THE RELATIVE CONTRIBUTION OF TOP-DOWN AND BOTTOM-UP INFORMATION DURING LEXICAL ACCESS.

Andrea Jane Lowe

PhD

Department of Artificial Intelligence
and
Centre for Cognitive Science
University of Edinburgh

ACKNOWLEDGEMENTS

I would like to thank CSTR for the use of equipment and facilities.

Many thanks to my supervisors Ellen Bard and Gerry Altmann for refraining from physical violence in getting me to submit this thesis! Thanks go to you both for your patience and interest and for helping me to formulate the ideas and put words into the right order. Also, a big thankyou to Gerry for reading and commenting on earlier drafts so speedily.

Special thanks are due to Joe (Joseph) and Nang. I would never have met this deadline without you two! Joe's formatting and Nang's diagrams at least make the thesis look good. Also, thankyou both for all general rescue operations throughout the time spent thesising. I'm very grateful to you.

ABSTRACT

The research reported in this thesis examines the relative contributions of top-down and bottom-up information during lexical access. I evaluate the Cohort Model of lexical access (Marslen-Wilson and Welsh, 1978; Marslen-Wilson and Tyler, 1980; and Marslen-Wilson, 1987) which specifies that the first stage in lexical access is fully autonomous and that during this stage all processing proceeds in terms of analysis of the acoustic-phonetic input.

Implicit in this model is the assumption that bottom-up processing is immune to any effects of contextual or top-down information. I examine the extent to which listeners ever rely exclusively on bottom-up information during lexical access and investigate this issue empirically, by measuring effects of context on both the production and the perception of words in various contexts.

I test the hypothesis that a word uttered in a constraining context will be acoustically indistinguishable from its competitors by, first, measuring one acoustic parameter (VOT) across constraining and non-constraining contexts and, then, examining the intelligibility of tokens of that parameter taken from the varyingly constraining contexts. The data from these experiments suggest that the realization of VOT is not an aspect of bottom-up information which would create problems for a bottom-up processor in terms of providing ambiguous acoustic-phonetic information.

I then investigate whether bottom-up processing during lexical access is immune to effects of context. Following Grosjean (1980) and Tyler (1984), I utilize the Gating Paradigm. Using incongruous contexts, I argue that direct assessment of the contributions made by different information sources during lexical access can be made. By presenting bottom-up information which is inappropriate to the contextual (top-down) information, I evaluate the extent to which one information source is given priority over the other. I vary both the contextual constraints available to the listener and the acoustic clarity of bottom-up information. The observed pattern of listeners' identifications of the words suggested that whilst bottom-up information was given priority, top-down information was available and was utilized during lexical access.

I present data which support the working *structure* of the Cohort Model of lexical access. I conclude, however, that the model places disproportionate emphasis on initial bottom-up processing. It appears that top-down information is *not* prohibited from contributing to processing during the initial stage of lexical access.

TABLE OF CONTENTS.

CHAPTER 1:

INTRODUCTION.

1. THE PROBLEM OF SPOKEN WORD RECOGNITION.

One of the fundamental goals in psycholinguistic research has been to explain how human listeners are able to relate spoken language to meaning. A major aspect of this problem is to explain how the human listener is able to convert the continuously varying speech waveform into discrete linguistic units (words) and how these units are then used to extract the linguistic message intended by the speaker.

For Marslen-Wilson 1987 word recognition lies at the core of the process of spoken language understanding: For the syntactic and semantic properties of a given word to be incorporated into a message level representation, the sensory input must make contact with those knowledge representations stored in the mental lexicon.

> "...it is the knowledge representations in the mental lexicon that provide the actual bridge between sounds and meanings, linking the phonological properties of specific word forms to their syntactic and semantic attributes." (Marslen-Wilson 1987, p. 72)

It is this duality of the lexical representations which makes word recognition such an important part of language understanding.

The time course of the process of spoken word recognition has attracted a great deal of attention in psycholinguistic research. One problematic issue has been *when,* during word recognition, the listener combines information from different knowledge sources. Although there is a general consensus that the listener uses phonetic, lexical, syntactic, semantic and pragmatic knowledge at some stage during the course of word recognition, theories differ widely as to when these knowledge sources might be invoked. The particular issue we will investigate is whether these higher level knowledge sources can be used during the *initial* stage of word recognition.

We shall be refering to this initial stage as the "access" stage, and shall use terms such as "perceptual processing", and "acoustic-phonetic" processing to refer to the processes which correspond to this stage. This is not intended as a claim on the precise nature of the input to the word recognition system. Many researchers remain neutral as to precisely what constitutes input to the system. For our purposes however, the "phonetic" label will be a convenient one. We shall also be using the terms "semantic", "contextual", and "top-down" information interchangeably. "Acoustic-phonetic" and "bottom-up" information will be similarily synonymous.

In this thesis we specifically examine one particular theory of word recognition, and we empirically investigate certain assumptions implicit in its definition. The model we examine is the Cohort Model (Marslen-Wilson and Welsh 1978, Marslen-Wilson and Tyler 1980, Marslen-Wilson 1987). Broadly, this model assumes that all initial processing during word recognition proceeds by reference only to the acoustic-phonetic information available in the speech stream. This bottom-up processing is assumed to activate a subset of the word candidates in the mental lexicon which share initial acoustic-phonetic properties. By assigning the initial impetus to bottom-up processing, the word recognition system increases its chances of finding the correct word and places tighter constraints on the size of the decision space. Under the Cohort Model, the initial bottom-up processing thus serves to constrain the number of word candidates which "compete" during the process of word recognition. An implication of allocating the burden of initial processing to bottom-up processing is that initial processing of the acoustic-phonetic information available in the speech waveform should be immune to any effects of context (e.g. the lexical, syntactic, semantic and pragmatic knowledge sources which might be available to a listener).

Because it assumes that initial processing proceeds bottom-up, the Cohort Model attatches great importance to the acoustic-phonetic information contained in the initial segments of words. This is problematic, however. As we will see in Chapter 2 (Sections 4.1 and 4.2) the acoustic quality of these segments may often be quite poor, or the initial segments may even be obliterated altogether by extraneous noises in the environment. Furthermore, it is known that word boundaries are notoriously difficult to locate. This, too, brings into question the usefulness of the constraints afforded by initial bottom-up processing during word recognition. If the acoustic-phonetic information is degraded, then a bottom-up model, such as the Cohort Model, would be required to tolerate whatever noise is present in the bottom-up information. As an extreme example, consider a situation where the acoustic-phonetic information is totally degraded. On the basis of such poor initial bottom-up information, the Cohort Model would be required to activate *all* word candidates in the mental lexicon. Clearly, in this extreme case, the constraints provided by bottom-up processing are minimal. The central aim of this thesis, therefore, is to test the assumption, explicit in the Cohort Model, that initial processing during word recognition proceeds purely in terms of the acoustic-phonetic input, and to consider whether top-down information might, in fact, influence this initial processing.

2. THE RESEARCH.

We shall describe two perception experiments which aimed to measure the effects of a constraining context on the recognition of the initial segment of target words differing only on their first segment (e.g. BEES/PEAS). We will then compare these with the effects of a non-constraining context on the recognition of the initial segments of the same target words. If we observe differences between recognition in the two contexts, then we assume that the constraints provided by the context somehow affect the perception of these word-initial segments. This observation would contradict the assumption in the Cohort Model, that initial processing proceeds purely bottom-up.

In the event of finding such differences, however, it could be argued that they are not attributable to the influence of contextual information on the initial processing of words *per se*, but rather that they are simply due to physical differences between the target tokens themselves. It could be argued that the contextual environment in which the target words were uttered affects the realization of the targets, and, that it is this difference in the *production* of the targets which affects perception.

To control for this possibility, we also look at the affects of context on the *production* of speech. More specifically, we look at effects of context on the production of the initial segments of the target words which we use in our perception experiments. Following Lieberman (1963), who demonstrated an inverse relationship between predictability from context and intelligibility, we might suppose that the production of speech is affected by context in a way which makes the physical realization of a word indistinguishable, on acoustic grounds, from the realization of similar a word. If this is the case, then the motivation for constraining the activation of word candidates by reference only to the bottom-up information available in word initial segments is weakened. If the acoustic-phonetic information is not sufficiently discriminable to be able to promote the activation of a particular set of word candidates over other sets of word candidates, then processing bottom-up information alone will not highly constrain the initial activation of any subset of word candidates in the mental lexicon.

Grice 1978 noted that in order for people to communicate accurately and efficiently, speakers and listeners try to cooperate with one another. It is thus plausible that the speaker might expect the listener to use whatever information is available to aid comprehension. We might, therefore, suppose that if the realization of the acoustic-phonetic information is made ambiguous by context, then, during *perception,* reference needs to be made to that contextual information in order to disambiguate the acoustic-phonetic material in question. The extent to which talkers control their production of speech is thus pertinent to the issue of whether initial processing during word recognition proceeds purely bottom-up.

The research to be reported in this thesis, thus aims to address the following questions:

a) Do talkers control for the amount of information they wish to give their listener? For example, if other factors, such as context, can carry meaning, then will a talker relax productions of utterances in a way that is (at least) acoustically measurable?
b) If adjustments are made by the speaker, do they serve to make speech more or less intelligible to the listener? Is it the case, for instance, that if the context is unconstraining, the utterance becomes more intelligible?
c) If such adjustments are made, how are they made: are they made phonetically and prosodically? Could there, for example, be any cue in the prosodic contour of an utterance that might encourage listeners to believe they were hearing to a constraining or a non-constraining utterance?
d) Are *listeners* able to make use of contextual information during the initial stages of recognizing words? More precisely, are listeners able to use the information provided by a constraining context to facilitate the process of word recognition?

As we noted above, these issues bear directly on models of lexical access which advocate that initial processing, during word recognition, proceeds bottom-up. To the extent that speakers productions of utterances might become ambiguous when other factors, such as context, can carry meaning, acoustic-phonetic information alone may not be useful in constraining the activation of a subset of the mental lexicon. Also, by measuring the extent to which listeners are able to utilize contextual information during the initial stage of word recognition, we test the assumptions, implicit in the Cohort Model, that initial acoustic-phonetic processing is immune to effects of context i.e. the assumption that initial processing during word recognition proceeds bottom-up.

3. THE METHODOLOGY.

Because we are interested in the time course of the availability of different information sources during the initial stage of word recognition we shall be utilizing the gating paradigm (Grosjean 1980) in 3 of the experiments to be reported in this thesis. Grosjean defines the gating paradigm as:

> " ... presenting a spoken language stimulus repeatedly and increasing its presentation time (duration from onset) at each successive pass ..."
> (Grosjean 1980, p.267)

The gating task presents subjects with increasing increments of some target stimulus and subjects are instructed to identify the target stimulus after each gated presentation (i.e. after each pass). This paradigm enables us to determine exactly how much of the acoustic stimulus is necessary for subjects to be able to identify the target stimulus. By varying the availability of other information sources it is possible to estimate the time course of their effects on processing. We assume that if the availability of other information sources produces identifications of the target stimulus at earlier gated presentations of the acoustic-phonetic input (than would have been observed in their absence), then the other information sources have influenced processing. [1]

In utilizing the gating methodology researchers have assumed that pooling the word candidates which different subjects propose at each gate reflects the words which would be candidates for recognition if we could tap word recognition under normal conditions. By analyzing the correct and incorrect guesses made by subjects at various gates throughout

[1] We consider this issue more fully in Chapter 2, Section 3.

the presentation of a word, insight may be gained into what types of information are being used at various stages during word recognition.

Three disadvantages with this paradigm are immediately apparent however: First, the subject is presented with an unnatural sounding stimulus; second, the responses, unless timed (cf. Tyler and Wessels 1985), may be subject to conscious judgement and thus not yield responses which reflect word recognition proper; and, finally, the repeated presentation of the same stimulus may somehow affect responding.

Grosjean 1980, however, demonstrated that the gating paradigm yielded the same robust effects that have been observed using a range of other tasks (e.g. phoneme monitoring, word monitoring, shadowing, and mispronunciation detection). He found, for instance, that less acoustic information was needed for words to be identified when they were short, frequent and appeared in context. Similar results were obtained by Tyler and Wessels 1983, in which speeded responses were required. The gating paradigm, however, may be preferable to these other tasks, because it does not require subjects to allocate any attention to analyzing subparts of the incoming stimulus (in the way that, for example, monitoring tasks do).

Forster 1981 argued that because the gating paradigm forces subjects to respond "unnaturally early", responses constitute no more than an intelligent guess. As Marslen-Wilson 1987 notes, however, to distinguish between perception and guessing

> " ... is to assume, as a theoretical a priori, a particular answer to the fundamental question at issue." (Marslen-Wilson 1987, p.76)

Finally, Cotton and Grosjean 1984, discredited the criticism concerning the consequences of repetition on subjects' performances. Cotton and Grosjean presented only one (different) gated presentation of the materials to different groups of subjects. The results were very similar to those obtained in the Grosjean 1980 study. It was therefore concluded that the successive presentation format did not influence subjects' performance in the gating tasks (cf. (Bard, Shillcock and Altmann 1988)).

The above discussion has aired some of the problems apparent with the gating paradigm. We have chosen to utilize it in the research reported in this thesis because it allows us to trace the time course of spoken word recognition and because the paradigm generates a large body of useful data.

4. THE MATERIALS.

The target stimuli which we shall use throughout this thesis will be minimal pairs which differ only on the voicing of the initial stop. It will be useful therefore to consider this choice of materials and to discharge certain assumptions concerning the voicing distinction.

It is widely held that the voicing contrast (i.e. the perception of voiced or voiceless segments) is cued by the differing values of Voice Onset Time (VOT) associated with the respective phonetic segments (see Abramson and Lisker 1970, Lisker 1978).

Voice Onset Time is a temporal parameter and, for our purposes, will be defined as the time between the release of a stop closure and the time at which glottal pulsing (i.e. voicing) begins. In English, VOT's for voiced and voiceless consonants are expressed respectively as short positive and long positive values. For example, the phoneme [b] is usually perceived when the VOT is less than 25 msec whereas [p] is the percept of higher values; [d] is perceived for values less than 35 msec and [t] at greater intervals, whilst the category boundary differentiating [k] and [g] falls at around a VOT of 42 msec (Abramson and Lisker 1970)

Although VOT is one of the major determinants of the voicing distinction, our perception of the voicing distinction does not rely solely on the interpretation of this temporal parameter. Rather VOT appears to be an ambiguous cue for voicing categories.

Research in acoustic-phonetics has characterized the speech waveform as:

> "providing the listener with *ensembles* of potentially informative acoustic cues to each of the phonemes in a given language." (Bernstein 1983, p.383).

and it appears that many parameters other than VOT are able to influence our perception of the voicing distinction.

Repp 1982 observed that the frequency of F1 at voice onset enters into a "trading relation" with the temporal component of VOT. The lower the F1 onset frequency, the longer the VOT interval required for perception of voiceless stops. Similarly, the VOT voiced/voiceless boundary appears to shift towards longer relative onset times when the duration of the first formant transition is increased or when the fundamental frequency at voicing onset is increased.

Wintz, Riviere and Herriman 1975 argues that aspiration is the primary cue to voicing. They report that the systematic lengthening and shortening of VOT does not result in an appreciable shift in subjects' perception of voicing whereas manipulating the amount of aspiration does.

It thus appears that we have an interaction of cues serving to create the voicing distinction. Whatever cue is primary, it is clear that many parameters can influence the perception of voicing. We have noted several cues here: frequency of F0; the transition length of F1; frequency of F1; relative intensity of aspiration and Voice Onset Time.

We acknowledge here that the voiced/voiceless distinction is not cued by simple contrasts of VOT values. Nonetheless, we choose to examine this particular parameter.

Due to its temporal nature, VOT is an ideal acoustic parameter to manipulate experimentally. It can be measured with relative ease by locating the boundaries between the release of the stop burst and the onset of voicing and can be synthetically varied by increasing or decreasing the time between these boundaries.

We choose only to study one of the possible cues to voicing [2] as it is outwith the scope of this thesis to examine all possible cues to the voicing distinction. VOT is clearly a very important cue to voicing and for reasons of simplicity which we noted above, we incorporate it as a variable in our materials.

[2] If we were to investigate all possible cues the study domain moves more into that of acoustic-phonetics and away from our immediate psycholinguistic interests.

5. OUTLINE.

In Chapter 2 we outline three approaches to the problem of lexical access. We present the experimental evidence which has been cited in support of the Cohort Model of lexical access and we review the empirical evidence which has been interpreted as demonstrating that bottom-up processing is immune to contextual effects. We consider the literature which points to problems inherent in assuming bottom-up priority and, in light of this, Chapter 2 concludes by asking how rigorously we need to adhere to the notion that initial processing proceeds purely in terms of bottom-up analysis.

Chapter 3 describes two experiments which aim to assess the extent to which the *acoustic realization* of an utterance is affected by context. In the first experiment we measure VOT's of voiced and voiceless stops which had been uttered in various contexts. This represents an attempt to assess the degree of constraint afforded by assuming that word recognition initially proceeds by reference to bottom-up information alone, and, also, serves as a pre-test on the materials to be used in subsequent experiments. We found that VOT was not significantly affected by context. The second experiment served to substantiate the findings of the first and tested the same issue from a perceptual stance. Although the VOT's gained in the first experiment did not show measurable differences across different contexts, the second experiment addressed whether or not there might have been any *perceptible* differences in the realizations of the VOT's. We did not observe any such differences.

Chapter 4 describes an experiment which was designed to investigate whether listeners were able to use information available in the prosodic contour of an utterance to discriminate between constraining and non-constraining contexts. This experiment served as a control for subsequent experiments. If the prosodic contour of an utterance could give sufficient information to listeners about the predictability of words within that utterance, then manipulations to the lexical/semantic context, which we planned for subsequent experiments, would have been confounded with inadvertent manipulations of prosody.

Chapter 5 presents two experiments which bear directly on the issue of bottom-up priority during initial processing in word recognition. Both experiments were designed to ascertain whether listeners are able to use contextual information during the processing of initial segments of words. We measure the perception of initial segments of targets presented following appropriate and inappropriate contexts and we observed that the targets were differentially perceived during what has been assumed to constitute the initial stage of processing during word recognition.

We conclude that initial processing during word recognition is *not* immune to effects of context. The research which we present in this thesis suggests that although bottom-up information is given priority during the early stages of word recognition, listeners *are* able to utilize contextual information. In Chapter 6 we describe the implications which our data have for the Cohort Theory of lexical access. Although the majority of our data do not impinge on the *structure* of the Cohort Model (indeed they *support* it), they reveal that the contribution of top-down information during the initial stage of word recognition is *not* prohibited.

CHAPTER 2:

BOTTOM-UP AND TOP-DOWN INFORMATION IN LEXICAL ACCESS: THE LITERATURE.

1. OVERVIEW.

There is a general consensus that the "mental lexicon" contains multi-layered descriptions of words known to an individual. It is assumed that these lexical representations hold information at all levels of description e.g phonetic, syntactic and semantic. In postulating models of word recognition, researchers have attempted to characterize the way in which *one* of these lexical representations in the mental lexicon is selected. It is often held that these representations are initially accessed by reference to the sensory input, [1] and that this initial stage somehow delimits a subset of the lexicon, making available these multi-layered descriptive representation for a "pool" or "search set" of potential word candidates to enter into the recognition process. This initial phase we shall refer to as "lexical access". Subsequently, processing is thought to involve the selection, isolation, elaboration and integration with context of the information associated with one of the word candidates made available during lexical access.

It is widely held that, during the process of Lexical Access, the "bottom-up" information available in the speech signal is somehow compared to the lexical representations to activate a subset of the lexicon. For the purposes of this Chapter we consider "bottom-up" information to be the acoustic-phonetic information available in the signal. [2] The phonological and contextual (lexical, syntactic, semantic and pragmatic) descriptions associated with the particular words activated comprise the "top-down" information sources. Top-down knowledge sources are commonly assumed to serve as "confirmation" of the "expectation" created by the bottom-up acoustic information (i.e. the

[1] But see the Dual Code Hypothesis, Foss and Blank 1980 for a different account.

[2] Note that prosody also comprises bottom-up information. We do not consider prosody here but we later assess its im-

subset of potential word candidates).

All theories of Lexical Access recognize the importance of both bottom-up and top-down information. They differ however, on the emphasis put on the amount, form and locus of the contribution of one or other knowledge sources. In this Chapter (and throughout the thesis) we consider the extent to which people ever rely *exclusively* on bottom-up processing to initiate the process of word recognition.

In the following section, we consider three models of word recognition. These models vary with respect to the importance they assign to the initial contribution of bottom-up information. In subsequent sections we critically examine the evidence which has, until now, been cited in support of the claims on the priority afforded to bottom-up processing and we present evidence which brings such claims into question.

2. BOTTOM-UP vs TOP-DOWN INFORMATION FLOW - THREE APPROACHES.

2.1. Introduction

Although researchers have investigated the use of top-down and bottom-up information in sentence processing, it is only recently that much attention has been devoted to spoken word recognition. Historically much significance was attached to the linguist's structural descriptions of language. Early psycholinguistic research focussed on the "autonomy of syntax" (Forster 1974) in *sentence* processing as opposed to *word* recognition (which was more often than not simply taken for granted). It was generally held that subsequent to phonetic and lexical analysis, the syntactic structure of a sentence was computed in order to retrieve the meaning of that sentence.

Understanding the meaning of a sentence necessitated analysing the grammatical relations that held between the individual words in the sentence. Therefore, semantic interpretation was presumed to take place only after syntactic analysis had been performed. The majority of the early models of sentence processing thus proposed that the flow of information through the system was strictly bottom-up, and analysis of an incoming sentence was assumed to move from the phonetic and lexical domain to the syntactic in strict sequence. Specifically, semantic analysis was not able to influence processing at other

portance in Chapter 4.

levels, since processing (at all levels) was assumed to be strictly modular i.e. "opaque" to the analyses at other levels (see, for example, Fodor and Garrett 1967, Bever, Garrett and Hurtig 1973, Forster and Olbrei 1973, and Fodor, Bever and Garrett 1974). [3]

Although some psycholinguists maintain that theories of sentence comprehension which adhere to the strong modularity hypothesis are now highly contentious, the modularity approach has, nonetheless, provided the basis for many current theories of sentence comprehension and, more recently, word recognition. The three models of Lexical Access that we consider in this section differ in their acceptance of modularity and in the degree to which they incorporate the concept of modularity.

2.2. Serial Search Model

Following Fodor, Bever and Garrett 1974, advocates of the strictly modular (bottom-up or autonomous) models of word recognition (Forster and Olbrei 1973, Forster 1979, Norris 1980) maintain that for each type of information there exists a domain specific processor. Processors operate independently and the output of one processor serves as input to another in a predetermined and predictable order.

Such autonomous approaches require that each distinct knowledge type be realized as a separate processing component, functioning as a computationally independent processing level within the system. The computations for any one module are opaque to other modules and it is only the final output of any one module that may be issued as input either to a higher level module or a central processor (this output being issued irrespective of the state of other modules).

Under Forster's framework (Forster 1976) stimulus features extracted by analysis of the acoustic-phonetic input are compared against a set of stored representations in a peripheral access file. These access files simply hold descriptions of the stimulus features of a word (the "access code") and a "pointer". Once the access codes for all words in a particular access file have been compared against the stimulus and the appropriate one located, the

[3] The evidence cited in support of this position demonstrated that the syntactic structure of an utterance appeared to have direct effects on the ease of comprehension. For example, people tend to impose a SUBJECT-VERB structure erroneously on centre-embedded sentences (Blumenthal 1967); people are able to comprehend self-embedded sentences more easily when the surface structure clues (e.g. pronouns) upon which reconstruction of the base structure depends are not obliterated (Fodor and Garrett 1967); that *after* clause boundaries people integrate preceding material into some sort of semantic representation (Jarvella 1971); that response latencies appeared to be contingent on the syntactic subcategorization of verbs (Fodor and Garrett 1968); active sentences are more quickly understood than their corresponding passive (Forster and Olbrei 1973) and

pointer directs the search to a "master file" (where all information we have about a word is stored) and subsequently compares entries in this master file to the incoming stimulus. The latter stage is known as the "post access check" and constitutes the earliest stage that the sentential context can be used in the analysis. The central requirement of this model is that contextual knowledge can only be brought to bear *after* the bottom-up processing has enabled the phonological access of a word candidate and has moved the search to the post access check procedure.

Such a modular account of word recognition requires that the bottom-up analysis of the acoustic-phonetic input necessarily continues until the word's acoustic offset regardless of the sentential context. As we shall see in section 3.1 however, there are copious data which call such assumptions into question.

2.3. The Logogen Model

A radically different approach to word recognition was proposed by Morton 1969 in the Logogen model. Unlike Forster's model, the Logogen model abandons the concept of modularity. Morton supposes that contextual information is available to influence processing along with the sensory analysis of a word. In this model, each word in the mental lexicon has a logogen (pattern recognition device) which specifies *all* defining characteristics of that word along various acoustic, syntactic and semantic dimensions. Logogens function primarily as counting devices: the internal count increases as a simple function of the number of inputs that fall into the logogens' defining set until a predetermined, criterial threshold is exceeded. If enough of these features are satisfied, recognition occurs.

Morton's model is a passive account of word recognition in that memory elements are accessed directly without an active search. In sharp contrast to Forster's autonomous model, top-down expectations can function to push a logogen over all its defining thresholds, even before any of the acoustic cues have been heard.

In comparison with the autonomous models, the Logogen model plays down the role of bottom-up information (in that it attributes equal importance to all information sources) and for this reason it predicts that words can be recognized before their acoustic offset. Also,

that anomalous strings are perceived more easily than scrambled word lists (Miller and Isard 1963).

the Logogen model accommodates the findings of failures in perception such as the restoration of disrupted speech. As we shall describe below however (Sections 3.2 and 3.3), there are other findings which it is unable to accommodate and it is possible that the model encounters such problems because it does not incorporate any degree of modularity. [4]

2.4. The Cohort Model

A major current theory of word recognition which combines certain aspects from both the Logogen and the Autonomous Search models is the Cohort Model (Marslen-Wilson and Welsh 1978, Marslen-Wilson and Tyler 1980, Marslen-Wilson 1987).

The Cohort theory of Lexical Access has attracted considerable attention recently because of its novel claim that all words in the mental lexicon sharing initial acoustic-phonetic information are activated; because of the priority it affords to the beginnings of words; and, importantly, because of its relatively precise claims about the temporal nature of word recognition.

In common with the autonomous models of word recognition, the Cohort Model assumes that the bottom-up analysis of the incoming acoustic-phonetic information is of crucial importance initially. As with Forster's model, the Cohort model relies on the analysis of the acoustic-phonetic information to activate some sort of subset of the mental lexicon in which potential word candidates compete for recognition.

As originally stated however, once this subset of candidates has been activated, *all* types of information may be processed simultaneously with each source of information being able to aid decisions at higher and/or lower levels:

> "...from the first word of a normal sentence, the analysis of the input is conducted at all available processing levels. In particular, the information at any one level of analysis can constrain and facilitate the decisions at any other level, so that the continuing phonetic and lexical processing of each word is directly influenced by its current semantic and syntactic context." (Marslen-Wilson 1975, p.)

[4] We shall not be discussing this issue explicitly but the argument is implicit in our review of the evidence supporting the Cohort Model of Lexical Access (Section 2.4 and Section 3).

The early version of the Cohort Model assumed two distinct processing stages. During the first phase, bottom-up information was obligatorily processed, and made contact with the lexical representations in the mental lexicon. This stage was completely autonomous. The bottom-up information was used to delimit a subset of the lexicon sharing initial acoustic-phonetic properties. It was not until after this initial pool of candidates had been accessed that context and higher level information could be brought to bear to select one of the candidates (this constituted the second phase).

Once the initial "cohort" had been activated by the sensory input it responded to *mismatches* with the higher order information associated with a particular word's recognition element (i.e. the multi-layered procedural descriptions in the mental lexicon). The selection phase under this model depended crucially on how the initial cohort was reduced and it was only during this second phase that contextual knowledge was presumed to take effect.

Essentially Marslen-Wilson maintained that the cohort was narrowed by pruning from it all those candidates whose requirements produced a mismatch with the input (whether phonetic, syntactic or semantic). The model thus required that the sensory information activated an "initial cohort" and then (and only then) the top-down information, in conjunction with the continuing acoustic input, was allowed to narrow the cohort to a single word candidate.

The latest statement of the model however (Marslen-Wilson 1987), incorporates a greater degree of modularity. It retains the concept of initial obligatory bottom up processing but, in addition to this form-based access, Marslen-Wilson now argues for form-based *selection*.

During selection, different sources of information are integrated together to give the perceptual output, but they do not interact in the conventional sense. Also, once the word initial cohort has been activated, the decision as to whether to exclude a given candidate is no longer all-or-none. Rather than simple inclusion or exclusion to the cohort, the concept of activation has now been incorporated into the model. Thus if a candidates' internal specification does not match the incoming signal, its activation level falls. Similarly, if the internal specifications match the incoming information, the activation level for that candidate rises. Once a candidate reaches a higher level of activation than its competitors, it is integrated in to the higher level representation of the utterance and word recognition is said to have occurred.

Both the early and the present statements of the Cohort model have emphasized the importance of bottom-up processing. Indeed, throughout the chronological development of the Cohort model, much significance has been afforded to the priority of the bottom-up analysis of the "first one or two phonemes" (Marslen-Wilson and Welsh 1978, Marslen-Wilson and Tyler 1980), the first 200 milliseconds (Marslen-Wilson 1973, Marslen-Wilson 1975), or, as it currently stands, the first 150 milliseconds of a word (Marslen-Wilson 1987, Tyler and Wessels 1983, Tyler 1984).

2.5. Section Summary.

We have reviewed three models of word recognition. The major distinction between all three lies in *when* each permits the higher level contextual knowledge (top-down information) to contribute to processing.

The Serial Search model prohibits the contribution of top-down information until *after* a word has been phonologically accessed. The Logogen model states that contextual information may give rise to recognition *before* any of the acoustic-phonetic input is received. The Cohort model does not allow top-down information sources to participate in processing until after a specified amount (approximately 150 msec) of initial bottom-up processing has activated a pool of word candidates.

Because of the Cohort model's relatively precise claims about the temporal nature of word recognition the following sections examine the experimental evidence which has been used to support the position that the initial portion of a word will be processed bottom-up and that this "access" stage is immune to the influence of higher-level information. As the stances of the three models are mutually exclusive on this point, evidence in favour of one model has implications for the other two.

3. EVIDENCE RELATING TO THE COHORT MODEL OF LEXICAL ACCESS.

3.1. The earliness of word recognition.

Clearly, an important result which lends support to the Cohort Model's claim that top-down information can be utilized during the process of word recognition is the earliness of word recognition (i.e. the recognition of a word *before* its acoustic offset).

Demonstrating the earliness of word recognition places strong restrictions on how the selection phase can be organized. The length of time taken to recognize a word presumably reflects the completion of the selection phase. If at this point in time the available acoustic-phonetic information is insufficient by itself to support correct identification of the word then we must assume that other factors (namely contextual information) are taken into account.

Typical tasks which have been used to demonstrate the earliness of word recognition relative to the available sensory information are shadowing (Marslen-Wilson 1973, Marslen-Wilson 1975, Marslen-Wilson and Welsh 1978), monitoring (Marslen-Wilson and Tyler 1980) and, more recently, the gating task - a technique which has enabled a more direct assessment of the sufficiency of the acoustic-phonetic information (Grosjean 1980, Tyler and Wessels 1983, Tyler 1984).

The shadowing task required subjects to repeat back, as rapidly as possible, prose which is binaurally presented. Marslen-Wilson 1973, Marslen-Wilson 1975, observed that people were able to repeat words before all of the word has been heard. Listeners were able to repeat normal prose materials accurately and clearly at response delays of 250 - 275 msec from the onset of a word. Marslen-Wilson suggested that 50 - 75 msec of this delay may be time taken to articulate the response and from this he estimated that subjects were "recognizing" (or at least starting to repeat) words only 200 msec after their acoustic onset. The prose that subjects shadowed included mispronunciations (e.g. **TOMMORANE* for TOMORROW) and the structure of the prose was varied to be either syntactically correct or anomalous. When the syntactic structure of the prose was normal, subjects were more likely to restore the mispronounced words to their original form than if the prose was anomalous. Furthermore, restorations were more frequent when the disruption occurred in the second or later syllable of words embedded in a normal prose context. Marslen-Wilson interpreted these findings as illustrating the existence of an on-line interaction between the structural, semantic, lexical and phonetic levels of sentence processing. It appeared that the availability of the syntactic, semantic and lexical information from the normal prose contexts and the initial syllables of words facilitated recognition of those words with the effect that close analysis of the incoming acoustic information was not undertaken and, thus, mispronunced words were restored as they were repeated.

Similarly, Cole and Jakimik 1978 found that people were better able to detect mispronunciations when the words preceding the target mispronunciation were more constraining, for example the mispronunciation of *NECKLACE* was detected faster in <GOLD *MECKLACE> than in <OLD *MECKLACE>. They also found that mispronunciations which occurred in later syllables were detected more slowly - a result which parallels Marslen-Wilson's finding that mispronunciations in later syllables are more likely to be restored in shadowing. To the extent that shadowing and mispronunciation detection reflect normal word recognition processes we can deduce from these results that both the lexical and the syntactic/semantic context is influencing word recognition.

In experiments designed to demonstrate that contextual information is used on-line during word recognition, Marslen-Wilson and Tyler 1980 illustrated that monitoring is facilitated by the presence of syntactic and semantic information and that response latencies increase when there is no such information available.

Over a series of experiments Marslen-Wilson & Tyler manipulated both the context in which target words were embedded and the task demands. Target words appeared in one of three types of context: Normal Prose, Syntactic Prose (grammatical but non-sensical) and Random Word Order. Subjects were required to monitor for targets which were specified in one of three ways: the actual target word was given as a cue (Identical Monitoring) or a word which rhymed with the target was given as the cue (Rhyme Monitoring) or a word which was semantically related to the target was cued (Category Monitoring). Reaction times for monitoring were taken and it was found that people were able to monitor for words faster in Normal Prose contexts than Syntactic Prose and they are slowest to monitor for words in Random Word Order. The reaction times for the Identical Monitoring task illustrated that subjects were able to monitor for targets in the Normal Prose conditions 94 msec before their acoustic offset. It was estimated that a further 50 - 75 msec could have been taken to execute the response. From this Marslen-Wilson & Tyler concluded that people were able to recognize words in Normal Prose conditions 170 msec *before* their acoustic offset.

Marslen-Wilson & Tyler interpreted these findings as confirmation of their predictions that:

> "...sensory and contextual inputs are combined ... the stronger the contextual constraint, then the less dependence of the word identification decision upon acoustic-phonetic information..."
>
> (Marslen-Wilson and Tyler 1980, p. 14)

3.2. The relationship between early word recognition and bottom-up priority.

The studies we have reviewed above demonstrate clearly that contextual information can be used during the selection phase of word recognition. Marslen-Wilson and co-workers have used these results to argue that demonstrating early word recognition places strong constraints not only on how the "selection" phase can be organized but also on how the "access" phase is structured. The early recognition results show that top-down information influences the selection of one word candidate from its competitors in the word initial cohort and because words appear to be recognized before their acoustic *offset* but never before their *onset* it has been assumed that top-down information does not influence the initial access stage. Under the Cohort framework therefore, accessing the initial cohort is assumed to be a strictly bottom-up process.

Consistent with the notion of initial bottom-up processing is the finding that mispronunciations are more frequently detected in word initial syllables (Marslen-Wilson and Welsh 1978). In an experiment which manipulated (among other things) the syllable position of mispronunciations, Marslen-Wilson and Welsh demonstrated the priority afforded to initial segments. Mispronunciations were often detected when they occurred in word initial position but remained unnoticed in later syllable positions. It was reasoned that this effect of syllable position was attributable to the availability of context aiding recognition, thus overshadowing bottom-up detection, after the initial cohort has been activated. If the mispronunciation occurred in initial syllable position, the activation of the cohort was disabled and this disruption enabled subjects to detect the mispronunciation. Similar position effects were obtained in Samuel's study (Samuel, 1981a and 1981b see Section 3.3) where better discrimination was found to altered segments in word initial position.

Tyler & Wessels' more recent study (Tyler and Wessels 1983), also claims to substantiate the Cohort model's assumptions about initial bottom-up processing. Tyler and Wessels used the gating paradigm to determine how a subject's percept of a target word developed at various stages throughout its acoustic lifetime. They presented increments of target words which were embedded in contexts with varying degrees of constraint: minimally constraining and anomalous. Subjects were required to identify target words after each gated presentation. Tyler & Wessels found that a word heard in a minimally constraining context will be recognized only when one member of the initial cohort is consistent with both sensory *and* contextual constraints although other members of the cohort may be consistent with only one source of information. They note:

> " ... semantic context did not make a specific target word predictable in the absence of any sensory input ... but did have a large facilitatory effect after some acoustic input had been heard. These data, then, are at least consistent with a view against preselection of word candidates and in favour of word recognition being the result of the intersection of sensory and contextual constraints, given an initial set specified solely from the bottom up". (Tyler and Wessels 1983, p. 418)

It is not too controversial that the Tyler and Wessels and Marslen-Wilson and Welsh results support a model which does not allow for the pre-selection of word candidates. What is of interest however, is that from these results inferences have been made concerning a model of lexical access which allocates a specific period of time to bottom-up processing and which prohibits the contribution of top-down information during this initial stage. Note that the aims of the experiments reported above (and also in the previous section) had been to determine the availability of top-down information sources. We should question why these results are then interpreted as providing support for obligatory bottom-up processing of (approximately) the first 150 msec of the acoustic input. For instance, Marslen-Wilson and Tyler 1980 were specifically interested in the nature and time course of the role of *Top-down* information during the process of word recognition. The precise question which Marslen-Wilson and Tyler addressed in these experiments was whether there is a strict order in the availability of different knowledge sources during word recognition (as specified by the autonomous model discussed earlier in section 2.2). They asked:

> "...[how] early in the analysis ... do the distinguishable types of processing information become available" (Marslen-Wilson and Tyler 1980, p. 7)

Although their results demonstrate that the analyses developed in terms of any one knowledge source can, in principle, be made available to affect the operations of any other knowledge source, Marslen-Wilson & Tyler exclude from this the acoustic-phonetic analysis of the sensory input:

> "It is implausible ... that an acoustic-phonetic analyzer should need to, or would be able to, communicate or interact directly with a semantic analyzer." (Marslen-Wilson and Tyler 1980, p. 6)

Yet they note that:

> "...it is clear that, as contextual constraints vary, so too will the amount of sensory information necessary to reach a unique word choice."
> (Marslen-Wilson and Tyler 1980, p. 29)

If it is clear that top-down knowledge sources will determine how much of the sensory input needs to be analyzed, then it is consistent to assume that when the contextual information is highly constraining, there need be no fixed amount of acoustic-phonetic information (such as the 150 msec) which is processed bottom-up. It may not, therefore, be appropriate to rule that contextual information is prohibited from contributing during this initial period of processing. We should not assume that because contextual information appears not to be available much before the first 200 milliseconds that it *cannot* become available during this initial period.

The issue here is that whilst there are copious data to support the Cohort model's assumptions about the role of top-down information during word recognition, the only evidence we have to support the claims for initial bottom-up processing and prohibition of top-down information is *negative* evidence. Because we find that contextual information is able to influence the process of word recognition 200 msec after a word's onset, we cannot be convinced that the first 200 msec of a word will always be immune to effects of context. Advocates of the Cohort model have interpreted their data as demonstrating that approximately 150 msec of initial acoustic-phonetic material must be processed bottom-up and in no other way. An alternative interpretation however, is that there is a latency on the recruitment of top-down information and this varies as a function of the context.

Having said this we should appreciate however, that the principle of bottom-up priority *is* appealing. By assigning the initial impetus to bottom-up processing, a word recognition system maximizes its chances of finding the correct word [5] while still placing more or less severe constraints on the size of the consequent decision space. Also, deciding which information to retain from a limited pool of alternatives is computationally less complex than marshalling various types of knowledge to restrict access initially. In the following section we consider various experiments which have been designed to examine explicitly the question of initial bottom-up processing.

3.3. Evidence against Top-down (sentential) influences during initial processing."

Tyler 1984 examined the structure of the word initial cohort in order to confirm that the role of top-down information in accessing the cohort is restricted. Tyler analyzed responses from a gating experiment according to whether they constituted phonetically accurate reports of the stimulus and whether they were appropriate or not given the context in which target stimuli had been presented. Tyler noted that if it is the case that a word initial segment may be processed bottom-up and in no other way:

> "...there should be a moment in time at which all members of the initial cohort are active - whether or not they are contextually appropriate."
> (Tyler 1984, p. 418)

Tyler found that at the early gates Subjects produced a number of responses which were compatible with only the *sensory* input. Tyler noted that if context works in advance of the sensory information, then only contextually appropriate candidates would ever have been activated and:

> "...to be consistent with this claim, subjects should not have produced any contextually inappropriate responses."
> (Tyler 1984, p. 423)

Her results show that contextually inappropriate candidates *were* sometimes proposed and that there was no difference between the mean number of word candidates produced at the first gate depending on whether the target appeared in isolation or in a minimal context. To the extent that a significant proportion of phonetically inaccurate (and semantically appropriate) responses would constitute a strong claim against bottom-up priority, Tyler notes that these data support the initial obligatory bottom-up processing account.

[5] This is assuming that the acoustic-phonetic input is robust.

These data do not however, provide grounds for rejecting the possibility introduced above (Section 3.2), that contextual information could enable less of a dependence on the initial acoustic-phonetic input - even to the extent that context may be useful within the first 150 msec.

Indeed, Tyler's results reveal that contextual information might be attended to within the first 150 msec. The number of incorrectly recognized target initial phonemes was higher for targets heard in isolation. Although not true of the first gate, at gates two and three (i.e. the first 100-150 msec) of targets which were presented in context, identification of target initial phonemes was significantly better than for corresponding targets presented in isolation. Tyler attributes this difference to the listeners' advantage for adjusting to the speaker's voice in the context conditions and suggests that acoustic-phonetic fragments heard in context constitutes

> "...a more appropriate stimulus with which to gain access to the mental lexicon." (Tyler 1984, p. 420)

Although the context conditions might have provided relevant acoustic-phonetic information (i.e. the transitions into the target-initial stop consonants) it may be argued that this difference in identification of initial phonemes between context and isolation conditions was attributable to the earlier role of contextual information in lexical access.

McAllister 1988 further questioned the interpretations of Tyler's data. As we noted above, if contextual information is used to preselect a class of likely word candidates then subjects in Tyler's experiment should not have reported hearing contextually inappropriate words. McAllister points out however, that the contextually inappropriate responses observed in Tyler's study may be artifactual. As subjects had to report target words only, Tyler had no measure of the degree to which the context in which target words were embedded had been correctly understood. McAllister demonstrated that subjects do parse contextual strings incorrectly thus giving rise to incorrect target identifications. In such cases, the response may appear to the experimenter to be semantically inappropriate, whereas the response *is* actually appropriate by virtue of the subjective mis-parsing of the context.

A phenomenon which superficially implicates the early role of top-down information during word recognition is the phoneme restoration effect (see Warren 1970, Warren and Obusek 1971. Warren and his colleagues spliced out segments in a word and replaced these portions with non-linguistic sounds such as a coughs. Listeners reported hearing the words intact with a cough in the background. Samuel (1981a and 1981b) developed a paradigm to determine the extent to which the phoneme restoration effect was, in fact, a *perceptual* effect. He was interested in whether phoneme restoration resulted from top-down influences during processing of the acoustic-phonetic information or whether it was merely attributable to a preference for listeners to report words as intact. Samuel compared the magnitude of restoration between words which had noise replacing segments (as in Warren's studies) and words which had noise superimposed over the segments. To the extent that top-down information influenced the *perception* of the words, listeners should have been unable to judge whether the segment had been replaced or had noise added to it. Low discriminability of these two stimulus types may, however, be an artifact of the noise. Therefore Samuel also compared listeners' discriminations of the noise and noise-added segments in isolation. He found that whilst subjects could clearly discriminate between the noise and noise-added *segments,* when the segments were heard within their lexical contexts discrimination was poor i.e. listeners restored the replaced phonemes. Samuel concluded from this that lexical context influences perceptual processing during word recognition.

Note that until now we have talked about "top-down" information very generally. Some models of word recognition which incorporate both top-down and bottom-up processing suggest that the influence of top-down, contextual information depends on which particular level of context we consider (see below, and also Frauenfelder and Tyler 1987, for a review). Samuel, for example, claims that only *lexical* context can directly influence processing.

The pattern found in Samuel's earlier experiment however, did not hold for a similar discrimination task which involved sentential, rather than lexical, contexts. In this second experiment Samuel replaced or added noise to segments of minimal pairs (e.g. TAVERN/CAVERN) which were embedded in contexts which were biasing (appropriate) or non-biasing. When the segments were replaced by noise the minimal pairs became ostensibly ambiguous (i.e. noise replaced the [t] and [k] segments of Tavern/Cavern). Samuel reasoned that to the extent that the top-down sentential context causes perceptual restoration of the replaced segments, discrimination of added and replaced segments should be poor in cases where the targets are predictable from the context. The results however, showed that whilst listeners reported the predictable words as intact (noise-added), discrimination of added and replaced segments was actually better when the targets were in appropriate rather than inappropriate contexts.

Samuel interpreted this data as evidence that sentential context does *not* influence acoustic-phonetic processing. Note however, that whereas Samuel included a control condition in measuring the effects of lexical context, there was no such control condition in the sentential context experiments. In the first experiment responses to the added/replaced segments in their lexical contexts were compared to the same segments presented in isolation. In the second experiment however, added/replaced segments in their appropriate lexical-sentential contexts are only compared to the same segments in their inappropriate lexical-sentential contexts. It is not clear whether the observed difference in discrimination between appropriate and inappropriate contexts is due to appropriate context initiating post-perceptual biases or to an artifact of the inappropriate contexts which give rise to poorer discrimination. It is interesting to compare responses across the two experiments and to note that the discriminability scores (d') for the targets in appropriate sentential contexts were actually lower than scores obtained for the segments in isolation.

Samuel 1988 notes that:

> " ... good discriminability of controls, coupled with poor discriminability of the same sounds in context, is evidence for perceptual restoration."
> (Samuel 1988, p. 18)

It is therefore surprising that no control condition was incorporated in his sentential context experiment. Because we observed that, across the two experiments, discriminability of added/replaced items is poorer for targets in appropriate contexts than the added/replaced segments presented in isolation, an alternative interpretation of Samuel's data is that sentential context *does* affect initial processing of the acoustic-phonetic information.

An alternative paradigm for studying the effects of context on acoustic-phonetic processing is the Boundary Shift Paradigm developed by Ganong 1980. Using continua of speech sounds Ganong demonstrated that the point at which listeners perceive either a voiced or voiceless segment varies as a function of the lexical context. He synthesized VOT continua ranging from voiced initial phonemes to voiceless and also from word to non-word (e.g. DASH - * TASH, *DASK - TASK). Ganong observed that people are more likely to classify the mid-continua stimuli as constituting words rather than non-words. Also, Ganong found that the category boundary (the point which separates the perception of the voiced from the voiceless segment) shifted depending on which end of the continuum resulted in a word rather than a non-word percept. [6] Lexical status appeared to influence Subjects' categorizations of the acoustically ambiguous (i.e. the boundary) stimuli but not of the acoustically unambiguous (endpoint) stimuli. Ganong concluded that because the lexical effect is greater at the category boundaries than at the end points of the continua, it must arise at a stage in processing which is sensitive to both lexical knowledge and auditory information. That is, lexical-contextual knowledge is able to interact with auditory processing. [7]

This paradigm was successfully exploited by Connine 1987 and Connine and Clifton 1987. Connine concluded, as Samuel (1981a, 1981b) had done, that whilst lexical context influences initial processing of acoustic-phonetic information, sentential context effects are attributable to post-perceptual mechanisms.

Connine's basic premise was that patterns of reaction times to identification judgements using the boundary shift technique would be different according to whether or not responses reflected post-perceptual processing. In an experiment which was designed to demonstrate the effects of lexical context, Connine and Clifton 1987 created a continuum from a voiced word (DICE) to a voiceless non-word (*TICE) and vice versa (e.g. *DYPE - TYPE). They found that when the voiced end of the continuum was a word, listeners reported hearing a voiced sound at points further into the continuum than when the voiced end constituted a non-word. Furthermore, reaction times at the category boundary were faster for word than non-word consistent responses. Reaction times at endpoints of the continua however were similar regardless of whether reports constituted words or non-words. These data were interpreted as consistent with the notion that lexical context influences perceptual processing. It was reasoned that the ambiguity of the acoustic information at the category boundary would result in more phonetic hypotheses being

[6] The boundary shifted away from the word end of the continuum with the effect that a greater range of VOT values were perceived as words rather than non-words.

[7] Note that Ganong's results do not necessarily implicate the influence of contextual information on analysis of the acoustic-phonetic information. Ganong's results are also explainable by assuming that the initial processing is purely bottom-up. It may be argued that a word bias in the ambiguous region occurs because more than one sound is consistent with the in-coming information. The presence of the lexical entry is used to bias the identification response to correspond to the word rather than the non-word whereas responses to the endpoint stimuli can be made on the basis of the acoustic-phonetic information alone. We shall not discuss this issue further.

developed relative to the unambiguous stimuli at the endpoints of the continua. The existence of the lexical information for word-consistent responses at the category boundary was assumed to speed reaction times by virtue of providing information by which to select amongst the competing phonetic hypotheses. There was no reaction time advantage at the continua endpoints because processing of the clear acoustic-phonetic input did not require disambiguation by lexical information.

In order to check that these results could not be attributable to a response bias, Connine & Clifton ran a similar experiment in which they manipulated monetary payoff (a standard method for manipulating postperceptual bias). Voiced - voiceless non-word continua were created (e.g. *DICEL - *TICEL) and monetary payoff was made contingent on correct responses for one end of the continuum. As in the first experiment, the boundary shifted according to whether the voiced or voiceless endpoint was bias consistent. The reaction times however, showed a very different pattern. At the category boundary there was no reaction time advantage for bias consistent responses, but there was an advantage at the continua endpoints. Connine & Clifton argued that because payoff was made contingent on *correct* responses, there was no reaction time advantage to category boundary stimuli (where the acoustic information was ambiguous: i.e. the subjects could not be sure they were correct) but there was a clear advantage for bias-consistent responses at the endpoints where the acoustic-phonetic information was unambiguous.

In a subsequent experiment Connine 1987 investigated the effects of sentential contextual information on acoustic-phonetic processing. Voiced - voiceless word continua (e.g. DENT - TENT) were created and the targets were embedded in contexts which predicted either the voiced or voiceless word. Subjects' task was to label the initial phoneme and say whether the word was congruous or incongruous given the preceding context (i.e. they were asked to make judgements at the word level). Once again, the category boundary shifted as a function of the context. The pattern of reaction times however, matched those obtained in the earlier monetary payoff experiment. That is there was a reaction time advantage for contextually appropriate responses to endpoint stimuli but not to the ambiguous category boundary stimuli. Connine concluded that sentential information was utilized in a post-perceptual manner in the way that monetary payoff is, i.e. that sentential information did *not* influence acoustic-phonetic processing.

It is not clear however, that this conclusion is justified. Connine compared reaction times for phoneme labelling in *non-words* (the monetary payoff experiment) to phoneme labelling and *word* identifications in sentential contexts. This is problematic because of the possibility that non-words were not processed in the same way as words. In the monetary payoff experiment subjects were aware that they were attending only to non-words. It is plausible that reaction times were affected because their attention was focused on phonetic, rather than lexical, processing. This hypothesis receives credit in noting that reaction times to these non-word only conditions were some 200 msec faster than the reaction times yielded to similar non-words when the continua permitted a word or non-word response. It thus appears that although the reaction times obtained in the monetary payoff conditions reflected a post-perceptual bias, a similar pattern of reaction times for a different set of task demands in the sentential context conditions does not necessarily permit the conclusion that sentential context serves to likewise affect post-perceptual strategies. Manipulating monetary payoff to words in neutral contexts would perhaps have been a tighter control experiment.

3.4. Evidence for top-down information influencing initial processing.

In Sections 3.1 and 3.2 we presented evidence that contextual information is used during the selection phase of word recognition. The debatable issue however, as we saw in Section 3.3, is whether such information is able to exert any effect on *initial* processing during word recognition.

In the previous section we discussed essentially three different paradigms which have been employed to demonstrate that top-down information (which we henceforth take to mean *sentential* context) does not influence processing at the acoustic-phonetic level and we noted problems inherent in the different approaches. In this section we shall consider evidence which has been cited as consistent with the view that top-down information *does* influence processing at the acoustic-phonetic level.

We noted earlier (Section 3.3) that the Boundary Shift Paradigm has been successfully used to demonstrate that *lexical* context can affect acoustic-phonetic processing (Connine and Clifton 1987, Ganong 1980). This paradigm has also been employed by other researchers who claim that their results demonstrate that acoustic-phonetic processing can also be influenced by *sentential* context.

Garnes and Bond 1976 created a series of stimuli which approximated the minimal triple BAIT through DATE to GATE and presented these at the end of sentences which predicted one of them. Identification judgements shifted systematically so that contextually appropriate judgements were given for the ambiguous stimuli. In order to illustrate that the bottom-up information was interacting with the contextual information, several control experiment were run in which it was observed that when the stimuli were unambiguous, no effect of context was obtained. This result was important. It demonstrated both the strength of clear acoustic-phonetic information and that the contextual effects observed with the ambiguous stimuli were not simply attributable to the subjects' preference to report meaningful as opposed to anomalous sentences.

Similarly, Miller, Green and Schermer 1984 created a voicing continuum ranging from BATH to PATH. Tokens were presented at the end of semantically biasing sentences and subjects were required to identify the word initial phoneme of the target and then report the sentence context. Once again, identification of the ambiguous stimuli was influenced by context whilst the endpoint stimuli remained immune to such effects.

Both Garnes & Bond and Miller et al concluded that semantic context directly interacted with acoustic-phonetic processing. This interpretation is compromised, however, as neither study incorporated a control condition to estimate the degree to which results may have been attributable to post-perceptual mechanisms. Whilst these results clearly demonstrate that sentential context affected subjects' responses, it is possible that context was not directly influencing *bottom-up processing*. It might also be argued that contextual information had no influence on *initial processing* because neither of these studies provided the means to trace the time course of the top-down contributions.

Grosjean 1980 varied the contextual constraints under which increments of target words were presented. He presented gated target words following short and long contexts and required listeners to identify the target words on the basis of whatever amount of acoustic information was available. He found that the words which listeners proposed were congruent with the contextual information even in the early gates and, moreover, that no contextually inappropriate candidates were ever proposed. Grosjean maintained that contextual information therefore played an early role in word recognition and had an effect on acoustic-phonetic processing.

Although Grosjean's data were interpreted as evidence that context plays an early role in word recognition it should be noted that the semantically appropriate responses which subjects proposed were also initially phonetically accurate. The results obtained in this study therefore do not necessarily motivate an account of processing whereby sentential context influences acoustic-phonetic processing. Grosjean's data do not preclude an obligatory bottom-up processing account because, at the early gates, subjects received enough acoustic-phonetic input to be able to initiate the process of word recognition by bottom-up processing and *subsequently* select a candidate by reference to the contextual constraints.

From her analysis of results obtained in a similar gating experiment, Tyler 1984 (see Section 3.3) concluded that context does not operate at the pre-selection stage. Tyler's conclusion hinged in part on her finding that 5.4% of the responses produced to stimuli in which minimal semantic and strong syntactic constraints were available were inappropriate. Tyler argued that if contextual information was used to pre-select word candidates, then *no* contextually inappropriate candidates should have been proposed (see Section 3.3). Tyler's experiment included long context conditions, short context conditions and a no context condition. McAllister 1988 notes however that Tyler compared only the semantically inappropriate responses which had been proposed following the short and long context conditions. Tyler did not include the No Context condition in her analysis. McAllister argued that the exclusion of the No-Context conditions from Tyler's analysis meant that the number of "inappropriate" responses that would have been produced by chance was not estimated and that this exclusion may have lead to erroneous conclusions concerning the numbers of inappropriate responses.

McAllister 1988 presented gated targets in three context conditions: no context, short context and long context. The numbers of semantically appropriate responses were analyzed. For the no context condition, appropriate responses were taken to be responses which would have been appropriate for the short context conditions. This allowed a better estimate of the number of inappropriate responses obtained to the contextual conditions which are due to chance factors. In a comparison between the No Context and the Short Context conditions, McAllister found that the number of semantically appropriate responses was significantly affected by context: 31% of responses to the No Context condition were appropriate compared against 83% to the Short Context conditions. Naturally, there was an even more marked difference between the number of appropriate responses in the No

Context and the Long Context conditions.

McAllister's results also replicate Grosjean's in that the amount of acoustic information needed before a word was isolated was significantly affected by context: words in long context conditions were recognized faster than those in short context and words in isolation were recognized slowest.

McAllister's interpretation of the data is not compromised in the way that Grosjean's was however (see above) because of her analysis of targets presented in the No Context conditions. Whilst McAllister makes no claims as to the influence of context on acoustic-phonetic processing *per se*, her data clearly support the view that contextual information is available and usable by listeners as early as 50 msec after the acoustic onset of a word.

3.5. Section Summary.

In this section we have reviewed the evidence both in favour of and against the view that contextual information is able to influence acoustic-phonetic processing. This issue was interesting because the point, during word recognition, at which contextual information is assumed to influence processing distinguishes the three models of word recognition that we considered in Section 2 of this Chapter.

The Cohort Model (Marslen-Wilson 1987) states that initial 150 msec of a word is processed purely bottom-up and that no reference to contextual information is made during this initial "access" phase. We have presented evidence which both agrees with and contradicts this claim. It appears to be the case that context is able to influence at least the selection phase of word recognition. The data are, unfortunately, less uniform on the role of contextual information during lexical *access*.

In Section 3.2 we noted why advocates of models of word recognition which place the burden of initial processing on analysis of the acoustic-phonetic input should want to define their models in such a way. There were notably two factors: computability and people's ability to recognize words in strange or anomalous contexts. Both of these notions are acceptable reasons for retaining an initial bottom-up component in models of lexical access. The data currently available however, allow no firm conclusions. On the one hand we discussed studies which claimed that the effects of sentential context were attributable to post-perceptual mechanisms and *not* to effects on bottom-up processing of the acoustic-phonetic information (Section 3.3). On the other hand, the studies reviewed in Section 3.4. implicated the early role of context. In the following section we consider evidence which illustrates that the principle of bottom-up priority may not, in fact, be quite appealing in terms of the constraints it affords and we discuss the reasons behind seeking an alternative account; one which specifies that the initial processing during word recognition may not be immune to contextual effects.

4. PROBLEMS INHERENT IN BOTTOM-UP PRIORITY.

4.1. The poor acoustic quality of speech.

A fundamental problem for the notion of obligatory bottom-up processing is the poor quality of conversational speech. During normal conversation, segments become elided or omitted altogether. Coarticulation gives rise to the problems associated with acoustic non-invariance in the speech wave form (see Liberman, Delattre and Cooper 1952; Liberman et al. 1957; Locke 1968; Liberman et al. 1967) and the natural environment provides extraneous noises which may distort or completely mask certain stretches of sound in the acoustic-phonetic input. Furthermore, Norris 1982, notes that it is very difficult to explain how words with distorted beginnings, such as "dwibble", might be recognized at all. This problematic state of affairs however, is not merely occasional. We are faced with such degraded speech stimuli in any normal conversation, yet somehow we are able to initiate the process of word recognition.

4.2. The order in which words in continuous speech are recognized.

A corollary of obligatory bottom-up processing is that words must be recognized in the order in which they are spoken (see Cole and Jakimik 1980). Until the first word has been successfully recognized, the system is unable to locate the initial segment of the second word upon which it relies for access to the lexicon.

This reliance on obligatory bottom-up processing of the initial segment of a word to initiate the recognition process clearly has severe drawbacks. Grosjean 1985, for example, points out that people are uncertain about the location of word boundaries. Also, several studies have demonstrated that many words are not actually recognized in strict sequence. It appears that a word may be recognized *after* its acoustic offset. In many cases this may be as late as three or four words later in the sequence (see for instance Pollack and Pickett

1963, Pollack and Pickett 1964; Grosjean 1985; Bard, Shillcock and Altmann 1988). If the first word of an utterance is not successfully recognized, it can be virtually impossible to locate subsequent word boundaries by way of reference to the acoustic signal alone as it is very rare to find discrete temporal segments separated by pauses in conversational speech (see Reddy 1976, and Klatt 1977). Second, experimental evidence suggests that we *are* able to initiate, or successfully recognize words from acoustic information later in the word (see Nooteboom 1981; and Salasoo and Pisoni 1985).

Clearly, the findings that words can sometimes be recognized after their acoustic offset and that word recognition can succeed without contemporaneous recognition of the initial segment of a word is problematic for theories of word recognition which allocate the initial burden to bottom-up processing.

4.3. Context effects in the production of speech.

The situation is complicated further still in light of studies which demonstrate an added dimension to the problem of non-invariance in the speech waveform; it appears that sentential context is able to effect the articulatory realization of an utterance.

Lieberman 1963 demonstrated that the predictability of an utterance (as measured by redundancy) is inversely related to intelligibility. He asked people to read aloud sentences which strongly or weakly predicted one of the words. He spliced targets from their original contexts and presented them in isolation for listeners to identify. When the targets were presented in noise, subjects were able to identify the predictable words significantly less often than they could identify unpredictable words. When the sentences from which the targets had been spliced were also presented, subjects identification of both predictable and unpredictable words was greatly facilitated.

Lieberman accounts for these findings by assuming that the speaker calls greater attention to the words that he/she thinks are unpredictable. He also notes that in the majority of cases, the unpredictable words received more stress (measured in terms of average amplitude). Lieberman notes:

> "The acoustic cues may be both necessary and sufficient for the identification of a word in the absence of any other information ... The acoustic cues may, however, supply information that is necessary but not sufficient for the identification of a word. The effect of predictable context is to reduce the acoustic cues from the primary level to the secondary level." (Lieberman 1963, p. 184)

Hunnicutt 1985 replicated and extended Lieberman's work. She criticized Lieberman's study and questioned whether it was general enough to allow the conclusions drawn. In particular, Hunnicutt noted that only 11 sentences and only 7 target words were used in his analysis. Because of the extensive use that had been made of the Lieberman results and the paucity of corroborative experiments examining this position, Hunnicutt attempted to replicate Lieberman's findings using a greater number of target words and a greater selection of predictable contexts.

She placed target word pairs in similar positions in two sets of sentences: sentence pairs which might be found in text (one with high and one with low predictability for targets) and sentence pairs consisting of a proverb and a matched, low predictable, sentence.

Hunnicutt found that for the text-type sentence pairs there was an intelligibility advantage for targets from the less predictive contexts. For the proverb and matched less predictive sentence pairs however, no such advantage for targets from the lower predictable contexts was observed. These results were interpreted as evidence that intelligibility and predictability are related in *some* instances (i.e. the text-type sentences). [8]

Using spontaneous speech, Bard and Anderson 1983 obtained similar results. They reported that speakers tended to adjust intelligibility according to the predictability of the words from their sentence contexts.

Similarly, Fowler and Housum 1987 measured the intelligibility of words taken from discourse contexts. They noted that words which had been previously uttered in the discourse ("old" words) were shortened by the speaker more than were words which were just being introduced into the discourse ("new" words). Furthermore, they found that old words were also less intelligible when presented in isolation and that listeners were able actively to use these differences in the production of old and new words to facilitate discourse comprehension. [9]

[8] It was argued that the proverb is not representative of highly predictable contexts.

[9] We shall consider this particular issue in greater depth in Chapter 4.

Unfortunately, not much other work has been undertaken on investigating the possible effects of context in speech production. It is clear however, that if the production of speech is controlled as a function of context, then the usefulness of constraining processing by bottom-up information alone may be reduced. If contextual information can serve to make words indistinguishable on acoustic grounds, then the acoustic-phonetic information will no longer provide the necessary bottom-up information to be able to promote the activation of a set of word candidates sufficiently above the activation of other candidates, e.g. the close competitors.

5. SUMMARY.

In this Chapter we have considered a model of lexical access which maintains that initial processing of during word recognition proceeds bottom-up. The Cohort model (Marslen-Wilson 1987) claims that top-down information can not be brought to bear during this initial period in processing.

In Section 3.1 we presented evidence which supported the Cohort model's claims that top-down information influences word recognition; in Sections 3.2 and 3.4 however, we reviewed evidence which suggested that the latency on the recruitment of contextual information may not be as great as 150 msec.

Section 4 noted the problems associated with the Cohort model's implicit assumptions regarding the disproportionate importance attached to word initial segments and reference was made to the problems the model faces in virtue of the poor acoustic quality of conversational speech.

We also briefly outlined two other models of lexical access: the Autonomous Search model (Forster 1979) and the Logogen model (Morton 1969) and it was explained how the evidence which we have cited in connection with the Cohort model bears directly on these other models. We chose to specifically examine the Cohort model because of its relatively precise claims about the temporal nature of word recognition. Clearly findings concerning the time course and relative contributions of top-down and bottom-up information during word recognition are as pertinent to the Autonomous Search and Logogen models as they are to the Cohort model.

The issue which will concern us throughout this thesis is *when* during word recognition top-down information can be brought to bear. More specifically , we shall empirically examine the assumption explicit in the Cohort model, that the first 150 msec of a word is immune to contextual effects. We ask how rigorously we need adhere to the notion that the first 150 msec is given bottom-up priority. We do not anticipate that context will operate in the manner precluded by Tyler 1984 i.e. by preselecting a class of likely word candidates to enter into the recognition process. Rather we predict that the role of context will be to aid processing during the first 150 msec of lexical access. In virtue of the poor quality of conversational speech, the noisy environment in which listeners typically have to operate and the possibility that the articulation of word initial segments may be controlled in some contexts, it seems desirable that contextual information might aid initial processing. We thus investigate the possibility of effects of context, both in perception and, (in light of the Lieberman 1963 studies), in production, and we pay special attention to responses made to the first 150 msec of the acoustic-phonetic stimuli.

The remainder of this thesis presents experimental evidence which bears directly on these issues. First, we investigate in the following chapter whether the clarity of articulation of speech is dependent on the contextual environment and consider the implications of this for bottom-up models. We measure differences in the physical stimulus and also discrepancies in intelligibility which may be attributable to contextual factors.

We also consider how prosody might contribute to the processing of bottom-up information and we examine the possibility that context may sufficiently influence the prosodic contour of an utterance to the extent that the listener is cued to expect a predictable or unpredictable target.

Finally, two gating experiments are presented. In these studies the intelligibility of the initial acoustic segment and the appropriateness of context is manipulated. Responses within the first 150 msec of target presentations are examined and the implications for the Cohort model of lexical access are discussed. We conclude that whilst bottom-up information is given priority during initial processing, top-down information can be brought to bear within the first 150 msec.

CHAPTER 3:

THE EFFECTS OF HIGHER LEVEL INFORMATION ON THE ARTICULATORY REALIZATION OF AN UTTERANCE.

1. INTRODUCTION.

Is the predictability of context inversely related to intelligibility? Previous studies have demonstrated that the predictability of the context in which a word is embedded trades against the intelligibility of that word to the extent that the more predictable a word is, the more difficult that word is to identify (see, for example, Lieberman 1963, Hunnicutt 1985, Bard and Anderson 1983, Fowler and Housum 1987).

If it can be demonstrated that the realization of an utterance is context dependent, then the acoustic-phonetic input during the process of word recognition is more complex than many psycholinguists have hitherto acknowledged. As we discussed in Chapter 2, Section 4.3, this issue bears directly on models of word recognition which emphasize the importance of initial bottom-up processing: To the extent that an appropriate context will make the realization of a word indistinguishable, on acoustic grounds, from other words (candidates in the mental lexicon), the constraints afforded by the notion of initial bottom-up processing are minimal.

In this Chapter we ask whether this predictability-intelligibility trade-off is realized in Voice Onset Time (VOT). More precisely, we ask whether voiced and voiceless stop consonants uttered in non-constraining environments are more clearly differentiated in terms of VOT than the same stops uttered in constraining environments, and we examine whether this predicted differentiation in production translates to perception, making the stops produced in non-constraining contexts easier to distinguish. [1]

We measure VOT's which have been uttered in appropriate (i.e. constraining) and inappropriate (non-constraining) contexts. We then examine whether these contexts affect the intelligibility of the syllables in question by presenting them in neutral contexts.

Before we describe the first of these experiments however, it will be useful for us to consider one or two points concerning the materials used. The following section describes how the materials were pre-tested.

2. PRE-TEST.

In order to relate the possible effects of context on the VOT of stop consonants to the issue of bottom-up priority, it is necessary to create materials which make possible the direct assessment of these effects. Materials are needed in which target words which begin with a voiced or voiceless consonant occur in contexts which predict the target words to the same extents (or, for the inappropriate conditions, contexts which never predict the target word). To achieve such control of preceding contexts, a pre-test was conducted which examined the predictability of chosen target words relative to the sentences in which they occurred.

Thirty five minimal pairs of words were found which differed only on the voicing of the initial stop consonant and which were matched for frequency within pairs (Kucera and Francis 1967). With the help of four independent judges, context sentences were created for each minimal pair. These contexts were matched within pairs as far as possible, for length, grammatical structure, and anticipated intonation pattern and were devised so that only one member of the minimal pair constituted an appropriate continuation of the string. For instance, for the minimal pair BEES/PEAS, the following matched contextual sentence pairs were created:

(1a) John hates hornets, wasps and

[1] We thus treat intelligibility as the discrimination of a word from its voiced or voiceless minimal pair counterpart as opposed to the discrimination of one word from all others.

(1b) John hates potatoes, carrots and _____.

and

(2a) The pollen was stuck all over the _____.

(2b) The gravy was poured all over the _____.

To verify that the contexts did in fact predict the target words, a cloze test was administered (see Marslen-Wilson and Welsh 1978). All context sentences [2] were printed without the target words and distributed amongst 37 subjects who were asked to write down what they considered to be an appropriate continuation of the string.

As in Marslen-Wilson and Welsh 1978, responses were scored as follows: responses which were identical to the (missing) target words scored 1; responses which were synonymous with the the target words scored 2; responses which were related to the target words scored 3 and unrelated responses scored 4.

Context sentences were selected according to the following criteria: The average score across the 37 subjects should be 2 or less - indicating that the context predicted at least a synonym of the target or the target itself, whilst the difference between the average scores for any pair of the context sentences (i.e. the strings predicting both members of a minimal pair) should not exceed 1. These constraints on the selection of materials ensured that all contextual strings predicted targets and predicted targets equally well. The scores from this pre-test and the lists of context sentences and associated targets are given in Appendix A.

3. EXPERIMENT 1.

In the Introduction to this Chapter we noted the implications of results which demonstrate that context affects the articulatory realization of an utterance. Studies to date which have typically been interpreted as evidence for such effects however (e.g. Lieberman 1963; and Hunnicutt 1985) were not designed to address this issue explicitly. We should not, therefore, evaluate the weakness of the constraints provided by bottom-up processing by citing the results to these earlier studies. They measured, primarily, the *intelligibility* of word tokens taken from varying contexts and do not provide sufficient evidence for

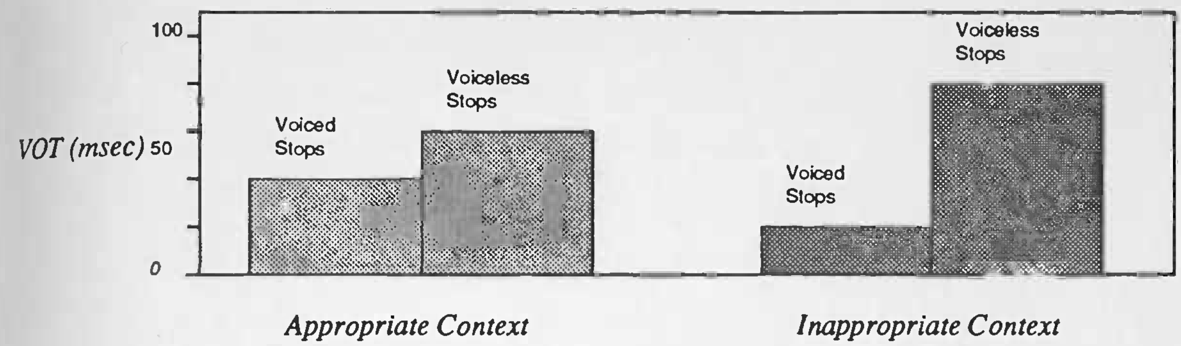
[2] There were more than 70 contextual strings in the cloze test; as many as 6 strings were associated with some target words.

describing how varying contexts may affect the *acoustic-phonetic output*. In measuring the Voice Onset Times of minimal pair target words uttered in different contexts, we isolate *one* such acoustic-phonetic parameter.

In the first of our experiments we ask whether the VOT value actually produced for a token of a voiced or voiceless stop is a function of the constraints provided by the context. If the control of VOT is relaxed when other environmental factors carry meaning, we predict that the distribution of VOT's for voiced stops will spread, making some stops more similar to the VOT values normally associated with their voiceless counterparts and, similarly, that the distribution of the VOT's for voiceless stops will spread (making some voiceless stops more similar to their voiced counterparts). In situations where the context is not highly constraining however, we hypothesize that the control of VOT is tightened to the extent that voiced and voiceless stops become more clearly differentiated, i.e. the voiced and voiceless stops will assume clearly distinct distributions. Whilst VOT is not the only parameter to influence the voicing distinction, to the extent that contextual constraints may allow VOT values for pairs of stops to overlap, stop pairs may tend to become indistinguishable on purely acoustic grounds. Figure 3.1 illustrates how the VOT values might be affected by context in this manner.

FIGURE 3.1

Hypothetical Voice Onset Times
Context by Voicing



In order to test this hypothesis, a simple production experiment was run. We measured the VOT's of word initial stops which appeared at the end of contexts for which the target words (containing the voiced or voiceless stops) were either appropriate continuations or inappropriate continuations. An inappropriate continuation was one in which the preceding context predicted a word with the opposite initial voicing feature to the word which actually appeared.

If context works to affect the realization of VOT on sentence-final word-initial stops in the way that we have hypothesized, then observing an interaction between context and voicing would be theoretically significant because it would reflect the differential spread of VOT values of the voiced and voiceless stops as a function of the context: For constraining contexts, we expect the distributions of the VOT values for voiced and voiceless stops to spread to the extent that they overlap, and, for non-constraining contexts, we expect the distributions of the VOT values to become more clear differentiated (see Figure 3.1). We also anticipate a main effect of voicing due to the nature of the voicing distinction. It is reasonable to suppose that voiced and voiceless stops will be generally distinguishable, although it is plausible that this affect be weakened by the existence of a strong effect of context on voicing.

4. METHOD.

4.1. Subjects.

Three female undergraduates of Edinburgh University were asked to participate in two recording sessions. The subjects were native speakers of British English and shared the same regional Edinburgh accent. All three were naive as to the purpose of the experiment. They received no payment or course credits for their help.

4.2. Materials.

Twenty-four minimal pairs differing only on the voicing of the initial stop were matched within pairs for frequency. The minimal pairs were balanced for place of articulation such that there were 8 bilabial stops ([p]/[b]), 8 alveolar stops ([t]/[d]), and 8 velar stops ([k]/[g]). These 24 pairs constituted the 48 target words that were placed in sentence final position of varyingly constraining contexts. Two levels of constraint were created for these contexts: appropriate and inappropriate contexts. These levels represented respectively high and low constraining contexts.

Appropriate contexts

For each target word a context sentence was created which predicted it. Context sentences were matched within pairs for length, rhythm, syntactic structure, expected intonation pattern and the extent to which it predicted the respective target word. Predictability was controlled for and measured by means of the cloze test described in Section 1.2 above, and examples of these "appropriate" context sentences are given in (3a) and (3b) below. This yielded 48 highly constraining contexts which we shall now refer to as the "Appropriate" contexts.

Inappropriate contexts

A second set of context sentences were then created which were inappropriate because they were *never* predicted the target word. The Inappropriate contexts were created by replacing the targets in the appropriate contexts by the opposite voiced or voiceless counterpart of the minimal pair. Thus the context which predicted the voiced member of the minimal pair was actually now associated with the voiceless member and vice versa. Examples of these inappropriate contexts are givin in (4b) and (5a) in Section 4.3 below. All materials used in Experiment 1 are given in Appendix B.

4.3. Design.

Twenty-four minimal pairs which differed on the voicing feature of the initial stop (2) and which were balanced for place of articulation were presented following appropriate and inappropriate contexts (2) thus giving 96 sentences in total.

Finally, two sets of 48 filler sentences were created. The final words of the filler sentences all began with fricatives, liquids, vowels or nasals and the sentences were matched to the experimental sentences for rhythm, length, syntactic structure, anticipated intonation pattern and the extent to which the filler target was predicted. [3]

Experimental and filler sentences were distributed across two groups with the constraint that a given context sentence, whether it was serving as an appropriate or inappropriate context for a given target, should only occur once within a group. A given target word however, was allowed to occur twice within one group: following its appropriate and its inappropriate contexts.

To illustrate the composition of materials across the two groups consider the following base sentences:

(3a) John likes hornets, wasps and bees.

(3b) John likes potatos, carrots and peas.

[3] The extent to which filler targets were predicted by the preceding context was measured informally. No cloze test was administered for the filler sentences.

These were distributed across the two materials groups in the following way:

Group 1.

Appropriate Context:

(4a) John likes hornets, wasps and bees.

Inappropriate Context:

(4b) John likes potatos, carrots and bees.

Filler matched to Appropriate Context:

(4d) Mary likes planes, trains and ships.

Filler matched to Inappropriate Context:

(4e) Mary likes cats, dogs and shops.

Group 2.

Inappropriate Context:

(5a) John likes hornets, wasps and peas.

Appropriate Context:

(5b) John likes potatos, carrots and peas.

Filler matched to Appropriate Context:

(5d) Mary likes golf, tennis and squash.

Filler matched to Inappropriate Context:

(5e) Mary likes bread, cheese and chess.

Sentences were pseudo-randomly distributed within each group. Two booklets were printed, one for each group of materials. Sentences were printed three times per page and only one context and target appeared on one page.

4.4. Procedure.

The three subjects participated individually in two recording sessions corresponding to the two material groups. There was an interval of one week separating each recording session and each session lasted approximately 50 minutes. [4]

The subjects were seated at a table in a sound attenuated recording studio. They were given as much time as they needed to record practice sentences. These practice sentences were similar to the filler sentences and were necessary so that the subjects could become accustomed to the experimental situation and to the microphone which was fixed to one side of the table. Subjects were also provided with a glass of water which they were allowed to sip between sentences.

Subjects were instructed to take as much time as necessary to read through the (three) sentences on any one page of the materials booklet before beginning to read aloud. They were asked not to turn pages whilst they were speaking into the microphone and to read all three repetitions of each sentence before continuing to the next page. No time restriction was put on reading speed, instead, subjects were encouraged to read at their own speed and to speak as "naturally" as possible.

4.5. Data Collection and Analysis.

Subjects spoke into an RF condenser microphone (Sennheeisser MKH815T) connected to a digital-audio processor (a Sony pcm-F1 system) which recorded the signal digitally on a Sony SLF1 Betamax video recorder.

These recordings were subsequently redigitized on a Unix Masscomp workstation at 1600 Hz and were analyzed using the speech analysis software package developed at Edinburgh University's Centre for Speech Technology Research ("Audlab").

VOT's for the target word initial stops in each of the experimental sentences were measured from time-amplitude, energy tracks and spectrographic displays. Boundaries corresponding to the release of the stop burst and the onset of voicing were located and the interval separating the two was calculated. As subjects had read aloud each sentence three times, there were three separate VOT values for each target word initial stop.

---

[4] In an effort to counterbalance the order of recording the 2 groups of materials, two of the subjects recorded the materials from Group 1 first and the the third subject recorded Group 2 first. Because of this imbalance, possible effects of order were not analyzed.

5. RESULTS.

The VOT values for a given sentence were fed into an Analysis of Variance with the following factors: Context, Voicing, Place of Articulation, and Speaker (subject). This yielded significant main effects of Speaker, $F1(2,2) = 146.81$, $p < .0001$, Voicing, $F1(2,2) = 1050.73$, $p < .0001$, [$F2(1,23) = 636$, $p < .05$]; and Place of Articulation, $F1(2,2) = 268.56$, $p < .0001$. Significant interactions were yielded for Place by Voice, $F1(2,4) = 8.00$, $p < 0.05$, and Place by Voice by Context, $F1(2,4) = 10.93$, $p < 0.05$, Note that the Voice by Context interaction was not significant ($p = 0.24$). The means for the Voice x Context x Place interaction are presented, by the 3 places of articulation, in Tables 3.1a, 3.1b and 3.1c below. These are illustrated in Graphs 3.1a, 3.1b, and 3.1c. For comparison, the Voice x Context means are presented in Table 3.2 and illustrated in Graph 3.2.

TABLE 3.1a

Mean Voice Onset Times (msec)

Context by Voicing by Place of Articulation: Bilabial Stops.

VOTs	Appropriate Context	Inappropriate Context
Voiced stops	8.46	8.43
Voiceless stops	39.74	56.35

TABLE 3.1b

Mean Voice Onset Times (msec)

Context by Voicing by Place of Articulation: Alveolar Stops.

VOTs	Appropriate Context	Inappropriate Context
Voiced stops	16.18	14.35
Voiceless stops	78.01	75.97

TABLE 3.1c

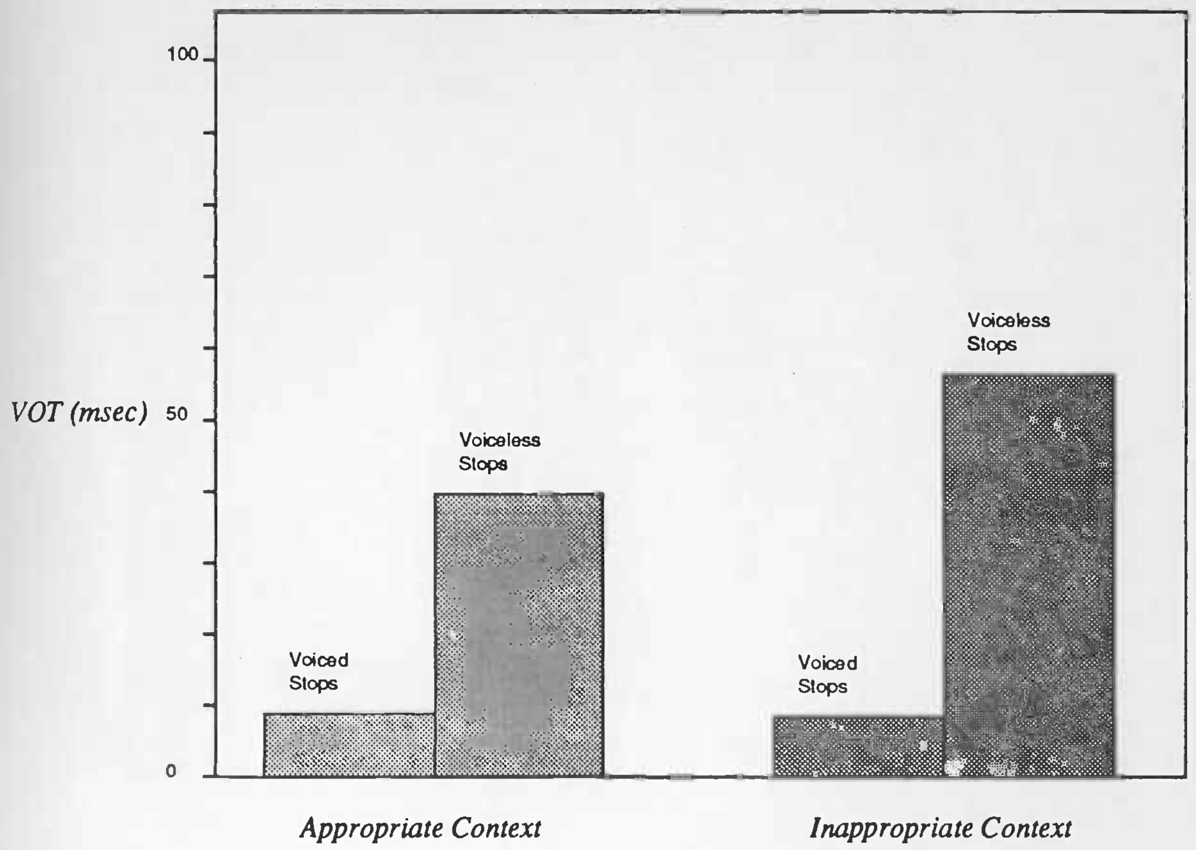
Mean Voice Onset Times (msec)

Context by Voicing by Place of Articulation: Velar Stops.

VOTs	Appropriate Context	Inappropriate Context
Voiced stops	19.76	17.62
Voiceless stops	81.35	83.84

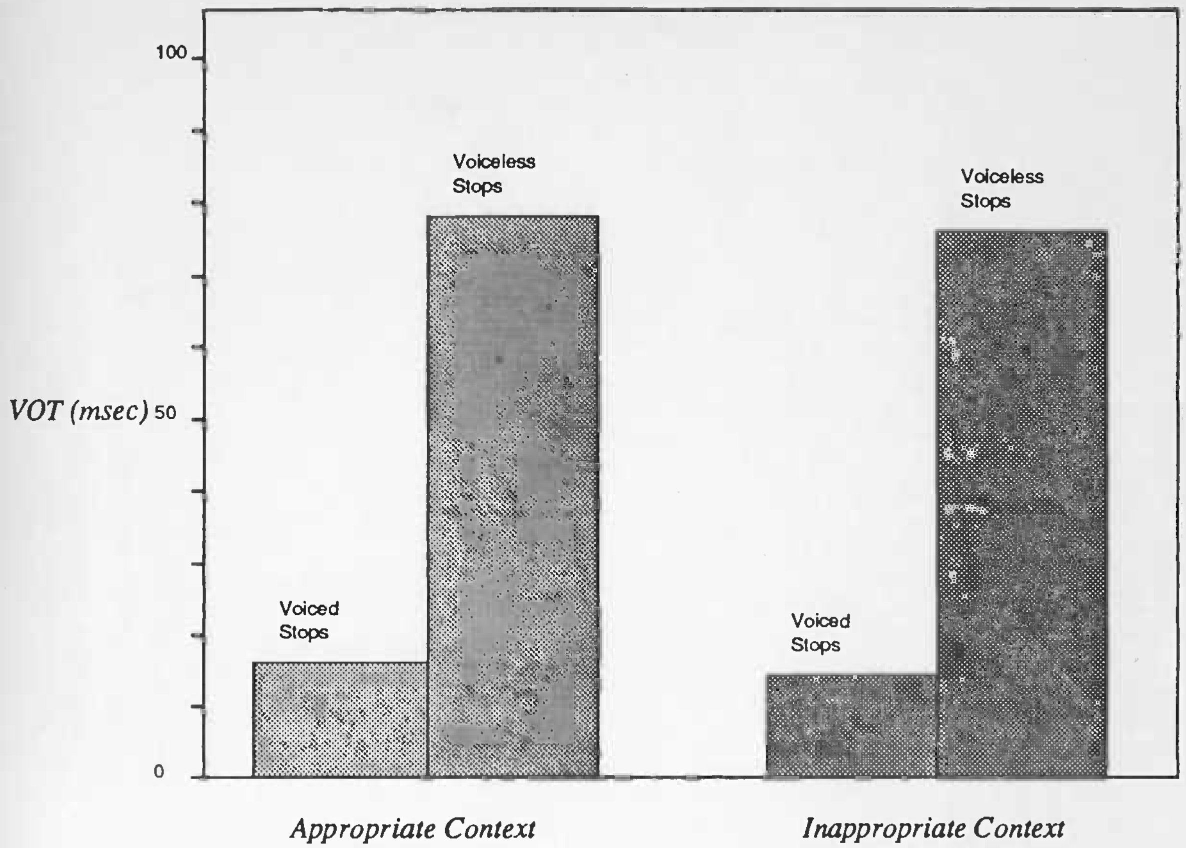
GRAPH 3.1a

Mean Voice Onset Times
Context by Voicing by Place of Articulation: Bilabial Stops



GRAPH 3.1b

Mean Voice Onset Times
Context by Voicing by Place of Articulation: Aveolar Stops



GRAPH 3.1c

Mean Voice Onset Times
Context by Voicing by Place of Articulation: Velar Stops

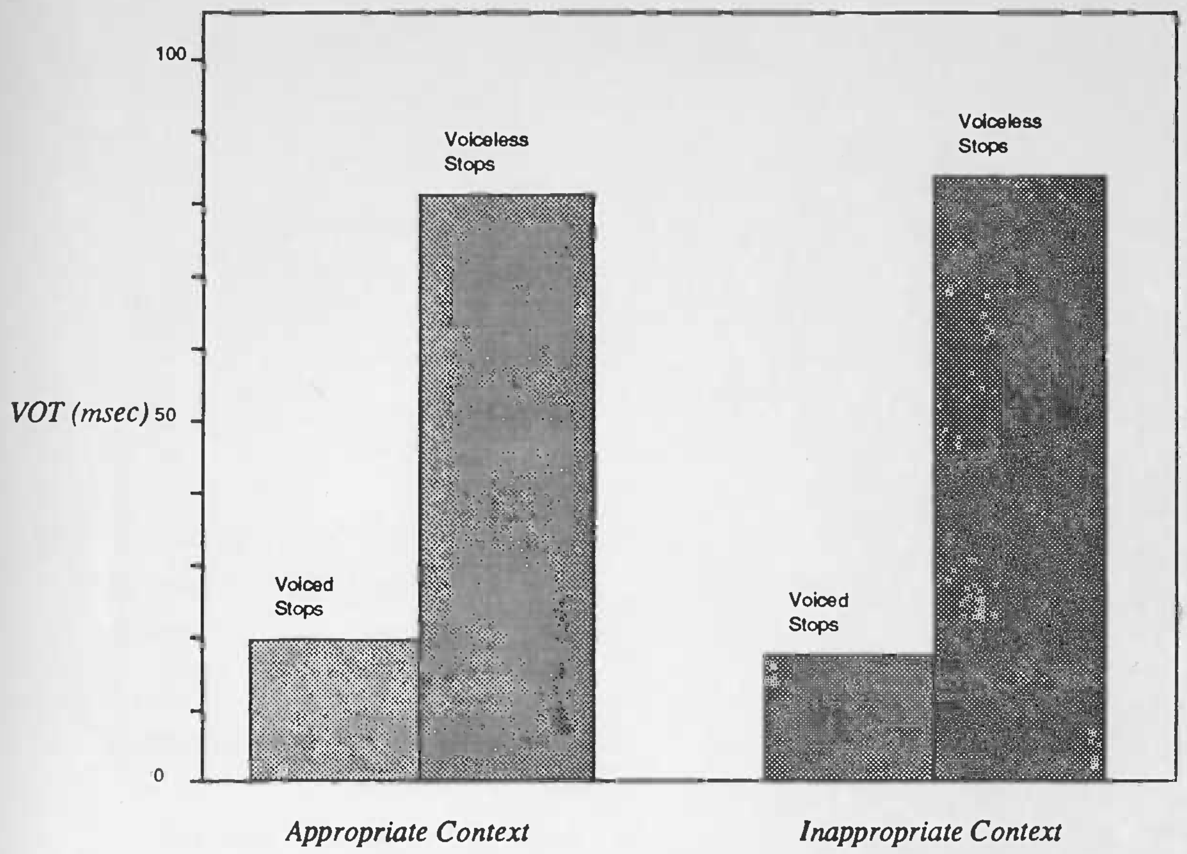


TABLE 3.2

Mean Voice Onset Times (msec)
Context by Voicing.

VOTs	Appropriate Context	Inappropriate Context
Voiced stops	14.79	13.46
Voiceless stops	66.37	72.06

The Tukey HSD test was computed (HSD = 12.25, p = 0.01). The only means to shift significantly apart from each other as a function of context were the voiceless bilabial means.

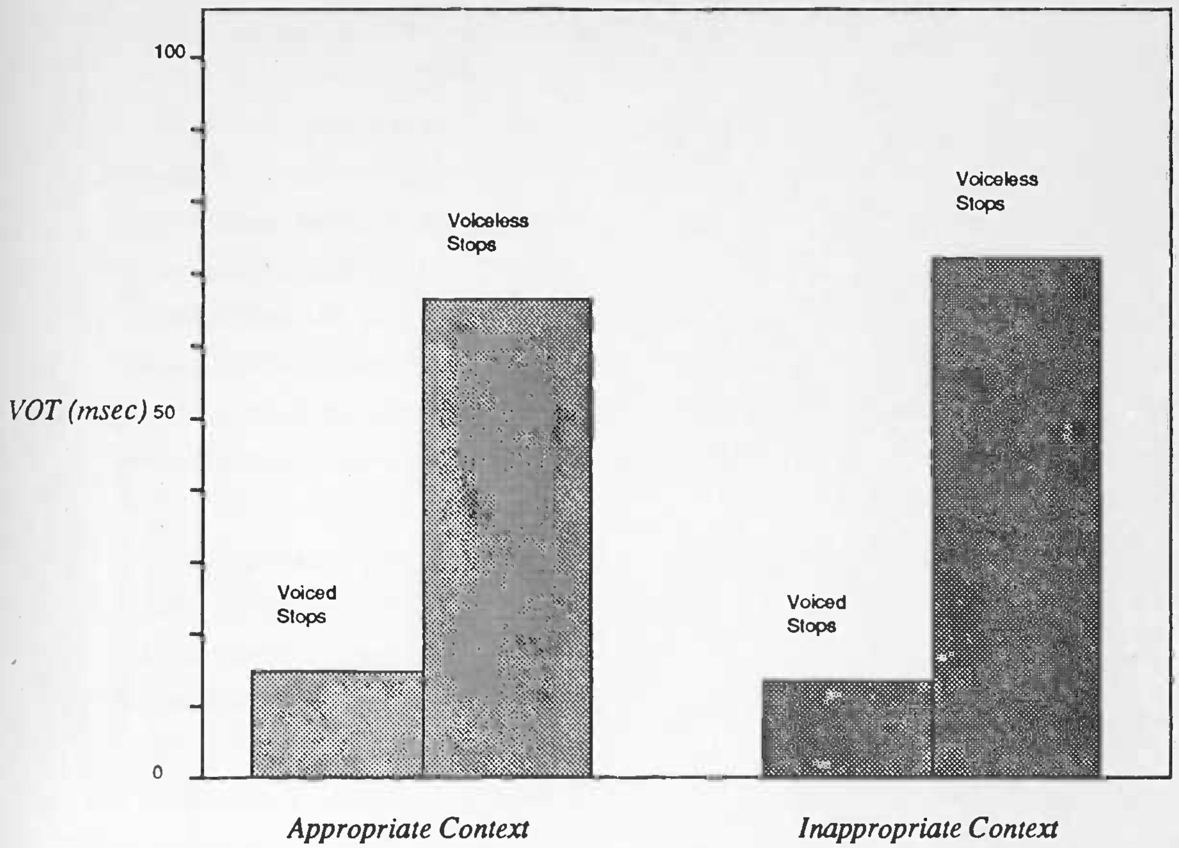
6. DISCUSSION.

The main effect of voicing is not surprising. Given that voiced and voiceless stops necessarily assume lower or higher VOT values respectively, we anticipated that the difference between voiced and voiceless stops should be significant. The aim of this experiment was to ascertain whether the voicing distinction might become less marked when the context was highly constraining.

The Voice x Context x Place of Articulation interaction appears to be the result of a large shift in the anticiapted direction for the voiceless bilabial stops and a shift in the *opposite* direction for the voiceless alveolar stops. The vital effect for our hypothesis was whether the distributions of VOT values for the voiced and voiceless stops spread as a function of context. Only the VOT values for the voiceless bilabials however, spread in the way which we predicted.

GRAPH 3.2

Mean Voice Onset Times
Context by Voicing



7. CONCLUSION.

It appears that under the experimental conditions which we have imposed, VOT is not controlled according to the context in which it is uttered. The lack of a statistically significant interaction of Voice x Context suggests that the VOT values for pairs of stops uttered in appropriate contexts do not tend to spread.

This result however, does not indicate that the pairs of stops uttered in appropriate contexts are not indistinguishable on acoustic grounds. We should be careful not to interpret these results as counter evidence to the suggestion that context and intelligibility are inversely related, as they do not contradict the data provided by Lieberman 1963 and Hunnicutt 1985. As we mentioned earlier, neither of these studies attempted to specify how contextual effects might be realized in the production of speech. Rather they (and many others) assumed that the results of their intelligibility studies gave implicit support to the notion that the articulatory realization of speech must somehow be affected by context.

In the present experiment we isolated one acoustic parameter which we studied for possible effects of context. Whilst we have observed that there is no statistically significant difference between the VOT's of stops uttered in appropriate and inappropriate contexts we can not conclude from this experiment that the *voicing contrast* is not affected by context.

Although we have demonstrated that *VOT* is not affected by context, it would be misguided to generalize from our results and conclude that no aspect of the speech signal has been affected by context. We have not attempted to determine the effects of context on the voicing distinction *per se,* (see Chapter 1, Section 4), nor have we looked further afield to examine the possible effects of context on the vast remainder of acoustic-phonetic parameters which comprise the speech signal. Although VOT is not affected by context, the real question of interest in this thesis is whether *perceptual* distinctions are affected. For this reason, we now study the intelligibility of the syllables which contain the VOT values in question.

8. EXPERIMENT 2.

In Experiment 1 we tested whether context is able to influence aspects of the articulatory realization of an utterance. More precisely, we asked whether the VOT's of voiced and voiceless stops are realized as a function of the contextual constraints. In line with the Lieberman 1963 and Hunnicutt 1985 studies, Experiment 2 now investigates whether the appropriateness of a context affects the intelligibility of an utterance or word. There are also other considerations which motivated this experiment.

We noted earlier that although VOT is not affected by context we should not rule out the possibility that other acoustic-phonetic parameters are varying as a function of the context. If we can demonstrate that the syllables containing the VOT's in question have no impact on the listener, i.e. they are not differentially perceived according to the origin of their context, then we will safely be able to rule out the possibility, in later experiments, that any observed effects of context are due to variations in physical aspects of the stimuli. Any effects of context which we may observe in later experiments, may, thus, be attributed to the influence of the preceding context on the *recognition* of the stimuli.

For the present Experiment, we selected voiceless target words from the recordings made for Experiment 1. We chose to use *voiceless* targets because they produced a greater range of VOT values, thus enabling us to measure the intelligibility of the most extreme values produced in appropriate and in inappropriate contexts.

We examine the possibility that context may affect the intelligibility of *word initial segments*. Previous studies have measured the intelligibility of word tokens by presenting whole words in isolation. If we are to ascertain, first, whether the voicing distinction is affected by context, and, second, whether the *initial 150 msec* of a word is processed bottom-up and in no other way, we need to present varying increments of the acoustic word tokens in order to determine *what* the listener's percept of the initial segment is and *when* this percept develops. The specific issue we address is whether the amount of sensory information needed to identify the target initial stops varies as a function of the context in which the target word was originally uttered.

To this end we use the experimental paradigm which was developed specifically for such purposes - the gating paradigm (Grosjean 1980). Also, in contrast to the previous intelligibility studies we do not present targets in isolation. We discussed in Chapter 2, Section 3.3 the possibility that a word in context, as opposed to a word in isolation, is a more appropriate stimulus with which to gain access to the mental lexicon (following Tyler 1984). Furthermore, we are ultimately interested in words *in context* rather than words in isolation. For these reasons we measure the intelligibility of words presented in *neutral* contexts in the present experiment.

Presentation of the targets in neutral contexts was achieved by two means. First, tokens were spliced from appropriate and inappropriate contexts into neutral contexts; and second, word tokens which had originally been uttered in neutral contexts were altered so that they approximated the VOT values for corresponding words which had been produced in both the appropriate and inappropriate contexts in Experiment 1. This comparison was made in order to determine whether the affect of replacing one word token by another from a different environment is disruptive in any way. As a further control we also contrast forced choice responding with free choice responding. The following section considers our reasons for incorporating these controls.

9. CONTROLS.

9.1. Waveform editing; two approaches.

Many studies employ the technique of splicing words into contexts other than the one in which they were originally uttered. In our case, this involves extracting the target word from the stretch of speech in which it was originally uttered and inserting it into a different (neutral) context. Such editing procedures allow the same word token to be studied in different environments. This is beneficial in the present circumstances as we can avoid the "citation"-like speech which might be produced if we only studied the target words which were spoken at the end of the unnatural carrier phrases (i.e. neutral contexts). [5]

[5] Note however, that many studies employ splicing precisely because it allows citation forms to be spliced into other environments. See for example, Tyler and Wessels 1983, Tyler 1984. In the experiments reported here, we attempt to use speech that is as "conversation"-like as possible although we acknowledge that we are still far from this ideal situation.

One of the problems with the splicing technique however, is that it may often be quite disruptive. If the prosody of the utterance from which a target word is spliced differs considerably from that which it is spliced into, the contour of the resulting sentence will be disrupted and this may, in turn, disrupt the intelligibility of the sentence as a whole.

An alternative to splicing is to alter some parameter in an original intact word (here, one which was uttered following the neutral context) to resemble the value of the word token of interest. This we shall term "manual-variation". For example, in Experiment 1 we observed that the average VOT for voiced stops uttered in appropriate contexts is 14.13 msec. We therefore manually vary the VOT value for the corresponding stop uttered in a neutral context either by lengthening or shortening the original value until it reaches 14.3 msec. Employing this technique means that other cues in the signal can ostensibly continue to affect perception. [6]

Whilst the manual-variation procedure leaves us with stimuli more closely approximating natural speech than, for instance, synthetically created stimuli, we should bear in mind the criticisms of Lisker 1978. He argued that in simulating VOT values in this way, we are creating a signal quite different from any signal which could have originated in the human vocal tract. This objection of course underlines the "control versus naturalness" dichotomy which is so prevalent in experimental psychology. By incorporating both splicing and manual-variation manipulations into the design of this second experiment however, we empirically evaluate how much naturalness might be lost at the expense of control.

9.2. Response Paradigms.

In virtue of the ease with which experimental data gained from forced-choice response paradigms can be collected and analyzed, it would seem beneficial if we were able to employ forced response choice paradigms in this, and later experiments. With forced choice paradigms, the subjects must select one item from a given, limited set of response possibilities. For our purposes, subjects would select from a set of printed words the word which they believe has just been aurally presented.

[6] This is a somewhat idealistic. We should be aware that stimuli which were manually varied also incorporate a degree of "cutback" of the first formant (see Liberman, DeLattre and Cooper 1958).

Objections to this paradigm are immediate. The prior knowledge of targets may bias subjects responses. Similarly, the restricted choice of responses may not conform to the subjects' percept. In such cases the forced choice paradigm is too constraining and may bury useful data that cannot be tapped because the subject is unable to report her actual percept.

In the present intelligibility study, we therefore compare results yielded from both the forced choice and free choice paradigms.

9.3. Summary of motivation for Experiment 2.

To summarize, the present gating experiment examines whether the intelligibility of voiceless word targets is affected by the contextual constraints which were available when the word was originally uttered. We ask whether stops are perceived differently depending on whether they were uttered in an appropriate or inappropriate context, and, in particular, whether the amount of sensory information needed to identify the stop varies as a function of the context in which it was uttered. We splice word targets from their original appropriate and inappropriate contexts into neutral contexts. To ensure that this procedure does not have any disruptive effect we also synthesize targets by manually varying the VOT values of targets *in situ* i.e. we alter intact targets which were uttered in neutral contexts so that they assume the VOT values of the appropriate and inappropriate targets. Finally, as a further control measure we compare the results for forced choice and free choice response paradigms.

10. METHOD.

10.1. Subjects.

Twenty-four subjects participated in this experiment. All were postgraduate students at the University of Edinburgh who had been exposed to the Edinburgh accent for a minimum of 2 years. They were all native speakers of British English with normal hearing and were naive as to the purpose of the experiment. Subjects received payment for their participation.

10.2. **Materials.**

All materials described below were created using the equipment referred to in the Method section of Experiment 1. Nine of the initial-voiceless target words which were obtained from Experiment 1 constituted the base materials for the present experiment. Three began with bilabial stops, three with alveolar and three with velar stops. Selection of these particular targets was based on the observed VOT ranges and the observed differences between these ranges across the appropriate and inappropriate contexts in the first experiment. The nine targets selected exhibited the largest range of VOT values (as produced by one speaker) both across the three repetitions within one context and between the appropriate and inappropriate contexts.

The nine target words were recorded, by the three speakers used for Experiment 1, in sentence-final position of neutral sentence contexts. These sentences were neutral with respect to what target word could appear in sentence final position. In theory, *any* target word could appear following these neutral contexts and the resulting sentence would not be anomalous.

These neutral contexts were loosely matched for length to the sets of appropriate and inappropriate sentences used in Experiment 1, and are presented in Example (6a) below.

(6a) The splodge on the typescript completely obscured the word ______.

(6b) John said he'd learned the Russian for ______.

(6c) The word at the end of this sentence is ______.

The three speakers used for Experiment 1 thus made recordings of the nine selected targets in sentence final position of the neutral contexts. Each speaker recorded a different three targets and targets appeared following only one of the three neutral carrier phrases. As with Experiment 1, sentences were recorded three times although ultimately only the repetitions with the highest and lowest extreme values were used in the presentation of materials.

For the splicing condition, copies were made of the neutral sentences and the target words were replaced by tokens from the appropriate and inappropriate contexts of Experiment 1. For the manual-variation condition, the VOT of the (neutral context) target words was systematically varied until it approximated the VOT values for the corresponding targets which had been uttered in the appropriate and inappropriate contexts of Experiment 1. To obtain these VOTs, a portion of the signal between the release of the stop burst and the onset of voicing was extracted or repeated by layering in one spliced section of this portion of the waveform. [7]

In order to sample adequately the VOT ranges observed in Experiment 1, the highest and the lowest VOT values for targets spoken in both the appropriate and the inappropriate contexts were used. That is, for both contexts, the repetitions of the targets which had yielded the longest and the shortest VOT's were spliced into the neutral contexts or were imitated in the neutral contexts by varying the VOT of the existing target which had been uttered in the neutral context. The VOT values of the neutral (intact) targets and the corresponding values which they were replaced by, or which they were made to approximate (i.e. the values of the appropriate and inappropriate context targets) are presented in Table 3.3 below.

[7] The length of the segment which was layered in varied according to requirements. These segments consisted of aspiration.

TABLE 3.3

Voice Onset Times (msec) to which Neutral targets were altered, or by which Neutral targets were replaced.

Word	Neutral	Appr Low	Appr High	Inapp Low	Inapp High
Cage	**108**	64	88	58	77
Cap	**100**	74	90	52	73
Class	**88**	66	83	72	114
Peas	**96**	57	85	75	101
Pound	**93**	77	109	45	78
Pull	**73**	58	77	58	89
Toe	**75**	57	99	50	68
Toll	**95**	68	91	60	77
Tusk	**74**	47	64	48	76
Above	*	0	3	0	4
Below	*	9	6	9	5

For each carrier phrase, 12 fillers were included; these were words which began with fricatives, liquids, vowels and nasals and all rhymed with one of the target words.

Target words (and fillers) were all gated in 50 msec increments. The onset of the first gate coincided with the onset of the closure prior to the release of the stop burst in the case of target words. For the fillers the onset coincided with the onset of frication. The gated

sentence final target words, together with the neutral carrier phrases, were recorded on to video tape. The first presentation of an utterance presented prior context only. The second and subsequent presentations consisted of the context plus the gated target increments. All presentations of any one utterance were presented consecutively.
All materials used in Experiment 2 are given in Appendix C.

10.3. Design.

Nine voiceless targets, balanced for place of articulation and speaker, were selected for the materials of this experiment. The Highest and the Lowest extremes of VOT values were included (2) from appropriate and inappropriate contexts (2) and were manipulated either manually or were spliced or were intact (3). This gave rise to 108 experimental sentences.

Sentences were blocked by neutral carrier phrase. This yielded three blocks which were subsequently arranged into a Latin Square for order of presentation to subjects. Subjects were yoked so that they received exactly the same stimulus tapes but one responded by way of forced choice and the other by free choice.

There were thus eight experimental versions for each target. In order to provide a control and balance conditions, 4 instances of the intact, neutral token were also included to yield 12 versions per token. Cells illustrating the design of Experiment 2 are given in Appendix C.

10.4. Procedure.

Subjects were seated in separate booths where they were unable to see either each other or the experimenter. They were given detailed instruction sheets to read which it was ensured they fully understood before the experiment began. Corresponding to each block of carrier phrase, Subjects had 48 response sheets (one per sentence rather than one per gated presentation). Free choice Subjects were instructed to write down whatever they thought they had heard after each gated presentation whereas Forced choice Subjects were asked to circle the word they thought they had heard after each presentation. The Forced choice was between the target, its word-initial stop voiced counterpart and a rhyming word which began with a fricative, or when this was not possible, a liquid. [8]

Subjects heard the stimulus tape over headphones. The stimuli were presented in 2 sessions, the first of which contained both the first and second block (with a short interval separating the two) and the second of which contained the final block of trials. The two sessions were administered on different days. The total time taken to complete all 3 blocks was approximately two hours and 10 minutes.

10.5. Data Analysis.

Scores were computed for each target and for each subject by calculating the mean number of gates taken before subjects began trying to identify the target (whether correctly or incorrectly). A calculation was then made to obtain a comparison of the experimental sentences (the sentences which had either spliced targets or altered VOT values) with the control sentences (the neutral sentence contexts with their intact target words). Mean scores for responses to the neutral intact contexts were subtracted from the scores to the contextual conditions. Each score thus represented a comparison of the experimental manipulations with the control condition.

Scores were then sorted and tabulated according to whether the response type represented a correct identification of the target, a voiced "mis-identification" or a fricated "mis-identification" Correct identifications constituted over 95% of the responses and an Analysis of Variance was computed for these data.

11. RESULTS.

There was no main effect of context $F1(1,22) = 1.32$, $p = 0.26$, and neither was there a main effect of response type (forced or free choice) $F1(1,22) = 2.38$, $p = 0.14$.

The mean number of gates taken to identify targets with the high VOT values was significantly higher than for targets with the low VOT values $F1(1,22) = 38.05$, $p < .0001$, [$F2(1,8) = .04, p=0.8$]; and also the mean number of gates taken to identify targets which had been spliced was significantly higher than the number of gates taken to recognize

[8] Fricatives were incorporated as a measure of the degree to which closure, rather than VOT might have been affected

targets which had been altered by varying the VOT $F1(1,22) = 34.57$, $p < .0001$, [$F2(1,8) = 1.63, p= 0.2$]. It was found that these two variables yielded a highly significant interaction $F1(1,22) = 37.01$, $p < .0001$, [$F2(1,8) = .47, p=0.5$]. The means for this interaction of manipulation and the extremity of VOT values are presented below in Table 3.4 and illustrated in Graph 3.4

TABLE 3.4

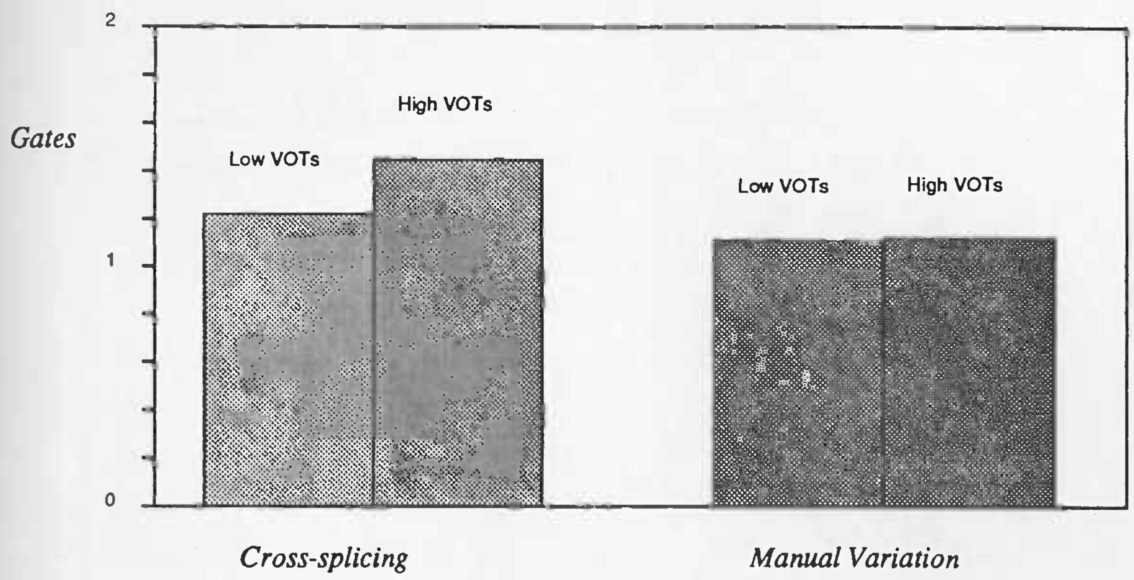
Mean number of gates taken to identify target initial segments.
(Means are transformed for comparison with the control, see section 10.5)

*	Cross-splicing	Manual-variation	Means
Low VOTs	1.22	1.11	1.17
High VOTs	1.45	1.12	1.29
Mean	1.34	1.18	*

by the contextual manipulations in Experiment 1. My thanks to Linda Shockey for discussion of this issue.

GRAPH 3.4

Experiment 2: Mean number of gates taken to identify target initial phonemes.



The Tukey HSD test was then computed (HSD = 0.09, $p = 0.05$) which revealed that when the targets are spliced those with lower VOT's are correctly identified at earlier gates than are the targets with the higher VOT values, and that the splicing manipulation yields significantly later identifications than does the manual-variation condition. The Tukey HSD test also revealed that the High VOT values under the splicing conditions required the highest number of gates for correct identification of the targets.

Tabulations were then made by the phonetic category of subjects responses. This analysis illustrated how often Subjects identified the target initial phoneme as anything other than a voiceless stop. We used the same method of comparison with the control within cells as described above. Only 10 of the possible 1592 responses were fricatives, and only 44 voiced. There were fewer than 10 liquid, vowel, nasal and other percepts. As these responses constituted such a small percentage of the data, further analysis was not considered to be worthwhile.

12. DISCUSSION.

The lack of a significant effect for response choice indicates that in both forced and free choice listeners are able to identify target initial segments on the basis of similar portions of the acoustic stimulus. It does not appear, therefore, that the prior knowledge of the target provided to the forced-choice subjects, affects the earliness of their response.

The observed difference between the mean number of gates at which high and low VOT values were identified was unsurprising. Incoming information about the following vowel was present earlier in the signal in the cases of short VOT values, and could be used to confirm the recognition of the stop at earlier gates than would be expected for longer VOT values. For the higher VOT values subjects waited proportionally longer until the transitions for the vowel could be used as confirmation in this manner. We should note that this difference between the two extreme values necessitates adequate sampling along this dimension in subsequent experiments.

The main effect of Manipulation shows that splicing yields relatively slower identifications of the target initial segments than manual-variation. As we suggested in the introduction, this difference may be due to the disrupting effect of the splicing paradigm. In the manual-variation conditions the VOT of the original signal is, theoretically, altered along only one dimension (time). Consequently the original carrier phrase along with the target remains ostensibly in tact. The signal may be smoother in the sense that no alterations (save a slight "stretching" of the aspiration phase after the stop) will be incurred. For targets which have been spliced however, there may be sudden inconsistencies between the carrier phrase and the target in the average amplitude, stress and pitch. In setting up the stimuli however, such contrasts were minimized. We ensured, for example, that target words which were spliced in to and out of contexts, were spliced at zero crossings and although the targets which had been spliced were identified later than the targets which had been subject to the manual-variation condition, note that they were nonetheless correctly identified within the first two gates. It does not appear that the splicing manipulation is ***disrupting*** the subjects' percept. Rather, splicing simply delays it. Also, there were no more voiced, fricated or other responses to targets which had been spliced than to targets which had been altered by varying the VOT. This too, illustrates that splicing is not behaving in a disruptive manner.

Post-hoc inspection of the lengths of VOT across all stimuli reveal that 29 (of the 36) VOT values which were to be approximated, were lower than the corresponding VOT values of the Neutral target. Table 3.3 (Section 10.2 above) illustrates this. The frequency with which the VOT values of Neutral targets were either increased or decreased is given at the bottom of this table. The majority of the alterations to targets by means of manual-variation thus required that the original stimulus be cut by taking out a certain portion of the aspiration (or silence). Consequently the transitions to the vowel which occur before the onset of voicing will be artificially premature giving additional cues which were not present in the spliced stimulus item. Thus manually varied targets will be identified earlier than their spliced counterparts. A corollary of this is that for stimuli which have been manually altered the High VOT values will require no more of the acoustic stimulus for identifications than the Low VOT values.

The observed interaction between the Extreme VOT values and Manipulation further supports this argument. The means presented above in Table 3.4, illustrate that for manual-variation, there is not much difference between the number of gates required to make a response whether the VOT value is High or Low. However, when splicing was employed, a difference between the High and Low values obtained. For spliced targets the High VOT values yield a higher mean number of gates than the low VOT values.

For the remaining 7 (of the 36) cases where the neutral intact target is manually varied to assume a higher value, all the information necessary to reach a decision as to what the initial phoneme is, has already been presented at the time when the "extension" to the signal begins. Thus whether the VOT value is high or low has little impact on the number of gates subjects take to identify targets. For spliced targets however, the length of the VOT is directly proportional to the time taken to identify the target. Thus low VOT's are identified significantly earlier than the high VOT's.

There was no evidence in support of the hypothesis that the context (either from which a target word is spliced, or to which the target word is assimilated) affects the mean number of gates taken to identify targets. The time taken to recognize a target initial segment which is presented in a neutral context, remains reasonably constant, whether the target be of appropriate or inappropriate extraction.

This result is important. The original context of utterance (appropriate versus inappropriate) had no significant effect on the number of gates taken to identify targets, and thus, the intelligibility of target words seems to be unaffected by the degree of constraint provided by their original context. At face value this result appears to contradict the conclusions reached in previous studies. There are, however, several reasons why present results differ from these studies.

Firstly, the targets utilized in the present experiments were more rigorously controlled than, for example, those employed in the Lieberman 1963 and Hunnicutt 1985 studies. For instance, we controlled for place of articulation, voicing features and the initial segments of targets whereas Lieberman controls for neither and Hunnicutt controls only for initial segments and manner. Moreover, the inverse relationship between predictability (or "redundancy") and intelligibility, as noted, for example, in the Lieberman 1963 and Bard and Anderson 1983 studies, is not a *robust* effect.

For instance, Lieberman noted that his intelligibility-redundancy relationship held only when subjects were presented with the redundant targets first. When the order of presentation was reversed so that subjects were exposed to the non-redundant ("clear") targets initially, the subsequent identifications of the redundant targets improved to over 90% correct.

Anderson & Bard found that this inverse relationship held when the target words were randomly sampled but it held to a much lesser extent for matched pairs.

Secondly, Lieberman's and Hunnicutt's conclusions are based on the intelligibility of targets which were presented at a signal to noise ratio of +4 DB. When the noise factor was withdrawn, intelligibility of targets rose to almost 100%. Also, the dependent variable used in the present study differs from that used in the previous studies. We examined the amount of sensory information required to identify targets, whereas the previous studies have measured the number of letter-perfect or fully homophonous identifications of the target word.

Finally, because a target presented in context is a more appropriate stimulus for accessing the mental lexicon, [9] our targets were presented in sentence-final position of neutral contexts as opposed to presenting the targets in isolation. In our experiments therefore, subjects receive a stretch of speech which enables them to normalize to the speaker's voice.

The null effect of context in the present experiment demonstrates that in a neutral sentential context word targets emanating from appropriate contexts are not perceived differently (in terms of requiring more or less of the sensory information) to those from inappropriate contexts. Thus the contextual origin of target words need have no bearing on future experiments. More importantly however, this null result may allow us later to conclude that any dissimilarity in the percept of these targets, when they are presented following varyingly constraining contexts, is due to the influence of the prior context and not the target word *per se*.

It is interesting to note that across all conditions, correct identification of target-initial phonemes was possible well within the first three gates. Note that this corresponds to the first 150 msec of the target words. If context does affect the realization of the first 150 msec of a word, the results of the experiments reported in this Chapter indicate that it does not do so to the extent that it is perceptible.

13. CONCLUSION.

The present chapter explored whether context might affect aspects of the articulatory realization of an utterance. We noted that if context controls the production of speech in a way that makes initial segments of words indistinguishable on purely acoustic grounds, then the motivation for obligatory bottom-up processing during lexical access is weakened, because processing ambiguous acoustic-phonetic information will not be sufficiently constraining to initially activate one set of word candidates more than other candidates in the lexicon. The results presented above, however, demonstrate that VOT, and other possible parameters pertaining to the voicing distinction, are *not* controlled according to context, at least not in a manner which is acoustically measurable or perceptible. It would thus appear that bottom-up processing of initial segments of words would serve to contsrain

[9] This is further supported by the studies which demonstrate that words are less intelligible when presented in isolation. See Locke 1968 and Pollack and Pickett 1963.

the initial activation of some word candidates over others.

The results of the two experiments presented in this Chapter are consistent with the notion that talkers do not control for the amount of information they wish to give the listener. Although we noted in Experiment 1, that talkers may make *slight* adjustments to make words more intelligible according to the perceived difficulty (notably in their production of the phoneme [p]), it is clear from Experiment 2 that such adjustments either do not serve their purpose, or are not sufficiently controlled that any differences are percpetible.

The data presented in this chapter force us to conclude that VOT and other possible factors pertaining to the voicing distinction do *not* constitute low level phenomena which are influenced by the higher level, contextual information. Given that the acoustic parameters we studied here coincided with the first 150 msec of word initial segments, we can conclude that the acoustic-phonetic information available to be processed by a bottom-up oriented models of word recognition, may not be as "insufficient" at providing contsraints as we anticipated. Context does not appear to affect the articulatory realization of voiced and voiceless stops to the extent that they may become indistinguishable on acoustic grounds. At least, we can safely say that context does not appear to affect the realization of voiced or voiceless segments which receive sentence primary stress and appear in sentence final position of contexts in which they are highly predictable or highly unpredictable. It remains to be seen however, whether the presence of a prior constraining context interacts with bottom-up processing during the *perception* of the first 150 msec of a word. Before we investigate this issue, however, we examine a further possible effect of context on the articulatory realization of an utterance, and measure listeners' ability to make use of prosodic information in discriminating between utterances which were constraining or non-constraining.

CHAPTER 4:

THE EFFECTS OF HIGHER LEVEL INFORMATION ON THE MANNER OF DELIVERY OF AN UTTERANCE.

1. INTRODUCTION.

The previous Chapter presented control data which suggested that contextual information did not affect the production of Voice Onset Time. If we had not already established that the realization of VOT in our target stimuli was not affected by context, then possible effects of context in *perception* [1] may have been due to differences in the targets themselves. Having established that context did *not* affect the realization of these target stimuli we might now attribute effects in perception to the use of contextual information during processing. In this Chapter however, we consider why we are not yet in a position to make such an assumption.

In Chapter 2 (Section 1) we noted that prosody, coarticulation, pitch contour and timing also constituted an important aspect of bottom-up information. We shall term these various types of information "delivery information". If listeners are able to use delivery information to distinguish whether or not a target is predictable then, in our later experiments, manipulations of semantic context may be confounded with the inadvertent manipulation of delivery information thus making difficult the interpretation of future experimental results. In the present Chapter therefore, we examine to what extent manner of delivery of an utterance, over and above the lexical/semantic context, can cue whether the

[1] In contrast to Chapter 3 where we ran a perception experiment to determine whether the *production* of VOT in different contexts was perceptible, here, and subsequently, "perception" denotes the actual process of recognition.

sentence final target word is predictable or unpredictable.

1.1. Evidence relating to the importance of delivery information.

Many researchers currently maintain that delivery information is of paramount importance during speech recognition. Darwin 1975 for example, claims that

> "Prosodic processing can be so efficient that listeners will attend to prosodic continuity at the expense of semantic continuity."
>
> (Darwin 1975,p.)

Indeed, many researchers have postulated models of lexical access which rely crucially on processing delivery information. Cutler and Norris 1988, for example, along with Grosjean and Gee 1987 argue that word recognition proceeds by stress based segmentation of the acoustic-phonetic input. Although discussion of such models is outwith the scope of this thesis, we shall, for completeness, consider some of the empirical evidence which has demonstrated the importance of delivery information in speech processing. Although we do not attempt to review the trade off between delivery and semantic continuity in this section, the evidence which we consider below is clearly consistent with the notion that delivery information has an important role in speech perception.

Cutler 1976 spliced two identical target words into two different sentences. The prosodic contour of the first sentence predicted that the target word would occur in accented position whereas the contour of the second sentence predicted that the target would appear in unaccented position. For example, the target [BOOK] appeared in the sentences:

(1a) The couple had quarrelled over a *book* they had read.

(1b) The couple had quarrelled over a book they hadn't *even* read.

In (1a) the target receives primary stress. In (1b) however, the word [EVEN] receives the primary stress. Cutler found that listeners were able to identify the target word faster when it occurred in accented position - even in conditions where the target which had been spliced in was not actually accented. Cutler concluded from this that listeners were making use of the prosodic contour. Similarly, Cutler and Foss 1977 reported that phoneme monitoring was facilitated when the target phoneme belonged to a word which bore sentence accent, including instances where the target was a function word.

The data presented here clearly illustrate that manner of delivery is a very important source of information and is available to aid processing. If listeners are not only able to distinguish between different prosodic contours but also able to actively use the distinctions they make then this has important ramifications for research on effects of context on lexical access. Fowler and Housum 1987 noted, also, that alterations in the delivery of some words was correlated to predictability. Whilst the other studies we have considered here have not dealt explicitly with the correlation of delivery and predictability, in light of the importance of manner of delivery generally we must ask to what extent delivery information might cue predictability.

1.2. Introduction.

The experiment described below involved listeners discriminating between two versions of the "same" semantic context. The contexts were the same by virtue of containing matched words but differed in that one version had been originally uttered as an appropriate precursor to a sentence final target word and the other had been uttered as an inappropriate context to a different target word, for example:

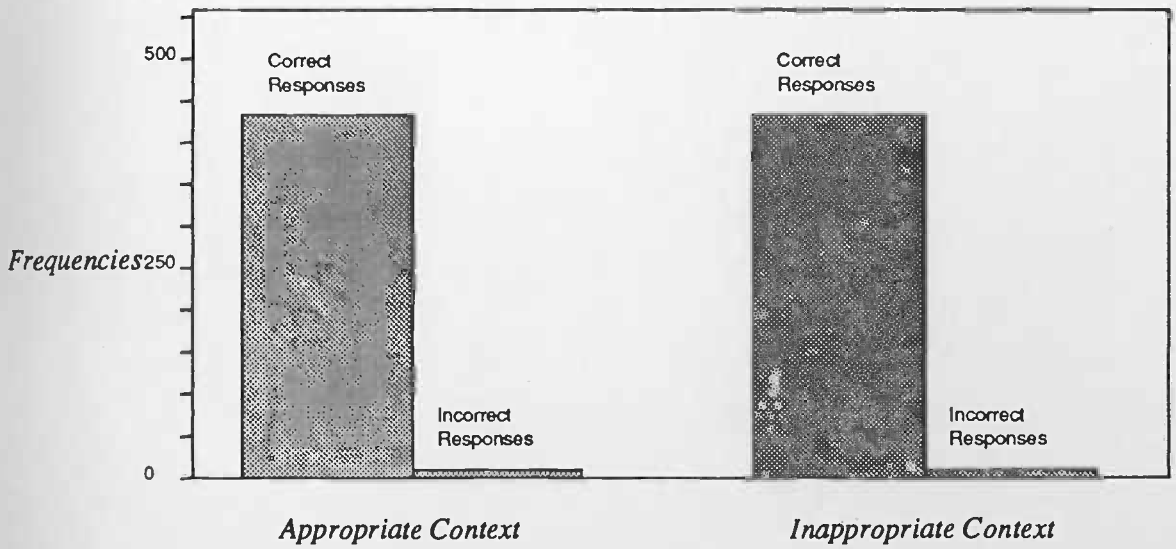
(2a) John hates hornets, wasps and ..... *(bees)*. [Appropriate context]

(2b) John hates hornets, wasps and ..... *(peas)*. [Inappropriate context]

To the extent that delivery information and co-articulatory information may be used by listeners to judge the predictability of the targets, listeners should be able to discriminate between the two versions of the context (the target words being omitted) and thus correctly identify which target word had been deleted (whether appropriate or inappropriate). If delivery information is available in this way we would expect that subjects will always be able to anticipate missing targets irrespective of whether the target is appropriate or inappropriate for the context they have heard. Note that it is outwith the scope of this thesis to be able to separate out possible effects of co-articulation from effects of delivery. What we might ultimately interpret as an effect of delivery may be due to, for instance, a tendency towards devoicing of segments preceding targets which had voiceless initial stops. For present purposes however, it will be sufficient for us to ascertain whether any unspecified acoustic information is given priority over and above the lexical/semantic context. The hypotheses outlined above are illustrated in Figure 4.1 below, we have termed these hypotheses respectively as "Prosodic Cuing" (i.e. delivery information alone is used by listeners), "Semantic Cuing" (only semantic information is attended to) and "Equal Cuing" (both delivery and semantic information are utilized equally).

FIGURE 4.1

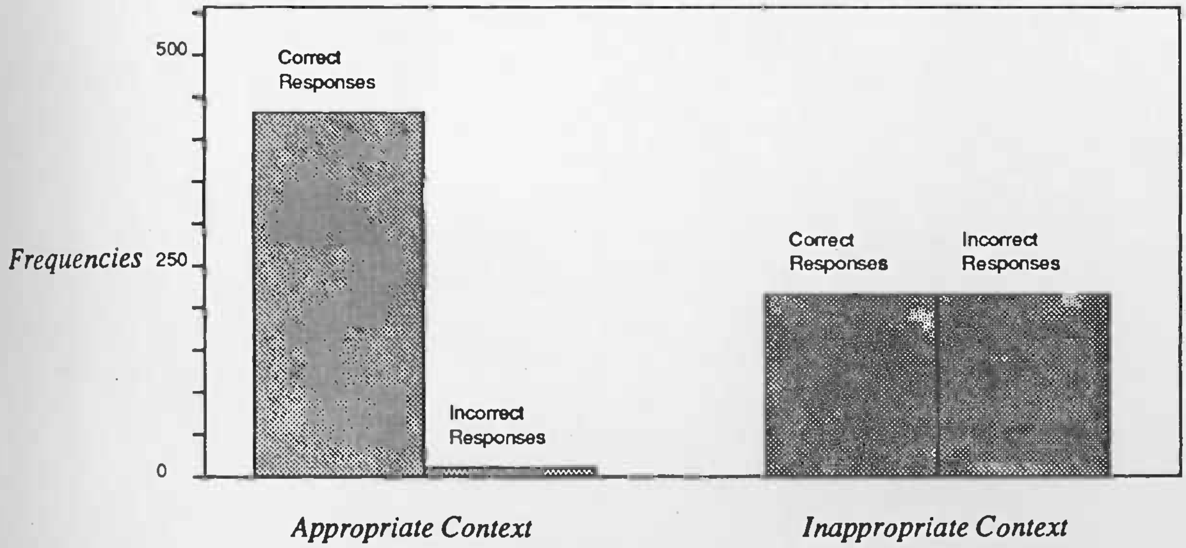
PROSODIC CUING HYPOTHESIS:
Frequencies of Correct and Incorrect Responses.



On the other hand, if both delivery and lexical/semantic cues were equally useful we predict that all *appropriate* targets would be correctly anticipated. Versions originally containing *inappropriate* final words however (see (2b) above), would yield some responses which were congruent with the lexical/semantic information and, therefore, incorrect. To the extent that the delivery and semantic information was equally useful to the listener, we predict an equal distribution of correct and incorrect responses to the inappropriate versions due to the conflict of information sources. We thus expect equal proportions of delivery-predicted (i.e. phonetically accurate) and semantically predicted (contextually appropriate) responses to targets which are inappropriate conclusions to the preceding context. Hence whatever numbers of phonetically accurate and semantically appropiate responses we observed to targets which were appropriate conclusions to the preceding context, we anticipate *half* that number of responses to the inappropiate targets (see Figure 4.2 below). [2]

FIGURE 4.2

EQUAL CUING HYPOTHESIS:
(Prosodic and Lexical-Semantic Information of Equal Importance):
Frequencies of Correct and Incorrect Responses.



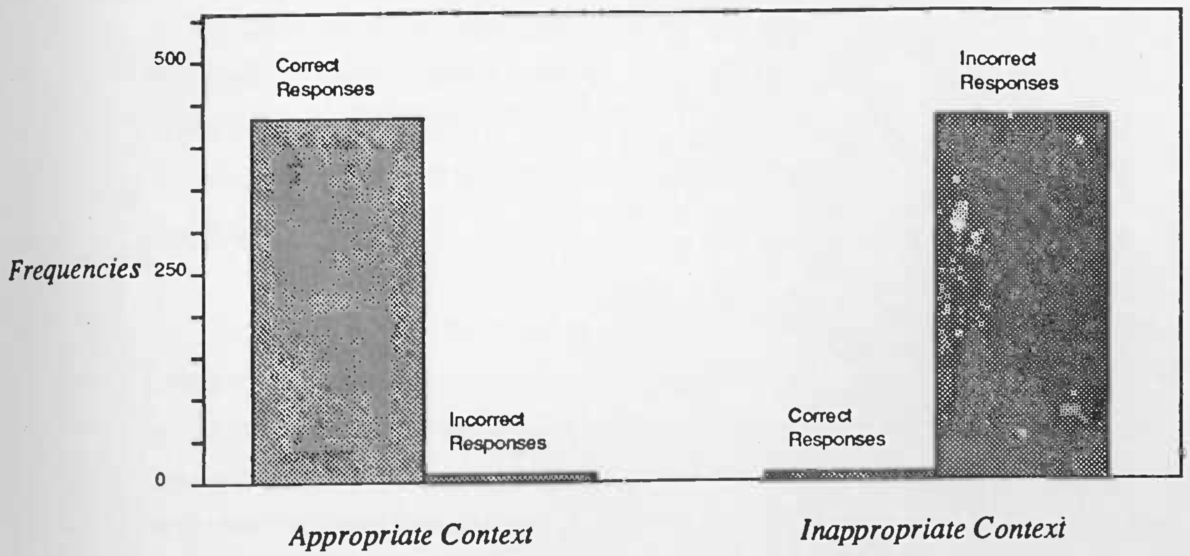
Finally, if the only cue to predictability available to the listener is contained within the lexical/semantic context then we anticipate that only the targets which constituted appropriate continuations of the sentences would be correctly anticipated. The non-availability of the delivery information in this case should be evidenced by the tendency towards semantically appropriate responses whether or not the target constituted an appropriate or inappropriate conclusion to the preceding lexical/semantic context. Figure 4.3 illustrates this hypothesis.

---

[2] Phonetically accurate and semantically appropriate responses are mutually exclusive when the target is contextually inappropriate.

FIGURE 4.3

SEMANTIC CUING HYPOTHESIS:
Frequencies of Correct and Incorrect Responses.



In order that we might obtain an effective estimate of the availability of delivery information to cue the predictability of sentence final target words, groups of trained phoneticians were used as subjects and we compared their discriminations to groups of linguistically naive listeners. Although it was not intended to use phoneticians as subjects in subsequent recognition experiments, they provided us here with a strong measure of the availability of delivery cues because if delivery information is helpful in cuing predictability, then the phoneticians should be able to make effective use of it.

2. METHOD.

2.1. Subjects.

Two groups of subjects participated. The first group consisted of six trained phoneticians whose special areas of research focussed on intonation and prosody. Six linguistically naive subjects comprised the second group of subjects. All subjects were native speakers of English and had been exposed to the Edinburgh accent for a minimum of two years. Subjects did not receive payments or course credits for participating.

2.2. Materials.

All stimulus sentences were taken from the digitized recordings made by 3 female Edinburgh speakers (see materials for Chapter 3). From the utterances of each speaker, 36 minimal pairs and their corresponding contexts were selected (i.e. 12 materials per speaker). As in previous experiments, minimal pairs differed only on the voicing of the initial stop and were balanced for place of articulation (bilabial, alveolar and velar). They appeared in sentence final position of matched appropriate and inappropriate contexts (e.g. (2a) and (2b) above) thus giving a total of 72 sentences. [3] Sentences were subsequently edited using the speech analysis software package developed at Edinburgh University's Centre for Speech Technology Research ("Audlab").

The final target words of all sentences were deleted by replacing all speech following the stop closure by silence. Sentences were sorted by speaker, phoneme and minimal pair and were stored and re-recorded accordingly. This ensured that listeners would be presented with all stimuli spoken by one speaker before hearing stimuli generated by the other speakers and thus maximized listeners' chances of being able to discriminate between appropriate and inappropriate versions of the same lexical string (see (2a) and (2b) above). All materials used in Experiment 3 are given in Appendix D.

2.3. Procedure.

Subjects participated individually and were seated at a computer terminal. Stimuli were presented over headphones. [4]

[3] For details of how the predictability of appropriate and inappropriate contexts was gained see the cloze test described in Chapter 3, Section 2.

[4] Stimuli presentation was controlled by means of a program structured around a "Playback" routine available on the Masscomp.

Subjects were told that they would hear sentences which lacked a final word and that the missing final word would appear as a menu item which would be displayed on the VDU. [5] They were asked to choose from the menu the word which they thought was most in keeping with the *manner of delivery* of the sentence they had just heard (as opposed to the lexical items constituting the sentence). Subjects were thus required to make forced choice responses on the basis of a two item menu: the correct target and its voiced or voiceless counterpart. [6] For instance, if the subject heard:

(2a) John hates hornets, wasps and ...

the menu illustrated in Figure 4.4 might be displayed:

FIGURE 4.4

1.	**BEES**
-	
2.	**PEAS**

Subjects were instructed to press the number corresponding to the menu item they had selected. Menu items were constructed such that the correct choice randomly appeared as the first or second item.

Subjects were thus made fully aware as to the purpose of the experiment and were instructed to try to base their decisions on whether the intonation of the sentences gave cues that the final word was "strange" in any way (as PEAS would be in (2b): "John hates

[5] My thanks to Nang Chan for writing the control and data collection programs for the present experiment.
[6] Forced choice responding was shown not to affect responses in Experiment 2.

hornets, wasps and ..."). All subjects heard all sentences.

2.4. Design.

Speaker (3) was crossed with voicing (2) and original context (2) in a between subjects design.

3. RESULTS.

A data collection program converted subjects' responses into 1's and 0's according to whether the response was correct or incorrect. A correct response was scored when the menu item which corresponded to the deleted sentence final target word was selected (regardless of whether the target was appropriate or inappropriate for the context which subjects heard). Frequency data was collated using the BMDP9D Multiway Description of Groups and was used subsequently in calculating Chi-squared and Binomial tests.

Chi-squared yielded no significant differences between the two groups of subjects and no significant differences in the responses of either group to the three different speakers.

In order to examine the distribution of correct responses to the appropriate and inappropriate targets the Chi-squared Test of Independence was calculated. This revealed that responses were contingent on whether or not the target was appropriate (Chi-squared = 70.6, DF = 1, p = < 0.001). Frequencies are presented below in Table 4.1 and illustrated in Graph 4.1 below.

TABLE 4.1

Frequencies of Correct and Incorrect Responses

Frequencies	Correct Responses	Incorrect Responses	Total
Appropriate	311	121	432
Inappropriate	189	243	432
Total	500	364	864

The Binomial test revealed that a significant proportion of the responses to inappropriate targets were incorrect ($z = -5.45$, $p < .00003$) and that the proportion of correct responses to appropriate targets was more than double the proportion of correct responses to inappropriate targets ($z = -19.18$, $p < .00003$). The Binomial test also forced rejection of the hypothesis that the proportions of semantically appropraite responses were equal across appropriate and inappropriate targets ($z = -2.85$, $p < .0002$).

A further Chi-squared Test of Independence was calculated over correct and incorrect responses to targets which had voiced or voiceless initial segments. It was found that responses were contingent on whether the initial stop of the target had been voiced or voiceless ((Chi-squared = 6.84, DF = 1, p = < 0.05). The frequencies presented below in Table 4.2 and illustrated in Graph 4.2 show that there were more correct responses to the voiceless targets even though the differences between cells did not reach significance.

GRAPH 4.1

Frequencies of Correct and Incorrect Responses

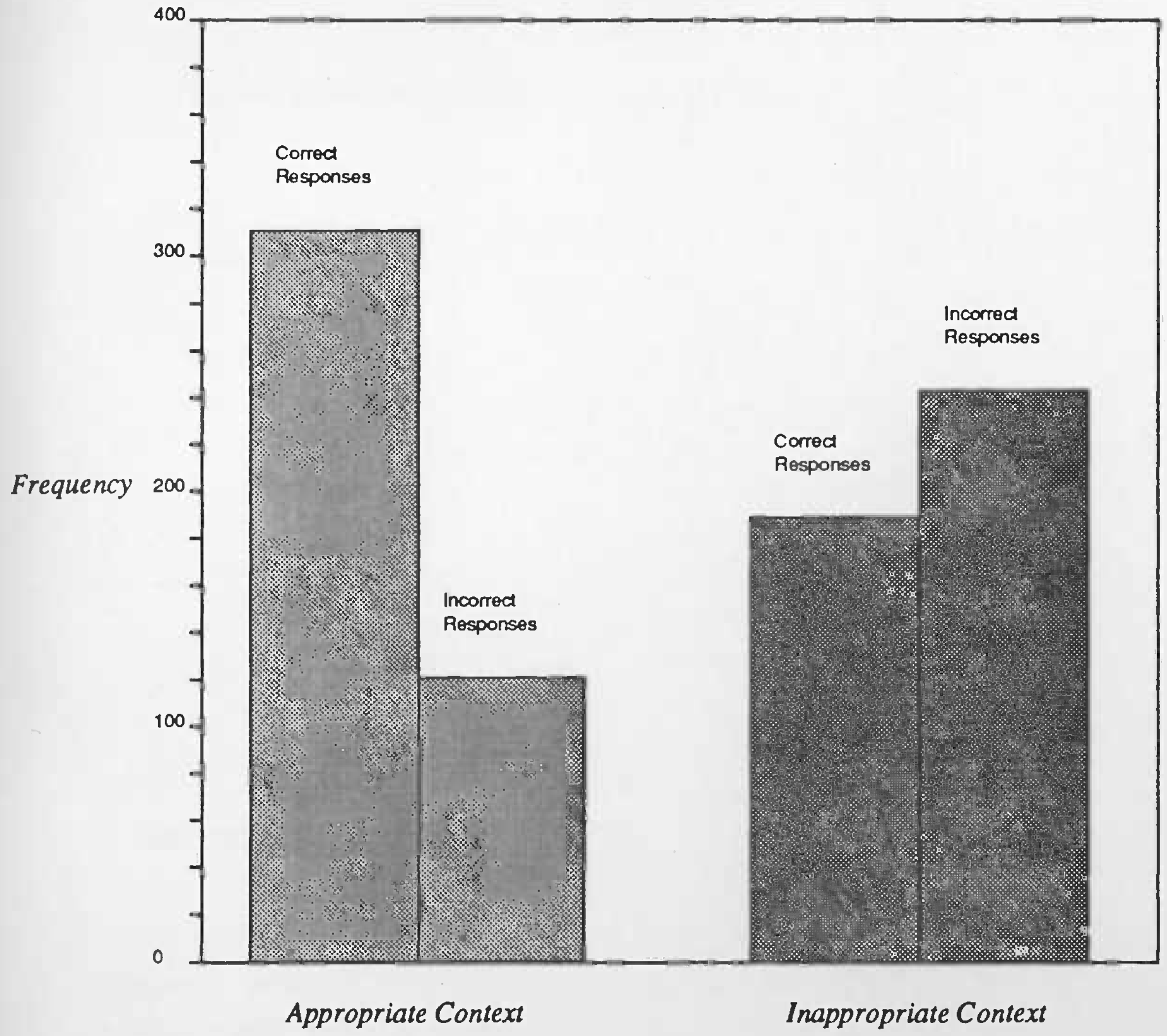


TABLE 4.2

Frequencies of Correct and Incorrect Responses; Voiced and Voiceless Targets.

Frequencies	Correct Responses	Incorrect Responses	Total
Voiced	231	201	432
Voiceless	269	163	432
Total	500	364	864

4. DISCUSSION.

In Section 1.2 above we described the different patterns of responses which, we suggested, reflected the differential availability of either delivery cues and/or lexical/semantic cues. For convenience we illustrate these hypothetical response patterns below in Tables 4.3, 4.4 and 4.5.

GRAPH 4.2

Frequencies of Correct and Incorrect Responses

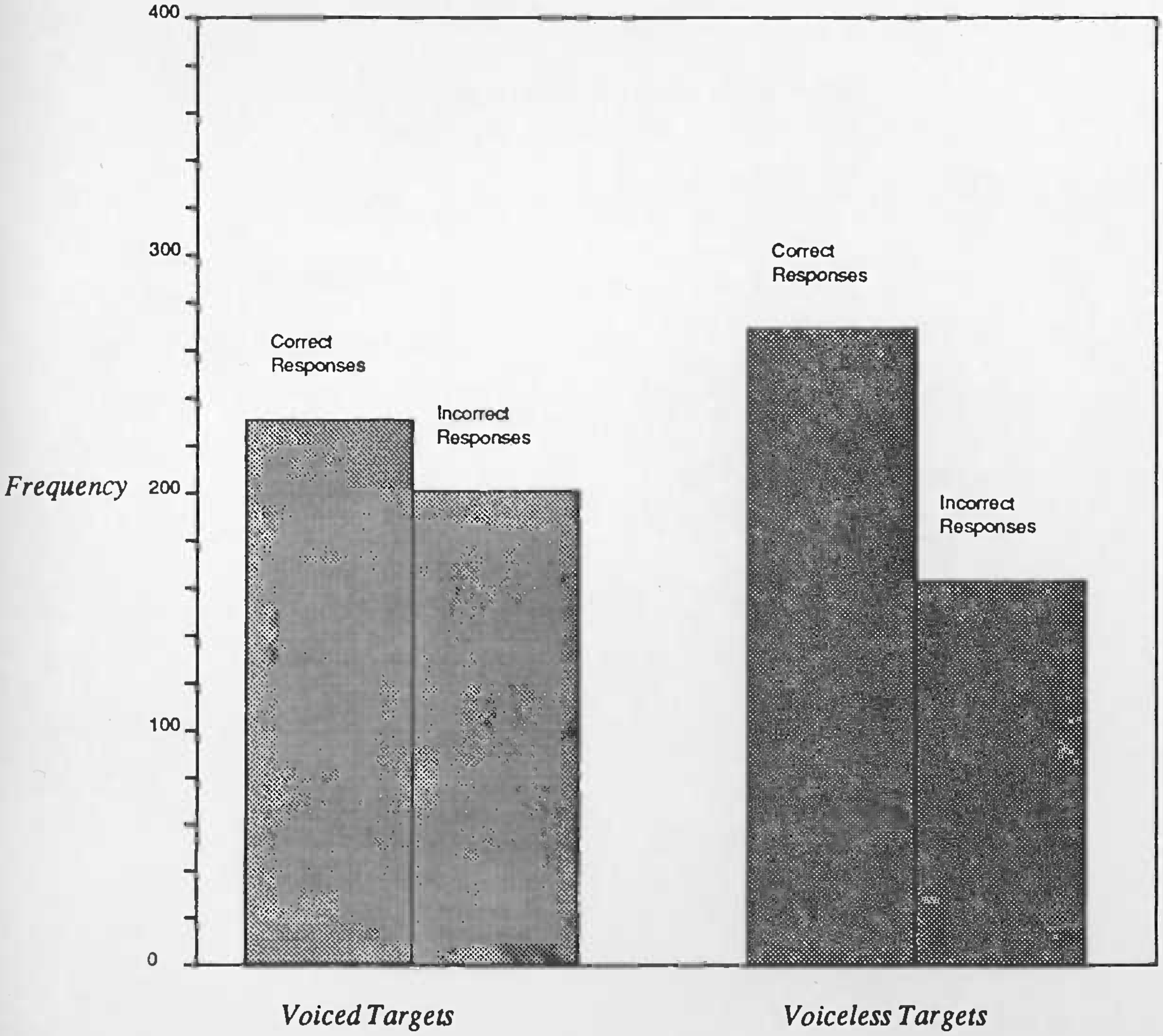


TABLE 4.3

PROSODIC CUING HYPOTHESIS:
Frequencies of Correct and Incorrect Responses.

Frequencies	Correct Responses	Incorrect Responses	Total
Appropriate	432	0	432
Inappropriate	432	0	432
Total%	864	0	864

TABLE 4.4

EQUAL CUING HYPOTHESIS
(Delivery and Lexical/Semantic Information of Equal Importance):
Frequencies of Correct and Incorrect Responses

Frequencies	Correct Responses	Incorrect Responses	Total
Appropriate	432	0	432
Inappropriate	216	216	432
Total	648	216	864

TABLE 4.5

SEMANTIC CUING HYPOTHESIS:
Frequencies of Correct and Incorrect Responses

Frequencies	Correct Responses	Incorrect Responses	Total
Appropriate	432	0	432
Inappropriate	0	432	432
Total	432	432	864

The observed data presented above in Table 4.1 were tested against the three possibilities presented above (and discussed in Section 1.2). In the following sections we consider the rationale for choosing between them.

4.1. Using only delivery information: Prosodic cuing hypothesis:

Although the observed data showed there were significantly more correct responses over incorrect responses as the Prosodic cuing hypothesis required, we rejected this possibility because the distribution of obtained responses to the inappropriate targets ran counter to that predicted if only delivery information were serving as a cue to sentence interpretation. In order to show a trend in the direction of delivery/prosodic-cuing it would have been necessary to demonstrate a greater than chance number of correct responses to the inappropriate targets. We found exactly the opposite however. There was a significantly larger proportion of *incorrect* responses to targets which were inappropriate continuations.

We should also consider the availability of *phonetic* information here, since sole reliance on phonetic cues to the predictability of the target word should yield the same distribution of responses as the availability of only delivery information.

Three of the subjects reported that some final segments of the presentation contexts were devoiced thus enabling them to deduce that the missing target word began with a voiceless stop. Indeed we noted above (Section 3) that there were more correct responses for the voiceless targets than the voiced. This result was not statistically upheld, however, and despite the possibility of phonetic information cuing targets, subjects still were significantly poorer at anticipating targets which were inappropriate continuations of the preceding context. To the extent that a phonetic advantage might exist, it is not able to significantly facilitate identification of targets which are inappropriate to the preceding context.

4.2. Equal semantic and prosodic cuing:

If lexical/semantic and delivery cues were equally available to the listener we would have expected an equal (chance) distribution of correct and incorrect responses to the inappropriate targets due to the conflict of two information sources. The inappropriate targets however, yielded significantly more incorrect responses. Furthermore, the proportion of correct responses to appropriate targets was more than double the propotion of correct responses to inappropriate targets. The available data do not, therefore, conform to the distribution pattern predicted by the equal availability of delivery and semantic cues.

4.3. Semantic cuing hypothesis:

Restricting ourselves to the limited possibilities which we presented above (see Tables 4.3, 4.4 and 4.5 and Section 1.2) the only remaining conclusion is that listeners were responding on the basis of the lexical/semantic cues available in the presentations. As we reported above however, lexical/semantic cues were not given priority to the extent that the delivery and co-articulatory information source were ignored altogether because the inappropriate targets did not yield as many semantically appropriate responses as the appropriate targets. Manner of delivery thus may have been minimally effective in cuing targets. If delivery did influence responses however, it was buried (at least 70% of the time) by the effects of the lexical/semantic cues - even for the trained listeners. The closest approximation to the observed distribution of responses, is the pattern presented in Table 4.5 (Graph 4.3) for the semantic cuing hypothesis. Although we rejected the possibility that delivery information is irrelevant and that targets are not cued *solely* by the lexical/semantic context, we noted that responses were contingent on whether the context was appropriate. Also, we noted that there were significantly more correct responses to targets which had been congruent with the context at the time of utterance (the "appropriate" targets) over the inappropriate targets, and a significantly greater proportion of *errors* for the inappropriate targets where lexical/semantic information was not congruent with targets at the time of utterance.

It thus appears that both semantic and delivery cues were available but not equally. Our results suggest that the lexical/semantic context provided by far the strongest cues to targets. It could be argued that although both sources of information are available to aid decisions, a weighting factor is involved. For example, semantic information might override delivery information for 70% of the time whilst delivery information "wins" for the remaining 30%. [7]

[7] This is an uncomfortable conclusion. Whatever ratios were obtained, it would be easy to argue that the ratios reflected weighting levels of some description. Because it is outwith the scope of this thesis to pursue this possibility empirically and because it does not impinge on subsequent interpretations of our data we shall not dwell on this issue.

5. CONCLUSION.

The aim of the experiment presented in this chapter was to examine the extent to which the delivery information might cue whether sentence final target words were predictable or unpredictable in a way which gave it *priority* over the lexical and semantic context. Given that the delivery information available in the presentation contexts did not exert strong enough influences for listeners, always, to correctly anticipate missing targets which were inappropriate continuations of the contexts, we must conclude that the data obtained from this experiment were attributable to the overriding influences of the preceding lexical/semantic context.

These data have implications for the experiments to be described in subsequent chapters. Whilst we have established that delivery information is not of primary importance in cuing whether or not a target is predictable, we have noted that it is, none-the-less attended to. [8] In future experiments it will be necessary, therefore, to be aware of the possible effects of delivery information. Examining utterances which were originally uttered with appropriate or inappropriate sentence final (target) words will be a potential variable (see Chapter 5).

A corollary of the results reported in the present chapter is that interpretation of future perception experiments will be made easier. If, in a later experiment, we establish that preceding semantic context does not affect perception we can not attribute this null result to the possible effects of delivery information cuing the listener to correctly anticipate targets words irrespective of the lexical/semantic context because the data presented above has demonstrated that the effects of delivery are minor in comparison with the lexical/semantic context. From the other side however, if we later observe that preceding context does affect lexical access - even though delivery information may be giving listeners some small chance to anticipate unpredictable target words - then we will have demonstrated the tenacity of contextual effects.

Wales and Toner 1979 observe that experiments claiming to test possible effects of intonation or prosody are difficult to analyze because we cannot be sure that a particular effect is due to the speaker or the listener. Although both speaker and listener must be

[8] Note that this is all we establish here. We do not question the importance of delivery information *per se* (see Section 1.1). Discussion on the relevance of the present experiment to the experimental literature reported earlier however, is beyond the scope of this thesis since the purpose of this experiment was simply as a control for subsequent experiments.

competent in order for communication to occur, we cannot know whether listeners are good at interpreting intonation but speakers are not competent at producing it "correctly", or whether both parties utilize the delivery information. For present purposes however, such characterizations of the speaker/listener are not important. It is sufficient that we have found that listeners do not make use of delivery information to the extent that they are able, significantly often, to discriminate between what would eventually be a predictable or an unpredictable utterance. We are now in a postition to be able to investigate whether semantic context is able to effect the initial processing of acoustic-phonetic during the first 150 msec of word recognition.

CHAPTER 5:

THE EFFECTS OF APPROPRIATE AND INAPPROPRIATE CONTEXT DURING THE INITIAL STAGE OF WORD RECOGNITION.

1. INTRODUCTION.

The main purpose of the experiments reported in Chapters 3 and 4 was to establish whether aspects of the articulatory realization of an utterance were affected by the appropriateness or inappropriateness of the preceding context. This issue was interesting because if appropriate contexts yielded ambiguous productions of, in these cases, voiced and voiceless stops, the speech input for a purely bottom-up model of lexical access would be more complex than has previously been acknowledged. If we had observed, for instance, that production of our target words were affected by context in a way which made them indistinguishable on acoustic grounds, then we would have questioned the degree of constraint provided by processing the bottom-up information alone. We concluded however, that context did not significantly affect the production of speech in this way and that the models of lexical access which rely crucially on initial processing of the acoustic-phonetic information would, therefore, usefully constrain the initial activation of a subset of the mental lexicon.

Our main concern in the present Chapter however, is whether the way in which people *recognize* words is consistent with the bottom-up approach. Aside from the problematic issue of context effects in the production of speech which we described in Chapter 2 (Section 4.1 and 4.2), we noted further problems in allocating all the processor's initial computation to obligatory bottom-up processing. One observation, for example, which might invalidate the assumptions of the bottom-up approach is that word recognition is known to succeed without contemporaneous recognition of the word's initial segment. Also, McAllister 1988 (see Chapter 2, Section 3.4) demonstrated that contextual information can be used by listeners as early as 50 msec after a word's acoustic onset.

2. EXPERIMENT 4.

The purpose of the experiment described below was to justify the assumption underlying Marslen-Wilson and Tyler's Cohort model of lexical access that the perception of the initial segment of a word is immune to any effects of context. To this end, the effect of appropriate and inappropriate contexts on the perception of sentence final target words was measured. Following Tyler 1984, sentence final target words were gated and the effects of context were gauged by noting whether the response offered constituted a phonetically accurate report of the target-initial phoneme and whether the response was semantically congruent with the preceding context.

As in the previous experiments, context was manipulated in such a way that voiced and voiceless initial stop minimal pairs were placed in matched appropriate and inappropriate contexts (see Chapter 3, Sections 2 and 4.2). By manipulating context in this way it was possible to estimate the relative contribution of top-down and bottom-up information during the recognition of the word initial stops. The inappropriate materials allowed us to measure explicitly these relative contributions. If the "inappropriate" targets were correctly identified despite the mis-match with the context, then we assume that only the acoustic-phonetic information was processed and that the contextual information was either not available at this stage in processing or was, in line with the predictions of the Cohort Model, prohibited from contributing to processing. We thus assumed that the inappropriate conditions gave an estimate of the listeners' bias towards using only the acoustic-phonetic information. The appropriate conditions, however, did not permit similar estimates of the listeners' bias towards using only the contextual information. If appropriate targets were correctly identified it was not possible to know whether recognition ocurred due to the contribution of only bottom-up; only top-down; or the contribution of both sources of information.

In order that the separate contribution of contextual-semantic information might be assessed, a further condition was included where the initial segment of the sentence final target word was replaced with shaped noise. [1] Initial segments, rather than whole words,

[1] My thanks to Gerry Altmann for writing the software which generated the shaped noise.

were replaced with the shaped noise because subjects were encouraged to believe they were listening to a word. Had the whole word been replaced with noise, the subjects may have adopted some sort of response strategy irrelevant to the process of speech recognition. We assumed that any correct identifications of the targets which had noise-replaced initial segments could be attributed to processing of the semantic information alone, because no clear initial acoustic-phonetic information was available for these materials.

According to the Cohort theory of lexical access, there should be no recognition of targets in the noise condition at the early gates, [2] because the crucial bottom-up information is not available to initiate the word initial cohort and contextual information, even if available, is not permitted to contribute to processing at this stage. The later gates however, should yield accurate identification of the targets because the availability of subsequent acoustic-phonetic information should successfully activate the target words. To the extent that initial computation is supposed to be purely data driven, there should also be no semantically appropriate responses offered for the inappropriate materials conditions. Rather only semantically *inappropriate* responses will be produced as this is what the acoustic-phonetic information specifies. [3] The Cohort Model also predicts that semantically appropriate responses will be observed only for the appropriate materials and that similar numbers of phonetically accurate reports of (at least) the target-initial phoneme should be found for both the appropriate and inappropriate contexts/targets.

3. METHOD.

3.1. Subjects.

Thirty-two subjects, who were naive as to the purpose of the experiment, participated and were divided into four groups (each of which was further divided for the purposes of counterbalancing the order of presentation of the materials). All were native speakers of British English. Subjects were staff and post-graduate students of Edinburgh University and had been exposed to the Edinburgh accent for a minimum period of two years. They

[2] The duration of the shaped noise did not exceed 3 gates.

[3] Note however that the correct recognition of the initial phoneme might yet give rise to a semantically appropriate response with the corresponding target phoneme. Although this situation was not apparent from the cloze test (see Chapter 3, Section 2) it is a possibility which we recognized in the later analysis of our results. See below.

received payment for their participation.

3.2. Materials.

All stimulus sentences were taken from the digitized recordings made by three female Edinburgh speakers (see materials for Chapter 3). From the utterances recorded by each of the speakers 12 minimal pairs, along with their matched appropriate and inappropiate contexts were selected (i.e. 4 materials per speaker). The minimal pairs and their corresponding contexts were balanced for how well each pair was predicted by the context (see the cloze test in Chapter 3) and for place of articulation of the target-initial stop so that there were 4 bilabial, 4 alveolar and 4 velar stop pairs. The appropriate and inappropriate contexts and associated target words (the minimal pairs) thus constituted the materials designed to evaluate whether the presence or absence of a congruent context prior to clear acoustic-phonetic information had any effect on the identification of the target-initial phonemes. Example (1) below illustrates these 4 conditions.

Example (1)

Appropriate contexts.

(1a) John hates potatoes, carrots and *PEAS*.

(1b) John hates hornets, wasps and *BEES*.

Inappropriate contexts.

(1c) John hates potatoes, carrots and *BEES*.

(1d) John hates hornets, wasps and *PEAS*.

In an effort to minimize any artificial or over-careful quality which the stimulus sentences might have, and also to randomly sample the high and low VOT values (see Chapter 3, Section 12), utterances were randomly selected on the basis of whether they constituted the first, second or third repetition of the original recordings. [4]

[4] Also, targets *originally* uttered in appropriate and inappropriate contexts were cross-spliced so that a given token of a target word appeared following both an appropriate and inappropriate context. This was achieved by cross-splicing targets from Examples (1a) with (1d) and (1b) with (1c), see above. In view of the data reported in Chapter 4, this served as a further control on possible effects of context on the articulatory realization of utterances but the contextual origin of the targets yielded no significant effects.

A further 12 minimal pairs and corresponding sentence frames were then selected, from the original recordings, for the materials. This second set was designed to assess the effects of presenting an appropriate context prior to noise-replaced initial segments of target words. These materials were matched to first set for frequency, place of articulation of target-intial stops and how well each was predicted by the preceding context. Because this condition was designed to estimate the extent to which listeners rely exclusively on semantic information during lexical access, only the appropriate contexts for the minimal pairs were used. [5] The target-initial stops of minimal pairs following appropriate contexts were replaced by shaped noise which was generated over frames of 5 msec. For every 5 msec of the target-initial stop 5 msec of noise was generated which took the mean power of the frame it replaced. Initial and final frames of noise were attenuated so that there were no sudden bursts of noise. The noise was shaped in this way to encourage listeners to believe that they were actually listening to speech like sounds and thus to minimize artifactual reponding. Figure 5.1 illustrates how this noise assumed the same shape as the segments it replaced and the noise conditions are illustrated in Example (2) below.

Example (2)

Noise-replaced target-initial segments; Appropriate contexts.

(2a) Although Bill likes animals, he's never kept a *(P)ET.*

(2b) Although Bill likes casinos, he's never placed a *(B)ET.*

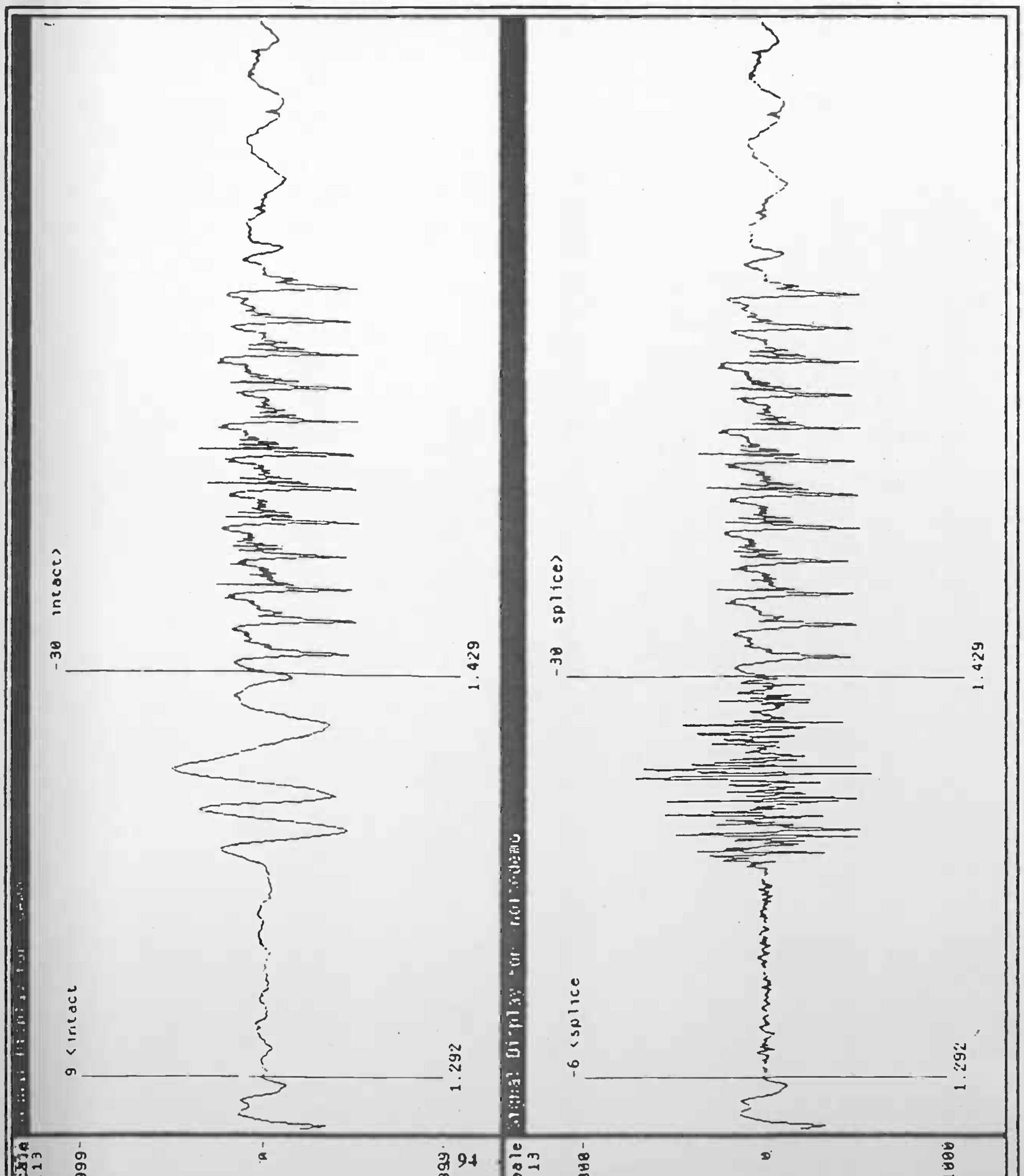
So that subjects would not hear a particular contextual string more than once, the first set of materials (Example (1)) were distributed across 4 groups, balanced for place of articulation and predictability and presented in such a way that subjects would hear only 2

---

[5] Only appropriate contexts were used because of the small possibility that some unspecified acoustic information might cue the inappropriate targets (see Chapter 4).
The noise materials were also informally pretested using a separate group of 14 subjects. The sentence final target words and the words immediately preceding them were gated in unit increments which coincided with the onset and the offset of noise for a given segment. We found that the co-articulatory information to the *right* of the noise segment aided identification of the noise-replaced segment, but that the information immediately to the *left* of the noise-replaced segment did not appear to help identifications. In the present experiment therefore, we could be sure that with each successive gate through the noise segments, the subjects would not have access to clear acoustic information.

FIGURE 5.1

Noise Generation



of the 4 conditions [6] associated with a given minimal pair but would receive all conditions (over 4 materials). Materials from the second set, i.e. with noise targets, were presented so that subjects heard both of the conditions associated with a minimal pair (e.g. Examples (2a) and (2b)).

All target words were gated by 30 msec increments. The first presentation of an utterance exposed subjects to the prior context only. The second and subsequent presentations consisted of the prior context plus the gated target increments. All gated presentations of any one utterance were presented consecutively.

Sixteen filler sentences were pseudo-randomly distributed throughout each group of materials. All filler targets began with fricatives, nasals, liquids and vowels in an effort to minimize subjects' hypotheses as to the voiced/voiceless contrast in the experimental conditions. All experimental materials used in Experiment 4 are given in Appendix E.

3.3. Design.

Voicing (2) was crossed with contextual presentation (appropriate or inappropriate) for 12 materials. This gave 4 conditions per minimal pair. Because subjects received 2 conditions of each of the 12 materials, subjects received 6 instances of each of these 4 conditions. For the further 12 materials (the noise targets), there were 2 conditions (appropriate contexts for the voiced and voiceless member of the minimal pair). Subjects received 12 instances of these 2 conditions. Subjects received twice as many instances of noise targets in order to balance the number of presentations between the 2 sets of (12) materials. [7] Thus subjects received 24 presentations of acoustically clear targets following appropriate and inappropriate contexts and 24 presentations of noise targets following appropriate contexts.

Materials were divided into 4 groups such that there were 3 materials (minimal pairs and associated conditions) per group. Subjects were assigned to one of 4 subject groups and the subject groups rotated through the 4 groups of materials according to a Latin square design such that all subjects received all conditions without ever being exposed to

[6] Subjects actually heard *one* of the four conditions presented in Example (1) along with the same target which had been cross-spliced across contexts. For example, a given subject would have heard (1a) and the cross-spliced version of (1d).

[7] In effect this divided the experiment into two smaller experiments.

the same contextual string twice.

3.4. Procedure.

Subjects were run in groups and were given written instructions asking them to write down everything they heard during the course of the experiment. They were able to ask questions once they had read the instructions and heard one of the gated filler sentences as a practice session. It was emphasized that subjects should only write down what they heard and should not attempt to guess at what the sentence final target word was if they had not clearly heard any of the initial portion of the target. This tactic was intended to minimize any artifactual responses. Thus subjects were allowed to note down syllables and phonemes if they were not sure of the whole word. Subjects were required to write down the sentence final target word (or portion thereof) after each gated presentation. They were not instructed to write out the contexts more than once however, unless they later changed their minds as to words in the contexts.

The experiment was run in 2 sessions each session lasting approximately one and a half hours. The sessions were counterbalanced so that the material presented in the first session for half the subjects, was presented in the second session for the other half.

Stimuli were presented over headphones and subjects wrote their responses on pre-numbered response sheets.

4. RESULTS.

As this experiment was designed to investigate the effects of prior context on word recognition it was considered pertinent to discard all those responses where the preceding context had not been correctly identified. In practice this represented only 8 cases across the 1536 responses of the 32 subjects.

The theoretically significant responses for our hypothesis were those to the first 150 msec of the sentence final target words. The results discussed below, therefore, reflect analysis of the first 5 gates. Later, however, we will consider subsequent responses. Two metrics were obtained: First we calculated the number of phonetically accurate reports of the target-initial phonemes and, second, we noted the number of responses which were semantically congruent with the preceding context (i.e. semantically appropriate).

Frequency data for responses to the first 5 gates were transformed into percentages. These proportions were then used in an Analysis of Variance with factors of information source (appropriate context with clear acoustic-phonetic material; inappropriate context with clear acoustic-phonetic material; and appropriate context with noise-replaced target-initial stops) and gate. In this and similar analyses reported below, "phonetically accurate" identifications of the noise target-initial phonemes were those which corresponded to the initial segments before they had been replaced with noise.

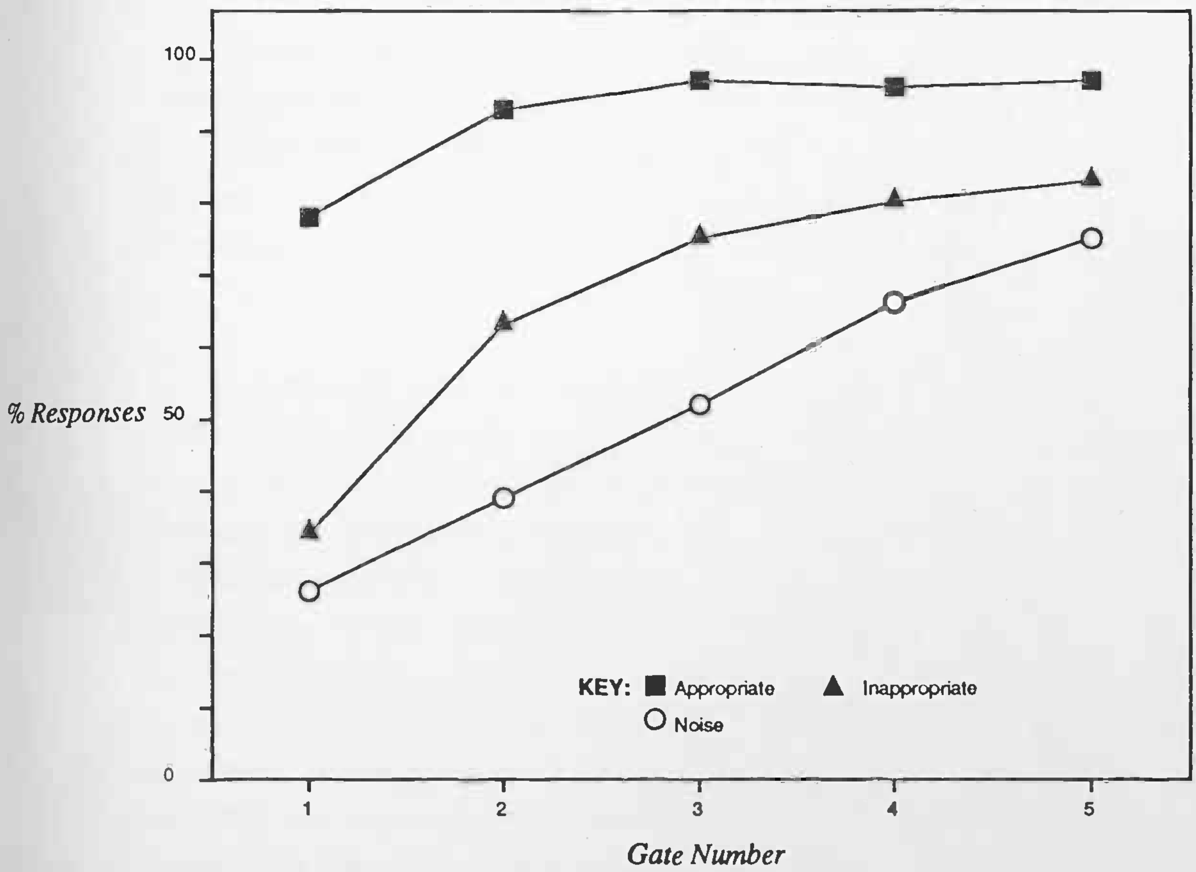
The proportions of phonetically accurate reports of the target-initial phonemes over source and gate yielded significant main effects of gate, $F1(4,124) = 164, p < .0001$; source, $F1(2,62) = 102$, $p < .0001$; and also a significant interaction of source by gate $F1(8,248) = 25.19, p < .0001$. The means for these phonetically accurate responses are presented below in Table 5.1 and illustrated in Graph 5.1 below.

TABLE 5.1
Experiment 4: Percentages of Phonetically Accurate Responses.

Gate Number	Appropriate	Inappropriate	Noise
1	78	34	26
2	93	63	39
3	97	75	52
4	96	80	66
5	97	83	75

GRAPH 5.1

Experiment 4: Percentages of Phonetically Accurate Responses.



Because applying the "phonetically accurate" criterion to the noise targets required a redefinition of "phonetically accurate", a further analysis of variance was run excluding this condition. This did not affect the patterns of significance nor the significance of the differences obtained for between mean comparisons. [8]

The Tukey HSD test was then computed to determine whether or not the source effect held at each gate. For the source by gate interaction the means for appropriate context in conjunction with acoustically clear targets were significantly higher than *all* other means and, at gates 2,3, and 4, the proportions of phonetically accurate responses to the inappropriate materials were significantly higher than those to the noise materials (HSD = 8.1, p = 0.05).

Proportions of semantically appropriate responses were then computed. Two judges were asked to judge whether or not a given responses was congruent given the context which was presented prior to the target word to which the response was made. If the response was semantically congruent with the preceding context, it constituted a semantically appropriate response regardless of whether it constituted a correct identification of the target. Responses were not only semantically appropriate or inappropriate however. Because of the experimental tactics noted above (Section 3.4) some of the responses were ambiguous because only a syllable or phoneme had been recorded by the subject. Although such ambiguities were often resolved quite soon after the first 5 gates it would have been unwise to interpret the responses during this initial stage as either semantically appropriate or inappropriate. For the purposes of this experiment it was crucial to ascertain the contributions of the different information sources during initiation of lexical access. It would, therefore, have been unsound to assume that later responses were representative of those produced at the earlier gates. For this reason we considered semantically appropriate responses in three ways. First we analyzed the number of semantically appropriate responses excluding all ambiguous responses. Second we examined the semantically appropriate responses which were appropriate by virtue of not contradicting the contextual information during the first 5 gates but which were resolved to a semantically appropriate response within the first 10 gates. Finally we considered

[8] A separate ANOVA was also run which included "target origin" as a variable (i.e. the cross-splicing of appropriately uttered targets into inappropriately uttered contexts and vice versa). This analysis excluded the noise materials. There was no significant effect of target origin. For completeness (see Clark 1973) a by items analysis was also run which excluded the noise materials (because they constitued different items). Note that the exclusion of noise conditions did not affect the by subjects analysis. The present by items analysis yielded the following results: [Source $F2(1,23) = 52.88$, $p < .01$; Gate $F2(4,92) = 84.29$, $p < .01$; Source x Gate $F2(4,92) = 23.37$, $p < .01$]. For reasons of experimental control, the materials were very restrict-

semantically appropriate responses which were appropriate because initial fragmented responses did not contradict the contextual information and the responses were later resolved at *any* time before the final gate to constitute a semantically appropriate response.

For the responses which were semantically appropriate at each gate percentage representations were fed into an Analysis of Variance with factors of source and gate as previously. [9] This analysis yielded significant main effects of source, $F1(2,62) = 44$, $p < .0001$; and gate, $F1(4,124) = 37$, $p < .0001$; and a significant interaction of source by gate, $F1(8,248) = 34$, $p < .0001$. [10] Mean proportions for this interaction are presented in Table 5.2 and illustrated in Graph 5.2

TABLE 5.2
Experiment 4: Percentages of Semantically Appropriate Responses, (excluding phonetically accurate identifications of inappropriate stimuli).

Gate Number	Appropriate	Inappropriate	Noise
1	35	20	13
2	48	16	18
3	52	13	27
4	58	12	36
5	61	11	46

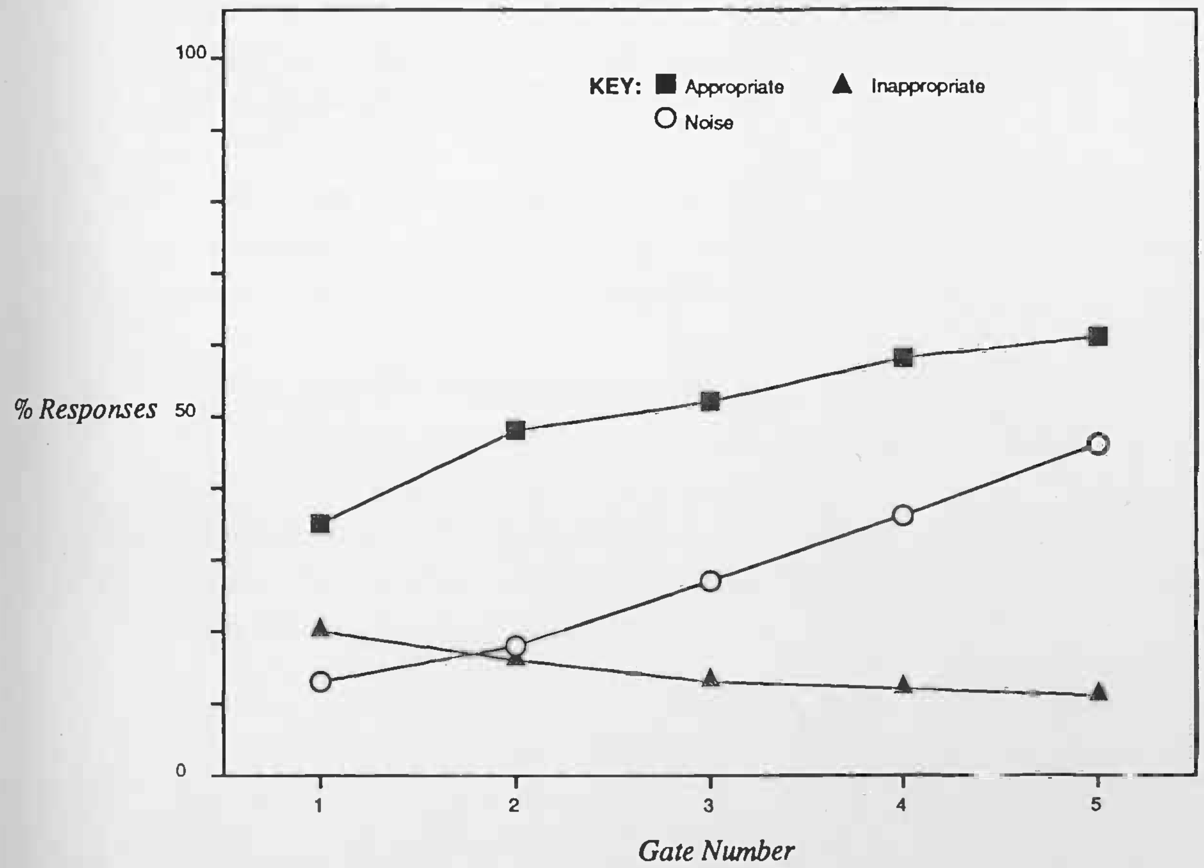
ed however. Generality by item analysis might, therefore, be difficult to obtain.

[9] Note that the proportions here denote percentages of *all* responses and is not restricted to the percentage of only *unambiguous* (whole word) responses.

[10] Again, a seperate ANOVA exluding noise materials was run in a by items analysis. [Source $F2(1,23) = 57.3$, p<

GRAPH 5.2

Experiment 4: Percentages Semantically Appropriate Responses.
(Excluding phonetically accurate identifications of inappropriate stimuli).



The Tukey HSD test revealed that the source effect held at all gates for the appropriate and inappropriate contexts with acoustically clear targets (HSD = 8.8, $p < 0.01$). Also, at the third, fourth and fifth gates the source effect held between the inappropriate contexts and the appropriate contexts with noise-replaced target-initial stops. Note that for the first two gates, the proportions of semantically appropriate responses offered to targets in the inappropriate contexts and the appropriate contexts with noise-replaced target-initial segments were not significantly different.

The second analysis of semantically appropriate responses differed from the first in that it incorporated responses to the inappropriate targets which also constituted correct identifications of the target-initial phoneme. [11]

These proportions also yielded significant main effects of information source $F1(2,62) = 38, p < .0001$; gate, $F1(4,124) = 45, p < .0001$; and a significant interaction of source by gate, $F1(8,248) = 25, p < .0001$. Means for these proportions are presented in Table 5.3 and are illustrated in Graph 5.3 below. For comparison the corresponding results from the analysis which excluded phonetically accurate responses of inappropriate targets (from Graph 5.1) are also shown.

[11] For example, the responses shown in Example (3) were produced to inappropriate targets. The targets actually presented are shown in square brackets.
(3a) The barman poured the whisky into the CUP *[class]*.
(3b) John hates potatoes, carrots and BEANS *[bees]*.

TABLE 5.3
Experiment 4: Percentages of Semantically Appropriate Responses, (Including phonetically accurate identifications of inappropriate stimuli).

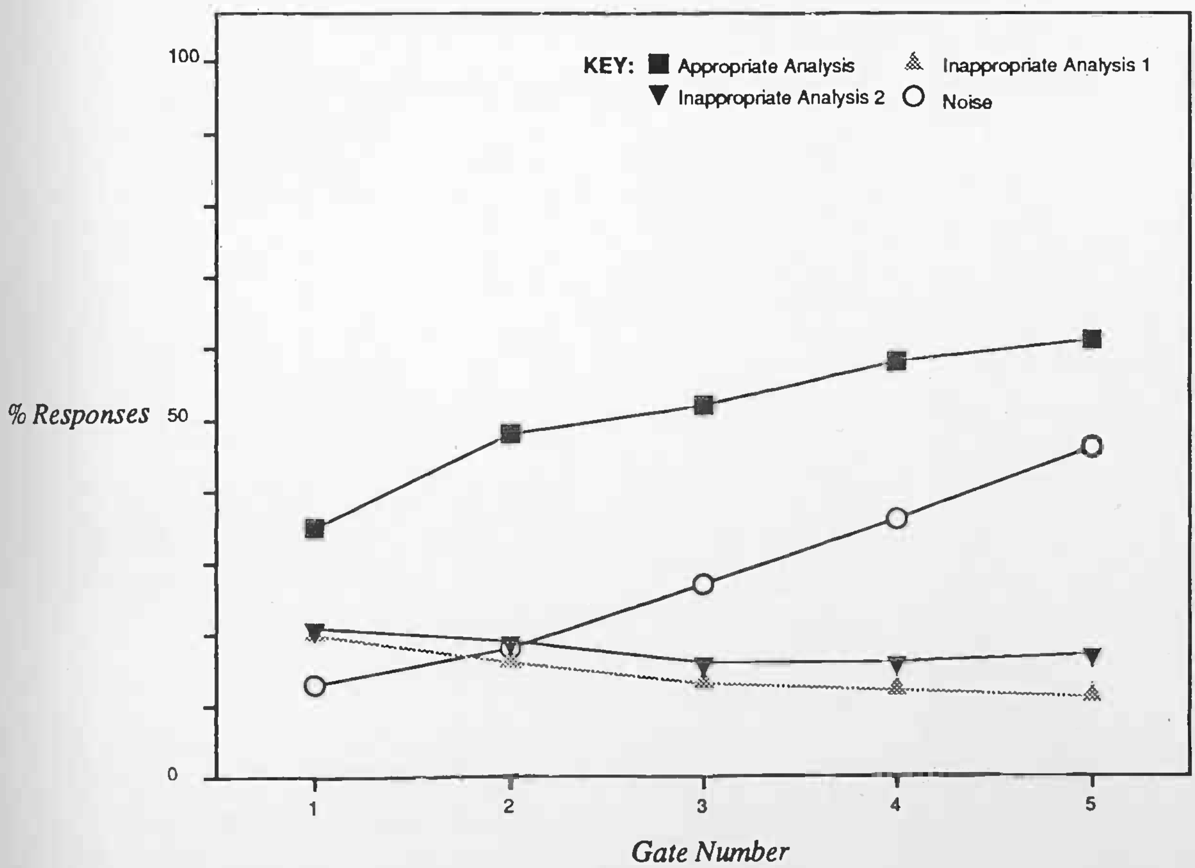
Gate Number	Appropriate	Inapp Analysis 1	Inapp Analysis 2	Noise
1	35	20	21	13
2	48	16	19	18
3	52	13	16	27
4	58	12	16	36
5	61	11	17	46

The Tukey HSD test was computed for this interaction (HSD = 9.1, p = 0.01). At all gates the source effect held between the appropriate context with acoustically clear targets and other two sources. At gates three, four and five the means between the appropriate contexts with noisy targets and the inappropriate contexts were significant.

Similar patterns of results were obtained for the syllabic and phonemic responses which became semantically appropriate word responses within the first 10 gates. This analysis excluded null responses. Thus responses to the first 5 gates such as illustrated in Figure 5.2 and 5.3 were included in this particular analysis.

GRAPH 5.3

Experiment 4: Percentages of Semantically Appropriate Responses, (Including phonetically accurate identifications of inappropriate stimuli).



FIGURES 5.2 and 5.3

Gate No	Response	*	Response	*
1	John hates hornets, wasps and	b	John hates hornets, wasps and	b
2	" " " "	b	" " " "	bea
3	" " " "	b	" " " "	b
4	" " " "	bees	" " " "	be
5	" " " "	"	" " " "	be
6	" " " "	"	" " " "	"
7	" " " "	"	" " " "	"
8	" " " "	"	" " " "	bees

Again two separate analysis were performed; Analysis 1 excluded phonetically accurate reports of the target-initial phoneme for inappropriate materials and Analysis 2 included such responses. The respective analyses yielded significant main effects of source $F1(2,62) = 128, p < .0001$, $F1(4,124) = 26, p < .0001$; gate, $F1(4,124) = 19, p < .0001$, $F1(2,62) = 91, p < .0001$; and an interaction of source by gate, $F1(8,248) = 47, p < .0001$, $F1(8,248) = 33, p < .0001$. Mean proportions are presented in Table 5.4 and are illustrated in Graph 5.4 below.

TABLE 5.4

Experiment 4: Percentages of Semantically Appropriate Responses Produced Before Gate 10. (Analysis 1 excludes phonetically accurate identifications of inappropriate stimuli, Analysis 2 includes them).

Gate Number	Appropriate	Inappropriate 1	Inappropriate 2	Noise
1	65	24	27	29
2	73	19	25	43
3	76	16	23	53
4	76	13	22	61
5	77	12	22	65

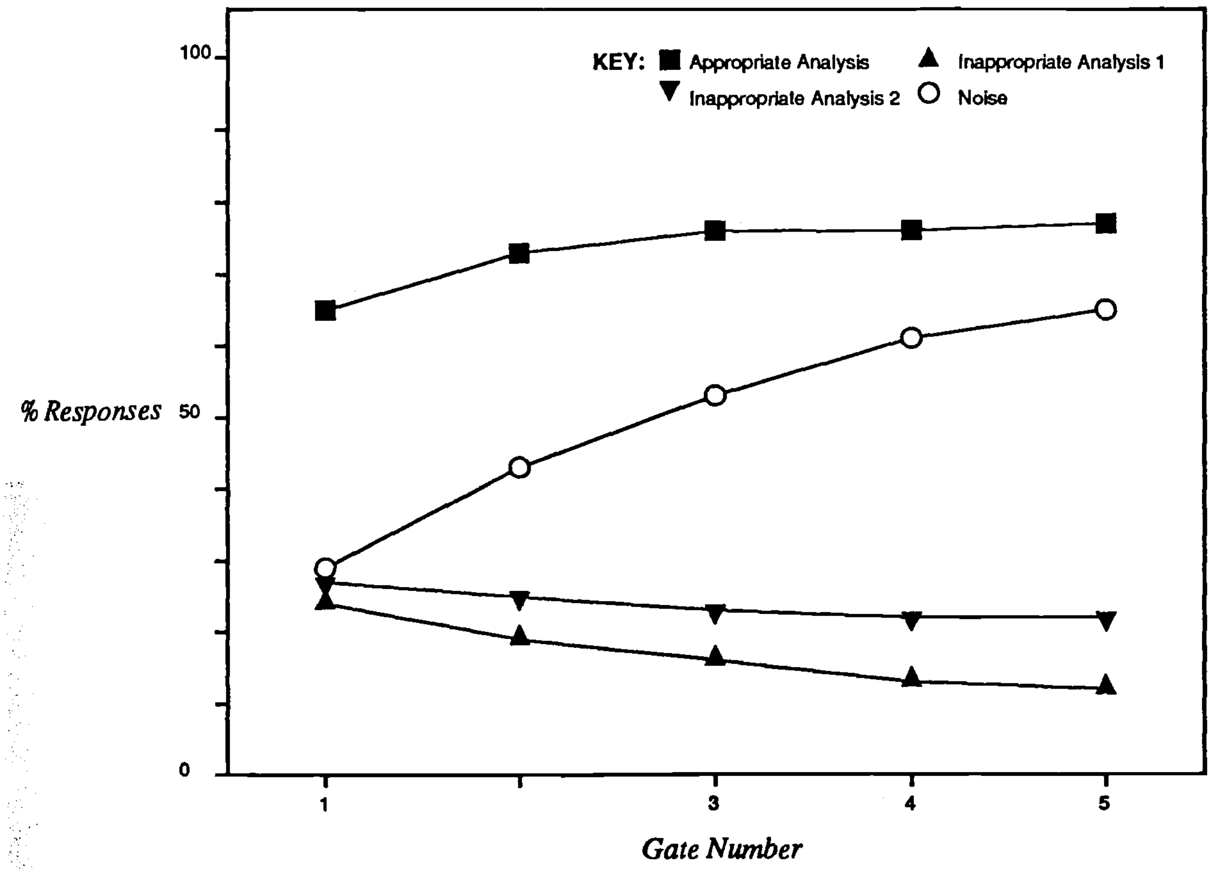
Tukey HSD (Analysis 1 HSD = 8.2, p = 0.01; Analysis 2 HSD = 7.9, p = 0.01) revealed that for both analyses the source effect held between the appropriate contexts with acoustically clear targets and all other conditions and that the means at all gates, except the first, between the inappropriate and the noise sources were significantly different. Exactly the same pattern of results was obtained when analysis of semantically appropriate responses also incorporated responses which were initially ambiguous but which were resolved at any time prior to the final gate.

4.1. Summary of Results.

Analysis of the accurate identifications of the target-initial phonemes yielded significant main effects of information source (i.e. whether the available information comprised appropriate contexts in conjunction with acoustically clear targets; inappropriate contexts in conjunction with acoustically clear targets; and appropriate contexts in conjunction with acoustically unclear targets) and gate (i.e. over time) and a significant interaction of information source by gate. There were more phonetically accurate responses to targets in

GRAPH 5.4

Experiment 4: Percentages of Semantically Appropriate Responses Produced Before Gate 10. Analysis 1 excludes phonetically accurate identifications of inappropriate stimuli, Analysis 2 includes them.



appropriate contexts which were acoustically clear than to targets in inappropriate contexts, and, at every gate after the first, inappropriate materials produced more phonetically accurate responses than the appropriate contexts with noisy targets.

Analysis of the proportions of semantically appropriate responses yielded significant main effects of information source and gate and a significant interaction of source by gate. Again the source effect held between the appropriate contexts with acoustically clear targets and the two other sources and, at all gates after the first, there were fewer semantically appropriate responses to inappropriate materials than to the noise materials (appropriate context with acoustically unclear targets).

5. DISCUSSION.

The present experiment aimed to study which sources of information were made available and subsequently utilized during the initial stage of lexical access. Acoustic-phonetic and prior contextual information were manipulated in such a way as to allow the processor access to one or other or both of these information sources.

The results presented above demonstrated, unsurprisingly, that when top-down and bottom-up information sources were jointly available, as they were for acoustically clear targets following appropriate contexts, the number of accurate reports of the target initial phoneme was much higher than when only one of the sources was made available (see Graph 5.1 above). Of more interest was the observation that when the acoustic information was available but incompatible with the preceding contextual information (i.e. when inappropriate contexts preceded acoustically clear targets) the number of phonetically accurate responses was significantly lower than where both information sources were present. This drop in the number of phonetically accurate reports directly opposes models of lexical access which claim that initial computation is purely bottom-up. If this were the case, then we should have obtained comparable numbers of phonetically accurate responses whether bottom-up information was available alongside compatible top-down information or not.

One explanation for the drop in the number of phonetically accurate responses to targets which were incongruous for the preceding context is that contextual information was being utilized within the first 150 msec of processing. If initial processing was effected by context, then the conflicting top-down and bottom-up information, present for inappropriate materials, would yield fewer accurate identifications of the target-initial phonemes.

This possibility is further substantiated by noting the gradients of the source curves in Graph 5.1. Because the prior context conflicts with the clear bottom-up information in the inappropriate conditions, the slope of the inappropriate curve climbs steeply throughout the whole 150 msec period. It would appear that the acoustic information available at the later gates is still useful in terms of increasing the number of correct identifications of the target-initial phonemes. The slope of the appropriate contexts with acoustically clear targets however, reaches a ceiling at the third gate. It is plausible that the compatible top-down information available in the appropriate condition has enabled earlier confirmation or identification of the bottom-up information. Where the top-down information was incompatible with the bottom-up information, however, the relevant curve in Graph 5.1 suggests that the incoming acoustic information was continually monitored and used to facilitate identification of target-initial phonemes at the later gates. It is possible that the number of phonetically accurate responses steadily increased through the later gates because compatible top-down information was not available to facilitate identification at the early gates in the way it was for appropriate targets. It would thus appear that the ceiling effect observed for the appropriate targets may be due to the prior context bringing forward the target word's recognition point.

An alternative explanation for the observed differences in the numbers of phonetically accurate responses between the appropriate contexts with acoustically clear targets and inappropriate contexts, is that subjects may have been responding on the basis of some strategy whereby they ignored the acoustic-phonetic information altogether. If subjects elected to attend only to the prior context then indeed we should expect fewer phonetically accurate reports of the target-initial phonemes when the target was inappropriate for the preceding context. For this account to be plausible however, we would not have predicted that the number of phonetically accurate responses be greater for the inappropriate conditions than conditions where the initial portion of bottom-up information had been replaced with noise. Our results clearly contradict this alternative account. We observed that the clear acoustic information present in the inappropriate conditions *was* able to significantly facilitate identification of target-initial phonemes over the conditions in which appropriate contextual information was made available alongside acoustically *unclear* targets in the noise conditions. We may thus conclude that subjects were not ignoring the acoustic information. [12]

Analysis of the semantically appropriate responses also yielded results compatible with our initial interpretation. Again unsurprisingly, a higher number of semantically appropriate responses were produced to targets from the appropriate context with acoustically clear targets. We also anticipated the increase at later gates in the number of semantically appropriate responses to targets from the appropriate and noise conditions. For the noise-replaced targets, this was due to the help from acoustic-phonetic information which was available in the fourth and fifth gates after the initial burst of noise.

The purely bottom-up models of lexical access however, predict that *few* semantically appropriate responses should have been obtained to targets from the bottom-up-only conditions. [13] We observed however, that some responses under these conditions were semantically appropriate. Even in our most conservative analysis which excluded semantically appropriate responses which were also phonetically accurate, the number of semantically appropriate responses was as high as 21%. Although it is debatable whether 21% constitutes a *high* proportion of semantically appropriate responses this does not compromise our interpretation. Compare this percentage with the 3% which we might expect by chance (McAllister 1988).

One curious aspect of our *first* analysis of these responses is that the proportions of semantically appropriate responses to targets in appropriate contexts are lower than the corresponding proportions of phonetically accurate responses. The explanation for this difference is simply that whilst all part-word responses could be included in the phonetically accurate response analysis, only *whole* word responses were included in the initial analysis of semantically appropriate responses. As we noted above (see Sections 3.4 and 4) although many of the fragmented responses quickly became resolved to whole words, it would have been unwise to interpret the early responses on the basis of later ones. Experimental tactics designed to minimize artifactual responses gave rise to many responses being part-word, and the exclusion of such responses from the first analysis of semantically appropriate responses thus produces dissimilarities between the proportions of semantically appropriate and phonetically accurate responses. Note however that subsequent

[12] Note that because the shaped noise never extended beyond the first 3 gates, we anticipate similarities as later gates between the noise replaced targets and the inappropriate targets due to noise replaced targets carrying subsequent coarticulatory information in the acoustic information which would have been available at later gates.

[13] That is unless the initial phoneme of the inappropriate targets might also constitute the initial phoneme of an ap-

analyses of semantically appropriate responses which include ambiguous (but later resolved) responses compares more directly with the analysis of phonetically accurate responses. That there were part-word responses, even in cases where the target word was highly predictable also supports the argument presented here that subjects were indeed basing their responses on acoustic information.

The important point to note from this analysis however, is that the proportions of semantically appropriate were always higher for targets in appropriate contexts than inappropiate contexts. If subjects were responding by electing to ignore the acoustic information, then we should have observed similar proportions of semantically appropriate responses across the two conditions.

The distribution of these responses over the 5 gates also disables the objection that this number of semantically appropriate responses is due to subjects electing to ignore the incoming acoustic-phonetic information altogether. As more of the bottom-up information is heard, the number of semantically appropriate responses to the inappropriate condition decreases. This exactly mirrors the inappropriate source curve in the analysis of the phonetically accurate responses. Furthermore, at the first 2 gates, the appropriate contexts with noise-replaced target-initial stops yielded the same number of semantically appropriate responses as the first 60 msec (2 gates) of incongruent but acoustically clear information available in the inappropriate conditions. For both these conditions, the numbers of semantically appropriate responses produced to the first 2 gates was significantly below those produced to targets which were acoustically clear *and* congruent with the preceding context. When the subsequent acoustic information became available after the intial segment of noise however, the acoustic information, in conjunction with compatible contextual information pulled the number of semantically appropriate responses produced to the noise conditions significantly *above* those offered to inappropriate targets.

propriate target.

6. CONCLUSION.

The interpretations which we have made on the data presented above lead us to conclude that contextual information *can* influence lexical access although acoustic-phonetic information is given priority.

When listeners are presented with conflicting top-down and bottom-up information (as in the inappropriate conditions) they gave priority to the acoustic-phonetic information. Thus, over time, we observed an *increase* in the number of phonetically accurate responses and a *decrease* in the number of semantically appropriate responses. [14] The low numbers of semantically appropriate responses for both the inappropriate and the noise materials at early gates suggests that bottom-up information was very important. These responses illustrate that subjects were not able to respond effectively on the basis of semantic information alone. Once acoustic-phonetic information was made available in the later gates however, the presence of the appropriate context was able to increase the number of semantically appropriate responses to noise targets relative to the inappropriate targets. Because the top-down information was incompatible in the inappropriate conditions however, the number of phonetically accurate responses produced was still far fewer than the numbers observed for conditions where the prior contextual information was compatible with the bottom-up information.

It thus appears that people *are* able to utilize prior contextual information during the first 150 msec of word recognition. When clear acoustic information is encounterd however, it is obligatorily processed. If the acoustic-phonetic information contradicts the semantic information, correct identification of the target-initial phoneme is not as efficient as when the two information sources are compatible. Yet when the two information sources clash in this way, identification of the target-initial phonemes is still more accurate than when appropriate contextual information is available alongside unclear acoustic information. The results presented above have demonstrated that whilst bottom-up information is given priority during the initial stage of lexical access, contextual information is *not* prohibited from contributing to processing.

[14] This inverse relationship also reflects the (uncontroversial) fact that the acoustic information is more constraining than the semantic information.

In reaching this conlusion however, we made much of the distribution of responses to targets which had their initial segments replaced with noise. These noise targets served as a control because they allowed us to estimate, at early gates, the extent to which listeners ever relied exclusively on the contextual information. As we mentioned earlier though (Section 3.3, footnote 6), the manner in which the noise materials were generated effectively divided the experiment in to two smaller experiments. A corollary of this is that the differences observed between the noise conditions and the appropriate and inappropriate conditons may have simply been due to the difference in materials. For this reason, a further experiment (Experiment 5) was conducted.

7. EXPERIMENT 5.

The experiment reported below was designed to replicate the findings and substantiate the conclusions drawn in connection with Experiment 4. Based on the results of Experiment 4, we suggested that although acoustic-phonetic information is given priority during lexical access, prior contextual information *can* be utilized during this early stage in processing. Broadly, this conclusion was based on the finding that listeners were better able to identify target initial phonemes when *appropriate contextual* information was available in conjunction with clear acoustic-phonetic information. When the clear acoustic-phonetic information was available following *inappropriate* contextual information, however, listeners were still better able to make phonetically accurate responses than when the context was appropriate but the acoustic-phonetic information was unclear (i.e. the noise condition). We argued that the difference between the listener's ability to identify target initial phonemes in appropriate and inappropiate contexts was due to the early influence of the contextual information. We also noted that this difference could not be attributed to the listener electing to ignore the acoustic-phonetic information altogether as 80% of listeners' responses to targets which were inappropriate for the prior context were phonetically accurate but semantically inappropriate. Presumably *all* the responses would have been semantically *appropriate* if listeners had not attended to the acoustic-phonetic information. Also, listeners were poor at identifying noise-replaced target initial phonemes even though they were highly predicted by the semantic context.

The noise-replaced targets were assumed to reflect a situation where only top-down information was made available to the listener. Results in the noise conditions were used to show that the difference in the number of phonetically accurate responses between the appropriate and inappropriate contexts was due to an early effect of context rather than an ad hoc response strategy whereby listeners paid attention only to the semantic information and ignored the acoustic-phonetic information altogether.

When we compared the patterns of responses produced to noise-replaced targets with acoustically clear appropriate and inappropriate targets, it appeared that listeners were utilizing both contextual and acoustic-phonetic information during lexical access. A flaw in the design of these noise conditions however, weakened this conclusion. Although the materials for the noise-replaced targets conditions were very closely matched to those of the appropriate and inappropriate contextual conditions they were nonetheless *different.* It is possible, therefore, that the observed patterns of results may simply have been due to differences in the materials. Furthermore, the method by which "inappropriate" noise-replaced targets were created gave rise to twice as many stimuli in these conditions as in either the appropriate or the inappropriate contextual conditions. We had not included the inappropriate versions of the noise materials in case there had been any unspecified acoustic information which had cued the listeners to expect unpredictable targets. It would have been interesting, however, to note whether there were any differences between responses to noise-replaced targets which followed originally appropriate or inappropriate contexts.

To resolve these problems, the experiment reported below was run as a follow-up to Experiment 4. The present experiment incorporated the basic design features of the previous one although all variables were fully crossed in order to accomodate the points noted above.

Again this experiment attempted to justify the assumption underlying some current theories of word recognition that all the processor's initial computation is allocated to obligatory bottom-up processing. Contextual information, therefore, was manipulated along with the clarity of the acoustic-phonetic information in order to ascertain the relative contributions of top-down and bottom-up information during lexical access. Following Tyler 1984, sentence final target words were gated and subjects responses were analyzed in terms of phonetically accurate reports of the target initial phonemes and semantic congruency with the contextual presentation.

As described previously, (see Section 1 above), the contribution of information sources was estimated by observing listeners' responses when appropriate and inappropriate contextual information was made available in conjunction with acoustically clear bottom-up information, and also when appropriate and inappropriate contextual information was available preceding acoustically unclear information. Acoustically unclear targets were created by replacing target initial segments with shaped noise. If subjects accurately reported targets which were inappropriate it is plausible that they were attending to the acoustic-phonetic information. If, instead, they had attended only to the contextual information then no phonetically accurate responses would be obtained [15] whilst any correct identifications of noise-replaced initial segments could be attributed to processing of only the semantic information because of the absence of clear acoustic-phonetic information.

8. METHOD.

8.1. Subjects.

Sixteen new subjects participated. They were all naive as to the purpose of the experiment. All were native British English speakers and had been exposed to the Edinburgh accent for a minimum period of two years. Subjects were students of the the University of Edinburgh and were paid for their participation.

8.2. Materials.

All stimulus sentences were taken from the digitized recordings made by three female Edinburgh speakers (see materials for Chapter 3). From these utterances 12 minimal pairs, along with their matched appropriate and inappropriate contexts, were selected (i.e. 4 minimal pairs per speaker). The minimal pairs and their corresponding contexts were balanced for how well each pair was predicted by the context (see the cloze test presented in Chpater 3) and for place of articulation of the target initial stop. There were thus 48 utterances (24 target words from the minimal pairs each appearing in an appropriate and an inappropriate context). In an effort to minimize any citation quality which the stimulus sentences might have, utterances were randomly selected on the basis of whether they constituted the first, second or third repetition of the original recordings.

[15] To the extent that the cloze test (see Chapter 3, Section 2) provided a reliable estimate of the mutually exclusive predictability of one member of a minimal pair from a particular context, we expect that no semantically appropriate responses

Appropriate and inappropriate materials, as originally uttered, were used in this experiment for estimating the effects of clear bottom-up information in conjucntion with either congruent or incongruent top-down information. Appropriate and inappropriate targets were not cross spliced in the present experiment, because in previous experiments, the contextual origin of the target had demonstrated no significant effects.

To estimate the extent to which listeners ever rely exclusively on top-down information, the appropriate and inappropriate contexts, along with all target words, were then copied. The initial stop segment for each target word was replaced by shaped noise as described previously. The durations of these noise segments ranged from 36 msec (the average length of noise replacement for the voiced stops) to 82 msec (the average length of noise replacement for the voiceless stops).

There were thus 96 utterances in total (2 contexts x 2 targets x 2 noise (present or absent) x 12 materials). Example (4) illustrates the 8 conditions per minimal pair.

---

to inappropriate targets will have the same initial phoneme as the inappropriate target.

Example (4)

Acoustically clear targets + Appropriate context

(4a) John hates potatoes, carrots and *peas*.

(4b) John hates hornets, wasps and *bees*.

Acoustically clear target + Inappropriate context.

(4c) John hates potatoes, carrots and *bees*.

(4d) John hates hornets, wasps and *peas*.

Noise-replaced target + Appropriate context.

(4e) John hates potatoes, carrots and *XX(P)eas*.

(4f) John hates hornets, wasps and *XX(B)ees*.

Noise-replaced target + Inappropriate context.

(4g) John hates potatoes, carrots and *XX(B)ees*.

(4h) John hates hornets, wasps and *XX(P)eas*.

Thirty-two filler sentences were pseudo-randomly distributed throughout the experimental stimuli. Filler target words began with fricatives so as to minimize subjects' hypotheses as to the limited set of initial phones in the experimental conditions. In an effort to discourage subjects from responding on the basis of the semantic information alone, one half of the targets in the filler sentences were inappropriate to the preceding context.

Target words were gated by 30 msec increments. For the conditions in which the acoustic-phonetic material had been replaced with noise, the onset of the first gate coincided with the onset of the shaped white noise. This point always corresponded to the closure prior to the release of the stop in targets which were acoustically intact. For any given target therefore, the onset and duration of the first and subsequent gates was constant across all 4 conditions. The first presentation of an utterance exposed subjects to the prior context only. The second and subsequent presentations consisted of the context plus the gated target increments. All gated presentations of any one utterance were presented consecutively.

In order that subjects did not hear a particular contextual string more than once materials were distributed across 4 groups balanced for place of articulation and predictability such that subjects would hear only 2 of the 8 conditions associated with a given minimal pair (e.g. (4a) and (4f) above) but would receive all conditions. Thus 4 materials tapes were recorded. Each tape contained a different 2 conditions from each of the 12 minimal pairs. Conditions were rotated through a Latin square. All experimental materials used in Experiment 5 are given in Appendix F.

8.3. Design.

Voicing (2) was crossed with contextual presentation (2) and clarity of acoustic-phonetic information (2) yielding 8 conditions per minimal pair (see Example (4) above). Due to the constraint that subjects should only hear a given contextual string once subjects were assigned to one of four subject groups which rotated throught the four groups of materials according to a latin square design (see Appendix F). Each subject thus received 3 instances of each of the 8 conditions.

8.4. Procedure.

The procedure was exactly the same as Experiment 4 (see Section 3.4 above). Because this experiment was designed to investigate the effects of prior context on word recognition it was considered pertinent to discard all those responses where the preceding context had not been correctly identified. In practice this represented only 1% of the total number of responses.

9. RESULTS.

The theoretically significant responses for our hypothesis were those to the first 150 msec of the sentence final target word. Therefore, the results discussed below reflect analysis of the first 5 gates. Two metrics were obtained. First we calculated the number of phonetically accurate reports of the target-initial phonemes and, second, we noted the numbers of responses which were semantically congruent with the preceding context.

Frequency data for responses to the first 5 gates were transformed into a percentage representation, which were then used in an Analysis of Variance with factors of contextual presentation, acoustic clarity and gate. In this and similar analyses reported below, phonetically accurate identifications of the noise-replaced targets were taken to be responses which corresponded to the initial segments before they had been replaced with noise. The proportions of phonetically accurate reports of the target initial phonemes yielded significant main effects of gate, $F1(4,60) = 65.15$, $p < .0001$, $F2(4,92) = 84.99$, $p < .0001$, [Min $F'(1,135) = 37$, $p<.01$];

contextual presentation, $F1(1,15) = 32.07$, $p < .0001$, $F2(1,23) = 28.57$, $p < .0001$, [Min $F'(1,37) = 15$, $p < .01$];

and acoustic clarity, $F1(1,15) = 57.03$, $p < .0001$, $F2(1,23) = 26.56$, $p < .0001$, [Min $F'(1,37) = 18$, $p < .01$];

and a significant interaction of these three variables, $F1(4,60) = 3.91$, $p = .0069$, $F2(4,92) = 2.19$, $p = .0767$ [Min F' not significant]. The means for these responses are presented in Table 5.5 and illustrated in Graph 5.5 below. [16]

[16] Because applying the "phonetically accurate" response criteria to the noise substituted segments required a redefinition of "phonetically accurate", an equivalent Analysis of Variance was performed which excluded the noise-replaced conditions. This did not affect the patterns of significance nor the differences obtained for the between mean compairsons.

TABLE 5.5

Experiment 5: Percentages of Correctly Identified Target Initial Phonemes.

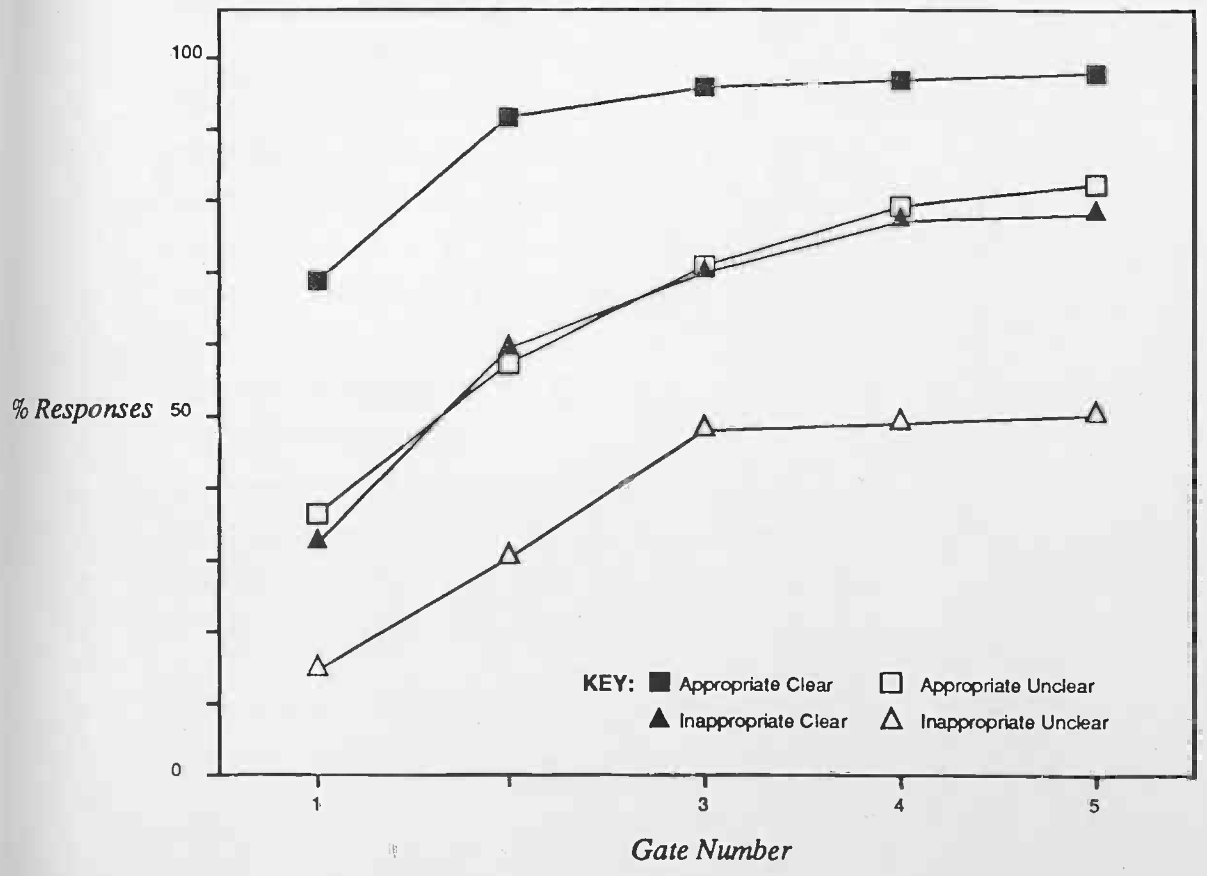
Gate Number	Appr Clear	Appr Unclear	Inappr Clear	Inapp Unclear
1	68.7	36.4	32.3	14.6
2	91.7	57.3	59.4	30.2
3	95.8	70.8	69.8	47.9
4	96.9	79.2	77.1	48.9
5	97.9	82.3	78.1	50.0

The Tukey HSD test was then computed to determine whether there were significant context or acoustic clarity effects at each gate. For the gate by contextual presentation by acoustic clarity interaction (see Graph 5.5 below) Tukey HSD = 20, $p = 0.05$. The means for the appropriate contextual presentations with clear (intact) acoustic-phonetic information were significantly above all other means at the first three gates. At the fourth and fifth gates the appropriate-clear means were also significantly greater than *all* inappropriate contextual presentation means. After the first gate, all means for the inappropriate contextual presentations with clear acoustic-phonetic targets were greater than the means for the inappropriate contextual presentations with unclear (noise-replaced) acoustic-phonetic targets. Finally, means for the unclear appropriate targets were significantly higher than inappropriate (clear and unclear) at all gates.

Proportions of semantically appropriate responses were then computed. As in the Experiment 4, two judges were asked whether or not a given response was congruent given the context which was presented prior to a given target word. Responses were either semantically appropriate, ambiguous, or semantically inappropriate. As noted in Section 3.4

GRAPH 5.5

Experiment 5: Percentages of Phonetically Accurate Responses



above, experimental tactics were responsible for creating fragmented responses such as syllabic or phonemic responses which were classed as ambiguous.

Since no interesting observations, in the analyses of Experiment 4, were gained from the separate analysis of ambiguous responses, we shall restrict the present analysis to the frequencies of responses which are "strictly" semantically appropriate by virtue of their being whole word responses. The only effect that this might have is to make our interpretation of the data more conservative. Proportions of semantically appropriate responses at each gate were fed into an Analysis of Variance with factors of gate, contextual presentation and acoustic clarity as previously. [17] This analysis yielded significant main effects of gate, $F1(4,60) = 8.14$, $p < .0001$, $F2(4,92) = 11.71$, $p < .0001$, [Min $F'(4,131) = 5$, $p < .01$];
contextual presentation, $F1(1,15) = 13.10$, $p = .0025$, $F2(1,23) = 18.67$, $p = .0003$, [Min $F'(1,33) = 7.69$, $p < .01$];
and acoustic clarity, $F1(1,15) = 4.61$, $p = .0486$, [$F2(1,23) = 1.5$, $p > .05$];
and a significant interaction of these three variables, $F1(4,60) = 3.02$, $p = .0245$, [$F2(4,92) = 2.3$, $p = .06$], (Min F' not significant). Mean proportions for the interaction are presented in Table 5.6 and illustrated in Graph 5.6 below.

[17] Note that the proportions here denote percentages of *all* responses and is not restricted to the percentage of only *unambiguous* (whole word) responses.

TABLE 5.6

Experiment 5: Percentages of Semantically Appropriate Responses, (excluding phonetically accurate responses for inappropriate stimuli).

Gate Number	Appr Clear	Appr Unclear	Inappr Clear	Inapp Unclear
1	22.9	20.8	26.0	17.7
2	35.4	25.0	16.6	16.6
3	43.7	31.2	20.8	18.7
4	50.0	33.3	17.7	18.7
5	57.3	46.9	16.7	19.9

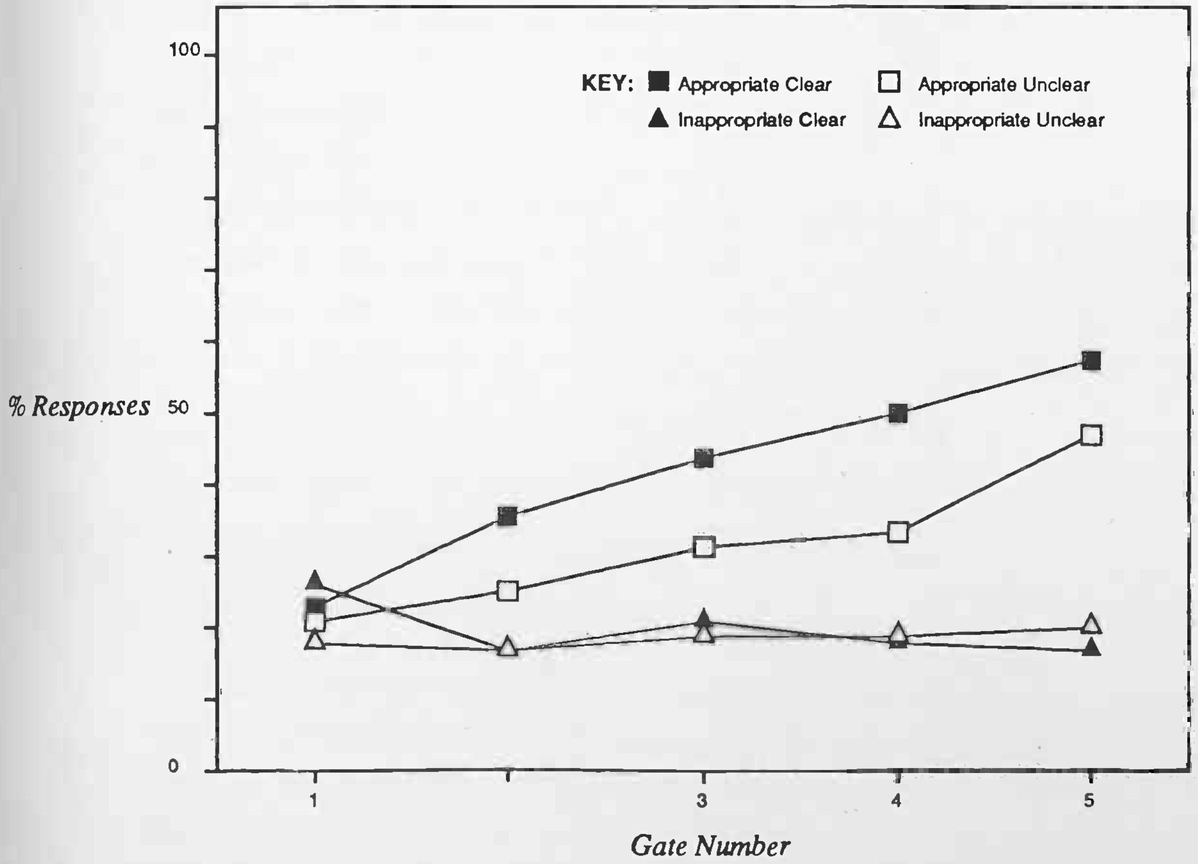
The Tukey HSD test (HSD = 18.5, p = 0.05) revealed that, after the first gate, the means for the acoustically clear appropriate targets were greater than all inappropriate targets. At the fifth gate the mean for the acoustically clear inappropriate contextual presentation was significantly lower than the mean for the acoustically unclear appropriate contextual presentations and the mean for the unclear appropriate context was significantly greater than both clear and unclear inappropriate context means

9.1. Summary of Results.

Analysis of the accurate identification of the target initial phonemes yielded significant main effects of contextual presentation, acoustic clarity and gate and a significant interaction of these three variables. The Tukey HSD test confirmed that, after the first gate, all means for acoustically clear appropriate materials were greater than both acoustically clear and unclear inappropriate context means. More phonetically accurate responses were gained to acoustically clear than acoustically unclear appropriate targets for the first three gates.

GRAPH 5.6

Experiment 5: Percentages of Semantically Appropriate Responses (Excluding phonetically accurate responses for inappropriate stimuli).



Analysis of the proportions of semantically appropriate responses yielded significant main effects of contextual presentation, acoustic clarity and gate and a significant interaction of these variables. The proportions of semantically appropriate responses were lower for acoustically clear and unclear targets presented after inappropriate contexts than for clear targets after appropriate contexts.

10. DISCUSSION.

The present experiment aimed to study the information sources made available and subsequently utilized throughout lexical access. The clarity of bottom-up information and the appropriateness of top-down information were manipulated in such a way as to make available to the listener one of the following combinations of information sources:

(a) Acoustically clear and appropriate contextual information.

(b) Acoustically clear and inappropriate contextual information.

(c) Acoustically unclear and appropriate contextual information.

(d) Acoustically unclear and inappropriate contextual information.

Many of the results presented above replicate the findings of Experiment 4. When clear acoustic-phonetic information was made available in conjunction with appropriate contextual information the number of correct identifications of target initial phonemes was significantly greater than for conditions where the listener was deprived of one of these information sources. As we argued previously, the drop in phonetically accurate responses to targets which followed inappropriate contexts has interesting implications for models of lexical access which emphasize that initial computation proceeds purely in terms of analysis of the acoustic-phonetic information.

We suggested that this pattern of results reflects the early role of contextual information in lexical access. The distribution of the phonetically accurate responses over the five gates is also similar to that observed in Experiment 4 (see Graph 5.5 above). Again, it would appear that, due to the conflicting acoustic and contextual information present for clear targets following inappropriate contexts, incoming acoustic-phonetic information is continually monitored thus continually increasing the number of accurate responses at later gates. In contrast, the gradient representing accurate responses to clear targets appropriate for the prior context reaches a ceiling as early as the second gate. We argued that this is because the compatible contextual information has enabled earlier identification of the target initial phoneme.

Yet again we are able to reject the possibility that the differences in the numbers of phonetically accurate responses across the appropriate and inappropriate contexts is due to a response strategy where subjects elected to ignore the acoustic-phonetic information altogether and attended only to the prior semantic information. Although we observed fewer phonetically accurate responses to targets which followed inappropriate contexts, as this strategy would require, [18] we also noted differences between the number of correctly identified initial phonemes to acoustically clear and unclear targets following appropriate contexts. If subjects were electing to ignore the incoming acoustic information, then the proportions of phonetically accurate responses to the clear and unclear appropriate materials should have been similar because *exactly the same* contextual information was available across the two conditions. The data presented however, yielded significantly more phonetically accurate responses to the *clear* over the unclear appropriate targets at the first, second and third gates. Finally, the observation that, at every gate after the first, there were also more phonetically accurate responses to acoustically clear as opposed to unclear targets following inappropriate contexts, indicates that the listeners were attending to the acoustic-phonetic information - a result which is incompatible with the possibility that the listener was operating an ad hoc strategy whereby they ignored the acoustic-phonetic information altogether.

Analysis of the semantically appropriate responses also yielded results which were consistent with the results of Experiment 4. Again we infer that subjects were processing the acoustic-phonetic information. After the first gate, for instance, there were always fewer semantically appropriate responses to inappropriate targets than there were for clear appropriate targets. The acoustic information available in these inappropriate targets specified semantically *inappropriate* responses and, therefore, fewer semantically *appropriate* responses were obtained. For targets which were acoustically unclear for the first few gates [19] but still appropriate to the preceding context there was a sharp increase in the number of semantically appropriate responses produced at the later gates. Presumably this reflected the use of the subsequent coarticulatory information. Also, at the later gates

[18] The increase at later gates in the number of phonetically accurate responses to inappropriate targets however, may pose a problem for the interpretation that subjects were ignoring the bottom-up information because this increase may be indicative of the listener attending to incoming acoustic information.

[19] Note that the duration of the first three gates is 90 msec and that the mean durations of noise replacements for

the clear acoustic-phonetic information available for inappropriate targets caused a drop in the number of semantically appropriate responses. If subjects had been attending only to the prior contextual information, then the number of semantically appropriate responses should have remained constant across all conditions. The observed differences between the proportions of semantically appropriate responses across the various combinations of information sources however, effectively rules out this possibility.

In conjunction with the observed differences between the number of phonetically accurate responses to appropriate and inappropriate contexts, the distribution of the semantically appropriate responses also supports the notion that contextual information was being utilized alongside the acoustic-phonetic information. The purely bottom-up models of lexical access predict that *few* semantically appropriate responses (which were not phonetically accurate) should have been obtained to clear targets following inappropriate contexts. Even though our analysis of semantically appropriate responses was conservative (we did not include ambiguous responses which later became semantically appropriate, nor did we incorporate semantically appropriate responses which were phonetically accurate) the number of semantically appropriate responses to inappropriate targets still ranged from 16.6% to 26%. [20]

Finally, the observed differences between the appropriate and inappropriate contexts for the acoustically unclear targets are also consistent with the notion that top-down information can contribute to the process of lexical access. These conditions yielded different proportions for both the phonetically accurate and, at later gates, semantically appropriate responses. Again, these differences may be explained in terms of the listener utilizing both acoustic-phonetic and contextual information. At all gates the unclear appropriate targets gave rise to more correct identifications of the target initial phonemes than unclear inappropriate targets Note that, at most, the noise lasted only 3 gates. It is possible, therefore, that the availability of the congruent contextual information in the former case facilitated interpretation of any coarticulatory information which may have been available at later gates. Similarly, because there was more coarticulatory information available at later gates we observed that there were fewer semantically appropriate responses to unclear inappropriate targets than there were to unclear appropriate targets.

---

voiced and voiceless segments were respectively 36 msec and 82 msec.

[20] Note that McAllister 1988 estimates that only 3% of responses should be semantically appropriate by chance.

11. CONCLUSION.

The main aim of the experiments presented in this chapter was to assess the relative contribution of top-down and bottom-up information sources during lexical access. The available data have lead us to conclude that although acoustic-phonetic information is given priority, contextual information *can* contribute to processing during the initial stage of word recognition (i.e. the first 150 msec).

When listeners are presented with compatible acoustic-phonetic and contextual information they produce the highest number of phonetically accurate and semantically appropriate responses. When they are deprived of one of these information sources however, either through degrading the quality of the bottom-up information or making the top-down information inappropriate, they produce fewer phonetically accurate and fewer semantically appropriate responses.

We suggested that the differences observed between the numbers of phonetically accurate responses to targets which were appropriate or inappropriate to the context was attributable the effects of context during early processing. We noted also that there were more phonetically accurate reports than semantically appropriate responses to targets which were inappropriate to the preceding context. This observation led us to conclude that the difference between the numbers of phonetically accurate responses to appropriate and inappropriate targets was not due to listeners electing to ignore the acoustic-phonetic information. In fact, this result demonstrated that listeners *were* attending to the acoustic-phonetic information.

To summarize, it appears that in the absence of clear acoustic information, people are able to utilize prior contextual information. When clear acoustic-phonetic information is encountered however, it is obligatorily processed although processing of the acoustic-phonetic information is not so efficient if the semantic information is incompatible.

CHAPTER 6:

GENERAL DISCUSSION AND CONCLUSION.

1. SUMMARY OF RESULTS.

It was the intention in this thesis to examine the assumption, explicit in the Cohort Model of lexical access, that initial processing of the first 150 msec of a word proceeds purely in terms of bottom-up processing and that top-down information is not permitted to contribute to processing during this stage. We addressed this issue in two ways: We asked whether aspects of the speech input to the word recognition system were sufficiently unambiguous that bottom-up processing would be able to provide useful constraints on the activation of an initial set of word candidates, and we asked whether listeners were ever able to use *top-down* information within the first 150 msec of word recognition.

The experiments described in Chapters 3 and 4 investigated the extent to which the spoken input itself might be affected by context. This question was pertinent for two reasons: If constraining contexts tend to influence the articulatory realization of speech in a way that makes words indistinguishable on purely acoustic grounds, then the motivation for a purely bottom-up model of lexical access may be weakened, because the acoustic-phonetic information would no longer be able to promote the activation levels of a particular set of word candidates over the activation levels of competitors. Also, if context affects the articulatory realization of utterances, then in our *perception* experiments, we would not have been able to distinguish contextual effects in word recognition from contextual effects on word production i.e. effects of context which we observed during perception may have been due to an artifact introduced by the pronunciations of the stimuli rather than by perception *per se*.

Chapter 3 described two experiments which directly tested the hypothesis that the Voice Onset Time of word-initial stops would be influenced by whether the word was uttered in an appropriate or inappropriate context. More precisely, we asked whether the VOT's of voiced and voiceless stops were more clearly differentiated when the context was *inappropriate*, and less clearly differentiated when other factors (viz. *appropriate* context) could carry meaning. The results of the first experiment suggested that VOT was not significantly affected by context. In light of the findings of Lieberman 1963 and Hunnicutt 1985, these results might appear somewhat surprising.

Shockey (personal communication) notes, however, that the potential for a reduction in the control of VOT in appropriate contexts may have been reduced in the particular environments in which our targets were embedded. All target words in this experiment had some degree of stress and were the semantic focus of the sentence. Also, the appropriate contexts which we used were not as highly constraining as the proverbs which were used, for example, by Lieberman 1963 and Hunnicutt 1985 . In Chapter 3, (Section 12), we discussed possible reasons for the discrepancies between our results and previous findings. We noted, primarily, that the effects observed by Lieberman and Hunnicutt were not ***robust*** effects; targets were presented in isolation, and at a signal to noise ratio of +4 DB (when the noise factor was withdrawn, intelligibility rose to almost 100%). Also, Lieberman and Hunnicutt measured intelligibility of whole word tokens in terms of the numbers of fully homophonous identifications of the target words, whereas we measured the amount of sensory information required to identify targets (cf. Chapter 3).

In Chapter 1 (Section 4) we noted that VOT was only one of the many cues to voicing. The second experiment reported in Chapter 3, however, confirmed that none of the cues which serve to create the voicing distinction were affected by context in a way that was *perceptible*. Listeners' identifications of the voiced and voiceless word-initial stops were similarly accurate whether the stops had originally been uttered in appropriate or inappropriate contexts.

The experiments reported in Chapter 3 demonstrated that the availability of a constraining context did not affect the articulatory realization of an utterance in terms of VOT. We noted, as a consequence of this conclusion, that purely bottom-up processing during lexical access could usefully constrain the initial activation levels of a set of word candidates which would be considered for recognition, and, that we could rule out the possibility, in later experiments, that any affects of context were due to the differential articulatory realizations of the target tokens. The experiments in Chapter 3 thus served as a control for later experiments and also allowed us to assess the value of assuming that lexical access proceeds bottom-up.

As a further control on the materials which we later used in our perception experiments, the next experiment (cf. Chapter 4) examined whether the context of an utterance affected the manner of delivery of that utterance. If the delivery of an utterance had correctly cued listeners to expect an appropriate or an inappropriate target in sentence final position, then we would have had to control for delivery in subsequent experiments. We found, however, that the effects of delivery on listeners were minor compared to the effects of the lexical/semantic context. Nonetheless, delivery information was contributing to the bottom-up information and so we controlled for possible affects of it in our first perception experiment.

The experiments described in Chapter 5 examined the extent to which listeners ever rely exclusively on bottom-up processing during lexical access. We manipulated the qualities of the top-down and bottom-up information available to a listener by varying the contextual constraints and the clarity of the acoustic-phonetic information in the materials. We assumed that responses observed across these qualitatively different conditions would reflect the differential use of top-down and bottom-up information during lexical access. For instance, we argued that by degrading the acoustic quality of the target word and presenting the target word in an appropriate context, any semantically appropriate responses which we observed, might reasonably be attributed to the listener using the available contextual information (because there was *no* clear acoustic-phonetic information on which subjects could base their responses).

In two similar experiments, we noted that when both top-down and clear bottom-up information was available to the listener, more phonetically accurate and semantically appropriate responses were produced during the first 150 msec of a word. When the listener was deprived of one of these information sources, however, significantly lower proportions of phonetically accurate and semantically appropriate responses were offered. Importantly, we observed that when prior contexts were inappropriate for the target words, there were significantly *fewer* correct identifications of the target initial phonemes than when targets were appropriate for the contexts they followed. We argued that this difference was due to the appropriate context facilitating identification of the target initial phonemes. This result held even though, for both the appropriate and inappropriate contexts, the targets provided acoustically clear bottom-up information. Because there were more phonetically accurate responses for inappropriate targets, however, than acoustically *unclear* targets, we ruled out the possibility that listeners were electing to ignore the bottom-up information altogether.

Note that this result also has implications for the validity of our methodology. In Chapter 1 (Section 3), we noted that one of the criticisms raised in connection with the Gating Paradigm was that subjects may be allowed to impose conscious judgements on responses. Although this was theoretically possible, the high numbers of phonetically accurate responses produced to targets which were inappropriate to the preceding context, demonstrated that subjects did not alter their responses to accord with the semantic context. This was further evidenced in the decrease of semantically appropriate responses to inappropriate materials in analyses of responses across all gates (cf. Experiment 4, Chapter 5).

We concluded that the difference in the number of phonetically accurate responses to targets which were appropriate or inappropriate for the preceding context was due to the availability of top-down information in the appropriate conditions. We also noted that the appropriate contextual information in these conditions was being used to facilitate recognition of target words well within the first 150 msec of the target words acoustic onset.

2. IMPLICATIONS FOR MODELS OF WORD RECOGNITION.

According to the latest statement of the Cohort Model (Marslen-Wilson 1987), bottom-up processing of the acoustic-phonetic input is assumed to activate some sort of search set in the mental lexicon. This period of bottom-up processing was assumed to continue for approximately 150 msec (cf. Marslen-Wilson and Tyler 1980). It was only after this fixed period of bottom-up processing, and after the word initial cohort has been activated by the bottom-up information that top-down information was permitted to contribute to the process of word recognition. When a word candidate's internal specifications matched the input, then the activation level of that candidate rose, and, conversely, when the specifications mis-matched with the input, then the activation level fell. Once a candidate's activation level was sufficiently greater than the activation levels of its competitors, it was integrated in to the higher level representation of the utterance, and word recognition occurred.

The majority of data which we presented in Chapter 5 are compatible with this general framework. It is possible to draw parallels between the proportions of responses (phonetically accurate and semantically appropriate) observed over 5 gates and the rise and fall in the activation levels which the theory ascribes to the correct candidates (target words) and their competitors. For example, as the number of phonetically accurate responses to inappropriate targets increased, the number of semantically appropriate responses decreased. Presumably, the simultaneous increase in phonetically accurate responses to the target stimuli and the decrease in semantically appropriate responses reflects the increasing lead in activation level of the (inappropriate) target over its competitors. We might view the decrease, over time, of semantically appropriate responses to inappropriate targets, as a reflection of the falling activation levels of all semantically appropriate candidates. We could thus argue that as the activation levels for semantically appropriate candidates fell, the activation level of the semantically inappropriate, and hence phonetically accurate, candidate rose.

The fact that recognition ultimately occurred for word targets which had initial segments replaced by noise, also supports the concept of activation. We argued that the continual monitoring of the input led, eventually, to successful recognition of these noise materials. With regard to the theory, therefore, we might argue that the continual monitoring of the input against the specifications of the candidate corresponding to the noise target successfully increased the activation level of that candidate, ultimately allowing recognition to occur.

Our results also support the notion, explicit in the Cohort Model, that bottom-up information is given priority. We observed fewer semantically appropriate responses for targets which were inappropriate to the preceding context. Presumably, this was because listeners were producing *phonetically accurate* responses. Furthermore, we also noted that the proportions of phonetically accurate responses were greater for materials which had *clear* acoustic-phonetic information than for materials which had unclear bottom-up information. This effect held even when the contextual information preceding the acoustically unclear targets was exactly the same as that preceding the clear targets. If we accept that bottom-up information is prioritized, then this would also account for why listeners sometimes failed to use the semantic information even though it was available.

Note that these results also bear directly on alternative processing systems. For example, Morton's Logogen Model (Morton, 1969 see Chapter 2 Section 2.3) predicts that context can be used to restrict the shape and size of the word initial cohort. To the extent that the present data demonstrate that bottom-up information is given priority during lexical access, they are incompatible with the Logogen model which prescribes that top-down information such as sentential context can be used to pre-select a sub-set of the lexicon.

In that our results support the notion of bottom-up priority during the lexical access, it might be argued that they also support the notion explicit in the Serial Search Model (Forster 1976; see Chapter 2 Section 2.2) that bottom-up processing is an autonomous process. Note however, that what our data also clearly demonstrated, was that top-down information *could* be used during the first 150 msec of processing. This violates the assumptions of autonomy in Forster's model and also clearly violates the assumption, explicit in the Cohort Model, that this period of initial processing proceeds in terms of bottom-up analysis alone.

Whilst the data we presented in Chapter 5 do not conflict with the motivation behind the basic framework of the Cohort Model, our observation that top down information affects responses within the first 150 msec of a word's acoustic onset requires us to re-evaluate the specification that top-down information is prohibited from contributing during this phase.

It would appear that there need be no fixed period of bottom-up processing. Clearly bottom-up information is given priority, but the available data suggest that the time allocated to bottom-up processing is context dependent: the more constraining the context, the earlier contextual information can be brought to bear. Consequently, less time needs to be allocated to processing the bottom-up information if the context is constraining. Whilst bottom-up information is given priority, we have presented evidence demonstrating that top-down information is *not* prohibited from contributing to processing during the first 150 msec. Neither the current version of the Cohort Model nor Forster's Serial search model allow for such flexibility.

A model with which our results are more compatible is the TRACE model (McClelland and Elman 1986). Broadly, this model states that any level of analysis can influence the state of any other level of analysis. Under this framework words are represented as nodes at various levels of linguistic abstraction, and all nodes are inter-connected. Bottom-up input to the system activates a set of feature nodes which most closely correspond to the input. Feature nodes activate phoneme nodes which, in turn, activate lexical nodes. The states of higher level nodes however, are also permitted to affect the activation of lower level nodes. For instance, if the input to the system is "dog", the lexical node "dog" feeds back information to the phoneme node "d" to increase its activation level. This arrangement of feed-back and feed-forward maximizes the system's chance of finding the correct word candidate at the earliest opportunity.

Note that in the current implementation of TRACE, effects of context are restricted to effects of *lexical* context and describe the situation in which lexical nodes feedback to increase the activation of lower level nodes. Although context was defined more broadly for the purposes of our experiments (including levels higher than simply the lexical level, for example syntactic, semantic and pragmatic), it should be theoretically possible to extend the scope of TRACE to incorporate these higher levels.

The present data can be interpreted as supporting TRACE's concepts of feedback and feedforward. In Chapter 5 (Section 5) we noted that there were more phonetically accurate responses than there were semantically appropriate responses for targets which were predictable from context. A possible explanation for this is that feedback from the lexical nodes had sufficiently activated the phoneme nodes for fragmented responses to be produced confidently. Furthermore, that recognition was possible for targets which had their initial segments replaced with noise supports the notion, discussed below, that, in TRACE, words can become activated from non-initial segments and ultimately recognized from the activation provided by these feedback and feedforward mechanisms.

In light of the poor acoustic quality of continuous speech (see Chapter 2, Section 4) such flexibility of the temporal courses of the contributions of top-down and bottom-up information would serve to create a highly efficient processing system. As the contextual constraints available in an utterance vary, so too will the amount of sensory information needed to recognize a word.

We thus conclude that the research conducted for this thesis yielded data which are compatible with the TRACE model of lexical access. Whilst these data do not contradict the *architecture* of the Cohort Model of lexical access, as stated by {.Marslen-Wilson parallel 1987.}, they do point to lifting the restrictions which the model imposes on the time course of the relative contribution of top-down and bottom-up information during the first 150 msec of word recognition. It would appear that there need be no *fixed* period of bottom-up processing, rather it seems that bottom-up information cycles through the system at some minimal (possibly even unobservable) level of discription.

3. FURTHER RESEARCH.

Although the research reported in this thesis has implications for the time course of the relative contributions for top-down and bottom-up information during word recognition, it offers no insights into how the respective contributions might be *represented,* nor into the *scope* of the top-down information source.

We consider first, processing of bottom-up information. Leaving aside the formidable question of *what* constitutes bottom-up input to the word recognition system (features, phonemes, syllables etc), the issue concerning how much noise the bottom-up processor might tolerate poses many problems. As it stands, for example, we have no explanation for how the targets in the noise conditions of Experiments 4 and 5 were initially activated. We observed earlier (see Chapter 5, footnote 5) that coarticulatory information in the segments immediately preceding the noise segments was insufficient to cue the noise-replaced target-initial phoneme. If we assume, in line with the Cohort Model, that word candidates were accessed from the noise-replaced initial segments, then *all* words in the lexicon should have been activated equally, as potential candidates for recognition.

Allowing the initial activation of all word candidates considerably weakens the notion of obligatory bottom-up processing. If all candidates are equally activated, then does it matter whether or not initial processing proceeds bottom-up? The question of which candidates should be active in the initial cohort is quite considerable. As McClelland and Elman 1986 note, it is not easy to liberalize the criterion either for entry into the initial cohort, or for increasing the activation levels of some candidates rather than others. On some occasions we will need to rule out certain items that mismatch the acoustic-phonetic information; on other occasions, however, we will need to activate candidates which mismatch on one or two dimensions. McClelland and Elman consider the example PLEASANT/BLACELET: In the first case, we need to activate PLEASANT over PRESENT, so the slight difference between the [l] and [r] must be sufficient to increase the activation level of only the PLEASANT candidate; in the second case, however, BRACELET must be activated because it provides the best fit with the input. In this second case, therefore, the difference between [l] and [r] must be "overlooked" to the extent that BRACELET can be activated.

Using the paradigms presented in Chapter 5, it might be possible to investigate the extent to which the bottom-up processor is able to discriminate between competing word candidates (and noisy candidates). It would be possible, for example, to replace the voiced/voiceless minimal pairs with minimal pairs containing increasing numbers of different features between the initial phonemes. The proportions of phonetically accurate and semantically appropriate responses, and the distribution of these responses over time may reveal certain properties of the bottom-up processor. We might expect, for instance, that whilst the minimal pairs differing on only one feature (voicing) yielded no difference at the early gates in the number of phonetically accurate responses, minimal pairs differing on more features (e.g. BEES/KNEES) might yield greater differences at the initial gate. It would also be interesting to test whether some of these featural differences resulted in the activation of all candidates in the way that we might suppose the noise-replaced targets in Experiment 5 did. Presumably such a result would reflect the differential discrimination of the bottom-up processor: whilst the voicing distinction was not sufficient to exclude the activation of a voiced or voiceless initial candidate, minimal pairs which differed on several features may yield results which suggest that only one of the candidates (the one specified by the acoustic-phonetic input) is activated. In order for us to be able to trace the activation levels in this way, i.e. trace the rates of increase and decrease in the activation levels of the correct candidate and competitors activated by the bottom-up input, it would, perhaps, be necessary to present very small increments of the acoustic-phonetic stimuli.

TRACE already provides us however, with a possible explanation as to how words with noise-replaced initial segments might be recognized. Under this framework bottom-up input activates nodes at various levels of linguistic description. Because all nodes at all levels are interconnected, feature nodes which are activated initially will (ultimately) activate all lexical nodes which contain the feature/phoneme specified in the bottom-up input regardless of that feature/phoneme postion within the word. Clear acoustic information which occurs later in a word can therefore activate the target lexical node as effectively as clear acoustic information which occurs in earlier segments of the word. It would thus appear that the bottom-up information needs to be represented at some minimal level such as the feature level and continually taken up. Under the Cohort framework, lifting the restriction on bottom-up processing for a fixed period of time also ammounts to this cycling of some minimal level of bottom-up information.

Now consider the issue of how top-down information might contribute to processing. Aside from the issue of which *level* of top-down information might be influencing processing (e.g. lexical, syntactic, semantic, pragmatic) which we did not explicitly deal with in this thesis, we are also left questioning the mechanism underlying the top-down contribution: How did the top-down information serve, within the first 150 msec, to effect earlier perception of the target materials? Also, how did top-down information contribute to the recognition of words which had noise-replaced initial segments? Did top-down information initially constrain the shape and size of the initial cohort during access? Is there, for example, some global mechanism which accepts inputs from both bottom-up and top-down sources at *all* stages in processing? When top-down and bottom-up information is available, perception is rapid. When top-down information is not available however, then the processing of the bottom-up information must continue until sufficient evidence has accrued for perception to occur. An alternative account is that top-down information was used to increase the activation levels of the target words during the first gates, subsequent to the bottom-up information having actually activated the word candidates. Or, as a further possibility, did the higher level representation, within the first 150 msec, select amongst the more active word candidates which were competing for recognition? For instance, if contextual information could serve to increase confidence in selecting one word candidate over another, then the absolute difference in activation levels between highly active competitors need not be as great as when there is no contextual information. In other words, the criteria for selection is context dependent. The more highly constaining the context is, the less confidence is needed to select between highly active candidates. As we noted above, the data presented in Chapter 5 does not permit us to distinguish between these three alternatives.

4. CONCLUSION.

The research reported in this thesis aimed to examine the relative contributions of top-down and bottom-up information during lexical access. We have presented evidence which demonstrated that aspects of the articulatory realization of speech did not vary as a function of context in a manner which would weaken the validity of assuming the lexical access proceeds by reference only to bottom-up information, as specified by the Cohort Theory. Furthermore, we presented data which illustrated effects of top-down information during the initial stage of word recognition. Further research is required to examine how both top-down and bottom-up information sources might be represented and the scope of their contribution . These outstanding issues do not, however, detract from the conclusions drawn here. The significance of the present research is that, in mapping the time course of the contribution of top-down information, we have demonstrated that there need be no fixed allocation of time for bottom-up processing during the first 150 msec of word recognition and, that, contrary to the predictions of the Cohort Theory, top-down information may contribute to processing during this period. As the contextual constraints vary, so too will the amount of time given to bottom-up processing.

BIBLIOGRAPHY.

Abramson, A. S. and Lisker, L. (1970) Discriminability along the voicing continuum: cross-language tests.. In *6th International congress of phonetic sciences*, Prague, 1970, pp569 - 573. Conference was in 1967 (Academia, Prague).

Bard, E. G. and Anderson, A. (1983) The unintelligibility of speech to children. *Journal of Child Language*, **10**, 265-292.

Bard, E. G., Shillcock, R. and Altmann, G. T. M. (1988) The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context.. *Perception and Psychophysics*. forthcoming.

Berinstein, A. (1979) A cross-linguistic study on the perception and production of stress.. In *UCLA Working papers in Phonetics*, Volume 47, pp1 - 59.

Bernstein, L. E. (1983) Perceptual development for labeling words varying in voice onset time and fundamental frequency.. *Journal of Phonetics*, **11**, 383 - 393.

Bever, T. G., Garrett, M. F. and Hurtig, R. (1973) The interaction of perceptual processes and ambiguous sentences. *Memory and Cognition*, **1**, 277-286.

Blumenthal, A. L. (1967) Prompted recall of sentences. *Journal of Verbal Learning and Verbal Behavior*, **6**, 674 - 676.

Clark, H. H. (1973) The Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research. *Journal of Verbal Learning and Verbal Behavior*, **12**, 335 - 359.

Cole, R. A. and Jakimik, J. (1978) Understanding speech: How words are heard.. In Underwood, G. (ed.) *Strategies of information processing.* London: Academic Press.

Cole, R. A. and Jakimik, J. (1980) A Model of Speech Perception. Chapter 6 in Cole, R. (ed.) *Perception and production of fluent speech,* pp133-163. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Connine, C. M. (1987) Constraints on Interactive Processes in Auditory Word Recognition: The Role of Sentence Context. *Journal of Memory and Language,* **26,** 527 - 538.

Connine, C. M. and Clifton, C. (1987) Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* **13,** 291 - 299.

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

Cutler, A. (1976) Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics,* **20,** 55-60.

Cutler, A. and Foss, D. J. (1977) On the role of sentence stress in sentence processing. *Language and Speech,* **20,** 1 - 10.

Cutler, A. and Fodor, J. A. (1979) Semantic focus and sentence comprehension. *Cognition,* **7,** 49 - 59.

Cutler, A. and Norris, D. (1988) The role of strong syllables in segmentation for lexical access.. *Journal of Experimental Psychology (Human Perception and Performance,* **14,** 113 - 121.

Darwin, C. J. (1975) On the dynamic use of prosody in speech perception.. In Cohen, A. and Nooteboom, S. (eds.) *Structure and process in speech perception..* Heidelberg: Springer.

Fodor, J. A. and Garrett, M. (1967) Some Syntactic Determinants of Sentential Complexity. *Perception and Psychophysics.*

Fodor, J. and Garrett, M. (1968) Some syntactic determinants of sentential complexity II: verb structure. *Perception and Psychophysics*, **3**, 453 - 461.

Fodor, J. A., Bever, T. G. and Garrett, M. F. (1974) *The Psychology of Language.* New York: McGraw Hill.

Forster, K. and Olbrei, I. (1973) Semantic heuristics and syntactic analysis. *Cognition*, **2**, 319 - 347.

Forster, K. I. (1974) Linguistic Structure and Sentence Production. In Cherry, C. (ed.) *Pragmatic Aspects of Human Communication.* Dordrecht: D. Reidel.

Forster, K. (1976) Accessing the mental lexicon. In Wales, R. J. and Walker, E. (eds.) *New Approaches to Language Mechanisms.* Amsterdam: North Holland.

Forster, K. (1979) Levels of processing and the structure of the language processor. In Cooper, W. and Walker, E. (eds.) *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Forster, K. (1981) Frequency blocking and lexical access. *Journal of Verbal Learning and Verbal Behavior*, **20**, 190-208.

Foss, D. and Blank, M. (1980) Identifying the speech codes. *Cognitive Psychology*, **12**, 1-31.

Fowler, C. and Housum, J. (1987) Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, **26**, 489-905.

Frauenfelder, U. and Tyler, L. (1987) The process of spoken word recognition: An introduction. *Cognition*, **25**, 1-200.

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

Games, S. and Bond, Z. S. (1976) The relationship between semantic expectation and acoustic information. In *Akten der 3 internationalen phonologie-taung*, Wein, Sept, 1976, pp285 - 293.

Grice, H. P. (1978) Further Notes on Logic and Conversation. In Cole, P. (ed.) *Syntax and Semantics*, Volume 9, pp113-127. New York: Academic Press.

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

Grosjean, F. (1985) The recognition of words after their acoustic offset: evidence and implications. *Perception and Psychophysics*, **28**, 299-310.

Grosjean, F. and Gee, J. P. (1987) Prosodic structure and spoken word recognition. *Cognition*, **25**, 135-156.

Hunnicutt, S. (1985) Intelligibility versus redundancy - conditions of dependency.. *Language and Speech*, **28**, 47 - 56.

Jarvella, R. J. (1971) Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, **10**, 409-416.

Klatt, D. H. (1977) Review of the ARPA speech understanding project. *Journal of the Acoustical Society of America*, **62**, 1345 - 1366.

Kucera, H. and Francis, W. N. (1967) *Computational Analysis of Present-Day American English*. Providence, Rhode Island: Brown University Press.

Liberman, Delattre and Cooper (1952) The role of selected stimulus variables in the perception of unvoiced stop consonants. *American Journal of Psychology*, **65**, 497 - 516.

Liberman, Harris, Hoffman and Griffith (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **54**, 338 - 368.

Liberman, DeLattre and Cooper (1958) Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, **1**, 153 - 167.

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

Lieberman, P. (1963) Some effects of the semantic and grammatical context on the production and perception of speech. *Language & Speech*, **6**, 172-175.

Lisker (1978) In Qualified Defense of VOT.. *Language and Speech*, **21**, 375 - 383.

Locke, J. (1968) Questionable assumptions underlying articulation research.. *Journal of Speech and Hearing Disorders*, **33**, 112 - 116.

Marslen-Wilson, W. D. (1973) Linguistic Structure and Speech Shadowing at Very Short Latencies. *Nature*, **244**, 522-523.

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

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

Marslen-Wilson, W. and Tyler, L. K. (1980) The Temporal Structure of Spoken Language Understanding. *Cognition*, **8**, 1-74.

Marslen-Wilson, W. (1987) Parallel processing in spoken word recognition. *Cognition*, **25**, 71-102.

McAllister, J. M. (1988) The use of context in auditory word recognition.. *Perception and Psychophysics*. forthcoming.

McClelland, J. and Elman, J. (1986) Interactive Processes in Speech Perception: The TRACE Model. Chapter 15 in *Parallel Distributed Processing*, Volume 2: *Psychological and Biological Models*. Cambridge, Mass.: MIT Press. a condensed version of the Cognitive Psychology paper.

McKoon, G. and Ratcliff, R. (1980) The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, **19**, 668 - 682.

Miller, G. A. and Isard, S. (1963) Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, **2**, 217-228.

Miller, J. L., Green, K. and Schermer, T. M. (1984) A distinction between the effects of sentential speaking rate and semantic congruity on word identification.. *Perception and Psychophysics*, **36**, 329 - 337.

Morton, J. (1969) The Interaction of Information in Word Recognition. *Psychological Review*, **76**, 165-178.

Nooteboom, S. G. (1981) Lexical retrieval from fragments of spoken words - beginnings vs endings. *Journal of Phonetics*, **9**, 407 - 424.

Norris, D. G. (1980) Serial and Interactive Models of Comprehension. PhD Thesis, Experimental Psychology, University of Sussex.

Norris, D. (1982) Reply to Marslen-Wilson and Tyler (1980). *Cognition*, **11**, 97-101.

Pollack and Pickett (1963) The intelligibility of excerpts from conversation. *Language and Speech*, **6**, 165-171.

Pollack, I. and Pickett, J. M. (1964) Intelligibility of excerpts from fluent speech: auditory vs structural context.. *Journal of Verbal Learning and Verbal Behavior*, **3**, 79 - 84.

Reddy, R. (1976) Speech recognition by machine: A review. In *Proceedings of the IEEE*, 1976, pp501 - 531.

Repp, B. H. (1982) Phonetic trading relations and context effects: new evidence for a phonetic mode of perception.. *Psychological Bulletin*, **92**, 81 - 110.

Salasoo, A. and Pisoni, D. B. (1985) Interaction of knowledge sources in spoken word identification. *JMEMLANG*, **24**, 210 - 231.

Samuel, A. G. (1981) The role of bottom-up confirmation in the phoneme restoration illusion.. *Journal of Experimental Psychology (Human Perception and Performance)*, **7**, 1124-1131.

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

Samuel, A. G. (1988) Using perceptual restoration effects to explore the architecture of perception. Paper delivered at the Workshop on Cognitive Models of Speech Processing, Sperlonga..

Taft, L. (1986) Lexical stress and lexical segmentation. (Unpublished).

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

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

Tyler, L. K. and Wessels, J. (1985) Is gating an on-line task? Evidence from naming latency data. *Perception and Psychophysics*, **38**, 217 - 222.

Wales, R. and Toner, H. (1979) Intonation and Ambiguity. Chapter 5 in Cooper, W. and Walker, E. (eds.) *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, pp135-158. Hillsdale, N.J.: Lawrence Erlbaum Associates.

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

Warren, R. M. and Obusek, C. J. (1971) Speech perception and phonemic restorations. *Perception and Psychophysics*, **9**, 358 - 363.

Wintz, H., Riviere, C. L. and Herriman (1975) Variations in VOT for English initial stops. *Journal of Phonetics*, **3**, 41 - 52.

APPENDIX A.

5. THE CLOZE TEST.

In order to tighten the definition of what would constitute an "appropriate" context, a Cloze Test was run. A small group of people were given lists of minimal pairs which were matched for frequency (within pairs) and were asked to produce sentences which predicted one member of the minimal pair (but not both) in sentence final position and further sentences which predicted the other member of the minimal pair (without predicting the first member). It was noted that sets of sentences should have the same syntactic structure, expected intonation pattern and rhythmn.

Written versions of each sentence were made whereby the contextual strings appeared but the sentence final word (i.e. one of the minimal pair target words) was missing.

The Cloze Test involved a group of 20 (and subsequently a group of 17) Subjects reading the sentences and writing down what they thought would have been the most likely final word. Responses were sorted into 4 categories: a) Subjects guessed the missing target word; b) Subjects guessed a synonym; c) Subjects guessed a related word; d) Subjects guessed a word unrelated to the target. Only those sentences which had yielded responses from the first 2 categories (a and b) were subsequently utilized as "appropriate" materials in our experiments. A further criterion was that the scores to the pair of sentences corresponding a given minimal pair should not differ by more than 1. The 24 sentences selected yielded a mean score of 1.6 while the mean difference for the between pairs scores was 0.3. The sentences actually used for the different experiments are presented in later appendices. Results of the cloze test are given in the table below. The numbers to the right hand side of each sentence represents the average score across all subjects for the word printed in upper case.

CONTEXT	CLOZE TEST SD
I fell and landed with a BUMP	1.75
I inflated the tyre with a PUMP	1
The teacher said the pupil was the best in the CLASS	1.1
The barman poured the whisky into the GLASS	1.1
The examiner realized the candidate was very TENSE	3.6
The firemen realized the smoke was very DENSE	2.6
It's lovely having the fire burning in the GRATE	2.3
Before he could light the fire, he put the logs in the GRATE	1.5
It's easier to carry bottles of wine in the CRATE	2.1
Outside the station the policeman arrested the DRUNK	3.5
Outside the station the porters collected the TRUNK	3.6
I needed to take all my luggage in a TRUNK	3.4
At the sound of danger the kangaroo started to BOUND	2.8
At the sound of danger the kangaroo began to leap and BOUND	2.6
It was across the meadows that the Dulux dog started to BOUND	2.3
At the sound of steps behind me, my heart began to thump and POUND	1.4

At the sound of steps behind me my heart started to POUND	1.9
To transport the wild animal he had to use a CAGE	2.3
To take an accurate measurement he had to use a GAUGE	2.3
When the Queen did her walkabout she only took one GUARD	2.7
Coming away from the prison, we were searched by a few GUARDS	1.4
It was his birthday but the postman only brought one CARD	1.9
On his birthday, the postman brought John one present and a few CARDS	1
You can tell the cake won't stick to the tin because the tin is GREASED	1.9
You can tell he hasn't ironed his shirt because it's so CREASED	1.6
To keep her hair in place the waitress wore a CAP	3
If you have a tooth out it sometimes leaves a GAP	1.5
When my tooth fell out it left a GAP	1.6
When I went to school I wore a CAP	3.4
If you play baseball, it's trendy to wear a CAP	2
After collecting the rubbish I took it to the TIP	2.45
Arriving at the beach I went in for a DIP	2.3
We stood around the pond throwing bread at some DUCKS	1.2

To make his shirt tighter, we sewed in some TUCKS	**2.3**
Cleaning the telephone the maid polished the DIAL	**3.3**
Cleaning the bathroom the maid polished the TILE	**3.8**
The Matador angered the crowd by refusing to kill the BULL	**1**
The climber tested the rope by giving it a PULL	**1.9**
Before shipping the antique he put it in a CRATE	**1.8**
After brining in the firewood he put it in the GRATE	**2.7**
The stag was alone and there was no sign of the DOE	**2.8**
The naturalist could see the stag, but there was no sign of the DOE	**1.5**
His car had broken down so we got ours and gave him a TOW	**3.4**
His car broke down so we got our rope and gave him a TOW	**1.4**
You show little Bill a cow and I'll show him a BULL	**2.9**
You give the trolley a push and I'll give it a PULL	**2.3**
My wedding ring is made of GOLD	**1.6**
As usual in winter I've caught a COLD	**1.1**
He'd lost his job and so signed on the DOLE	**1.2**
She crossed the bridge and so paid the TOLL	**1.5**
John hates hornets, wasps and BEES	**1.6**

I like potatoes, carrots and PEAS	2.2
It's important to serve champagne in the right kind of GLASS	1.1
In order to learn quickly one has to sign up for the right kind of CLASS	1.4
Although James likes a few drinks, I wouldn't call him a DRUNK	1.6
Although I have large suitcases, I generally use a TRUNK	3.5
The pollen was stuck to the legs of the hairy BEES	1.1
The gravy was poured over the potatoes, the carrots and the PEAS	2.1
Dusting the telephone, the maid pushed round the DIAL	2.5
The 999 call is a number which often needs to be DIALED	2.2
The bathroom walls are crooked so it's difficult for them to be TILED	2.8
Fixing the roof the workman slipped on the TILE	2.2
Chiseling at the bathroom wall, the man accidently cracked the TILE	2.2
The bathroom can't be papered or painted, it needs to be TILED	1.8
The most dangerous part of an elephant is the TUSK	2.1
The children were horrified to discover that ivory is made of TUSK	1.7
The most mystical time of evening is the DUSK	2
To be back before it got dark, the lovers parted at the fall of DUSK	1.7

On the dodgems, it was the cars that we managed not to get BUMPED	2.1
On the bikeride, it was the tyres that we forgot to get PUMPED	2.5
Although the forest was bright the foilage was very DENSE	3.1
Although the opponents smiled at one another the moment was very TENSE	1.1
The girl who shaved her head did it as a BET	3.1
The man who ate the poison did it as a BET	2.8
The man who found the dog kept it as a PET	1.1
Although John likes casinos, he's never placed a BET	1.2
Although John likes animals, he's never kept a PET	1.8
In order to learn Karate he went on a kind of COURSE	1.7
The hills over there are yellow as they're covered in a kind of GORSE	2.9
The dentist starting to drill was the moment I'd been DREADING	1.2
The stains on my feet were from the grapes that I'd been TREADING	1.7
We know you won all those prizes but there's no need to BOAST	1.3
I know you have some stamps so I'll give you this letter to POST	1.1
It was during the service that the groom kissed the BRIDE	1
He wanted to say sorry but what stopped him was his PRIDE	1.9
The atmosphere in the exam was so TENSE	1.8

THe mixture for the cake was so DENSE	3.8
Although I'd spun-dry the shirt, it continued to DRIP	1.6
Although they'd fixed the stairs, the step caused me to TRIP	1.9
The groom at the alter turned round to see his BRIDE	1.9
The winner of the race couldn't hide his PRIDE	2.1
The dog wanted to be stroked but the child only knew how to PAT	2.6
He'd done really well and it was his back that the teacher tried to PAT	1.6
It was the dog's head that the girl tried to PAT	2
The bowler was ready and it was John's turn to BAT	1.7
The plumber came to fix the tap that had been DRIPPING	1.4
The junkie had been scared when he had when had been TRIPPING	2.4
When people walk their dogs here I always watch where I TREAD	2
I don't mind the fillings it's the injections that I DREAD	2.8
John watched the baker knead his DOUGH	1.5
John said Fred had stubbed his TOE	1.1

APPENDIX B.

The materials for Experiment 1 are presented below.

John hates hornets, wasps and BEES

John hates hornets, wasps and PEAS

I like potatoes, carrots and PEAS

I like potatoes, carrots and BEES

Although John likes casinos, he's never placed a BET

Although John likes casinos, he's never placed a PET

Although John likes animals, he's never kept a PET

Although John likes animals, he's never kept a BET

I fell and landed with a BUMP

I fell and landed with a PUMP

I inflated the tyre with a PUMP

I inflated the tyre with a BUMP

To transport the wild animal, he had to use a CAGE

To transport the wild animal, he had to use a GAUGE

To take an accurate measurement, he had to use a GAUGE

To take an accurate measurement, he had to use a CAGE

Coming away from the prison, we were searched by a few GUARDS

Coming away from the prison, we were searched by few CARDS

On his birthday, the postman brought John one present and a few CARDS

On his birthday, the postman brought John one present and a few GUARDS

You can tell she didn't want the cake mixture to stick to the tin because the tin is GREASED

You can tell she didn't want the cake mixture to stick to the tin because the tin is CREASED

You can tell he hasn't ironed his shirt because it's so CREASED

You can tell he hasn't ironed his shirt because it's so GREASED

It was across the meadows that the Dulux dog started to BOUND

It was across the meadows that the Dulux dog started to POUND

At the sound of steps behind me my heart started to POUND

At the sound of steps behind me my heart started to BOUND

On the dodgems, it was the cars that we managed not to get BUMPED

On the dodgems, it was the cars that we managed no to get PUMPED

On the bikeride, it was the tyres that we forgot to get PUMPED

On the bikeride, it was the tyres that we forgot to get BUMPED

We know you won all those prizes, but there's no need to BOAST

We know you won all those prizes, but there's no need to POST

I know you have some stamps, so I'll give you this letter to POST

I know you have some stamps, so I'll give you this letter to BOAST

The groom at the alter turned 'round to see his BRIDE

The groom at the alter turned 'round to see his PRIDE

The winner of the race couldn't hide his PRIDE

The winner of the race couldn't hide his BRIDE

It was the dog's head that the girl tried to PAT

It was the dog's head that the girl tried to BAT

The bowler was ready and it was John's turn to BAT

The bowler was ready and it was John's turn to PAT

After collecting the rubbish, I took it to the TIP

After collecting the rubbish, I took it to the DIP

Arriving at the beach, I went in for a DIP

Arriving at the beach, I went in for a TIP

The naturalist could see the stag, but there was no sign of the DOE

The naturalist could see the stag, but there was no sign of the TOW

His car broke down so we got our rope and gave him a TOW

His car broke down so we got our rope and gave him a DOE

He'd lost his job and so signed on the DOLE

He'd lost his job and so signed on the TOLL

She crossed the bridge and so paid the TOLL

She crossed the bridge and so paid the DOLE

The 999 call is a number which often needs to be DIALED

The 999 call is a number which often needs to be TILED

The bathroom can't be papered or painted, it needs to be TILED

The bathroom can't be papered or painted, it needs to be DIALED

The children were horrified to discover that ivory is made of TUSK

The children were horrified to discover that ivory is made of DUSK

To be back before it got dark, the lovers parted at the fall of DUSK

To be back before it got dark, the lovers parted at the fall of TUSK

The dentist starting to drill was the moment I'd been DREADING

The dentist starting to drill was the moment I'd been TREADING

The stains on my feet were from the grapes that I'd been TREADING

The stains on my feet were from the grapes that I'd been DREADING

Although I'd spun-dry the shirt, it continued to DRIP

Although I'd spun-dry the shirt, it continued to TRIP

Although they'd fixed the stairs, the step caused me to TRIP

Although they'd fixed the stairs, the step caused me to DRIP

John watched the baker knead his DOUGH

John watched the baker knead his TOE

John said Fred had stubbed his TOE

John said Fred had stubbed his DOUGH

The teacher said the pupil was the best in the CLASS

The teacher said the pupil was the best in the GLASS

The barman poured the whisky into the GLASS

The barman poured the whisky into the CLASS

Before he could light the fire, he put the logs in the GRATE

Before he could light the fire, he put the logs in the CRATE

Before shipping the antique, he put it in the CRATE

Before shipping the antique, he put it in the GRATE

If you have a tooth out, it sometimes leaves a GAP

If you have a tooth out, it sometimes leaves a CAP

If you play baseball, it's trendy to wear a CAP

If you play baseball, it's trendy to wear a GAP

My wedding ring is made of GOLD

My wedding ring is made of COLD

As usual in winter I've caught a COLD

As usual in winter I've caught a GOLD

APPENDIX C.

6. BASE MATERIALS.

The materials presented in this section were used to generate the appropriate and innappropriate versions of target tokens following neutral carrier phrases for Experiment 2. The materials and design used in Experiment 2 are presented in the following section.

To transport the wild animal, he had to use a CAGE

To take an accurate measurement, he had to use a CAGE

If you play baseball, it's trendy to wear a CAP

If you have a tooth out, it sometimes leaves a CAP

The teacher said the pupil was the best in the CLASS

The barman poured the whisky into the CLASS

I like potatoes, carrots and PEAS

John hates hornets, wasps and PEAS

At the sound of steps behind me my heart started to POUND

It was across the meadows that the Dulux dog started to POUND

The climber tested the rope by giving it a PULL

The Matador angered the crowd by refusing to kill the PULL

John said Fred had stubbed his TOE

John watched the baker knead his TOE

She crossed the bridge and so paid the TOLL

He'd lost his job and so signed on the TOLL

The children were horrified to discover that ivory is made of TUSK

To be back before it got dark, the lovers parted at the fall of TUSK

7. ACTUAL MATERIALS.

The following carrier phrases and targets constituted the materials for Experiment 2.

John said he'd learned the Russian for CLASS

John said he'd learned the Russian for PULL

John said he'd learned the Russian for TOE

The splodge on the typescript completely obscured the word CAGE

The splodge on the typescript completely obscured the word POUND

The splodge on the typescript completely obscured the word TUSK

The word at the end of this sentence is CAP

The word at the end of this sentence is PEAS

The word at the end of this sentence is TOLL

The following filler target words also appeared following these carrier phrases: SPEND; SCRATCH; FRIENDS; FLIRT; RUSH; LUNCH; SNAKE; SPRINT; LATE; SUPPER; WASH; CHESS; SAVED; STEEL; FRIDGE; FISH; FLOOR; RIVER; FIRE; WET; SWEETS; WALKING; RUN; WINE; SCARED; SLEEVES; FERNS; FRUIT; SHINING; HELP; FAST; WIPED; JELLY; SHOP; WINDOW; INTERFERE.

Cells illustrating the design of Experiment 2.

Context Manipulation	Appr Splice	Appr Manual	Inapp Splice	Inapp Manual	Neutral
High VOT	1	1	1	1	2
Low VOT	1	1	1	1	2

APPENDIX D.

The following appropriate and inappropriate contexts were used in Experiment 3. Apart from the random presentation of appropriate and inappropriate versions within sets of materials, the order of presentation is maintained below. Although target words are presented here for convenience, they were not heard by the subjects.

8. SPEAKER 1.

I fell and landed with a BUMP

I fell and landed with a PUMP

I inflated the tyre with a PUMP

I inflated the tyre with a BUMP

John hates hornets, wasps and BEES

John hates hornets, wasps and PEAS

I like potatoes, carrots and PEAS

I like potatoes, carrots and BEES

Before he could light the fire, he put the logs in the GRATE

Before he could light the fire, he put the logs in the CRATE

Before shipping the antique, he put it in the CRATE

Before shipping the antique, he put it in the GRATE

Coming away from the prison, we were searched by a few GUARDS

Coming away from the prison, we were searched by few CARDS

On his birthday, the postman brought John one present and a few CARDS

On his birthday, the postman brought John one present and a few GUARDS

The dentist starting to drill was the moment I'd been DREADING

The dentist starting to drill was the moment I'd been TREADING

The stains on my feet were from the grapes that I'd been TREADING

The stains on my feet were from the grapes that I'd been DREADING

After collecting the rubbish, I took it to the TIP

After collecting the rubbish, I took it to the DIP

Arriving at the beach, I went in for a DIP

Arriving at the beach, I went in for a TIP

9. SPEAKER 2.

The Matador angered the crowd by refusing to kill the BULL

The Matador angered the crowd by refusing to kill the PULL

The climber tested the rope by giving it a PULL

The climer tested the rope by giving it a BULL

We know you won all those prizes, but there's no need to BOAST

We know you won all those prizes, but there's no need to POST

I know you have some stamps, so I'll give you this letter to POST

I know you have some stamps, so I'll give you this letter to BOAST

To transport the wild animal, he had to use a CAGE

To transport the wild animal, he had to use a GAUGE

To take an accurate measurement, he had to use a GAUGE

To take an accurate measurement, he had to use a CAGE

My wedding ring is made of GOLD

My wedding ring is made of COLD

As usual in winter I've caught a COLD

As usual in winter I've caught a GOLD

The 999 call is a number which often needs to be DIALED

The 999 call is a number which often needs to be TILED

The bathroom can't be papered or painted, it needs to be TILED

The bathroom can't be papered or painted, it needs to be DIALED

He'd lost his job and so signed on the DOLE

He'd lost his job and so signed on the TOLL

She crossed the bridge and so paid the TOLL

She crossed the bridge and so paid the DOLE

10. SPEAKER 3.

Although John likes casinos, he's never placed a BET

Although John likes casinos, he's never placed a PET

Although John likes animals, he's never kept a PET

Although John likes animals, he's never kept a BET

It was across the meadows that the Dulux dog started to BOUND

It was across the meadows that the Dulux dog started to POUND

At the sound of steps behind me my heart started to POUND

At the sound of steps behind me my heart started to BOUND

If you have a tooth out, it sometimes leaves a GAP

If you have a tooth out, it sometimes leaves a CAP

If you play baseball, it's trendy to wear a CAP

If you play baseball, it's trendy to wear a GAP

Although I'd spun-dry the shirt, it continued to DRIP

Although I'd spun-dry the shirt, it continued to TRIP

Although they'd fixed the stairs, the step caused me to TRIP

Although they'd fixed the stairs, the step caused me to DRIP

The naturalist could see the stag, but there was no sign of the DOE

The naturalist could see the stag, but there was no sign of the TOW

His car broke down so we got our rope and gave him a TOW

His car broke down so we got our rope and gave him a DOE

Although I'd spun-dry the shirt, it continued to DRIP

Although I'd spun-dry the shirt, it continued to TRIP

Although they'd fixed the stairs, the step caused me to TRIP

Although they'd fixed the stairs, the step caused me to DRIP

APPENDIX E.

11. NOISE MATERIALS.

The following materials were used to generate the noise materials described in Experiment 4.

John hates hornets, wasps and BEES

I like potatoes, carrots and PEAS

Although John likes casinos, he's never placed a BET

Although John likes animals, he's never kept a PET

It was across the meadows that the Dulux dog started to BOUND

At the sound of steps behind me my heart started to POUND

I fell and landed with a BUMP

I inflated the tyre with a PUMP

After collecting the rubbish, I took it to the TIP

Arriving at the beach, I went in for a DIP

He'd lost his job and so signed on the DOLE

She crossed the bridge and so paid the TOLL

The dentist starting to drill was the moment I'd been DREADING

The stains on my feet were from the grapes that I'd been TREADING

The children were horrified to discover that ivory is made of TUSK

To be back before it got dark, the lovers parted at the fall of DUSK

If you have a tooth out, it sometimes leaves a GAP

If you play baseball, it's trendy to wear a CAP

My wedding ring is made of GOLD

As usual in winter I've caught a COLD

You can tell she didn't want the cake mixture to stick to the tin because the tin is GREASED

You can tell he hasn't ironed his shirt because it's so CREASED

Coming away from the prison, we were searched by a few GUARDS

On his birthday, the postman brought John one present and a few CARDS

12. INTACT MATERIALS.

The following materials were used as originally uttered, and as cross-spliced for the remaining materials in Experiment 4.

The groom at the alter turned 'round to see his BRIDE

The groom at the alter turned 'round to see his PRIDE

The winner of the race couldn't hide his PRIDE

The winner of the race couldn't hide his BRIDE

We know you won all those prizes, but there's no need to BOAST

We know you won all those prizes, but there's no need to POST

I know you have some stamps, so I'll give you this letter to POST

I know you have some stamps, so I'll give you this letter to BOAST

The Matador angered the crowd by refusing to kill the BULL

The Matador angered the crowd by refusing to kill the PULL

The climber tested the rope by giving it a PULL

The climer tested the rope by giving it a BULL

It was the dog's head that the girl tried to PAT

It was the dog's head that the girl tried to BAT

The bowler was ready and it was John's turn to BAT

The bowler was ready and it was John's turn to PAT

John watched the baker knead his DOUGH

John watched the baker knead his TOE

John said Fred had stubbed his TOE

John said Fred had stubbed his DOUGH

The 999 call is a number which often needs to be DIALED

The 999 call is a number which often needs to be TILED

The bathroom can't be papered or painted, it needs to be TILED

The bathroom can't be papered or painted, it needs to be DIALED

The naturalist could see the stag, but there was no sign of the DOE

The naturalist could see the stag, but there was no sign of the TOW

His car broke down so we got our rope and gave him a TOW

His car broke down so we got our rope and gave him a DOE

Although I'd spun-dry the shirt, it continued to DRIP

Although I'd spun-dry the shirt, it continued to TRIP

Although they'd fixed the stairs, the step caused me to TRIP

Although they'd fixed the stairs, the step caused me to DRIP

Coming away from the prison, we were searched by a few GUARDS

Coming away from the prison, we were searched by few CARDS

On his birthday, the postman brought John one present and a few CARDS

On his birthday, the postman brought John one present and a few GUARDS

To transport the wild animal, he had to use a CAGE

To transport the wild animal, he had to use a GAUGE

To take an accurate measurement, he had to use a GAUGE

To take an accurate measurement, he had to use a CAGE

The teacher said the pupil was the best in the CLASS

The teacher said the pupil was the best in the GLASS

The barman poured the whisky into the GLASS

The barman poured the whisky into the CLASS

Before he could light the fire, he put the logs in the GRATE

Before he could light the fire, he put the logs in the CRATE

Before shipping the antique, he put it in the CRATE

Before shipping the antique, he put it in the GRATE

APPENDIX F.

13. DESIGN FOR EXPERIMENT 5.

All materials used in Experiment 5 are presented in Section 2 below. The design for Experiment 5 is illustrated below.

* Subject Groups	Voiced Appropriate	Voiced Inappropriate	Voiceless Appropriate	Voiceless Inappropriate
1	a	b	a	b
2	c	d	c	d
3	b	a	b	a
4	d	c	d	c
1	c*	d*	c*	d*
2	a*	b*	a*	b*
3	d*	c*	d*	c*
4	b*	a*	b*	a*

* denotes materials with noise-replaced segments

Thus a subject in Group 1 might hear the following:

a) John hates hornets, wasps and bees.

b) John hates potatoes, carrots and bees.

a) I inlated the tyre with a pump.

b) I fell and landed with a pump.

c*) John watched the baker knead his XXdough.

d*) John said Fred had stubbed his XXdough.

c*) The teacher said the pupil was the best in the XXclass.

d*) The barman poured the whisky into the XXclass.

14. MATERIALS FOR EXPERIMENT 5.

The groom at the alter turned 'round to see his BRIDE

The groom at the alter turned 'round to see his PRIDE

The winner of the race couldn't hide his PRIDE

The winner of the race couldn't hide his BRIDE

We know you won all those prizes, but there's no need to BOAST

We know you won all those prizes, but there's no need to POST

I know you have some stamps, so I'll give you this letter to POST

I know you have some stamps, so I'll give you this letter to BOAST

The Matador angered the crowd by refusing to kill the BULL

The Matador angered the crowd by refusing to kill the PULL

The climber tested the rope by giving it a PULL

The climer tested the rope by giving it a BULL

It was the dog's head that the girl tried to PAT

It was the dog's head that the girl tried to BAT

The bowler was ready and it was John's turn to BAT

The bowler was ready and it was John's turn to PAT

John watched the baker knead his DOUGH

John watched the baker knead his TOE

John said Fred had stubbed his TOE

John said Fred had stubbed his DOUGH

The 999 call is a number which often needs to be DIALED

The 999 call is a number which often needs to be TILED

The bathroom can't be papered or painted, it needs to be TILED

The bathroom can't be papered or painted, it needs to be DIALED

The naturalist could see the stag, but there was no sign of the DOE

The naturalist could see the stag, but there was no sign of the TOW

His car broke down so we got our rope and gave him a TOW

His car broke down so we got our rope and gave him a DOE

Although I'd spun-dry the shirt, it continued to DRIP

Although I'd spun-dry the shirt, it continued to TRIP

Although they'd fixed the stairs, the step caused me to TRIP

Although they'd fixed the stairs, the step caused me to DRIP

Coming away from the prison, we were searched by a few GUARDS

Coming away from the prison, we were searched by few CARDS

On his birthday, the postman brought John one present and a few CARDS

On his birthday, the postman brought John one present and a few GUARDS

To transport the wild animal, he had to use a CAGE

To transport the wild animal, he had to use a GAUGE

To take an accurate measurement, he had to use a GAUGE

To take an accurate measurement, he had to use a CAGE

The teacher said the pupil was the best in the CLASS

The teacher said the pupil was the best in the GLASS

The barman poured the whisky into the GLASS

The barman poured the whisky into the CLASS

Before he could light the fire, he put the logs in the GRATE

Before he could light the fire, he put the logs in the CRATE

Before shipping the antique, he put it in the CRATE

Before shipping the antique, he put it in the GRATE