33,39,47,53,60,66,73,79,87
GOAL/COAL	19,25,32,36,41,45,52,57,62,66
GOLD/COLD	30,36,39,43,48,53,57,62,67,72

TABLE 7. VOICE-ONSET-TIME VALUES OF STIMULUS CONTINUA. EXPERIMENT 3.

WORD PAIR	CONTEXT DURATION VOICED	VOICELESS	DURATION 'THE' + WORD
DENT/TENT	1466	1299	560
DIME/TIME	3319	3436	642
DRIP/TRIP	3142	2931	730
GUARD/CARD	1434	1397	812
GOAT/COAT	1346	1402	660
GOAL/COAL	3158	2999	513
GOLD/COLD	1760	1792	605

TABLE 8. SENTENCE CONTEXT AND TARGET WORD DURATIONS. EXPERIMENT 3.

each subject heard 8 repetitions of each context/stimulus combination (three of the randomizations were presented twice on separate days of testing). Trials were separated by 3 seconds of silence and a rest break was available after every 10 trials. Instructions were identical to Experiment 2.

Results and Discussion

Identification functions. The percentage of voiced responses as a function of point along the continua for the seven voicing series is shown in Figure 5. The identification functions clearly show a greater percentage of voiced responses in the voiced sentence context compared with the voiceless sentence context and this effect is localized in the boundary region. The VOT value at the boundary region was determined employing the linear regression analysis used in the previous experiments. This analysis confirms the observed difference in the identification functions in that a relatively longer boundary value was found for the voiced sentence context than the voiceless sentence context (60.92 msec vs. 58.71 msec, respectively). The effect of sentence context (2.21 msec) was highly significant by subjects ($F(1,19) = 19.09$, $p < .001$) and by items ($F(1,19) = 19.1$, $p < .001$). The introduction of a larger stimulus set was somewhat successful in increasing the magnitude of the identification function shift (compared to Experiment 2).

The items analysis also showed a significant interaction between

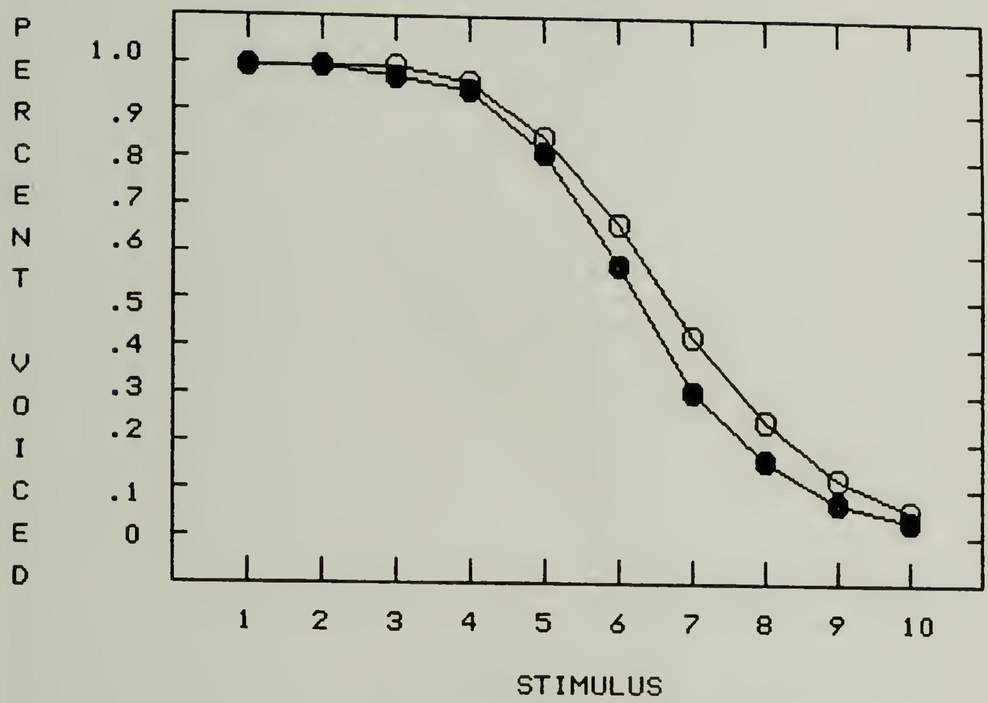


FIGURE 5. PERCENTAGE OF VOICED RESPONSES AS A FUNCTION OF STIMULUS AND SENTENCE CONTEXT. OPEN CIRCLES INDICATE VOICED BIAS (/D/, /G/); FILLED CIRCLES INDICATE VOICELESS BIAS (/T/, /K/). EXPERIMENT 3.

WORD PAIR	SENTENCE BIAS VOICED	VOICELESS	ISOLATION
DENT/TENT	44.9	43.2	38.86
DIME/TIME	43.0	41.7	37.17
DRIP/TRIP	105.2	104.0	92.7
GUARD/CARD	74.3	70.3	60.0
GOAT/COAT	52.3	52.6	44.5
GOAL/COAL	47.6	45.6	41.8
GOLD/COLD	58.9	53.3	51.9

TABLE 9. IDENTIFICATION FUNCTION BOUNDARY VALUES FOR SENTENCE CONTEXTS AND IN ISOLATION. EXPERIMENT 3.

stimulus series and sentence context ($F(6,114) = 6.27$, $p < .001$). The boundary values are listed separately for each continuum in Table 9. Table 9 also lists the category boundary value obtained from presentation of the isolated stimuli in the warmup session. As in Experiment 2, the isolated boundary values are consistently shifted toward the voiced endpoint of the continua relative to the boundary values found in sentential context. The explanation for this phenomonon is not known.

In order to determine which items contributed to the sentence context main effect, an individual one-way ANOVA was performed on each of the seven continua separately. A significant effect of sentence context in the predicted direction was found for GUARD/CARD, ($F(1,19) = 21.81$, $p < .001$), GOAL/COAL, $F(1,19) = 7.02$, $p < .01$), GOLD/COLD, $F(1,19) = 44.17$, $p < .001$). The effect of context was in the predicted direction but not reliable for DENT/TENT, ($F(1,19) = 3.03$, $p < .10$) and DRIP/TRIP ($F(1,19) < 1$). The remaining continuum, GOAT/COAT, showed a small effect of sentence context in the unpredicted direction but was unreliable ($F(1,19) < 1$).

<u>Reaction time functions</u>. The reaction time data are plotted as a function of point along the stimulus continuum in Figure 6. Appendix I shows the identification function and reaction time function for each continuum separately. The reaction time data were analyzed using the quadratic function method described in the preliminary experiment. The subset data analysis resulted in 19 cases (6%) in

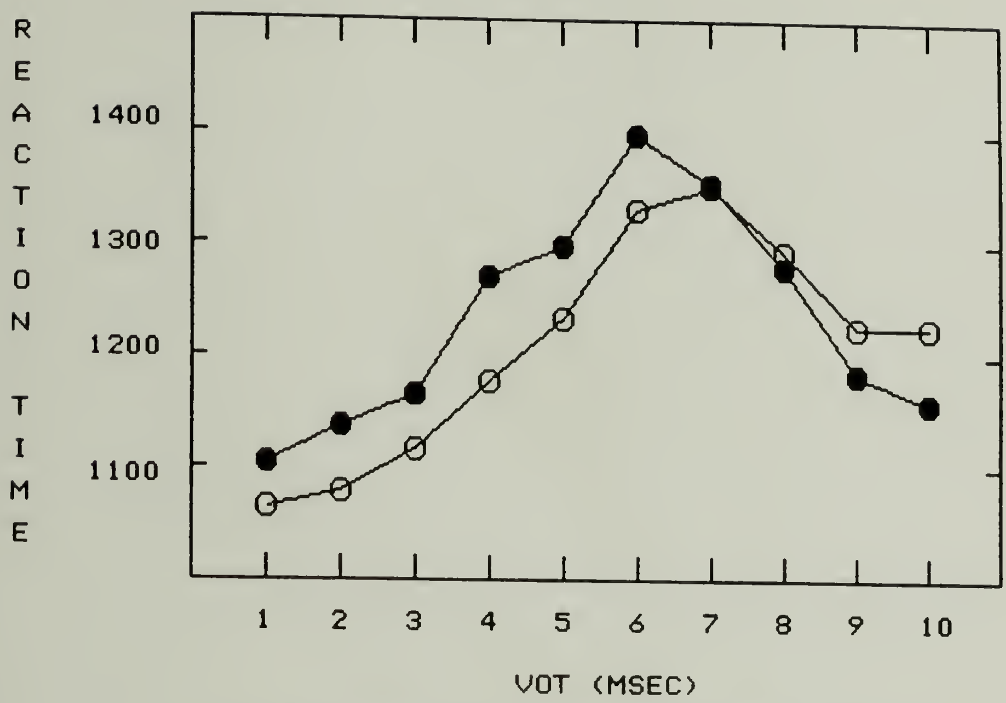


FIGURE 6. REACTION TIMES AS A FUNCTION OF STIMULUS AND SENTENCE CONTEXT. OPEN CIRCLES INDICATE VOICED BIAS (/D/, /G/); FILLED CIRCLES INDICATE VOICELESS BIAS (/T/, /K/). EXPERIMENT 3.

which spurious true peaks were eliminated. The eliminated true peaks averaged 4 steps from the boundary and ranged from 7 to the left and 6 to the right of the boundary stimuli. Similarly to the previous experiments, some reaction time peaks could not be sensibly estimated based on the obtained quadratic function. A total of sixty (20%) of the peaks were estimated using the alternative procedure described in the preliminary experiment. Forty-four of these cases involved quadratic functions with positive _a_ coefficients and the remaining cases predicted reaction time peaks beyond the range of stimulus VOT values used in the experiment.

An ANOVA performed on the predicted VOT values associated with the slowest reaction times (voiced bias = 60.6 msec, voiceless bias = 58.5 msec) showed a significant effect for subjects ($F(1,19) = 14.3$, $p < .001$) and by items ($F(1,19) = 14.3$, $p < .001$). However, the items analysis also showed an interaction between stimulus-series and sentence context ($F(6,114) = 6.17$, $p < .0001$). The peak VOT values are shown separately for each stimulus series in Table 10.

Inspection of these values indicates that the main effect of sentence context may be due in large part to the very large effect in one item (GOLD/COLD). A second analysis was performed in which the contribution of this item was removed. Figure 7 shows the reaction time function without the contribution of GOLD/COLD. The effect of sentence context (voiced bias = 60.5 msec, voiceless bias = 59.7 msec) was no longer reliable by subjects ($F(1,19) = 1.5$, $p < .25$) or

WORD PAIR	SENTENCE BIAS	
	VOICED	VOICELESS
DENT/TENT	44.4	43.5
DIME/TIME	44.7	42.7
DRIP/TRIP	105.3	104.5
GUARD/CARD	73.5	70.0
GOAT/COAT	50.7	53.5
GOAL/COAL	44.4	44.2
GOLD/COLD	61.7	51.4

TABLE 10. REACTION TIME PEAK VALUES AS A FUNCTION OF SENTENCE CONTEXT. EXPERIMENT 3.

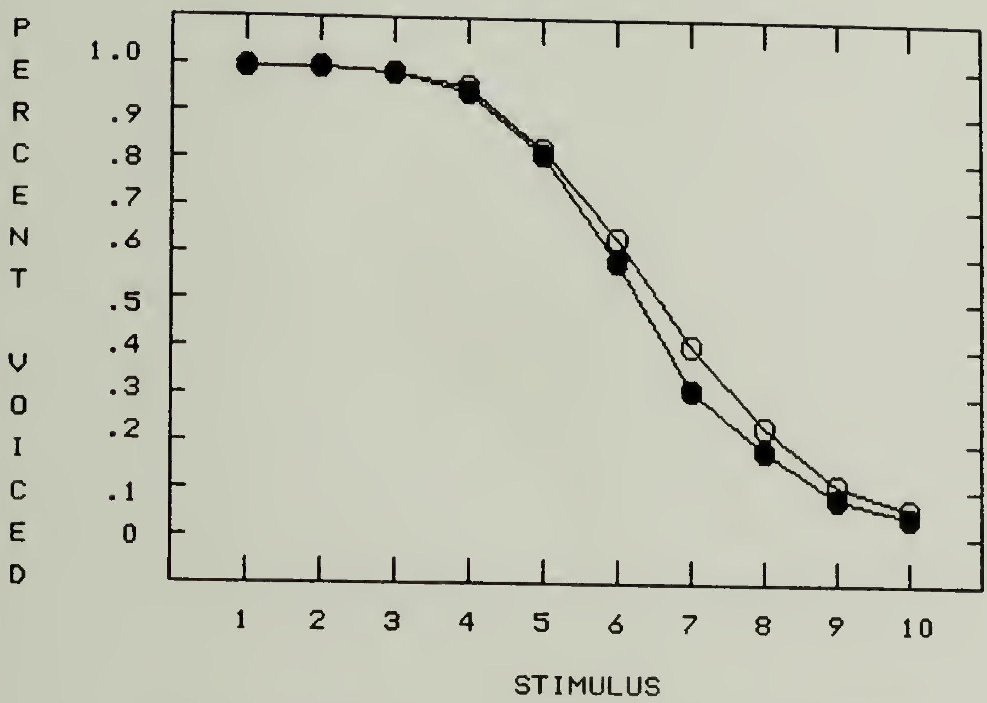


FIGURE 7. REACTION TIMES AS A FUNCTION OF STIMULUS AND SENTENCE CONTEXT WITH THE CONTRIBUTION OF GOLD/COLD REMOVED. OPEN CIRCLES INDICATE VOICED BIAS (/D/, /G/); FILLED CIRCLES INDICATE VOICELESS BIAS (/T/, /K/). EXPERIMENT 3.

by items ($F(1,19) = 1.52$, $p < .23$). In addition, the items analysis no longer showed a reliable items x sentence context interaction ($F(5,95) = 1.69$, $p < .15$). It should be noted that in an analysis of the boundary values from the identification functions with this item removed, the ANOVA still shows a highly reliable context effect by subjects (voiced bias = 61.62 msec, voiceless bias = 59.6 msec, $F(1,19) = 11.94$, $p < .01$) and by items ($F(1,19) = 11.86$, $p < .01$). Figure 8 shows the identification function for the stimuli without the contribution of the GOLD/COLD stimuli.

The pattern of data seen in the GOLD/COLD stimuli deserve further comment. Inspection of the reaction time data for this continuum (see Appendix 1) indicates that the voiced sentence context reaction time function did not show a reliable peak. Rather, the reaction time function for the stimuli in the voiced sentence context was relatively flat and elevated toward the voiceless endpoint compared with the voiceless sentence context. While the peak reaction time analysis shows an apparent peak shift, inspection of the reaction time function for GOLD/COLD indicates in fact no apparent peak in the voiced bias sentence context.

Consistent and inconsistent reaction times. The reaction time data at the category boundary were partitioned into consistent and inconsistent stimulus responses for comparison and are shown in Table 11 for each sentence bias condition. Consistent responses are only 7 msec faster than inconsistent responses. In both sentence biases, an

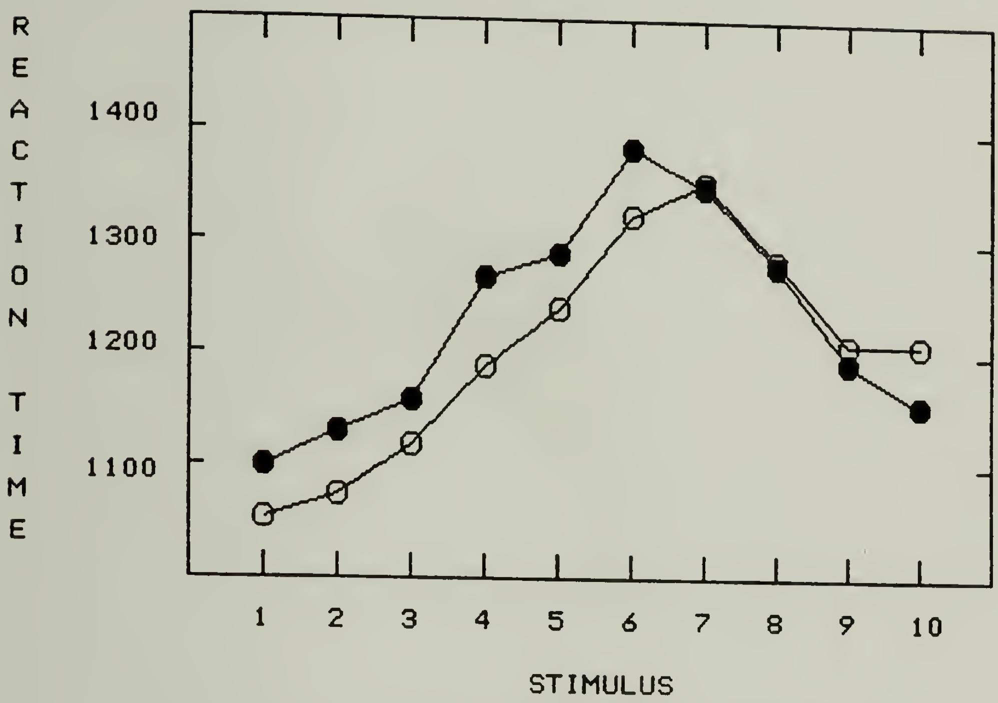


FIGURE 8. PERCENTAGE OF VOICED RESPONSES AS A FUNCTION OF STIMULUS AND SENTENCE CONTEXT WITH THE CONTRIBUTION OF GOLD/COLD REMOVED. OPEN CIRCLES INDICATE VOICED BIAS (/D/, /G/); FILLED CIRCLES INDICATE VOICELESS BIAS (/T/, /K/). EXPERIMENT 3.

		SENTENCE BIAS	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1458 (46)	1563 (52)
	INCONSISTENT	1496 (54)	1538 (48)

TABLE 11. CONSISTENT AND INCONSISTENT REACTION TIMES AT THE CATEGORY BOUNDARY AS A FUNCTION OF SENTENCE CONTEXT. EXPERIMENT 3.

advantage for voiced responses (31 msec) is observed: a small advantage for a consistent response (voiced response) in a voiced bias sentence (38 msec) and an inconsistent response (voiced response) in a voiceless bias sentence (25 msec). The consistency main effect was unreliable ($F < 1$) in a three way ANOVA (continuum X sentence bias X consistency) as was the interaction of bias X consistency ($F(1,19) = 1.9$, $p < .18$). The ANOVA did show, however, a reliable advantage (73 msec) for responses made in the voiced bias contexts relative to the voiceless bias context ($F(1,19) = 5.66$, $p < .03$). There was a reliable main effect of continuum ($F(6,114) = 8.9$, $p < .001$). Continuum did not interact with sentence bias or consistency, nor was there a three way interaction.

Overall, the data indicate an advantage (unreliable) for voiced responses at the category boundary in both voiced and voiceless sentence biases. Inconsistent responses at the category boundary in the voiceless bias context suggest an advantage for responses favored by the underlying acoustic information at the expense of the sentential bias. As described in Experiment 1, the advantage for voiced responses is attributed to an artifact of the technique used to create the stimuli. An analysis of the endpoint reaction times supports this conclusion.

The reaction time to stimuli at the endpoints of the continua for each sentence bias are shown in Table 12. A three-way ANOVA (endpoint X consistency X continuum) shows that the advantage for voiced

		SENTENCE BIAS	
		VOICED	VOICELESS
	VOICED	1061	1157
ENDPOINT			
	VOICELESS	1102	1215

TABLE 12. ENDPOINT REACTION TIMES AS A FUNCTION OF SENTENCE CONTEXT. EXPERIMENT 3.

responses (104 msec) was highly significant ($F(1,19) = 21$, $p < .001$). The effect of consistency (49 msec advantage for consistent responses) was also reliable ($F(1,19) = 6.4$, $p < .01$). The advantage for consistent responses was quite similar for the voiced and voiceless sentence biases (41 and 58 msec, respectively). The main effect of continuum was reliable ($F(6,114) = 16.7$, $p < .001$) and continuum interacted with endpoint ($F(6,114) = 4.07$ $p < .001$) and consistency ($F(6,114) = 4.08$, $p < .001$). No other interactions were reliable.

Summary. The influence of sentence context on phoneme identification was replicated in Experiment 2. In addition, a reliable reaction time peak shift was found. However, the reaction time peak shift was primarily due to one item. Endpoint response times showed an advantage for consistent responses, however, reaction times to stimuli at the category boundary showed an advantage for voiced responses, regardless of sentence bias. The reaction time data support a post perceptual explanation for the observed context effect on phoneme identification.

Descriptive materials analysis

Although the effects of the boundary and endpoint stimulus as a function of response category did not interact with continua, it is useful to examine the patterns of data exhibited for the separate continua. The endpoint and boundary values are shown in Appendix II separately for the seven word pairs. In this section, the boundary and endpoint patterns are considered jointly for each sentence bias/stimulus continuum.

Inspection of the patterns of data exhibited for the endpoint and boundary reaction times reveal three groups of effects. In the first group, endpoint and boundary patterns taken together show an advantage for consistent responses at the endpoints in conjunction with an advantage for inconsistent responses at the boundary. This group includes 5 voiced bias sentence contexts (DRIP-TRIP, GUARD-CARD, GOAT-COAT, GOLD-COLD and DIME-TIME) and 3 voiceless sentence contexts (DTENT, GUARD-CARD and GOAL-COAL). For example, the DRIP-TRIP continuum showed a 45 msec advantage for a voiceless endpoint response in a voiceless context, that is a consistent response was facilitated. The boundary responses showed a quite different pattern of data however, in that voiced responses were facilitated (56 msec) relative to voiceless responses.

Interestingly, the boundary pattern of data appear to reflect the acoustic information in the stimulus. The boundary stimulus for a voiced context is, in fact, slightly toward the voiceless end of the continuum. The pattern of reaction times suggest sensitivity to the acoustic information in the stimulus at the boundary despite a contextual bias in the opposite direction.

A second group of joint boundary and endpoint reaction time effects include 3 cases in which a consistent response is faster at the boundary and at the endpoint. Two voiceless sentence biases (DENT-TENT and GOAL-COAL) and one voiced sentence biase (GOLD-COLD) show this pattern. For example, the DENT-TENT continuum shows an advantage for consistent responses at the boundary (177 msec) and at the endpoint (119 msec) when this continuum is heard in a voiceless bias sentence context. The size of the consistent response advantage is fairly symmetrical for DENT-TENT. This is less the case for GOLD-COLD and GOAL-COAL. These continua show a much larger consistent response advantage at the boundary than at the endpoint. The general trend, however, points to fairly symmetrical effects at the boundary and the endpoint. This is reinforced by the observation that the same pattern is seen in a fourth case, the voiced sentence bias responses for BATH-PATH (Experiment 1).

The remaining three continua (GOAT-COAT, DRIP-TRIP and DIME-TIME, voiced sentence bias) show facilitation for consistent responses at the boundary but an advantage for inconsistent responses at the endpoints. The effect at the category boundary in conjunction with the lack of an effect at the endpoint may constitute some evidence for an interactive influence of sentence context on speech analysis for ambiguous stimuli . The advantage for a consistent response at the boundary cannot be attributed to acoustic factors or by a general advantage for consistent responses shown also at the endpoint. Consideration of the particular biasing contexts do not point to any obvious characteristics to distinguish these three contexts from the those in the entire set. In any case, the bulk of the data appear to be more naturally explained by a simple acoustic bias (group 1) or a general bias for a response consistent with the context that is independent of the quality of the acoustic information (group 2).

CHAPTER V

SUMMARY AND CONCLUSIONS

In this chapter, the major results of the three experiments reported here will be reviewed. In addition, the interactive mechanism supported by the lexical effect data and the modular mechanism supported by the sentence context effect data will be outlined. A second section will consider the implications of structuring the auditory input for models of auditory word recognition and lexical access.

Summary of Results

Experiments 2 and 3 found that sentential context biases the identification of ambiguous word initial phonemes. The sentence context effect was demonstrated by the observed shift in the identification function category boundary in a voicing series, that was dependent upon the pragmatic sentential context. In a sentence context pragmatically biased toward the voiceless end point, the identification function boundary tended to shift towards the voiced end point of the continuum relative to the identical speech series heard in a sentence context pragmatically biased toward the voiced end point.

The observed influence of sentence context on phoneme identification is attributed to a non-interactive mechanism. This is based on a comparison of the reaction time functions associated with the identification responses in the voiced and voiceless sentence bias. An analysis of the difference in location of the peaks (slowest reaction times) of the voiced and voiceless sentence bias showed a smaller (non-significant) shift relative to the reliable identification function boundary differences. Response times to voiced and voiceless stimuli at the category boundary showed an advantage for responses consistent with the context that was also found at the continuum endpoints (Experiment 2) or an advantage for voiced responses regardless of sentence context (Experiment 3). This pattern of data is inconsistent with a mechanism in which sentential context directly influences analysis of the incoming sensory evidence.

The sentence context pattern of results contrasts with Experiment 1 in which the role of lexical status in speech perception was investigated. In this experiment, identification of ambiguous sounds from a voicing continuum was biased such that a word, as opposed to a pronounceable non-word was formed. The lexical effect was demonstrated by comparing the identification boundary value for a voicing series in which the voiced end point formed a word (DICE-TICE) with a series in which the voiceless end point formed a word (DYPE-TYPE). The boundary value for the voiced lexical anchor identification function was shifted away from the voiced end point relative to the voiceless lexical anchor identification function. The reaction time functions associated with the identification functions showed a clear and significant shift in the location of the peak reaction time. Moreover, the reaction time peak shift was slightly larger than the identification function shift. An analysis of word and non-word responses showed a reliable word advantage only for the category boundary stimuli. Virtually no advantage for a word response was found at the continua end points.

The lexical effect results are attributed to an interactive network between word and sound level representations. This network allows lexical knowledge, specifically the prestored phonological form of an item, to influence the interpretation of an acoustic event. The specific mechanism for this interaction is cast in terms of a connectionist model framework. It is assumed that a positive feedback connection exists between a word level node and component sound level nodes. This mechanism allows positive evidence at one level to influence nodes at another level.

Other researchers working within the connectionist model framework have proposed that positive feedback connections exist between a representation of sentence meaning and sound levels (cf. Feldman & Ballard, 1975). The evidence presented here suggests such feedback is not available at the perceptual level. This considerably constrains interactive theories of auditory word recognition by allowing only a small class of linguistic knowledge to influence perceptual processing. Sentence context can be used to decide among lexical hypotheses but may not directly contribute to the development of the lexical hypotheses. This contrasts with prestored phonological form of an item which may indeed contribute positively to a sound level representation.

Word Level Structure

The evidence presented in Experiment 1 in support of the lexical storage hypothesis focussed on a restricted class of lexical items in which the target phoneme occured at the beginning of a single morpheme lexical item. In this section, the implications of syllable and lexical structure are explored. First, the sequential order relationship between sounds and the evidence for hierarchical organization of syllables is discussed. It is argued that syllable structure may be used to maintain position information of sounds during processing.

In a second section, the implications of stored lexical units other than single syllable, monomorphemic words, such as those used in Experiment 1, are examined. The items considered include both larger lexical units (idioms) and sublexical units (prefixes and root morphemes). The consequence of enriched structural assumptions concerning lexical representations are outlined within the lexical storage hypothesis framework. Finally, a model of auditory word recognition incorporating syllable and morpheme structuring is outlined.

Sequential order. The predictions of the lexical storage hypothesis were tested using a limited case in which the target phoneme occurred at the beginning of a word. As described, the positive feedback connections for this class of lexical items predicted a shift in the identification function as well as a peak reaction time shift. Additional predictions included no reaction time advantage for word responses at the end points of the continua. At the end points, the quality of the auditory information is sufficient for initiation of a response prior to any positive feedback from the lexical level.

A different reaction time prediction emerges for the end point stimuli if the target phoneme occurs in the final position of a word. Under these conditions, positive feedback from the lexical level will have an opportunity to accrue at the sound level prior to analysis of the input relevant to that sound. Here, facilitation for word relative to nonword responses would be expected at the end points.

A consideration of the sequential position of a sound segment in a word implies that sounds are represented in a form that preserves the relationship among the sounds in the string. One possible way to encode positional information in words and syllables is to assume a linear sequence of sounds. Under this assumption, a sound is assigned a label simply indicating its sequential position. However, theories of syllable structure postulate a hierarchical representation in which phoneme sequences are organized into subunits (cf. Halle & Vargnaud, 1980). According to this analysis of syllable structure, the syllable is composed of two major constituents, the onset, which is optional, and the rime. These constituents correspond respectively to the initial consonant or consonant cluster and the remainder of the syllable. The rime is composed of two additional units, the peak or obligatory vowel, and the coda, the optional final consonant or consonant cluster. These units organize the sequential order of the string since each unit occupies an identifiable and unique position in the syllable.

A representation of syllable structure could be used quickly to begin to structure the incoming signal, even under conditions in which the acoustic information is not sufficient to uniquely identify it. Conditions of insufficient acoustic information include simply ambiguous input, as was the case in the experiments reported here, as well as cases in which a decision rests on acoustic information occurring later in the syllable. The latter case applies to instances of co-articulation in which, for example, information in a vowel is relevant to the interpretation of an immediately preceding consonant. Under these conditions, the initial segment cannot be labeled immediately and the processing system must wait for additional information. Similar conditions exist for incorporating rate of production information in speech perception. It has been found that the rate of production of a syllable influences the interpretation of temporal cues for speech sounds (cf. Miller, 1981). Thus, category decisions concerning syllable initial sounds in which temporal parameters (i.e. voice-onset-time) are relevant to the sounds identity, must await information later in the syllable.

There exists some psychological evidence for this view of syllable structure. Treiman (1983) found evidence for hierarchical syllable structure in adults learning novel word games. Subjects preferred rules that respected the onset and rime as units. Rules that maintained the onset and rime constituents were learned more readily than rules that break up the constituents. These results are consistent with a view of syllable structure as hierarchically organized rather than a representation consisting of a concatenated string of phonemes.

Morpheme structure. A major assumption of the dominant view of auditory word recognition, cohort theory, is that sequential left to right analysis of the input results in activation of compatible lexical entries. An alternative view has been developed by researchers in visual word recognition. Taft & Forster (1976) propose a model of lexical access in the visual domain in which the internal lexicon is accessed via stem morphemes. Taft & Forster adopt the peripheral access file system assumed in Forster's (1976) autonomous model of lexical access. Each lexical entry listed in the access file system corresponds to a stem morpheme. Lexical access of prefixed words proceeds by isolating the stem from the prefix. For example, DETACH would be accessed via the stem TACH.

In contrast to cohort theory, Taft & Forster claim that the input is structured prior to access of lexical items. The lexicon is organized in a fashion that captures the regularities in lexical structure, in particular, morphemic structure. Morphological organization of the lexicon would facilitate lexical access particularly of prefixed words which begin with frequent phoneme sequences (Knuth, 1973). Instead of locating DETACH in the list of entries beginning with the frequent phoneme sequence DE, DETACH could be located more quickly under the less frequent sequence TACH. Support for this view of morphological structuring in the visual domain is the finding that prefixed words show an advantage over pseudoprefixed words in a number of tasks (lexical decision, naming, eye fixation duration, see Lima, 1985 for an extensive review). This effect is attributed to an initial unsuccessful access attempt via the stem for pseudoprefixed words.

It is possible that frequently occurring syllables such as prefixes require only low level pattern matching operations. Some models of lexical access have suggested a similar processing distinction between open class (nouns, verbs ect) and closed class (articles, conjunctions, prepositions) vocabularies. In the visual domain, the grammatical distinction between open and closed class items is emphasized. The division between the two classes is assumed to be reflected in the organization of the lexicon (Bradley, 1978). Closed class items are accessed more efficiently perhaps because the relatively few number of items in this class can be searched quickly.

In the auditory domain, Grosjean & Gee (1985) have cast the distinction between open and closed class vocabularies in terms of relative degree of stress (or emphasis). Typically, closed class words, at least monosyllabic items, receive a lesser degree of stress than open class words. Similarly, prefixes typically receive less emphasis than stem morphemes. They argue that an initial attempt is made to analyze all low stress syllables via a low level, pattern matching process. Stressed syllables are used to activate items in the lexicon.

Morphological structuring, particularly into prefix and stem components, has a number of implications for models of auditory word recognition. This view suggests that access to the lexicon does not proceed simply on the basis of a left to right sequential analysis of the input. Rather, an initial attempt is made to match the input against a list of frequent syllables. Prefixes as well as closed class items which are relatively unstressed in on-going speech may be quickly identified in this fashion. If the identified syllable is a prefix, the remaining root morpheme is used to activate the entry containing complete information concerning the lexical item.

It is useful to consider the consequence of a morphologically organized lexicon for the lexical storage hypothesis. If morpheme units are the crucial units for lexical access and represented as such in the lexicon, then this suggests that the unit of feedback from the lexicon is in fact the morpheme.

Idioms. The preceding section presented evidence consistent with structure within words and syllables. Lexical structure above a single word level is also evident in language. Consider, for example, common, frozen phrases such as idioms (e.g. kick the bucket) which have meanings that do not consist of an integrated interpretation of the individual word meanings. It would be efficient to exploit the predictability of the words in such phrases by representing these phrases as separate entities in the lexicon.

Assuming idioms are stored as entries in the lexicon, idiomatic expressions present a means to explore and generalize the lexical storage hypothesis. The lexical storage hypothesis predicts that interactive principles will apply between identification of auditorily presented idiomatic expressions and acoustic analysis. In a fashion similar to monomorphemic lexical items, positive feedback from the phonological representation stored with an idiom contributes to the evidence for a component sound segment.

Swinney & Cutler (1979) have argued that idiomatic expressions are stored as lexically represented phrases in the lexicon. Idioms are stored as separate entries and retrieved in the same fashion as other words. This hypothesis was investigated by presenting idioms and matched non-idiom counterparts (e.g. lift the bucket) for a meaningfulness judgement. It was found that subjects were faster to make a meaningfulness judgement for the idiomatic phrases relative to the non-idiom controls. Swinney & Cutler reasoned that the non-idiomatic phrases required additional time for lexical access of each item in the phrase and the computations involved in establishing the syntactic and semantic relationships between the words.[10]

It is useful to consider the consequence of an interactive mechanism in identifying idioms for the identification and reaction time function paradigm developed here. One possible means to test the hypothesis is to compare the identification of a voicing series in a context forming an idiom at the voiced end point with a context forming an idiom at the voiceless end point of the voicing continuum. This design is comparable to the word and non-word series employed to demonstrate the lexical effect. The positive feedback mechanism predicts a shift in identification functions such that a voiced idiom end point series would show relatively more voiced responses than a voiceless idiom end point series. The positive feedback mechanism also predicts a shift in the peak reaction time values for the two continua comparable to the size of the identification function shift.

If the lexical item containing the target phoneme occurs as the final word in the idiomatic expression, then additional predictions emerge for unambiguous stimuli (stimuli at the end points of the continuum). Under these conditions, feedback from the lexical level is able to accrue sufficiently for portions of the sequence occurring temporally late in the phrase. This predicts facilitation for continuum-endpoint idiom responses. Further, since feedback accrues for targets occurring late in the sequence of words in the idiom, facilitation is predicted for phoneme targets in both word-initial and word-final (within word) positions for words occurring late in the sequence.

Implications for Auditory Word Recognition

The present results demonstrate the influence of contextual factors (lexical status and sentence context) in speech perception. These findings join a large body of research that focus on the role of linguistic context in perceptual processing. The experiments extend the research exploring context effects by identifying a principled class of information that is exploited directly during auditory word recognition. It is proposed that prestored lexical knowledge is used directly to influence analysis of speech. This contrasts with a post-perceptual use of sentential context to choose among lexical hypotheses. In addition to exploiting prestored phonological knowledge, it is suggested that prestructuring the incoming information will facilitate recognition processes. Emphasis is given to arguments for organizing the auditory input into morphemes (in particular, prefixes and stems) and syllable constituent structure.

A model of auditory word recognition. A mechanism that accounts for an interactive use of prestored phonological form is outlined in terms of connectionist models. Lexical status influences the rate of accrual of evidence for a given sound segment. Stored with each lexical entry is a phonological representation. The prestored phonological representation is organized according to syllable constituent structure. As the acoustic information is received, sound segments are hypothesized and assigned labels according to legal syllable constituent structure. The structured information is reinforced or de-activated based on the feedback available from the prestored structure of hypothesized lexical items.

If the syllable is unstressed, an initial attempt is made to match it against a subset of frequently occurring syllables. A stressed syllable begins activation of a set of lexical hypotheses. The lexical hypotheses, in turn, are used to reinforce hypothesized sound segments.

This sketch of a model of auditory word recognition incorporates word structure and a constrained interactive mechanism. Many of the assumptions of the model are untested. However, existing evidence for morphological structuring in the visual domain, syllable constituent structure and processing distinctions between stressed and unstressed syllables points to the utility of incorporating enriched lexical representations.

FOOTNOTES

1. Additional assumptions by the categorical perception model could incorpoprate the relative degree of acoustic information in the signal for a sound to account for the lexical effect seen only at the category boundary. Such assumptions, however, are *ad hoc* given the categorical model as stated by the major proponents of the model.

2. Some recent word (Fox, 1983) has used reaction time to argue for hierarchical models as explanatory mechanisms for the lexical effect. Fox replicated the lexical effect and looked at identification functions for slow, medium and fast response times. He found that for a given continuum, the slowest response times corresponded to an identification function in which relatively more word responses are reported compared to the fastest time range. Fox argued that an interactive model would result in a comparable lexical effect in all response time ranges. If, however, we assume that the extreme pressure on subjects to respond quickly results in some proportion of the trials in which subjects simply guess, very fast responses will include more guesses and will result in an overall smaller lexical effect. This predicts that even the fast reaction times would show some lexical effect compared to a baseline boundary series without a lexical bias. Although Fox did

include a series with non-words at both endpoints, the appropriate statistical comparisons were not performed to determine if a lexical effect did exist for lexically biased continua even for the fast reaction times. Inspection of these data does indicate a lexical effect compared with the unbiased continuum.

3. The data obtained for one stimulus (40 msec VOT from the DYPE-TYPE series) had to be discarded. An error was made in specifying the VOT value during synthesis of this stimulus and was not discovered until after the experiment was completed. The data for both lexical anchor series at this VOT value are not included in any of the analyses reported here.

4. Derivation of closed form expressions.

p_i = predicted point i

o_i = observed point i

$$p_i = ax_i^2 + bx_i + c$$

SS = Sum of squared deviations

$$= E_i\,(o_i - p_i)^2$$

$$= E_i\,[o_i - (ax_i^2 + bx_i + c)]^2$$

Taking partial derivative with respect to a:

$$SS/a = 0 - 2Eox^2 + 2Ex^2(ax^2 + bx + c)$$

Setting equal to 0:

$$Eox^2 = Ex^2p$$

Taking partial derivative with respect to b:

$$SS/b = 0 - 2E(ox) + 2Exp$$

Setting equal to 0:

$$Eox = Exp$$

Taking partial derivative with respect to _c_:

$$SS/c = 0 - 2Eo + 2Ep$$

Setting equal to 0:

$$Eo = Ep$$

Replace o

$$\begin{aligned} Eox^2 &= Ex^2(ax^2 + bx + c) \\ &= 2Ex^4 + bEx^3 = cEx^2 \\ Eox &= Ex(ax^2 + bx + c) \\ &= aEx^3 + bEx^2 + cEx \\ Eo &= E(ax^2 + bx + c) \\ &= aEx^2 + bEx + cn \end{aligned}$$

5. The peak reaction time VOT value was derived from the obtained quadratic function using the following procedure. The equation that describes a quadratic function is $Y = aX^2 + bX + c$. The peak of the function is defined as the point at which the slope of the function is equal to 0. The value of X at this point is solved by differentiating with respect to Y and with respect to X. (differentiating a quantity squared (X^2) = 2 times the quantity; differentiating a quantity = 1). This yields $2ax + b = 0$. Solving for X: $X = -b/2a$.
6. An analysis in which the entire reaction time function was performed but this analysis grossly misrepresented the shape of

the reaction time functions. This was primarily because the tails of the reaction time functions tend to flatten, unlike quadratic functions.

7. The isolation boundary is based on the data from 18 subjects. Due to experimenter error, data from the warmup session for two subjects were lost.

8. An eighth word pair, DRAIN-TRAIN, was used in the experiment but was not included in any data analysis. We were not successful in creating stimuli from this word pair that were clearly perceived as DRAIN and TRAIN.

9. It is important to note that the stimuli used in the lexical effect experiment were synthesized voicing continuum. The data discussed here are based on stimuli created by manipulating natural speech. Differences in locating the peak reaction time of a continuum based on this superficial aspect of the stimuli will require future research. The most obvious test would be to replicate the lexical effect with natural speech continua.

10. The lexical storage hypothesis would appear to predict faster response times for idioms in this experiment due to facilitated processing of the acoustic information. However, it is likely that the task used in these experiments is not sensitive enough or the response is initiated at a late stage in which these effects are masked by other factors.

REFERENCES

Abbs, J.H. & Sussman, H.M. (1971). Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. _Journal of Speech and Hearing Research_, _14_, 23-36.

Best, C.T., Morrongello, B. & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. _Perception & Psychophysics_, _29_, 191-211.

Blumstein, S.E., & Stevens, K.N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. _Journal of the Acoustical Society of America_, _67_, 648-662.

Bradley, D.C. (1978). Computational distinction of vocabulary types. Unpublished doctoral dissertation. Massachusetts Institute of Technology.

Cairns, H.S., Cowart, W., & Jablon, A. (1981). Effects of prior context upon integration of lexical information during sentence comprehension. _Journal of Verbal Learning and Verbal Behavior_, 20, 445-453.

Cairns, H.S. & Kamerman, J. (1975). Lexical information processing during sentence comprehension, _Journal of Verbal Learning and Verbal Behavior_, _14_, 170-179.

Cole, R.A. (1973). Listening for mispronunciations: A measure of what we hear during speech, _Perception and Psychophysics_, _11_, 153-156.

Cole, R.A. & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), _Strategies of Information Processing_. London: Academic Press.

Cole, R.A. & Jakimik, J. (1979). A model of speech perception. In R.A. Cole (Ed.), _Perception and Production of Fluent Speech_. Hillsdale, N.J., LEA.

Cole, R.A. & Perfetti, C.A. (1980). Listening for mispronunciations in a childrens story: The use of context by children and adults. _Journal of Verbal Learning and Verbal Behavior_, _3_, 297-315.

Cole, R.A. & Rudinicky, A. (1983). What's new in speech perception? The research and ideas of William Chandler Bagley, 1874-1946. _Psychological Review_, _90_, 94-101.

Cotton, S. & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. Perception & Psychophysics, 35, 41-48.

Connine, C. & Clifton, Jr, C. (in preparation). Interactive use of lexical information in speech perception.

Delattre, P.C., Liberman, A.M., & Cooper, F.S. Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of Americal, 68, 769-773.

Dell, G.S. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. Cognitive Science, 9, 3-23.

Dorman, M.F., Raphael, L.J., & Liberman, A.M. (1979). Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society, 65, 1518-1532.

Feldman, J.A., & Ballard, D.H. (1982). Connectionist models and their properties. Cognitive Psychology, 6, 205-254.

Fischler, & Bloom, (1979). Automatic and attenticnal processes in the effects of sentence contexts on word recognition. Journal of Verbal Learning and Verbal Behavior, 18, 1-20.

Fodor, J.A. (1983). The Modularity of Mind. MIT Press, Cambridge, Massachusetts.

Fodor, J.A., Bever, T. & Garrett, M. (1974). The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. New York: McGraw-Hill.

Forster, K. (1978). Accessing the mental lexicon. In E. Walker (Ed.), Explorations in the Biology of Language. Vermont: Bradford Books, 139-174.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper & E.C.T. Walker (Eds.), Sentence Processing: Psycholinguistic studies presented to Merrill Garrett. Hillsdale, N.J. LEA.

Forster, K. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. Quarterly Journal of Experimental Psychology, 33A, 465-495.

Fox, R.A. (1984). Effect of lexical status on phonetic categorization. Journal of Experimental Psychology: Human

Perception & Performance, 10, 526-540.

Frazier, L (1979). On comprehending sentences: Syntactic Parsing Strategies. (Indiana University Linguistics Club, Bloomington, IN).

Frazier, L., Clifton, C. & Randall, J. (1981). Filling gaps: Decision principles and structure in sentence comprehension. Cognition, 13, 187-222.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences, Cognitive Psychology, 14, 178-210.

Frazier, L. & Rayner, K. (in preparation). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences.

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception & Performance, 6, 110-125.

Garnes, S. & Bond, Z.S. (1975). The relationship between semantic expectation and acoustic information. Phonologica, 295-293.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. Perception & Psychophysics, 28, 267-283.

Grosjean, F. & Gee, J. (in press). Another view of auditory word recognition. Cognition.

Halle, M. & Vergnaud, J.-R. (1980). Three dimensional phonology, Journal of Linguistic Research, 1, 83-105.

Harris, K.S., Bastian, J., & Liberman, A.M. (1961). Mimicry and the perception of a phonemic contrast induced by silent interval: Electromyographic and acoustic measures. Journal of the Acoustical Society of America, 33, 842. (Abstract)

Hogaboam, T.W. & Perfetti, C.A. (1975). Lexical ambiguity and sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 14, 265-274.

Isenberg, D., Walker, E. & Ryder, J. (1980). A top-down effect in the identification of function words. Presented at the Acoustical Society of America Meeting, Nov, Los Angeles, California.

Jakimik, J. (1979). Word recognition and the lexicon. Paper presented at the 97th Meeting of the Acoustical Society of America.

Jusczyk, P. (1984). A review of speech perception. In Kaufman, Thomas, & Boff (Eds.), Handbook of Perception & Performance, New York: Wiley.

Kewley-Port, D. (1980). Representation of spectral change as cues to place of articulation in stop consonants. Unpublished doctoral dissertation, City University of New York.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. Research in Speech Perception, Technical Report 3, Indiana University.

Klatt, D. (1980). Software for a cascade/parallel formant synthsizer. Journal of the Acoustical Society of America, 67, 971-995.

Klatt, D. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R.A. Cole (Ed.), Perception and Production of Fluent Speech, Lawrence Erlbaum, 243-288.

Knuth, D.E. (1973). The art of computer programming., Vol. 3.: Sorting and Searching. Reading, MA: Addison-Wesley.

Liberman, A.M. (1957). Some results of research on speech perception. Journal of the Acoustical Society of America, 29, 117-123.

Liberman, A.M. (1982). On finding that speech is special. American Psychologist, 37, 148-167.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, A.M., Harris, K.S., Kinney, J.A., & Lane, H. (1965). The discrimination of relative onset time of the components of centain speech and nonspeech patterns. Journal of Experimental Psychology, 61, 379-388.

Lima, S.D. (1985). Morphological encoding in the reading of sentences. Unpublished doctoral dissertation. University of Massachusetts at Amherst.

Lisker, L. (1976). Stop voicing production: Natural outputs and synthesized inputs. In Haskins Laboratories: Satus reports on speech research, 1976, SR-47. (ERIC Document Reproduction service No. ED-128-870; NTIS NO. AD A031789).

MacNeilage, P.F. (1970). Motor control of serial ordering of speech. Psychological Review, 77, 182-196.

Marslen-Wilson, W.D. (1973). Linguistic structure and speech shadowing at very short latencies. Nature (London), 244, 522-523.

Marslen-Wilson, W.D. (1975). Sentence perception as an interactive parallel process. Science, 189, 226-228.

Marslen-Wilson, W.D. (1983). Function & process in spoken word recognion: A tutorial review. In Attention & Performance X: Control of Language Processes, LEA.

Marslen-Wilson, W.D. & Tyler, L.K. (1980). The temporal structure of spoken language understanding. Cognition, 8, 1-71.

Marslen-Wilson, W.D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10, 29-63.

Massaro, D. (1981). Sound to representaiotn: An information processing analysis. In The Cognitive Representation of Speech, T. Myers, J. Laver & J. Anderson (Eds.), North-Holland Publishing Co., 181-192.

Meyer, D.E. & Schaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 90, 227-234.

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

Miller, J.L. (1981). Effects of speaking rate on segmental distinctions. In P.D. Eimas & J.L. Miller (Eds.), Perspectives on the Study of Speech, Hillsdale:Erlbaum Associates.

Miller, J.L., Dexter, E.R. & Pickard, K.A. (1984). Influence of speaking rate and lexical status on word identification. Paper presented at the 1984 Fall meeting of the Acoustical Society of America, Minneapolis, Minnesota.

Miller, J.L., Green, K. & Schermer, T. (in press). On the distinction between prosodic and semantic factors in word identification, Perception & Psychophysics.

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

Nooteboom, S.G. (1982) Lexical retrieval from fragments of spoken words: beginnings vs. endings. Journal of Phonetics, 9, 407-424.

Norris, D. (1980) Serial and interactive models of comprehension. Unpublished doctoral dissertation, University of Sussex, England.

Pickett, J.M. & Pollack, I. (1964). Intelligibility of excerpts from fluent speech: effects of rate of utterance and duration of excerpt. Language & Speech, 6, 151-164.

Pickett, J.M. & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: auditory vs. structural context. Journal of Verbal Learning and Verbal Behavior, 3, 79-84.

Pisoni, D. (in press). Acoustic-phonetic representations in word recognition. Cognition.

Pisoni, D. (1971). On the nature of categorical perception of speech sounds. Unpublished doctoral dissertation, University of Michigan.

Pisoni, D. (1978). Speech Perception. In W. Estes (Ed.), Handbook of Learning and Cognitive Processes, Vol. 6, Hillsdale, N.J.:LEA.

Pisoni, D. & Tash, J. (1975). Reaction times to comparisons within and across phonetic categories. Perception & Psychophysics, 15, 285-290.

Pollack, I. & Pickett, J.M. (1964). Intelligibility of excerpts from conversation. Language & Speech, 6, 165-171.

Posner, M.I. & Snyder, C.R. (1975). Attention and cognitive control. In R. Solso (Ed.), Information Processing and Cognition. Hillsdale: Erlbaum Associates.

Rayner, K., Carlson, M. & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. Journal of Verbal Learning and Verbal Behavior, 22, 358-374.

Repp, B. (1983). Categorical perception: Issues, Methods Findings. In Speech and Language: Advances in Basic Research and Practice, Vol. 9, N.J. Lass (Ed.), Academic Press.

Salasoo, A. & Pisoni, D.B. (1982). Sources of knowledge in spoken word identification. Research On Speech Perception Progress

Report, *no. 8*, Indiana University.

Samuels, A.G. (1981a). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474-494.

Samuels, A.G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 1124-1131.

Schuberth, & Eimas, P.D. (1977). Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology: Human Perception & Performance*, *3*, 27-36.

Seidenberg, M., Tanenhaus, M., Leiman, J., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*, 489-537.

Seidenberg, M.S., Waters, G.S., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word perception. *Memory & Cognition*, 1*2*, 315-328.

Shattuck-Hufnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production. *Journal of Verbal Learning and Verbal Behavior*, *18*, 41-55.

Simpson, G.B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*, 120-136.

Spencer, N.J. & Halwes, T. (1981). Phonetic categorization shifts for four continua embedded in fluent semantic contexts. *Journal of the Acoustical Society*, *69*, S115.

Stanovich, K.E. & West, R.F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition*, *7*, 77-85.

Stevens, K.N. & Blumstein, S.E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *55*, 653-659.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)Consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645-659.

Swinney, D. (1982). The structure and time course of information during speech comprehension: Lexical segmentation, access and interpretation. In J. Mehler, E. Walker, & M. Garrett (Eds)., Perspectives on Mental Representation : Experimental and Theoretical Studies of Cognitive Processes and Capacities, Hillsdale, N.J.: Lawrence Erlbaum Associates, 151-167.

Swinney, D. & Cutler, A. (1979). The access and processing of idomatic expressions. Journal of Verbal Learning and Verbal Behavior, 18, 523-534.

Swinney, D., Onifer, W., Prather, P. & Hirschkowitz, M. (1979). Semantic facilitation across sensory modality in the processing of individual words and sentences. Memory & Cognition, 7, 159-165.

Taft, M., & Forster, K. (1976). Lexical storage and retrieval of prefixed words. Journal of Verbal Learning and Verbal Behavior, 14, 638-647.

Tannenhaus, M.K., Leiman, J.M. & Seidenberg, M.S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. Journal of Verbal Learning and Verbal Behavior, 18, 427-440.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. Cognition, 15, 49-74.

Treiman, R. (1984). On the status of final consonant clusters in English syllables. Journal of Verbal Learning and Verbal Behavior, 23, 343-356.

Tyler, L.K. (1980). Serial and interactive-parallel theories of sentence processing. Theoretical Linguistics, 8, 29-65.

Tyler, L.K. (1984). The structure of the initial cohort: Evidence from gating. Perception & Psychophysics, 36, 417-427.

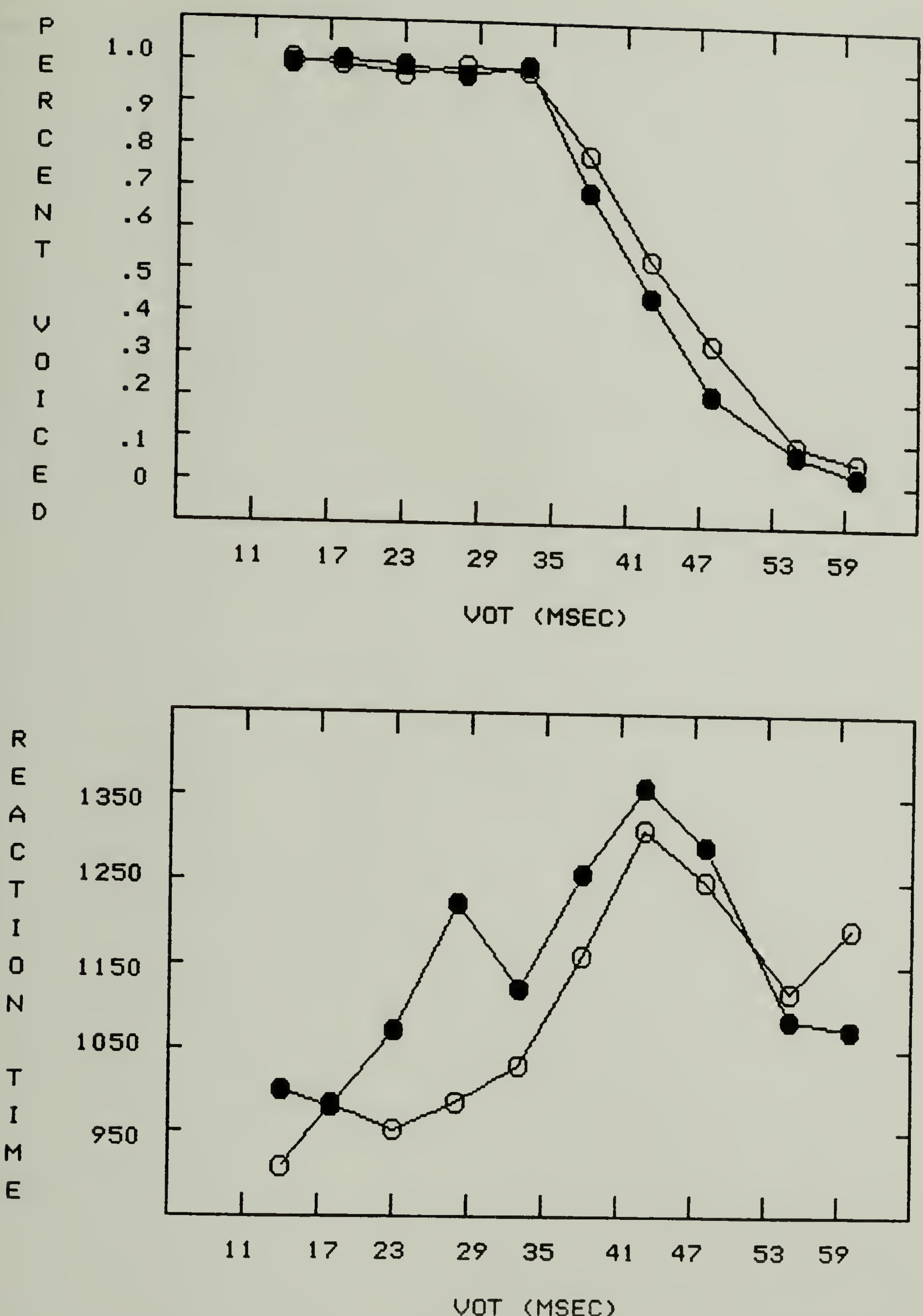
Tyler, L.K. & Wessels (1983). Quantifying contextual contributions to word recognition processes. Perception & Psychophysics, 34, 409-420.

Warren, R.M. (1970). Perceptual restoration of missing speech sounds. Science, 167, 392-393.

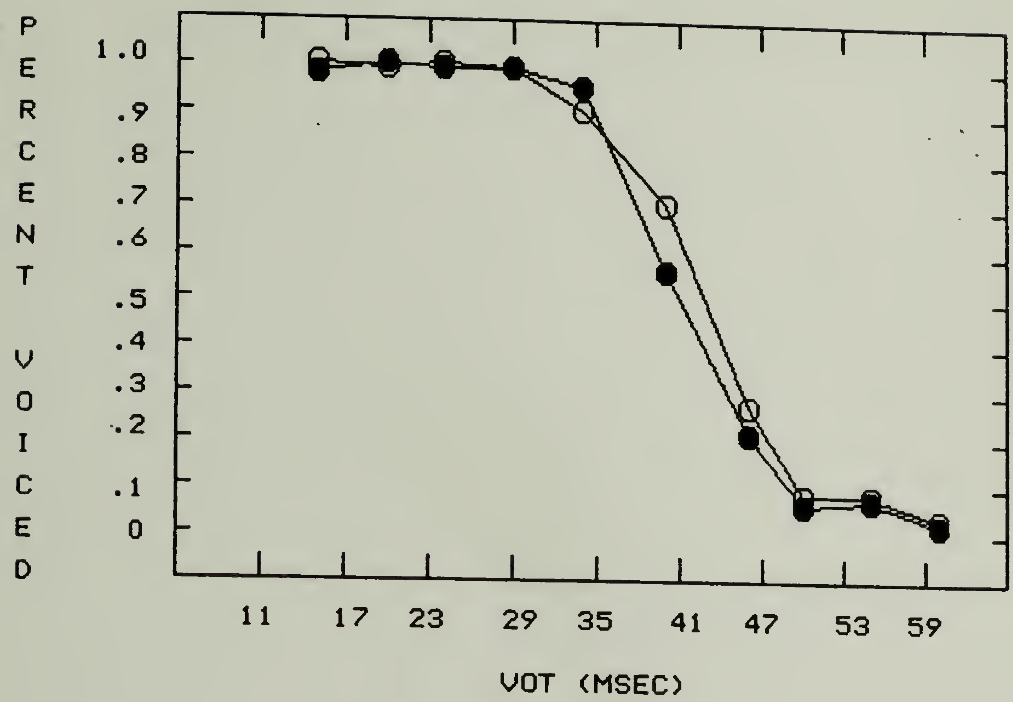
Warren, R.M. & Obusek, C.J. (1971). Speech perception and phonemic restorations. Perception & Psychophysics, 9, 358-362.

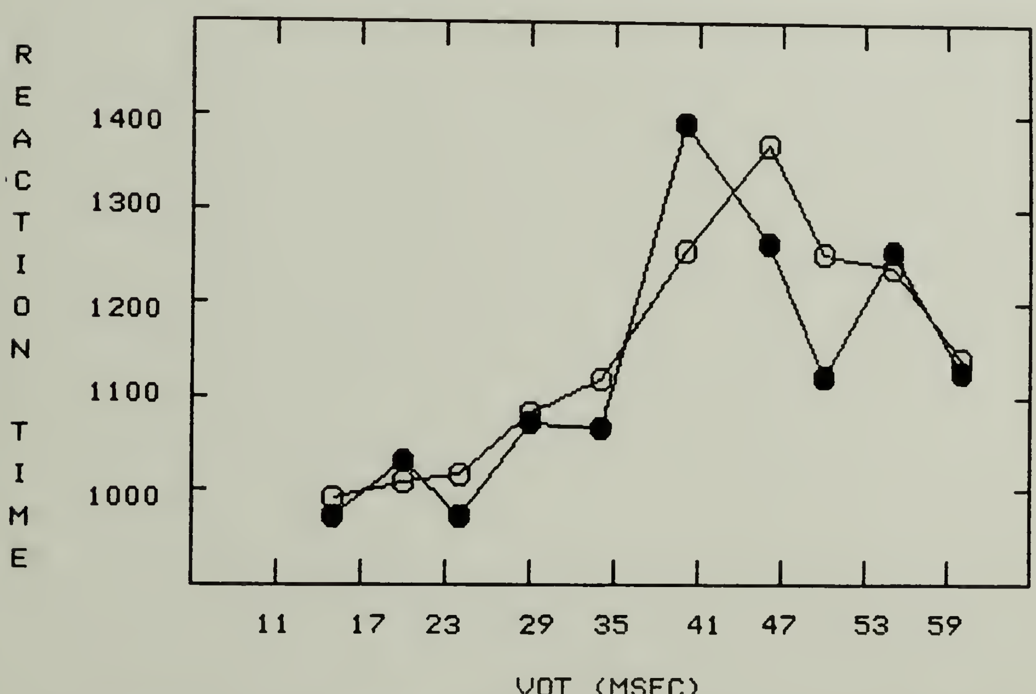
Yates, J. (1978). Priming dominant and unusual senses of ambiguous words. Memory & Cognition, 6, 636-643.

APPENDIX 1

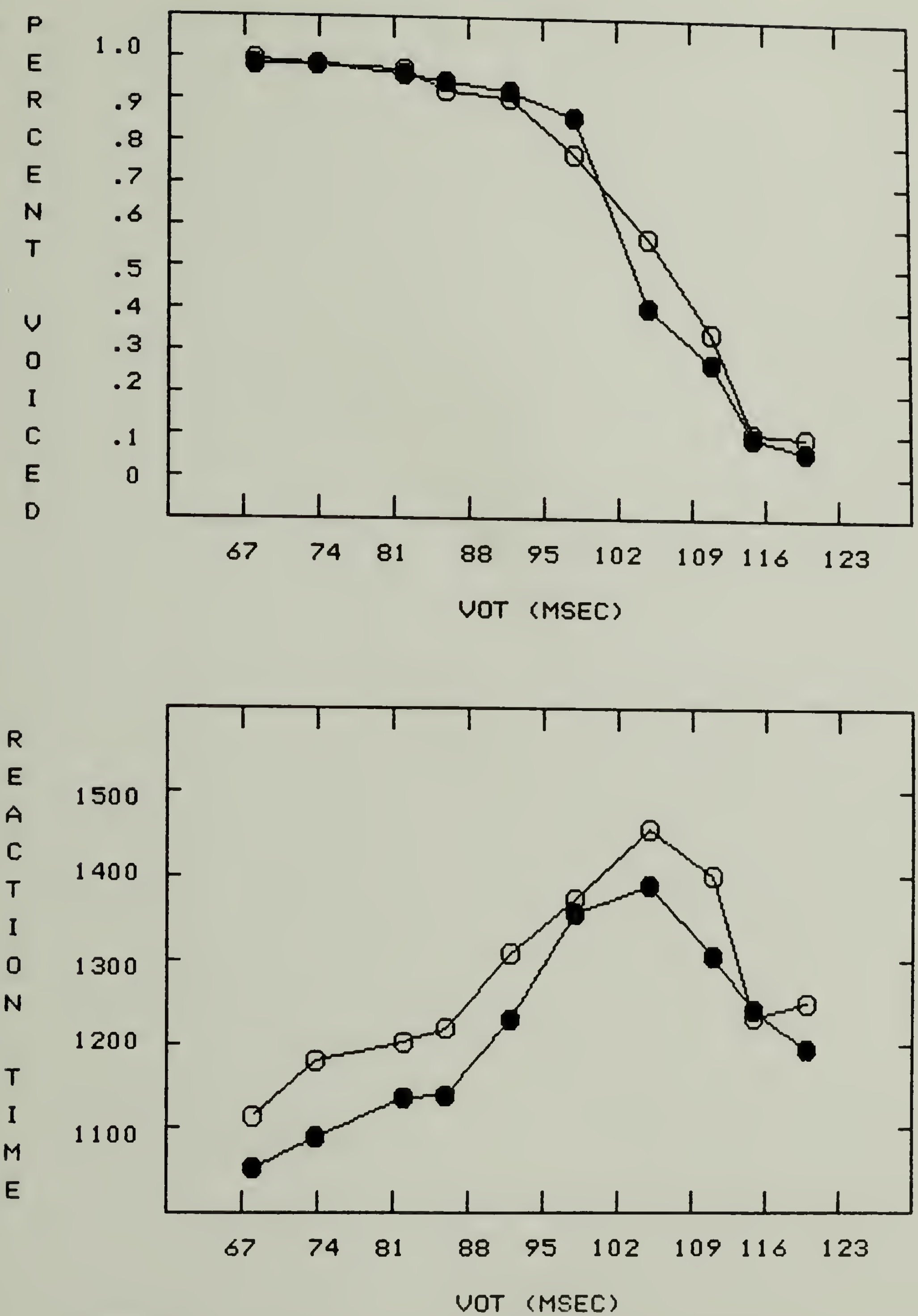


FIGURES 9 & 10. DENT-TENT IDENTIFICATION (TOP) AND REACTION TIME FUNCTIONS. OPEN CIRCLES - 'D' BIAS; FILLED CIRCLES - 'T' BIAS.

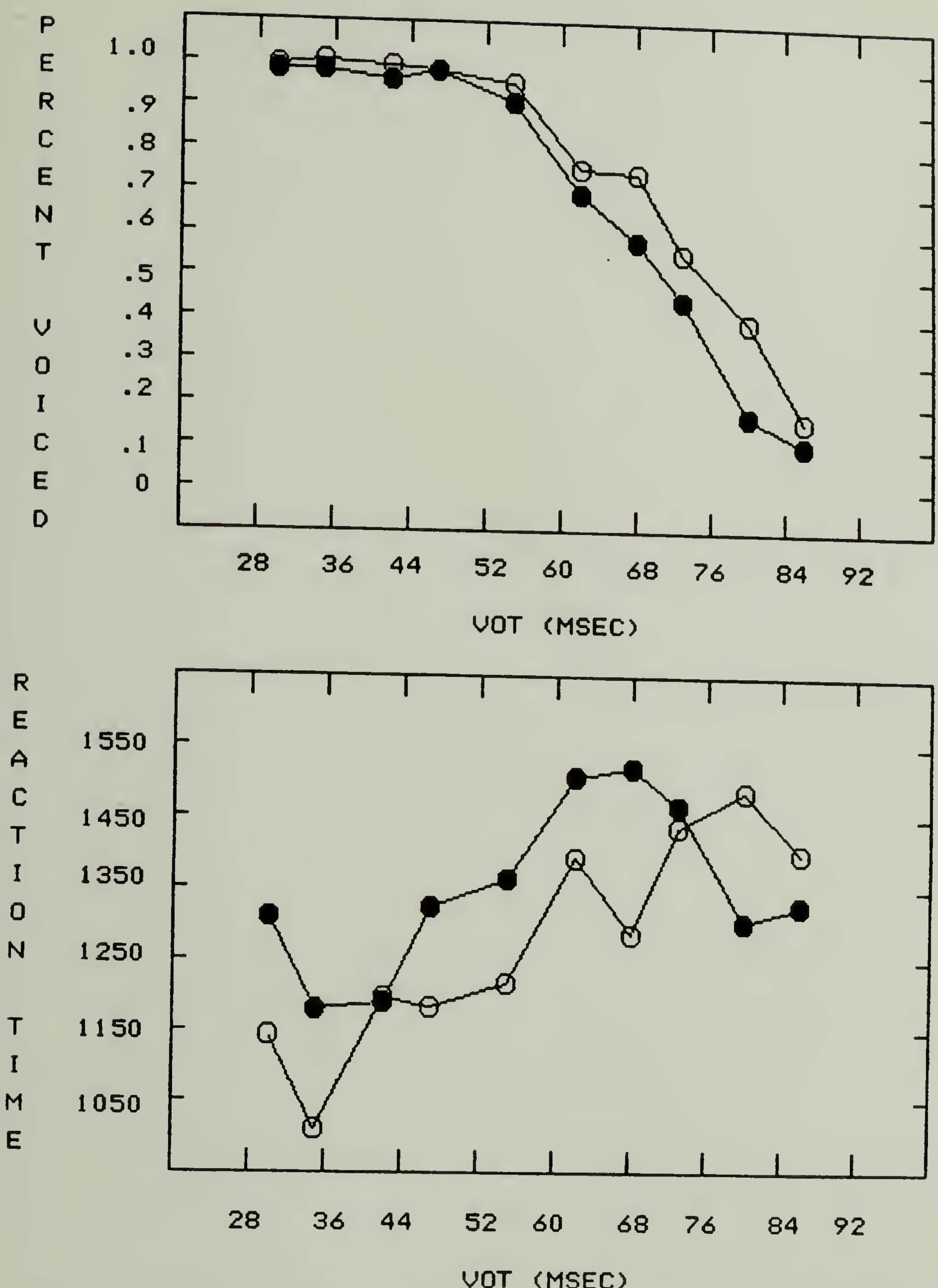




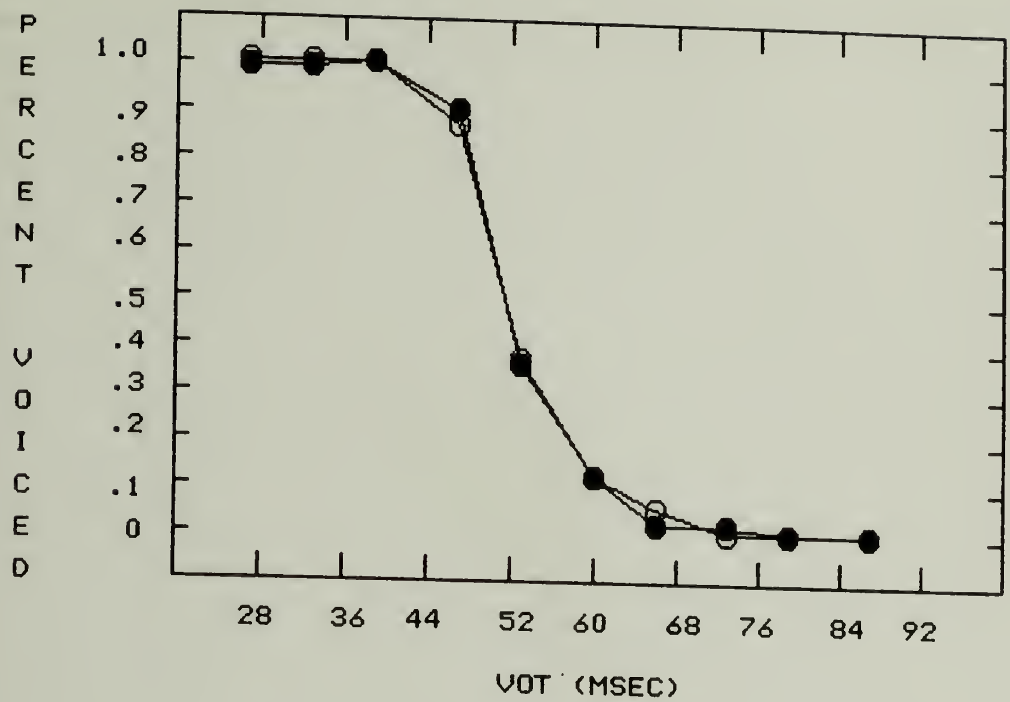
FIGURES 11 & 12. DIME-TIME IDENTIFICATION (TOP) AND REACTION TIME FUNCTIONS. OPEN CIRCLES - 'D' BIAS; FILLED CIRCLES - 'T' BIAS.

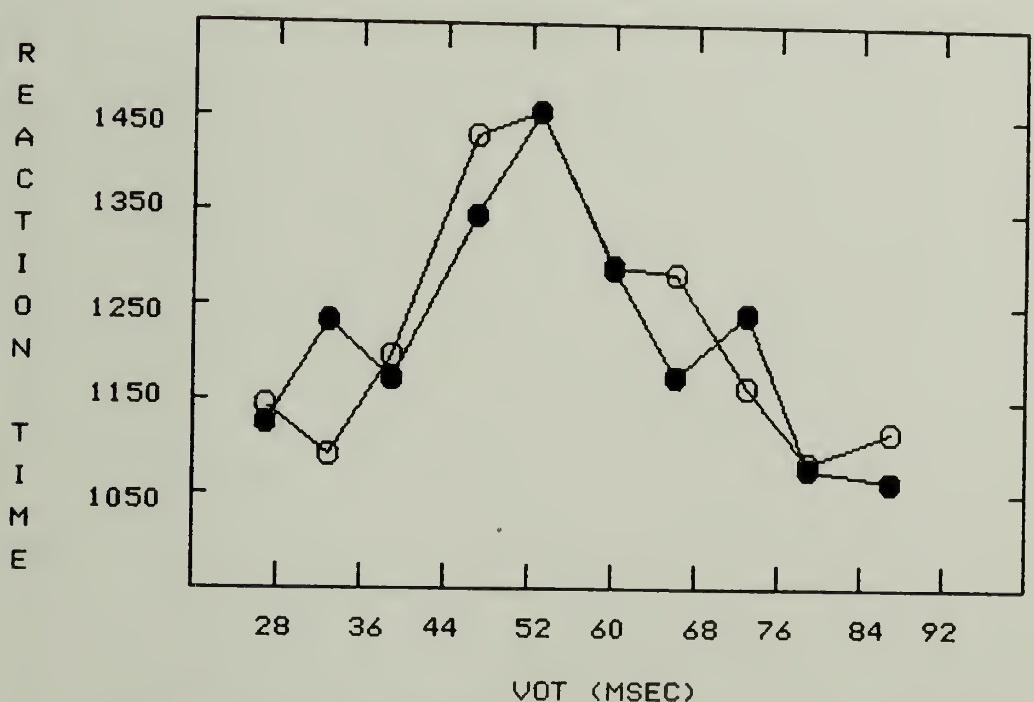


FIGURES 13 & 14. DRIP-TRIP IDENTIFICATION (TOP) AND REACTION TIME FUNCTIONS. OPEN CIRCLES - 'D' BIAS; FILLED CIRCLES - 'T' BIAS.

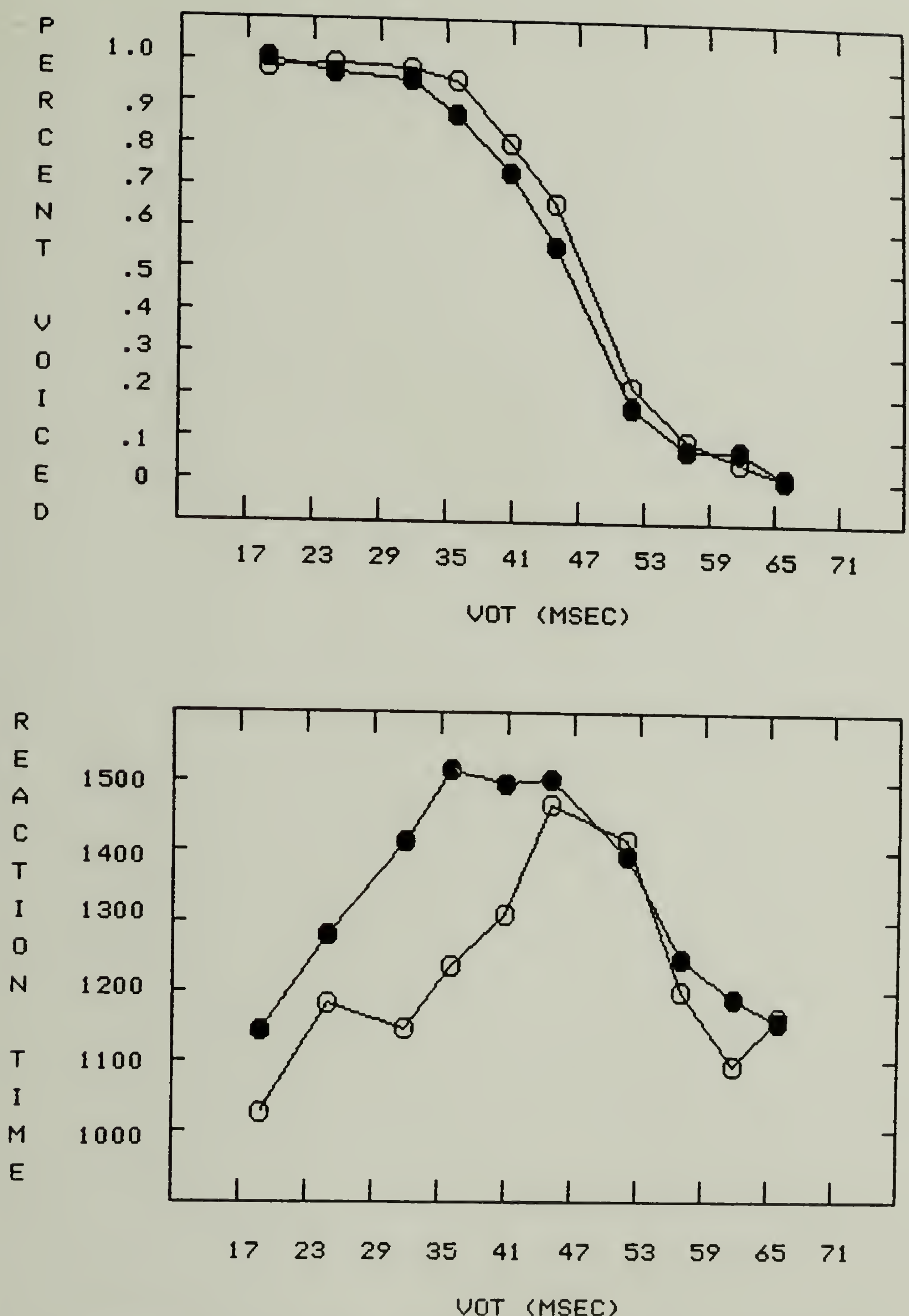


FIGURES 15 & 16. GUARD-CARD IDENTIFICATION (TOP) AND REACTION TIME FUNCTIONS. OPEN CIRCLES - 'G' BIAS; FILLED CIRCLES - 'C' BIAS.

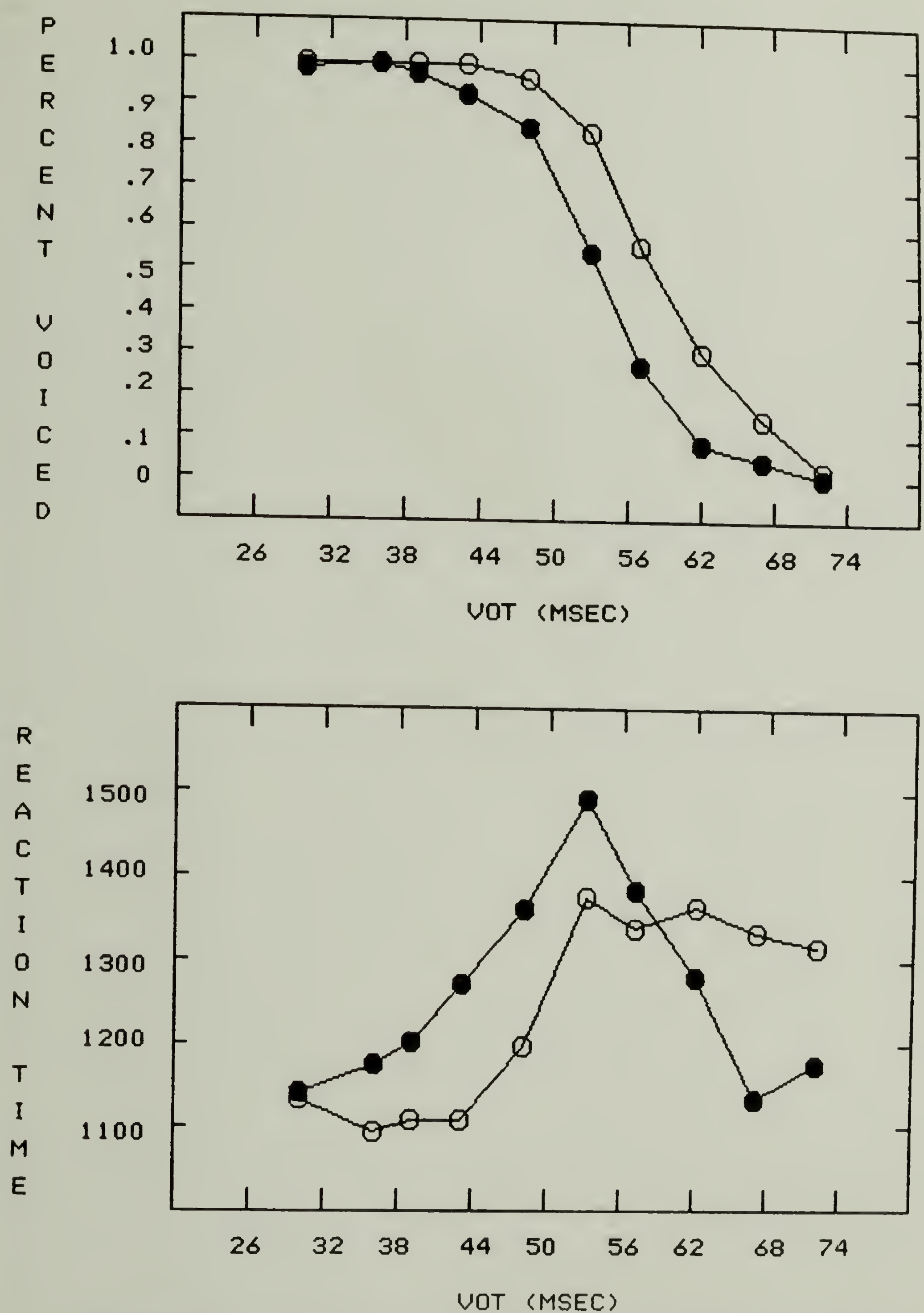




FIGURES 17 & 18. GOAT-COAT IDENTIFICATION (TOP) AND REACTION TIME FUNCTIONS. OPEN CIRCLES - 'G' BIAS; FILLED CIRCLES - 'C' BIAS.



FIGURES 19 & 20. GOAL-COAL IDENTIFICATION (TOP) AND REACTION TIME FUNCTIONS. OPEN CIRCLES - 'G' BIAS; FILLED CIRCLES - 'C' BIAS.



FIGURES 21 & 22. GOLD-COLD IDENTIFICATION (TOP) AND REACTION TIME FUNCTIONS. OPEN CIRCLES - 'G' BIAS; FILLED CIRCLES - 'C' BIAS.

APPENDIX 2

APPENDIX II. BOUNDARY AND ENDPOINT REACTION TIMES FOR VOICED AND VOICELESS RESPONSES SEPARATELY FOR EACH CONTINUUM/SENTENCE CONTEXT COMBINATION. EXPERIMENT 3.

DENT/TENT

BOUNDARY

		SENTENCE BIAS	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1317 (45)	1307 (43)
	INCONSISTENT	1277 (55)	1484 (57)

ENDPOINT

		ENDPOINT	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	906	1082
	INCONSISTENT	1000	1174

DIME/TIME

BOUNDARY

		SENTENCE BIAS	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1287 (45)	1515 (48)
	INCONSISTENT	1415 (55)	1353 (52)

ENDPOINT

		ENDPOINT	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	988	1127
	INCONSISTENT	972	1139

DRIP/TRIP

BOUNDARY

		SENTENCE BIAS	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1468 (45)	1521 (51)
	INCONSISTENT	1473 (55)	1476 (49)

ENDPOINT

		ENDPOINT	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1111	1169
	INCONSISTENT	1048	1127

GUARD/CARD

BOUNDARY

		SENTENCE BIAS	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1633 (48)	1617 (51)
	INCONSISTENT	1495 (52)	1565 (49)

ENDPOINT

		ENDPOINT	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1133	1332
	INCONSISTENT	1305	1347

GOAT/COAT

BOUNDARY

		SENTENCE BIAS	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1530 (47)	1648 (45)
	INCONSISTENT	1651 (53)	1611 (55)

ENDPOINT

		ENDPOINT	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1144	1063
	INCONSISTENT	1124	1113

GOAL/COAL

BOUNDARY

		SENTENCE BIAS	
		VOICED	VOICELESS
	CONSISTENT	1558 (44)	1626 (46)
RESPONSE			
	INCONSISTENT	1540 (56)	1785 (54)

ENDPOINT

		ENDPOINT	
		VOICED	VOICELESS
	CONSISTENT	1008	1155
RESPONSE			
	INCONSISTENT	1142	1150

GOLD/COLD

BOUNDARY

		SENTENCE BIAS	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1416 (44)	1705 (47)
	INCONSISTENT	1624 (56)	1490 (53)

ENDPOINT

		ENDPOINT	
		VOICED	VOICELESS
RESPONSE	CONSISTENT	1136	1170
	INCONSISTENT	1125	1313