PROSODY, SYNTAX AND THE LEXICON IN PARSING AMBIGUOUS SENTENCES

NIVEDITA MANI

St Catherine's College

Thesis submitted to the Committee for General Linguistics, Comparative Philology and Phonetics in partial fulfillment of the requirements for the degree of Doctor of Philosophy

University of Oxford

Trinity term 2005 [i.e. 2006] (LTS 28.ii.06)

Prosody, Syntax and the Lexicon in Parsing Ambiguous Sentences

Nivedita Mani, St Catherine's College

Degree of Doctor of Philosophy, Trinity term 2005

This thesis tests the early incorporation of prosodic information during on-line processing of ambiguous word pairs such as *Packing cases*. The word pair is syntactically ambiguous between a noun or verb phrase interpretation. However, the two interpretations are prosodically distinct. An on-line, cross-modal, response-time task found that subjects disambiguated the word pairs using prosodic information. Experiment 2 swapped the timing, f_0, and amplitude of the noun phrase versions with the verb phrase versions. If prosodic information were guiding parsing, swapping the prosody of the alternatives should change subjects' parses of the word-pairs. Subjects interpreted the cross-synthesised noun phrases as verb phrases and the cross-synthesised verb phrases as noun phrases. This provides additional evidence in favour of early prosodic processing. Experiment 3 tested whether subjects' ability to differentiate the two forms would be affected by flattening the f_0 of the word pairs. Subjects' ability to disambiguate the word pairs was reduced by flattening the f_0 of the stimuli. Again, this provides evidence in favour of f_0 guiding parsing. Experiment 4 investigated the perceptual salience of prosodic information in the absence of lexical information, by testing parsing of delexicalised versions of the same word-pairs. Subjects continued to disambiguate the stimuli. This indicates that prosody can guide parsing even without lexical information. The results of the four experiments provide strong evidence in favour of the early incorporation of prosodic information in parsing: prosodic information can influence on-line parsing even in the presence of contradictory syntactic and spectral preferences; and in the absence of lexical information. This thesis concludes that the results of the experiments support strong interaction models of processing.

ACKNOWLEDGEMENTS

I am grateful to my supervisor Dr John Coleman for his help and expert guidance throughout this research.

I would also like to extend my thanks to members of the Phonetics Laboratory and the Committee for General Linguistics, Comparative Philology and Phonetics, the University of Oxford for their help. In particular, I would like to thank Mr. Andrew Slater for his technical guidance and for permitting me to use his response-time program for my experiment.

Thanks are also due to Petra Wagner for the delexicalisation program used in Experiment 4 and to my supervisor, John Coleman for the centroid program used in all experiments and the cross-synthesis program used in Experiment 2. I would also like to thank friends who participated in numerous pilot studies and subjects who took part in the experiments.

Special mention is due to all my friends, especially my housemates, who were there during the final months of labour before the birth of this thesis and for enduring numerous practice talks and proof reads of the contents of the thesis and another special note of thanks to Michael Muentefering, for putting up with me through all the crankiness at the end.

This DPhil would not have been possible without the Clarendon Scholarship and the St Catherine's College Great Eastern Scholarship for which I am grateful. I could not have finished without additional help from a Radhakrishnan Scholarship and the Vice-Chancellor's fund.

Of course, any errors of fact or interpretation are entirely my own.

This thesis is dedicated to my mother, Mrs Uma Mani, for her love and unending support all through my life, and especially during the writing of this thesis, for her faith in me, and, more importantly, her faith in the importance of prosody!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

A LITERATURE REVIEW

1.1. INTRODUCTION

This thesis assesses the contribution of prosody to the parsing of syntactically ambiguous word pairs. The question addressed here is whether listeners can use prosodic cues *alone* to differentiate between the two interpretations of ambiguous sentence-initial word pairs such as *Packing cases* in (1.1) and (1.2):

(1.1) Packing cases are always newsworthy.

(1.2) Packing cases is always newsworthy.

In (1.1), the words *Packing* and *cases* form a noun phrase similar in meaning to 'crates'. However, in (1.2), they form a verb phrase, denoting the act of packing. The hypothesis of this research is that listeners will be able to use their knowledge of the prosodic characteristics associated with these two structures to differentiate between them, even if they are given only the first two words of either sentence. A sequence of four experiments gauge listeners' ability to use prosodic cues — such as f_0, word duration and amplitude — towards choosing one interpretation of the word pair over the other.

This research also contributes to the ongoing theoretical debate between the proponents of serial vs. parallel models of speech processing. Most advocates of serial processing models argue that the initial parsing of ambiguous information is conducted using solely syntactic information (Carlson et al: 2001; Clifton and Ferreira: 1987; Forster: 1976; Raynor et al: 1983). Re-analysis of the ambiguous stimuli is subsequently conducted using prosodic and pragmatic information. This is why it is called serial processing. Conversely, proponents of strong interactive parallel processing models argue that prosodic and pragmatic information can guide initial parsing. Moreover, in a parallel processing model, this is the only stage of parsing, since prosodic, syntactic and semantic information simultaneously guide parsing (Marslen-Wilson and Tyler: 1980; MacDonald et al: 1994; Reisbeck and Schank: 1978Crain and Steedman: 1985). However, for reasons I present in this chapter, despite extensive research, it remains uncertain whether prosodic information can guide initial structural assignment.

Therefore, the four experiments presented here test the role of prosodic information in the parsing of sentence-initial ambiguous word pairs. Experiment 1 tests whether listeners can use the prosodic differences between the noun and verb phrase versions of the sentence-initial word pairs in (1.1) and (1.2) to tell them apart. The results of this experiment show that listeners can propose an *initial* parse of the stimulus that is consistent with the prosodic contrast between the alternative forms of the word pairs. The second to fourth experiments attempt to discover what acoustic cues listeners use in parsing. Experiment 2 swaps the f_0, amplitude and duration of one of the alternatives of the word pairs with the other in order to see whether cross-synthesising the prosody reverses subjects' parsing of the stimuli. I found that conflicting prosodic information – f_0, duration and amplitude – made subjects predict noun phrase parses of the cross-synthesised verb phrases and verb phrase parses of the cross-synthesised noun phrases. This argues for a more influential role for prosodic information in initial parsing than has hitherto been proposed. Experiment 3 narrows the focus to ascertain whether subjects' parsing is affected by removing f_0

differences between the word pairs. In Experiment 3, I present subjects with f_0-flattened versions of the stimuli. All the word pairs have constant f_0 at 115 Hz without any of the differences in f_0 between the words that may have guided listeners' interpretations in the first two experiments. I found that subjects' ability to differentiate between the two forms was attenuated by flattening the f_0 of the word pairs. This suggests a leading role for f_0 in parsing. The results of the first three experiments establish that subjects were not being guided by syntactic information in parsing. However, there remains the possibility of lexical information influencing parsing. Corpus analysis suggested that the likelihood of the word pairs occurring as noun phrases was much greater in some of the stimuli than in the others: The other words pairs were more likely to occur as verb phrases. Therefore, the fourth experiment removes any lexical influence that might have been asserted by the probability of the word pairs occurring as verb or noun phrases. Experiment 4 examines whether solely prosodic cues can affect listeners' parsing decisions, by presenting subjects with delexicalised versions of the word pairs. Subjects disambiguated even the delexicalised word pairs accurately. The results of the four experiments in this thesis present strong evidence for the possibility of exclusively prosodic initial analysis of speech stimuli in the absence of syntactic and lexical information. This thesis concludes that subjects can use solely prosodic information towards initial disambiguation. Finally, I analyse whether the results of the experiments reported here support serial or parallel models of processing. I argue that there is strong evidence supporting interactive activation models of processing that allow prosodic information to *initiate* parsing independent of syntactic, lexical, and segmental information.

In the remainder of this chapter, I review the prior research testing the early incorporation of prosodic information during initial parsing. I present the procedure and results of Experiment 1 in Chapter 2. In Chapter 3, I detail the procedure and results of the second experiment. In Chapter 4, I present the procedure and results of the third experiment. In Chapters 5 and 6, I present the fourth experiment. Finally, in Chapter 7, I review different models of speech processing and the implications of the results of the four experiments presented here on prior research in the field.

1.2. REVIEW OF LITERATURE

The aim of this research is to provide evidence for initial analysis using solely prosodic input. Most of the earlier literature on speech processing tends to focus on the possibility of initial semantic or contextual analysis in the place of syntactic analysis (Fodor: 1983; Forster: 1976; 1974; Frazier: 1979; 1987; 1990; Frazier and Raynor: 1982; Raynor, Carlson and Frazier: 1983; Raynor and Frazier: 1987; Clifton & Ferreira: 1987). Later work has started to focus on the nature of the interface between prosodic and syntactic information in discussions of "syntax-first" or garden-path models (Frazier et al: 1983; 1990; Carlson et al: 2001; Carlson: 2001). In this thesis, I concentrate on literature from Frazier and Carlson (Frazier: 1989; Frazier: 1990; Carlson: 2001; Clifton, et al: 2002; Frazier et al: 2004) and Marslen-Wilson et al (1992), though there are significant differences, not only between different syntax-first models, but also between the "syntax-first" and parallel processing models that have been suggested to date.

1.2.1. "Syntax-first" in processing?

Frazier and colleagues (Frazier: 1978; Frazier and Raynor: 1982; Carlson, Clifton and Frazier: 2001; Clifton, Carlson and Frazier: 2002; Frazier, Clifton and Carlson: 2004) argue that ambiguous speech stimuli are parsed serially: construction of an initial parse is followed by reanalysis of the parse constructed using other information. This initial parsing is called 'first pass analysis'. According to the serial model of speech processing that they advocate, first pass analysis is always conducted using the syntactic information provided by the input. This first pass is only subsequently checked for compatibility with the other information (for example, the thematic relations identified by semantic analysis) provided by the input:

> Constituent-structure analysis of an input is accomplished by initially assigning just one analysis, presumably on the basis of general structural principles (Frazier: 1989: 304)

Certainly, first pass analysis does not consider the pragmatic content of the input:

> [A] "Thematic Processor" is responsible for choosing the pragmatically most plausible thematic frame for each head of a phrase, with the benefit of discourse and real-world knowledge. If the thematic frame selected by the Thematic Processor is consistent with the initial constituent-structure analysis assigned by the syntactic processor, the sentence should be relatively easy to process. However, if the chosen thematic frame is inconsistent with the initially assigned constituent structure, this will serve as an error signal alerting the syntactic processor to the presence of a locally more plausible analysis of the input (Frazier: 1989: 301-302)

This proposal allows the consideration of pragmatic information only after the creation of an "initially assigned constituent structure" using "general structural principles" (Frazier: 1989: 304). These general structural principles include at least two mandatory syntactic parsing guidelines: The principles of Minimal Attachment and late closure. The principle of minimal attachment states that as each new item is presented to the parser, the position of the new item in the syntactic structure is computed in such a way as to add the fewest number of nodes *at the time of its inclusion* in the structure constructed so far. The number of nodes is counted from left to right and does not consider the possibility of increased complexity later in the construction.

Marslen-Wilson et al (1992: 74) illustrate the condition of minimal attachment with the following two sentences:

(1.3) Karen knew the schedule by heart.

(1.4) Karen knew the schedule was wrong.

Following the principle of minimal attachment, both sentences would be initially parsed with *the schedule* as the direct object of the verb *knew*, since this adds the fewest number of nodes to the clause (1.5). However, *the schedule* is the direct object of the verb only in the first sentence. In the second sentence, *the schedule* is the subject of the subordinate clause. (1.3) would be parsed as

(1.5) [**S** *Karen* [**VP** *knew* [**NP** *the schedule*] [**AdvP** *by heart*]]].

(1.4), on the other hand, would be parsed to contain a subordinate clause:

(1.6) [**S** *Karen* [**VP** *knew* [**CP** Null comp [**NP** *the schedule*][**VP** *was* [**AdjP** *wrong*]]]]].

According to serial processing models advocating minimal attachment, (Frazier: 1989), *the schedule* would always be initially parsed as the object of the verb *knew* and not as the subject of the subordinate clause as in (1.6). However, this minimal attachment parse is objected to by the parser when the verb *was* is presented to be added to the structure. The initial parse has to be restructured to contain an embedded clause. According to Frazier, this reanalysis of input after the failure of the initial parse is also independent of pragmatic information. The addition of pragmatic information is only permitted after the syntactically autonomous component has produced a parse that it is satisfied with.

Frazier's second principle, Late Closure, holds that the clause currently being processed by the parser is kept open to additions for as long as possible. Each new item presented to the syntactic processor is added within this clause during initial parsing. Following the principle of late closure, the sentence in (1.4) would be initially parsed with the main clause kept open for as long as possible, favouring the structure in (1.5). This necessitates that any new component that needs to be added into the structure would have to be inserted within this main clause. An anomalous parse – without the introduction of a subordinate clause – would be initially produced, which conflicts with the presentation of the next word: *was*. The input has then to be re-parsed to contain a subordinate clause. The re-parsed structure is only *subsequently* checked for compatibility with the other properties of the input after an initial parse has been satisfactorily constructed using only 'general structural principles'. Some research argues that late closure is restricted to English and not as universal as earlier work has supposed. Speakers of Spanish and Dutch, among other languages, have been found to favour ending the main clause early (Cuetos & Mitchell, 1988; Gilboy et al, 1995; Mitchell & Brysbaert, 1998). Japanese also shows high attachment in some circumstances (Kamide & Mitchell, 1997). However, Fodor (1998) argues that subjects prefer to end the main clause early in Spanish only when parsing an adjunct with ambiguous attachment in a complex NP (e.g., 'the servant of the actress who was on the balcony'). She argues that subjects consistently prefer to keep the main clause open for as long as possible in the case of other instances of attachment in Spanish.

The two principles of late closure and minimal attachment may not be syntactically valid anymore, given recent developments in syntactic theory. However, the main aim of the two principles is to ensure that the parser constructs the simplest structure first, before paying attention to other information, since this would be the structure required in a number of cases.

While Frazier (1989b) suggests that initial parsing considers these two structural guidelines, Frazier (1987) argues that prosodic information can reverse the effects of minimal attachment and late closure during parsing:

> [I]f a string is locally disambiguated (e.g. by punctuation or by clear prosodic effects) then by definition there will be only one permissible analysis of the input and we would expect perceivers to construct that analysis" (Frazier: 1987: p. 563).

Similarly, Carlson, Clifton and Frazier (2001) argue that prosodic information *is* a part of the grammatical information that listeners use to assign structure:

> We assume that listeners use their grammar, including prosodic constraints, to assign structure to a sentence ... Throughout this paper we have assumed that the prosodic hierarchy is a part of the grammar. (Carlson, Clifton and Frazier: 2001: 76-77)

However, Frazier (1987) does not make any attempt to outline how prosodic information can influence parsing. Moreover, Carlson et al (2001) argue that the prosodic structure of speech stimuli is too variable to consistently relate it to syntactic structure. It is unclear whether they are arguing for the inclusion of prosodic information in initial structural assignment or not. Certainly, by saying that prosodic information is a part of the grammatical organisation of an input[1], prosodic information is incorporated in parsing earlier than pragmatic information. But it is

[1] In personal communication, Frazier argues that she does believe that "prosody can influence syntactic processing: the input to the syntactic parser is a phonologically (and prosodically) structured representation."

unclear whether the "general structural principles" that are actually used to produce a first pass of ambiguous stimuli include information of the prosodic content of the stimulus. Consider, for instance, Carlson, Clifton and Frazier's concluding argument in the same paper in 2001:

> The listener is choosing between two grammatical analyses of a sentence, because on either structure a prosodic boundary may but need not occur ... Long or syntactically heavy constituents will increase the possibility that the prosodic boundaries surrounding the heavy phrase are triggered by the internal properties of the phrase rather than by the relation of the phrase to the larger syntactic structure. This allows the listener to justify the existence of an optional prosodic boundary without assuming that the prosodic boundary signals closure of the preceding phrase (Carlson et al: 2001: 77).

Firstly, this suggests that prosodic cues are variable and therefore not as reliable during the assignment of structure as more invariant syntactic principles might be. Secondly, the argument that the listener chooses between two grammatical analyses of a sentence leaves open the question as to whether these two grammatical analyses were created with regard to the prosodic content of the stimulus. Finally, the conclusion of the quotation suggests that the listener does have access to prosodic information during initial parsing – while at the same time insisting that prosodic boundary information is not strictly adhered to. I understand from this concluding argument that prosodic information is used in assigning structure, but the lack of a close correlation between prosodic boundaries and syntactic structure may prompt listeners to sometimes ignore any prosodic information. This suggests that the invariant structural principles proposed by Frazier seem to take priority over the structures presented by prosodic boundaries – especially as these boundaries may be included merely because the input was "syntactically heavy". It is uncertain whether the authors intend that prosodic boundaries guide first pass analysis or not. For instance, Clifton, Carlson and Frazier (2001) claimed that

> At present, it is unclear whether the informative prosodic boundary effects established here occur early in processing, influencing the initial parse of constituents or later in some postsentence interpretive judgment process (p. 106)

As I have mentioned above, Frazier (see footnote 1 also) had argued that prosodic structure is a part of the general structural principles guiding initial parsing. However, close reading of work by Carlson et al (2001) highlights an uncertainty about the possibility of prosodic boundaries guiding initial parsing that sits uneasily with Frazier's suggestion.

Frazier and Fodor (1978) propose a no-bookkeeping approach to sentence processing:

> At each point in the ongoing analysis of a sentence, a decision will not be made within any module without checking that the decision is consistent with *all* within module information (Frazier: 1990: 415: my italics)

Assuming that prosodic information is still a part of the general structural principles guiding parsing, this would imply that parsing would not be conducted without taking prosodic information into account. Further, Clifton, Carlson and Frazier (2002) argue that:

> Imagine this sort of parser systematically begins a new package when it receives a prosodic boundary larger than any prosodic boundary at the beginning of the current package, and systematically does not begin a new package where no prosodic boundary occurs or only a smaller boundary does....Given this packaging routine, we need only assume...that attachments within the package are more readily available than attachments outside the current package. (p. 106-7)

This is entirely in keeping with the "sausage parser" advocated by Frazier and Fodor (1978), according to which speech is divided into little chunks, the edges of which are marked by prosodic boundaries. However, there is considerable inconsistency between a parser that cannot provide an initial parse without utilising all the information in the syntactic-prosodic module (Frazier: 1990) and the uncertainty of the on-line access of prosodic information during initial parsing (Carlson, Clifton and Frazier: 2001; Carlson: 2001). This leaves open the possibility of initial analysis using solely syntactic cues with prosodic cues coming into play only during reanalysis. Carlson (2001) underlines this contradiction. Following a series of off-line experimental tasks, she argues that prosodic differences can affect processing of gapping and non-gapping structures. However, she goes on to propose that:

> The simple structure hypothesis predicted that the simplest syntactic structure would be the preferred analysis, and this was clearly the dominant factor in the auditory processing of potentially gapping sentences as well. Other factors like prosodic parallelism only partially influenced the choice between structures. (Carlson: 2001: 19).

The standpoint that seems to be shared by Frazier, Carlson and Clifton is demonstrated by their outline of the constraints on the parser as one that "establishes "sausages" (packages) based on prosodic information when possible and on other considerations such as length when necessary" (Clifton, Carlson and Frazier: 2002: 106). Although Frazier, Carlson and Clifton seem to lean towards an initial parsing mechanism that accommodates prosodic information, their retraction from this inclusive process seems to be motivated by two reasons. Firstly, they are yet to provide strong evidence in favour of solely prosodic information guiding initial parsing. As I shall argue in the next section, this is a problem common to research to date on initial prosodic processing. Although they argue that prosodic information is a part of the general structural principles guiding initial parsing, they remain uncertain whether prosodic information *is* consistently accessed during initial parsing. This manifests itself in the contradictions and retractions I have outlined above. Claiming that the general structural principles guides initial parsing does not entail that prosodic structure does – irrespective of whether prosodic information is a part of the general structural principles.

Secondly, the lack of a self-evident correlation between prosodic and syntactic structure necessitates that any parsing mechanism suggesting that prosodic structure can guide syntactic structure has also to outline the nature of this interface. Clifton, Carlson and Frazier (2002) attempt a sketch of this interface by arguing that speech is divided into chunks by prosodic boundaries. However, prosodic boundaries are not the only characteristics of the prosodic contour of a speech stimulus. Prosodic boundaries are one-step removed from the simpler acoustic characteristics of the stimulus, such as f_0 and duration. Conversely, independent measures of f_0, amplitude and timing are more stable as far as correlating syntactic boundaries with prosodic information. However, although there has been considerable research on the effect of prosodic boundaries on syntactic structure (Lehiste: 1973; 1979; Cooper & Paccia-Cooper: 1980; Nespor and Vogel: 1983; Price et al: 1991; Grabe et al: 1994; Speer et al: 1996; Schafer: 1997; Schafer et al: 2000; Carlson et al: 2001), and on the correlations between pitch accents, focus and interpretation (Hirschberg and Ward: 1992; Schafer et al: 1996; Schafer: 1997; Maynell: 1999; Schafer et al: 2000), for reasons I present in the next section, none of these have shown that these prosodic cues (f_0, amplitude and timing) can actually *guide* initial parsing. The possibility of correlating prosodic and syntactic structure necessitates further research into the possibility of initial prosodic processing – and not just on the effect of prosodic boundaries on parsing. The work of Frazier and colleagues proposes that prosodic information can guide initial parsing. However, they have not provided strong evidence to this effect and the cracks in the argument appear in the apparent selfcontradictions in their definition of the interface between prosody and syntax during initial parsing. The "syntax-first" serial processing models and the "sausage packing" model (Frazier and Fodor: 1978) fail to provide an adequate description of the mechanism of initial prosodic processing.

If prosodic input were as influential in guiding initial parsing as syntactic input, then assignment of structure to ambiguous sentences should be guided, and not merely facilitated, by their prosodic characteristics. This is the question the research in this thesis addresses. In other words, I ask whether autonomous, prosodic "first pass" analysis is possible. Can prosodic information provide an initial parse of speech stimulus independent of syntactic, pragmatic, segmental and lexical information?

1.2.2. Off-line tests of prosodic structure assignment

There has been so much work on this subject that it is almost inevitable to find contradictions in assumptions correlating prosodic structure to syntactic structure. More often that not, the various factions have used different experimental procedures and criteria for relevance, causing a discrepancy in theories of the role of prosodic information in initial parsing of ambiguous stimuli. As I argued in the last section, proponents of one and the same model have unsurprisingly retracted earlier arguments because of later research using different procedures. The experiments that I mention in this section employ off-line tasks such as comprehension tests, paraphrasing tasks or contextual assignment (Speer, Shih and Slowiaczek: 1989; Beach: 1991; Price et al: 1991; Speer, Crowder and Thomas: 1993; Ferreira et al: 1996; Kjelgaard and Speer: 1999; Schafer et al: 2000; Cohen et al: 2001). These off-line tasks test subjects' reactions to prosodic structure after subjects have heard the entire speech stimulus, typically a complete sentence. They do not necessarily test initial processing of the incomplete input and are more pertinent to charting the final stage of processing when the listener has constructed a "unified representation of the input using all the information provided" (Marslen-Wilson et al: 1992: 85). Therefore, off-line tasks are unsuitable to test whether prosodic information is being accessed in the initial stages of processing. Results of off-line tests have, in many cases, been used to make indiscriminate assertions in favour of or against the inclusion of prosodic information in immediate parsing. This is not to suggest that these experiments have provided evidence against the possibility of initial parsing using prosodic information. However, they cannot be used to conclude in favour of immediate prosodic processing.

Previous research argues that the influence of prosodic information on parsing might depend on whether the ambiguous syntactic alternatives are produced by deep structure ambiguity of the sort in (1.7).

(1.7) Flying planes can be dangerous.

In (1.7), the first word could either be an adjective or a gerund. Liberman et al (1967) argue that prosodic cues cannot guide initial disambiguation of deep structure ambiguity of stimuli such as in (1.7). Conversely, they propose that prosodic cues could influence surface structure disambiguation of sentences like:

(1.8) I'll move / on / Saturday

In (1.8), *Saturday* could either refer to the day that the speaker plans to move houses, or the day that the speaker will contribute more to a current topic of discussion. It is immediately clear that it is more difficult to fit the latter interpretation to an appropriate context. However, Liberman et al's arguments were based on their expectations of the data rather than on experimental observations. Therefore, they do not consider comparing the contextual plausibility of the two interpretations. The prosodic differences between the two interpretations are, consequently, a result of greater pause after *move* in the first version, and after *on* in the second. Liberman et al's suggestion of the prosodic differences between the two forms has been corroborated by later experiments using different stimuli reporting surface structure prosodic differences between similar syntactic alternatives (Lehiste, Olive and Streeter: 1976; Klatt: 1976; Cooper & Paccia-Cooper: 1980).

Wales and Toner (1979) tested Liberman's hypothesis that prosodic cues were not influential in resolving deep structure ambiguity in sentences such as in (1.7). Their experiment tested the ability of subjects to choose one of two contexts from which the sentences that were played to them had been taken. They found that conflicting prosody reduced subjects' preference for the unintended interpretation, while contributing prosodic cues did not have a strong effect on subjects' choosing the intended interpretation. They concluded that prosodic information could only be expected to reduce a syntactic preference, but not to reverse it. However, I argue that listeners would not be able to use prosodic cues to differentiate between the two alternatives of the sentences-initial word pairs in (1.7). The noun phrase form of the sentence-initial ambiguous word pair refers to 'planes that are flying' (i.e. planes that are currently being flown) and the verb phrase form refers to 'the act of flying planes'. However, I do not expect there to be adequate prosodic differences between the two forms of *Flying planes*. I do not think it would be natural to assign a nuclear accent to *Flying* in the noun phrase version of (1.7). There are few differences in the 'plane-ness' of flying and stationary planes. However, there might be differences in, for e.g., the 'case-ness' of packing cases and 'crates' in (1.1) and (1.2). Given that there should be no major prosodic differences between the two alternatives of *Flying planes*, it is unsurprising that listeners cannot differentiate between the two interpretations of the sentence: the remainder of (1.7) provides no further prosodic cues for disambiguation. Furthermore, Wales and Toner (1979) concede that the material provided to subjects might also have been lexically biased towards one interpretation. The likelihood of the word pair occurring as a noun phrase might have been greater than the likelihood of it's occurrence as a verb phrase. Unfortunately, the only instance of the phrase 'Flying planes' in the British National Corpus is the sentence '*Flying planes may be dangerous* is an instance of semantic ambiguity'! Therefore, it is not possible to conclude whether the word pair occurs more frequently as a verb phrase or a noun phrase. Irrespective of putative lexical influences, the absence of prosodic cues differentiating the two alternatives in Wales and Toner's experiment prevents conclusions regarding the possibility of prosodic information guiding immediate parsing.

Ferreira et al (1996) also conclude against the possibility of prosody influencing parsing, using an Auditory Moving Window task. Subjects listened to successive fragments of sentences. Upon comprehending each fragment, subjects were asked to press a button in order to hear the next fragment of the same sentence. The task is claimed to be relatively on-line in that it measures comprehension times at different points in the processing of an utterance. Ferreira et al report that conditions of mismatched prosody were no more difficult to parse than conditions of matched prosody. However, the lack of effect of prosody on parsing could be attributed to the disruption of the prosodic consistency of the stimulus caused by fragmenting the sentences. Conversely, Ferreira et al argue that the prosodic consistency could not have been disrupted since comprehension times were the same for complete as for fragmented sentences. Nevertheless, it is unclear whether comprehension times for complete and fragmented sentences include reading time differences between the two, which might affect overall comprehension time.

Beach (1991) argues that the combined effect of differences in duration and pitch might help subjects differentiate between direct-object and complement clause interpretations of ambiguous sentence-initial fragments. She recorded four different kinds of sentences varying between direct object and complement clause parses of the fragments in (1.9) and (1.10):

(1.9) Jay believed...

(1.10) Jay believed the gossip...

She tested the effect of contrastive duration and pitch on listeners' interpreting the fragments as part of a direct object or a complement clause structures. The f_0 of the last syllable of the verb *believed* falls steeply in both the short and long version in (1.9) and (1.10). Additionally in (1.10), f_0 of the first stressed syllable of the post verbal noun phrase (*the gossip*) rises steeply in the complement clause version. The duration of the main verb in both the long and short version also increased when the sentence was read as a complement clause structure compared to a direct object reading. The effect found was not very large, with subjects disambiguating the fragments as direct object interpretations only 56% of the time when prosody was consistent with this interpretation. However, they also chose the direct object reading 42% of the time when the prosody was consistent with the complement clause reading. In a second experiment, subjects were informed of the correct interpretation of the stimuli. This improved subjects' disambiguation of the stimuli considerably: subjects chose the correct interpretation of the stimuli significantly more often than they chose the unintended interpretation. Beach argues that this indicates that subjects could not correlate prosodic structure to syntactic structure adequately in her first experiment. When subjects were informed of the interpretation intended of the ambiguous stimuli, as in her second experiment, this facilitated the correlation between prosody and syntax. In testing subjects' paraphrases of the stimuli in (1.9) and (1.10), Beach found that there was a combined effect of durational and pitch changes on parsing. Subjects did not provide the appropriate paraphrase when the duration of the verb *believed* was kept constant and f_0 varied between the two versions as described above. However, when duration and f_0 were appropriate to one of the interpretations, subjects correctly parsed this interpretation of the stimuli. Beach's results are encouraging for further research into the effects of prosodic cues on disambiguation, although she concludes that listeners could not correlate prosodic structure with syntactic structure without prompting. Her results indicate that prosodic cues such as f_0 and duration might help resolve deep structure ambiguity between direct-object and complement clause interpretations of syntactically ambiguous stimuli.

Price et al (1991) also conclude in favour of prosodic boundaries guiding parsing. They employed a context-matching task, similar to Wales and Toner (1979), in which listeners had to match the stimuli to one of two contexts presented to them. The input was syntactically ambiguous in one of the following ways:

i. Parenthetical vs. non-parenthetical subordinate clauses (*Mary knows many languages, you know/ Mary knows many languages you know*)

ii. Appositions vs. attached noun/prepositional phrases (*Wherever you are, in Romania or Bulgaria, remember me/ Wherever you are in Romania or Bulgaria, remember me*)

iii. Main-main vs. main-subordinate clauses (*Mary was amazed and Dewey was angry/ Mary was amazed Ann-Dewey was angry*)

iv. Tags vs. attached phrases (*Ben would never leave, would he?/ Ben would never leave Woody*)

v. Far vs. near attachment of phrases (*Raoul murdered the man with a gun*)

vi. Left vs. right attachment of middle phrases (*Although they did run in the woods, they were uneasy/ Although they did run, in the woods, they were uneasy*)

vii. Particles vs. prepositions (*They may wear down the road*)

They found that major syntactic boundaries almost always co-occurred with major prosodic breaks. Acoustic analysis of the stimuli suggested that lengthening of items and greater pauses seem to consistently mark the syntactic boundaries. They found that listeners correctly disambiguated around 84% of the sentences presented to them. However, they argue that prominence of one word over another played only a supporting role in disambiguation, with subjects not always using relative prominence as a reliable cue. Conversely, they found that larger prosodic breaks did seem to guide parsing. Items with larger prosodic breaks were more reliably and consistently disambiguated than items with smaller prosodic breaks.

Warren (1985) also presents evidence in favour of a prosodic effect on parsing. Subjects were presented with the ambiguous fragments of sentences that were either

i. Initially ambiguous between direct object and complement clause interpretations: *The actor learnt the text amused the cast/ The actor learnt the text and knew his role.*

ii. Ambiguous between far and near attachment of PP: *He climbed the peak with snow on top/ He climbed the peak with Pete and David.*

iii. Ambiguous between early and late closure: *Before the king rides his horse takes ages to groom/ Before the king rides his horse Ted gives it a groom.*

He found that the parses chosen by subjects were consistent with the parses indicated by the prosodic content of the ambiguous stimuli. Consistency was measured using the continuations chosen by subjects for the ambiguous fragments. However, consistency between the parse indicated by prosody and the parse chosen by subjects was greater in the case of sentences such as in (iii) than in sentences such as in (i). Prosody seemed least influential in parsing sentences such as in (ii). His results indicate that subjects could use prosodic cues towards disambiguation in some cases. This suggests a more influential role for prosodic information than reported in Wales and Toner (1979), Beach (1991) and Ferreira et al (1996).

The studies reviewed here present contradictory evidence for the effect of prosodic boundaries on parsing. Other off-line studies not discussed here also present similarly contradictory results in favour of (Wingfield and Klein: 1971; Kjelgaard and Speer: 1993; Lehiste et al: 1976; or against (Stirling and Wales: 1996 – employing an experimental task similar to Beach (1991) with naïve speakers) prosodic information guiding parsing. However, these and the work reported in this section have all used off-line measures of the influence of prosody on parsing. Consequently they provide very little information with regard to the possibility of the early incorporation of prosody in parsing. Initial parsing might have already taken place using solely syntactic information. Additionally, the absence of prosodic influences on parsing might be because the tasks did not test immediate reactions and the delay might have entailed a lessening of prosodic effect. More appropriate evidence for the possibility of the early incorporation of prosodic cues in parsing can be obtained from data of on-line tests of the influence of prosody that I shall present in the next section.

1.2.3. On-line tests of prosodic structure assignment

On-line tasks are much more pertinent to research on the initial stages of processing to be studied here. On-line tasks judge subjects' reactions to incomplete sentences or clauses. They attempt to ascertain which of several alternative parses of an ambiguous stimulus subjects seem to be leaning towards *during initial parsing.* For instance, Marslen-Wilson et al (1992) employed an on-line naming task. They recorded the time taken by subjects to name a visual probe following auditory stimulus. In most naming tasks, listeners' interpretations of ambiguous stimuli are tested by subjects' either calling out the probe word or pressing a button that corresponds to the probe word. The latter option was used in Marslen-Wilson et al (1992). This task was designed to find out whether listeners use only the syntactic strategies outlined by Frazier (Frazier: 1979, 1987, 1990; Frazier and Raynor: 1982) or whether they also take prosodic cues into consideration during initial analysis. Subjects were played the italicised parts from sentences like (1.11), (1.12) and (1.13):

(1.11) *The workers considered that the last offer from the management* was a real insult

(1.12) *The workers considered the last offer from the management* was a real insult

(1.13) *The workers considered the last offer from the management* of the factory

At the end of each of the fragments, a visual probe appeared on the computer screen in front of them. For the three sentences above, the probe was almost always the verb *was*, which is an appropriate continuation of (1.11) and (1.12), but not of (1.13) – given the prosodic and morphosyntactic differences between them. Sometimes, the sentences were also followed by the control probe *were* to check subjects' response times to violation of number agreement. This was to check whether subjects were paying attention to the auditory stimulus and not merely focussing on automatic button pressing. Delayed reaction times were expected in subjects' naming of the *were* probe. The absence of this would indicate that subjects had either not been concentrating on the auditory stimulus or that the task is unsuitable to the demands made of it.

Marslen-Wilson et al argue that if subjects were being guided by prosody, they would analyse *the last offer from the management* as the direct object of *considered* in (1.13). However, the direct object structure would conflict with the complement clause continuation confirmed by the probe *was*. Therefore, it was expected that their responses would be delayed due to the conflict between the continuation confirmed by the probe and the continuation suggested by the prosodic cues. Listeners would be forced to do further processing in order to resolve the conflict. Conversely, they would take less time to respond to the probe *was* following (1.12), as the prosodic input would have already cued the complement clause reading. The probe does not conflict with the complement clause interpretation of the stimuli.

However, if prosodic input does not have an effect on initial structural assignment, subjects would take the same amount of time to respond to the probe *was* following (1.12) and (1.13). In both cases, they would first construct a direct object minimal attachment parse of the structure, irrespective of prosodic input cueing otherwise in (1.12). This parse would then be re-considered on presentation of the visual probe because of the complement clause interpretation confirmed by the probe. The results were in keeping with Marslen-Wilson et al's hypothesis that prosodic information does influence structural decisions. Subjects took about the same amount of time to name the probe *was* following (1.11) and (1.12), but took longer to name the same probe following (1.13). In all cases, they took longer to name the control probe *were*. Subjects also completed an appropriateness evaluation task following their response times to the visual probe, where they indicated whether they thought the probe was a 'good' or 'bad' continuation of the auditory stimulus. They approved of the non-minimal attachment probe following (1.13) in appropriateness evaluation tasks; despite prosodic cues to the contrary. This may have been due to subjects' preference for morphosyntactic cues (the *was* continuation indicating a non-minimal attachment parse) over prosodic cues, despite their initial on-line response.

Their results indicate that prosodic cues *may be* incorporated during initial parsing, either independently of or simultaneously with syntactic information. However, their results do not overrule a 'weaker form of interaction': where prosodic cues are accessed only after all plausible parses of the input, including the minimally attached parse, have been constructed using only syntactic information:

> The Frazier-type parser could still apply the minimal attachment strategy, but this would be blocked at a *later* point in the processing if the resulting assignments conflicted with those indicated by prosodic cues. (Marslen-Wilson et al: 1992: 84; my italics)

Marslen-Wilson et al argue that, using the mechanism of weak interaction, the parser would produce a parse using the minimal attachment strategy, while simultaneously producing a complement clause parse of the input. Both parses would then be checked for compatibility with the prosodic characteristics of the input. Both initial parses of the stimulus, however, are produced using syntactic cues and prosodic information is accessed only subsequent to the production of all plausible parses. As Marslen-Wilson et al later comment:

> It is difficult to evaluate this possibility without probing earlier in the spoken string than we were able to here. (Marslen-Wilson et al: 1992: 84)

Marslen-Wilson et al's interpretation of their results allows for prosodic information to play a merely inhibitory role: If the parse constructed is inconsistent with the prosodic content of the stimuli, then prosody can block further processing of the suggested structure.

Marslen-Wilson et al's work (1992) lays the foundation for research on the influence of prosody on on-line sentence parsing. There are, however, a few objections to Marslen-Wilson et al's research that prevent a conclusive argument in favour of prosodic first pass analysis independent of putative syntactic input. I shall outline these problems in greater detail next, while reviewing other experiments that have been conducted using on-line tasks.

Most research employing on-line tasks has tested stimuli beyond the point when syntactic ambiguity has become apparent – therefore, preventing stronger claims in favour of prosodic information guiding parsing at the stage of realisation of syntactic ambiguity. (Marslen-Wilson et al: 1992; Nagel, Shapiro, Tuller and Nawy: 1995; Shapiro and Nagel: 1995; Speer, Kjelgaard and Dobroth: 1996; Kjelgaard and Speer: 1999; Watt and Murray: 1996). Consider, again, the sentences that have been used in some of these studies:

(1.14) (The workers considered ^{H*} the last offer $^{H*\ L\ \%}$)$_{IPh}$ (from the management of the factory $^{H*\ L\ \%}$)$_{IPh}$

(1.15) (The workers considered $^{H*\ L\ \%}$)$_{IPh}$ (the last offer from the management $^{H*\ L\ \%}$)$_{IPh}$ (was a real insult $^{H*\ L\ \%}$)$_{IPh}$

The probe word *was* was presented to subjects immediately after auditory stimulus up to and including the word *management.* I suggest that the prosodic contour of these sentences might be as given in (1.14) and (1.15). There might be other more prominent prosodic characteristics differentiating the two stimuli that I have not considered. For instance, the duration of the main verb in the two sentences might be different in the two readings. However, duration might be interpreted as one of the phonetic correlates of the pitch accents marked in (1.14) and (1.15). Generally, however, speakers might place a major boundary after *offer* and *factory* in (1.14) and after *considered* and *management* in (1.15). They might also place a minor boundary after *considered* in (1.14), represented above by the absence of an IPh but the presence of an H*.

Marslen-Wilson et al claim:

> We know from the results of [our] experiment that prosodic cues do come into effect before the probe point [*was*], as otherwise we would not get the NMA – Comp/MA differences[2]. But there still may be a period of two or three words early on, where an attachment ambiguity has emerged on morphosyntactic grounds and where there are still insufficient prosodic cues to resolve it. (1992: 84)

Testing parsing at the word *management* does not adequately assess how parsing progressed following the verb *considered,* at which point syntactic ambiguity first becomes noticeable. The delay between the point of testing for parsing of syntactic ambiguity (*management*) and the point of realisation of syntactic ambiguity (*considered*) admits the possibility of a period when syntactic ambiguity has become apparent but the resolution of ambiguity using prosodic cues has still not been tested for. What is required is a test "earlier in the string" (Marslen-Wilson et a: 1992: 84). I propose that it might be more appropriate to test subjects' response times to the probe word *that* following both versions of the string up to the presentation of the verb *considered.* The probe word *that* would be incompatible with the prosodic characteristics of (1.15), but not of (1.14). Currently, the results of Marslen-Wilson et al's experiment could also be explained using an autonomous syntactic parser that puts forward a direct object parse at the point of presentation of *considered* (following minimal attachment). This parse might then be reconsidered two or three words further into the sentence, perhaps after the realisation of a prosodic boundary after *offer*. Again, this would only allow a secondary role to the prosodic input provided to the parser. Prosody would come into play only after syntactic assignment has already taken place at the point of recognition of ambiguity (*considered*). This is not to suggest that mandatory syntactic processing does take place at the point of

[2] NMA – Comp: Non minimal attachment sentence without complementizer as in 1.12.
MA: Minimal attachment sentences as in 1.13.
Subjects were faster to respond to the probe word *was* following NMA-Comp sentences than MA sentences.

recognition of ambiguity, but merely to illustrate that Marslen-Wilson et al do not test whether prosodic processing has taken place at this point.

Marslen-Wilson et al's results are also consistent with a parsing mechanism that compiles all plausible parses of the sentence using merely syntactic information, i.e. at the point of presentation of the verb *considered.* The parser could still have produced both direct object and complement clause readings of the verb. Only if subjects show a preference for one reading over another at an earlier point could we infer that first pass analysis has taken place using prosodic input.

Schepman (1997) argues that there is an additional problem with Marslen-Wilson et al's study. The fragments presented to subjects were clipped from the original sentences given in (1.11), (1.12) and (1.13). The visual probes presented to subjects following the fragments taken from (1.11) and (1.12) were also the words following the fragments in the sentences initially recorded. Schepman argues that co-articulation between the last syllable of *management* and the first syllable of the following word in (1.11) and (1.12) might have quickened subjects' response times to the visual probe. However, the probe was not the word following the fragment in the sentence in (1.13) – the sentence the fragment was clipped from. Therefore, the absence of co-articulation in the last syllable of *management* in (1.13) might have led subjects away from responding to the probe as quickly as their response times to (1.11) and (1.12). This would lead to results similar to those reported by Marslen-Wilson et al. However, the motivation behind subjects' response times would be segmental, and not prosodic.

Watt and Murray (1996) used a similar paradigm to test the conclusions reached by Marslen-Wilson et al (1992). The sentences used by Watt and Murray were temporarily ambiguous between direct object and reduced complement structures as given below:

(1.16) *The teacher noticed* [*one girl from her class*] $_{CP}$... (Complement) **was**

(1.17) *The teacher noticed* [*one girl from her class*] $_{NP}$ ... (Direct object) **in**

As soon as the subjects heard the italicised parts of the utterances above, they were asked to read aloud the visual probe that appeared before them. In stimuli such as (1.16) and (1.17) above, this involved the subjects repeating either the word *was* or *in* as soon as they saw it on a screen in front of them. The time taken by subjects to respond to the probe was measured. Contrary to Marslen-Wilson et al (1992), Watt and Murray's results showed longer response times for appropriate probes (*was* following (1.16) and *in* following (1.17)) than for inappropriate probes (*in* following (1.16) and *was* following (1.17)), although this difference was not significant. This indicates that subjects could not construct parses of the stimuli using the prosodic contour of the fragments: they took longer to name probes that confirmed the interpretation consistent with the prosody of the stimulus than to name probes that conflicted with this interpretation. However, Watt and Murray's first experiment did not use an appropriateness-rating task. Marslen-Wilson et al (1992) had included this task to focus subject's attention on the meaning of the utterance as a whole: Each time subjects named the probe, they were given a few seconds to decide whether they thought the probe was an appropriate or inappropriate continuation of what they heard. The task ensures that subjects pay attention to the auditory and visual stimuli as a cohesive whole. Without this, subjects might focus on robotic button pressing and not pay attention to the auditory stimulus. This is particularly important in the case of reaction time experiments, where subjects' response times to numerous stimuli are tested. For instance, the experiments in this thesis present subjects with 384 stimuli, which subjects could easily tire of. Watt and Murray's second experiment included an appropriateness-rating task and the results show a marginal effect of prosody. Watt and Murray argue that prosody could be expected to influence subject's response times in parsing. However, Watt and Murray questioned the effectiveness of prosodic cues. Although naming times were longer for probes that conflicted with the interpretation intended by the prosody than for probes that confirmed this interpretation, this difference was not significant.

In order to test whether this was because of the naming task not tapping into subjects' parsing, they performed a follow-up experiment using a lexical decision task, instead. Subjects pressed a 'yes' button with their left hand if the cross-modal target was a word or a 'no' button with their right hand, if it wasn't. The time taken by them to press the buttons was recorded. They found that subjects' response times to the lexical decision task were similar, irrespective of whether the probe was consistent with the parse indicated by the prosodic contour. The stimuli used in the lexical decision task were prepared by interrupting the sentence with the presentation of the probe immediately after the ambiguous NP (*one girl from her class*). The authors suggest that the absence of an effect of prosody might have been caused by the presentation of the probe word so soon after the ambiguous NP. This might have violated the prosodic consistency of the stimulus. The task was therefore repeated using different stimuli. More importantly, the fourth experiment used the same probes for both interpretations of the ambiguous stimuli in a position that could not interfere with the prosodic consistency of the fragment. The stimuli used in this last experiment were as follows:

(1.18) The tutor understood that the problems the student was having had...

(1.19) The tutor understood the problems the student was having had...

(1.20) The tutor understood the problems the student was having with...

The stimuli presented to subjects were cross-spliced just before the disambiguating word (*had/with*). The sentences with cross-spliced prosody were all followed by the same probe *various* (after *had/with*). This was to ensure that there could be no effect of the probe word on listeners' reaction times. Again, they did not find any evidence of an effect of prosody on parsing decisions. They argue that the lack of significant effect of prosody on parsing could be because their experiments did not include an appropriateness-rating task. They argue that an appropriateness-rating task might have created an effect of slowing down subjects' response times in Marslen-Wilson et al. Conversely, I argue that tasks that do not include appropriateness rating do not focus the subject's attention on extracting meaning from the stimulus and are subject to the criticism that subjects could be concentrating solely on finishing the task without actually listening to the fragments and seeing the probe as a continuation of the auditory stimulus.

Schepman (1997) argues that there are a number of reasons for the contradictions in the results reported by Marslen-Wilson et al (1992) and Watt and Murray (1996). She argues that the former have greater statistical power, because of their recording almost double the number of subjects' response times. Watt and Murray recorded subjects' response times following 540 presentations. Marslen-Wilson et al recorded subjects' response times following 960 presentations.

Furthermore, Marslen-Wilson et al did not take into consideration any response times over 1000 milliseconds, while Watt and Murray considered response times up to 4000 milliseconds. Once all response times over 4000 milliseconds were excluded, Watt and Murray also replaced all response times that were over 2 standard deviations from the mean with the 2 standard deviations value. Schepman argues that these factors give Watt and Murray's results much less statistical power. Furthermore, Schepman does not agree with the argument that the lack of the appropriateness-rating task slowed down subjects' response times. Conversely, she notes that subjects' response times in Marslen-Wilson et al (1992) were actually 125 milliseconds faster than subjects' response times in Watt and Murray (1996). She also finds problems in Watt and Murray's explanation of the lack of significance of difference found in the results of their second experiment. Watt and Murray had argued that the lack of effect of prosody might have been because of subjects' responding to the probes before the integration of prosodic and syntactic information. However, Schepman points out that subjects' response times to the probes in Watt and Murray (1996) were around 768 milliseconds, while subjects' response times to the probes in Marslen-Wilson et al (1992) were around 400 milliseconds. Finding an effect of prosody earlier than later argues against Watt and Murray's explanation for the lack of significance reported in their second experiment. Finally, Schepman notes that Watt and Murray report that their stimuli were spoken slowly and hesitantly (Watt and Murray: 1995). Schepman argues that stimuli that sound hesitant might guide listeners to overriding prosodic cues, since the breaks might not actually coincide with grammatical boundaries as they do in fluent speech. Altogether, Schpeman concludes that Watt and Murray's results cannot be used to argue for a lack of effect of prosody on parsing, since there are a number of alternative reasons for the pattern of results reported.

Schepman (1997) also reports the results of her own work on the possibility of prosody guiding parsing. Five of her experiments concentrate on production studies and off-line tasks, which I shall not review here. The on-line studies reported all test for subjects' parsing of sentences such as (1.21) and (1.22), taken from Schepman and Rodway (2000). Using Marslen-Wilson et al's cross modal priming task (1992), Schepman and Rodway (2000) presented subjects with the italicised parts of the sentence given below:

(1.21) *The lawyer greeted the powerful barrister and the wise judge who* was walking to the courtroom

(1.22) *The lawyer greeted the powerful barrister and the wise judge who* were walking to the courtroom

They suggest that native speakers signal the contrast between the two fragments by providing a prosodic break between *barrister* and *and the wise judge* if the relative clause is to be assigned solely to the *judge* as in (1.21). This prosodic break is absent from the stimulus if the relative clause is to be assigned to *the barrister and the wise judge* as in (1.22). The probe *was* would be appropriate to the inclusion of a prosodic break between the two conjuncts and inappropriate in the absence of the same prosodic break. Subjects' response times to appropriate and inappropriate probes were measured. If subjects were using prosodic information to parse the structures, then they would take longer to respond to inappropriate probes. Schepman and Rodway found that subjects did take longer to respond to inappropriate probes. This, again, argues for an integration of prosodic cues into the information used to construct a parse of the stimulus. However, Schepman and Rodway also check for parsing of ambiguity at a point much later in the structure beyond the point of recognition of ambiguity – the presentation of the word *barrister*. Similarly, the stimuli presented in

Watt and Murray (1996) is also subject to my criticism of Marslen-Wilson et al (1992). With the exception of Grabe et al (1994) and Warren et al (1995), all the experiments reported in this section are subject to the same criticism. Therefore, I shall review these two experiments next.

Grabe, Warren and Nolan (1994) tested for the effect of stress on initial parsing of ambiguous clauses such as (1.23) and (1.24) below. The placement of main stress on '*Chi'* indicates that *Chinese* is an adjective modifying the noun in (1.23). Conversely, the placement of main stress on *nese* indicates that *Chinese* and *teacher* are both nouns and form a compound structure in (1.24).

(1.23) A CHInese TEAcher (adjective + noun)

(1.24) A ChiNESE teacher (compound noun)

In a production experiment, they established that similar phrases (e.g *Chinese fan*[3]) were produced with a nuclear accent on *Chi* and *fan* in the adjective + noun reading. However, the compound noun reading was produced with stress on *nese* and no stress on *fan*. Consequently, in the adjective + noun reading, this word provides an earlier cue to its' syntactic category in the form of stress on *Chi*. This is contrasted with the prosodic realisation of the two interpretations of *Malay teacher*: The stress on *Malay* does not change between the two readings. Therefore, the realisation of Malay in the adjective + noun reading does not provide an earlier cue to its' syntactic category. They studied cross modal naming tasks, similar to Marslen-Wilson et al's task above (1992), using sentences such as:

[3] The adjective + noun reading could refer to a fan of the Chinese language, while the compound noun reading refers to a fan of Chinese origin.

		Visual probes
(1.25)	It's hard to find reliable Chinese	servants/ teachers
(1.26)	It's hard to find reliable Malay	servants/ teachers

The compound reading of *Chinese/ Malay servants* does not exist. Therefore, the probe word *servants* would be inappropriate as a continuation of the compound reading of both (1.25) and (1.26), i.e. when nuclear accent is on *ChiNESE* (compound reading) and *Malay*. Therefore, we would expect subjects to take longer to name the probe in these cases. The probe word *teachers* would be an appropriate continuation of the compound and adjective + noun reading of (1.25) and (1.26). The probe offers access to both interpretations of 'teachers of the Chinese/Malay language' and 'teachers of Chinese/ Malay origin'.

Subjects' response times to the different probes following the different conditions were measured. They found that subjects could predict the adjective + noun reading earlier in items like *Chinese* (with a difference in stress on *Chinese* between the adjective + noun and the compound reading) than in items like *Malay* (with no difference in stress on *Malay* between the two readings). This was indicated by the fact that the difference in response times to *servants* following the two readings of *Chinese* was more significant than following items like *Malay*. Grabe et al argue that this establishes that subjects can indeed use their knowledge of stress patterns to parse input. Note that the stimuli tested in this experiment were such that the point of testing for subjects' parsing of syntactic ambiguity was immediately after the point of recognition of ambiguity – *Chinese* in (1.25). Subjects' response times are given in the table below:

	Compound (prenuclear stress) (ms)	Adjective + noun (nuclear stress) (ms)	Difference in response times (ms)
Item type: Malay	669	638	31
Item type: Chinese	651	588	63

Table 1.1 Mean naming latencies to test probes (servants)

It is clear that the difference between responses times to test probes following adjective + noun and compound readings is greater in the case of items like *Chinese* than items like *Malay*. However, subjects' responses to the adjective + noun reading also appear to be faster than subjects' responses to the compound reading following *Malay*. Grabe et al comment that the significance of this difference was marginal. However, this raises doubts of another influence being established simultaneously with prosodic influences on parsing. Are there significant structural distinctions between the two readings to cause a syntactic preference? Grabe et al's proposed structures of the word pairs do not suggest significant syntactic preferences on parsing. Conversely, there might be a lexical bias towards the adjective + noun reading of the word pair that might affect subjects' parsing? Further experiments by Grabe et al on non-ambiguous stimuli such as *Torquay* also established a preference for the adjective + noun reading. While Grabe et al's work provides strong evidence of a prosodic influence on parsing; it also raises doubts of an additional influence on parsing. It would be worth exploring the possibility of additional influences to check whether there is any interaction between the influences working on parsing. Chapter 5 of this thesis, for instance, looks at the interaction between prosody and lexical bias on parsing.

In addition, in a subsequent study, Warren et al (1995) did not find a corroborative effect of word stress guiding initial parsing, although they do report an effect of prosodic boundaries on parsing. They used sentences such as:

(1.27) Whenever parliament discusses Hong 'Kong | 'problems

(1.28) Whenever parliament discusses 'Hong Kong 'problems |

(1.29) Whenever parliament discusses 'Hong Kong | 'problems

(1.30) Whenever parliament discusses Hong 'Kong 'problems |

where ' indicates the location of the peak accented syllable and | indicates a prosodic boundary. In (1.27), *problems* is a part of a following clause. However, in (1.28), it is part of the object noun phrase including *Hong Kong*. In (1.28), the main accent in *Hong Kong* is on *Hong*, while in (1.27), the main accent is on *Kong*. (1.29) and (1.30) present subjects with conflicting prosodic information within the same utterance. Warren et al argue that the absence of the prosodic boundary between *problems* and *Hong Kong* motivates the placement of the nuclear accent of *Hong Kong* on *Hong* in (1.28). Therefore, the presence of nuclear accent on *Hong* and the boundary between *problems* and *Hong Kong* in (1.29) presents subjects with conflicting prosodic information. Similarly, the absence of the boundary in (1.30) and the presence of nuclear accent on *Kong* would also present subjects with conflicting prosodic information. Subjects were played the four different stimuli. Their response times to the probe word *arise* were measured. *Arise* is an appropriate continuation of (1.27) because of the presence of the prosodic boundary between *Hong Kong* and *problems* and the nuclear stress on *Kong*. *Arise* cues in the interpretation that *problems* is a part of a following constituent clause. (1.29) also contains a prosodic boundary between *problems* and *Hong Kong* with a nuclear accent on *Hong*. The accent on *Hong* is inconsistent with Warren et al's suggestions for the stress pattern of the reading that includes *problems* within the same noun phrase as *Hong Kong*. If subjects were using prosodic boundaries to parse the stimuli, then Warren et al predicted that response times would be faster following (1.27) and (1.29) since the boundaries between *problems* and *Hong Kong* might cue the interpretation where *problems* is not a part of the direct object noun phrase containing *Hong Kong*. Similarly, if subjects interpreted the presence of a nuclear accent on Kong in (1.30) as indicative of the interpretation where problems is not a part of the same noun phrase as Hong Kong, then subjects' response times to arise following (1.30) should be equally fast. However, if subjects interpreted the presence of the nuclear accent on *Hong* in (1.29) as indicative of the interpretation where *problems* is a part of the same noun phrase as *Hong Kong*, then their response times to the probe might be delayed. Finally, subjects' response times to the probe following (1.28) should be delayed, since both the nuclear accent and the boundary cue are inconsistent with the probe. Therefore, Warren et al predict that response times would differ in a gradient with the lowest response times for (1.27), followed by the inappropriate lower response times for (1.32) followed by the delayed response times to (1.28) and (1.29). However, they found that subjects' response times were the shortest for (1.27), followed by (1.29); followed by (1.30) and then (1.28). Their results are in favour of prosodic boundaries guiding parsing. However, they are not in favour of stress guiding parsing. If subjects were using the presence of a nuclear accent on *Hong* in (1.29) to guide parsing, this should have delayed their response times to the probe. As we can see from the results, this was not the case. Response times were slower following (1.30) than (1.29), although stress in the former was consistent with the interpretation confirmed by the probe. Conversely, stress in the latter (1.29) was inconsistent with the interpretation confirmed by the probe. Warren et al showed that prosodic cues such as stress were useful in deciding structure in a post-experiment off-line acceptability-rating test. However, the results of their on-line test were not sufficient to attribute a decisive role in parsing to the perception of stress. While they do provide evidence that prosodic boundaries can guide parsing, it would also be interesting to look at the acoustic characteristics of these prosodic boundaries. It is possible that phrase-final stress will play a part in the acoustic characteristics defining the prosodic boundaries between *Hong Kong* and *problems.*

Measuring brain activity seems to be the most direct method of uncovering the sequence of the process of comprehension. In addition, this has proved effective in attributing more than just a supporting role to prosody in sentence processing. For instance, Steinhauer et al (1999) compared the shape of the waveforms obtained while listening to sentences such as (1.31) and (1.32). Previous research indicates that ERP measures recorded during syntactic misanalysis have usually taken the form of delayed activity between 500 and 1200 ms: the P600 response (Osterhout and Holcomb: 1992). Similarly, ERP measures recorded during semantic misanalysis are marked by earlier centroparietal activity between 300 and 900 ms: the N400 response (Kutas and Hillyard: 1980). When the sentences were presented to subjects, ERP measures showed positive going waveforms distinct from earlier measures recorded for either semantic or syntactic misanalysis. Importantly, these waveforms coincided with the prosodic phrase boundaries in the two sentences:

(1.31) Peter verspricht Anna zu arbeiten... und das Büro zu putzen

([Peter promises Anna to work] $_{IPh}$ [and to clean the office] $_{IPh)}$

(1.32) Peter verspricht Anna zu entlasten... und das Büro zu putzen

([Peter promises] $_{IPh}$ [to support Anna] $_{IPh}$ [and to clean the office] $_{IPh}$)

While sentences such as (1.31) produced positive going waveforms after *arbeiten*, sentences such as (1.32) correctly produced two positive going waveforms at *verspricht* and *entlasten*. These positive deviations were termed 'closure positive shift', as they coincided with the closure of the intonational phrase being parsed. Steinhauer et al also presented subjects with cross-spliced sentences and tested for event related responses related to garden path effects. They found P600 and N400 patterns at the garden path points. They argue that these two measures represent syntactic and semantic re-analysis caused by the cross-splicing. Comparison of the amplitude differences of the N400 and P600 ERP measures in the syntax-prosody mismatch condition and those displayed by the closure positive shift in (1.31) and (1.32) showed increased activity in the mismatch condition – suggesting that the ERP measures in (1.31) and (1.32) were caused by prosodic processing. However, it could still be argued that this positive deflection was caused by syntactic parsing and not, indeed, by the perception of the end of the phonological phrase. Therefore, in a later experiment, Steinhauer (2003) played subjects delexicalised versions of the stimuli. These consisted of stimuli which had their lexical content obscured, but which maintained the prosodic content of the speech with respect to pitch, amplitude and rhythm. This was followed by one of the two sentences in (1.31) and (1.32) being visually presented to the subjects word-by-word on a computer screen. Subjects were told to read the sentences silently using the melody of the delexicalised stimuli played to them earlier. The occurrence of a closure positive shift at two boundary positions in (1.31), but not in (1.32) was explained by the fact that the shift was related to prosodic processing and not to the syntax, as might have been suggested of their earlier experiment (1999). A closure positive shift related to syntactic information could not have been caused by delexicalised stimuli. This lends support to Frazier and

Fodor's sausage parser, as it suggests that subjects can recognise prosodic boundaries. It does not provide evidence that subjects are actually using these boundaries to break speech up into constituent units.

Jeschniak et al (1998) present additional ERP results in support of the possibility of prosodic processing at the initial stages of parsing using syntactically inappropriate sentences such as

(1.33) Die Ente wurde im gefüttert

The duck was being in the fed

(1.34) Die Ente wurde IM gefüttert

The duck was being IN THE fed

Jeschniak et al argue that (1.33) and (1.34) are both syntactically inappropriate. However (1.33) is also prosodically inappropriate with main stress on IM (denoted by the capitals). (1.34) caused an early negativity in ERPs due to syntactic inappropriateness. Jeschniak et al tested to see whether the deviant prosody in (1.34) affected this late negativity. They found that deviant prosody effectively stopped the occurrence of late negativity while listening to (1.34). Jeschniak et al argue that this indicates that inappropriate prosody blocks further syntactic processing of (1.34).

Jeschniak et al conducted a follow-up experiment to test the latter claim. They presented subjects with questions that would necessitate focus on the prepositions in response ('Wurde die Ente VORM Teich gefüttert?' translated as 'Was the duck IN FRONT OF THE pond being fed?'). The questions would make (1.34) prosodically appropriate by warranting focus on the preposition in the answers. If the absence of the negativity in (1.34) was caused by prosodic inappropriateness blocking further syntactic analysis, then the questions should cause reappearance of the early negativity. The authors report finding a similar but not identical early negativity followed by a late positivity. They are currently exploring the causes of the slight change in wave patterns. They do conclude, however, that the presence of an early negativity following (1.34) after question prompts does lend credence to the possibility of prosodic inappropriateness blocking further syntactic analysis in their first experiment.

These ERP studies are yet to be corroborated with evidence from the on-line tasks described above. Grabe et al (1994) and Warren et al (1995) raise interesting questions for future research in initial prosodic processing. Typical of the contradictions inherent in research in the subject, however, almost all of the experiments reported so far have their analogues in similar experiments reporting results to the contrary. Marslen-Wilson et al (1992) was contradicted by Watt and Murray (1994). Grabe et al (1994) and Warren et al (1995) present details of similar experiments with different results. Grabe et al's results also raise interesting questions of additional information influencing on-line parsing. For instance, this information might concern the lexical bias of the stimuli tested. Warren et al could not find a similar influence of stress on parsing, however, it is possible that phrase-final stress might help cue in the influence of prosodic boundaries that Warren et al report. These are only a few of the numerous studies testing the influence of prosody on initial parsing that have been conducted to date. The majority of the studies have reported results that admit the possibility of prosodic information influencing initial parsing. However, in many cases, authors have been forced to retreat from earlier, stronger claims of early incorporation of prosody because their results could be interpreted by weaker interaction between prosodic and syntactic structure in initial processing. For instance, Speer, Kjelgaard and Dobroth (1996) argue that prosodic cues were 'fundamental to the auditory sentence comprehension process' (1996: 251). In later work in 1999, discussed in detail in a later chapter, they argue that cooperation between prosodic and syntactic boundaries assists rather than guides syntactic decision-making. There is a need for further strong evidence in favour of early incorporation of prosodic information in processing. I argue that the four experiments I present in this thesis fulfil this need and present strong evidence in favour of solely prosodic processing of ambiguous stimuli. This, in turn, suggests that prosodic information can be incorporated during initial processing.

1.3. CONCLUSION

I have reviewed the main literature in the field in this chapter. Obviously, there has been more research completed on the subject. However, I have not reported them as the bulk of the assumptions made by them are shared by other experiments mentioned here. The question asked at the beginning of this chapter was whether prosodic information could guide parsing of ambiguous stimuli independent of syntactic and lexical preferences. I argue that this would provide evidence in favour of the early incorporation of prosodic information in parsing. Howver, there are a number of problems with the research that has been conducted.

Firstly, the experiments outlined in Section 1.2.2 all employed off-line tests of the ability of prosodic information to guide initial parsing. However, off-line tests typically test the processing of speech at the pace required by the listener to feel confident enough to propose an interpretation of it. Finding an effect of prosody during reanalysis is an unsurprising result. There is little argument that listeners use the prosodic characteristics of speech to contribute to their understanding of what they have heard. Conversely, what is at issue here is the ability of prosodic cues to guide parsing prior to reanalysis. These experiments do provide an essential foundation to the wealth of research on on-line tests of the parsing of prosodic information. However, they are not adequate to claim that prosodic information can guide initial parsing of ambiguous stimuli.

Secondly, in Section 1.2.3, I have outlined the experiments that employ on-line tests of the ability of prosody to guide initial parsing. With the exception of Grabe et al (1994) and Warren et al (1995), all of them have used stimuli that provide too much lexical and consequently too much syntactic information to the parser. This is compounded by testing for parsing ambiguities at a much later point in processing – not as late as the off-line tests in parsing, but almost definitely, at a point much after syntactic ambiguity has become apparent. This is not to say that these results indicate that prosodic information does not guide initial parsing, but merely that they do not adequately test whether prosodic information can guide initial parsing of ambiguous stimuli. I argue that, so far, there has not been enough evidence that prosodic information can guide on-line parsing of stimuli at the point of presentation of syntactic ambiguity.

In the next chapter, I provide the details of the first experiment of this thesis. This experiment avoids the criticism that I have outlined of previous literature on the possibility of prosodic first pass analysis. There are four requirements of an adequate test for initial prosodic analysis that were highlighted in reading the prior literature in the field. First, the task must test for immediate reaction of subjects to intermediate stages of processing – the task must be an on-line test of subjects' parsing decisions. Secondly, the stimuli must test for subjects' parsing decisions at the point of realisation of syntactic ambiguity – and not, as most on-line tasks have been wont to do – at the point of resolution of syntactic ambiguity. The stimuli must aim to provide adequate prosodic information, but inadequate syntactic information, in an attempt to judge the ability of almost solely prosodic cues to guide parsing. Thirdly, the experimental task must include an appropriateness-rating task to focus subjects' attention on the acquisition of meaning from the stimuli and not merely on responding to the stimuli as quickly as possible. This was one of the problems with Watt and Murray (1996). As I have argued in Section 1.2.3, an appropriateness-rating task focuses listeners' attention on getting meaning from the stimuli played to them, instead of robotically responding to the experimental task. The aim is for listeners to focus on the cohesiveness of the auditory and visual stimuli presented to them, rather than merely reacting to the demands of the visual probe.

Finally, I discuss a stipulation that I have not already considered. Most of the work on prosody in parsing has relied on the ability of prosodic cues such as major prosodic boundaries marking sentence-internal or, at best, clause-internal ambiguity (Carlson et al: 2001; Cooper & Paccia-Cooper: 1980; Grabe et al: 1994; Lehiste: 1973; Nespor and Vogel: 1983; Price et al: 1991; Schafer et al: 2000; Schafer: 1997; Speer et al: 1996; Steinhauer et al: 1999; Steinhauer: 2003). Although a number of experiments have tested the ability of duration and pitch to guide parsing, not many have tested the influence of these prosodic cues at the phrase-internal stage of processing at the beginning of a sentence. I argue that this tests initial parsing adequately. I will bring this latter point up in much greater detail in Chapter 5. Next, in chapter 2, I present the details of the first experiment of this thesis that incorporates the four stipulations mentioned above to record a suitable test for the possibility of subjects' initial parsing of ambiguous stimuli using prosodic cues.

CHAPTER TWO
PROSODY AND PARSING AMBIGUITY: EXPERIMENT ONE

2.1. INTRODUCTION

At the end of the last chapter, I argue that the four experiments in this thesis must fulfil four requirements before claiming to test the possibility of prosodic information guiding initial parsing. Therefore, all experiments in this thesis employ on-line response time tasks, testing subjects' parsing of sentence initial word pairs that contrast in prosodic information immediately before the point of recognition of syntactic ambiguity. Consequently, all experiments in this thesis test the parsing of phrase-internal ambiguity. Three experiments include a subsidiary appropriateness-rating task in order to ensure that subjects concentrate on the cohesiveness of the auditory and visual stimuli presented to them. The appropriateness-rating task also serves as an off-line measure of subjects' parsing decisions. Therefore, the experiments truly assess the ability of prosodic cues to guide initial parsing of ambiguous stimuli.

The contrasting word pairs were taken from sentences from Tyler and Marslen-Wilson's experiment (1977), e.g. (2.1) and (2.2):

(2.1) Although they are no longer used to carry cargo, sailing ships are....

(2.2) Although you need a well-trained crew to do it, sailing ships is....

Experiment 1 focuses on ambiguous word pairs such as – *sailing ships*. The word pairs contrast syntactically in that the word pair in (2.1) is a noun phrase – the first word of the pair (*sailing*) is an adjective and the second (*ships)* is a noun. Conversely, the word pair in (2.2) is a verb phrase – the first word (*sailing*) is a verb and the second (*ships*) is a noun. 'Are' would be an appropriate continuation of the noun phrase interpretation of the word pair and 'is' would be an inappropriate continuation. 'Is' would be an appropriate continuation and 'are' would be an inappropriate continuation of the verb phrase interpretation of the ambiguous word pair. Moreover, I expected that there would be consistent prosodic differences between the noun and the verb phrase versions of some of the word pairs. I show in Section 2.2 that this prosodic difference consists of the first word being stressed in the noun phrase reading and the second word being stressed in the verb phrase reading. Experiment 1 tests whether listeners can use the prosodic contrast between the two interpretations of the word pairs to disambiguate them on-line. In testing the possibility of prosody guiding initial parsing, I hope to provide evidence for the early incorporation of prosodic information in parsing.

2.2 STIMULI

The test material was recorded using naïve speakers. Two male native speakers of British English sat facing each other in a booth. Both speakers had their own parts of a dialogue on cue sheets. An example of the material given to the speakers is provided below:

Speaker 1: Modern ships are never interesting.

Speaker 2: Sailing ships are always newsworthy.

Speaker 1: Sailing boats is no longer very newsworthy.

Speaker 2: Sailing ships is always newsworthy.

Each speaker was given only the parts he was intended to speak. The speakers read the sentences out at a normal conversational rate. Speaker 2 was told to listen to what Speaker 1 said before responding with the rebuttal on his sheet. The sentences on both sheets were randomised, but the order was maintained so that the order of what Speaker 1 said always cued what Speaker 2 was to say. The aim was for the context provided by the first speaker to cue the second speaker's interpretation, and thus, possibly also the pronunciation – of the ambiguous words. The dialogue also ensured that the speaker pronounced the stimuli reasonably naturally. They were also told to repeat the entire dialogue if they did not feel comfortable with a particular reading. Speakers were not given any other instructions. Any differences in prosodic patterns in Speaker 2's productions were naturally induced rather than being forced by explicit instructions.

I recorded twelve different pairs of syntactically ambiguous word pairs – 24 stimuli altogether. Each was repeated five times in the recording. Of the twelve paired sentences, six were pairs of sentences whose noun and verb phrase versions were prosodically contrastive as well as syntactically distinct (experimental stimuli). The other six pairs were prosodically non-contrastive, in that the change from a noun phrase to the verb phrase reading made no difference prosodically. These were control sentences to test the hypothesis that listeners could use prosody to differentiate the word pairs. The different stimuli are listed in Table 2.1.

Prosodically contrastive (Experimental Stimuli)	Prosodically non-contrastive (Control Stimuli)
Cooking apples	Breaking glasses
Cutting boards	Burning trees
Packing cases	Flying kites
Playing cards	Melting glaciers
Racing cars	Ringing bells
Sailing ships	Visiting relatives

Table 2.1 Word pairs presented to subjects

The recordings were made to recordable audio CD (CD-R) and were subsequently transferred to the computer with a sampling rate of 16 kHz at a resolution of 16 bits. The sentences were then edited so that each fragment consisted of just the word pair. Consequently there were few effects of editing in the stimuli that were finally played to subjects. There might still be co-articulation effects between the initial vowel of the auxiliary verb and the last syllable of the noun, which I address in Experiments 3 and 4. Since the material presented to subjects was restricted to only one speaker, there was no between-speaker variability. We used a linear prediction coding analysis of the stimulus. LPC spectral analysis of the data in this experiment was performed using a window of 160 samples, which overlapped every 80 samples – i.e. every 5 ms – to derive 22 filter parameters of which 16 were reflection coefficients. Once we obtained the waveforms corresponding to the different fragments, the LPC parameters of the fragments were loaded into an algorithm (Coleman: 2003) in order to find significant differences, if any, between the five tokens of the sentences. The sum of differences between each token and the other four informs us how different each token is from the other. The 'find centroid' program finds the token that has the lowest distance from the other four tokens. This was called the centroid of the five tokens.

As I have mentioned above, six pairs of the ambiguous fragments were prosodically contrastive, i.e. subjects could use prosodic information to differentiate between the noun and verb phrase interpretations of these stimuli. In this study, I shall concentrate on the acoustic features that might bias the listener towards a noun or verb phrase reading of the ambiguous word pairs. A number of previous studies have concentrated on the prosodic boundaries in speech, and how they might affect subjects' parsing of ambiguous stimuli (Carlson et al: 2001; Schafer et al: 2000; Schafer: 1997; Speer et al: 1996; Grabe et al: 1995; Price et al: 1991). However, prosodic boundaries are a step removed from the prosodic cues that this thesis focuses on – such as duration, f_0, and amplitude – since prosodic boundaries are themselves marked by differences in duration, amplitude, and f_0. Furthermore, since this thesis tests the parsing of ambiguous phrases, there were no major prosodic boundaries in the stimuli presented to subjects.

The word pair stimuli were three or four syllables long, with two syllables in the first word, and one or two syllables in the second word. There would obviously be considerable variation in how the word pairs were produced in different situations. However, I expected that the first word would be stressed in the noun phrase reading of the prosodically contrastive word pairs. Conversely, I expected that the second word would be stressed in the verb phrase reading of the word pairs. Again, there could be a number of different ways of producing the prosodically non-contrastive stimuli. However, I expected emphasis to be placed on the second word of both the noun and verb phrase versions of these word pairs. The results of analysis of the prosodic characteristics of the stimuli suggested that the speaker maintained this pattern in most cases.

Previous studies have argued that duration usually contributes towards marking the prosodic contrast between alternative interpretations of ambiguous stimuli (Lehiste: 1973; Klatt: 1976; Beach: 1991; Price et al: 1991). I expected that a similar trend would be observed in the analysis of the stimuli used in Experiment 1. Therefore, I measured the duration of the first word and second word of the noun and verb phrase versions of the word pairs (where possible). These values are given in Table A.1 in Appendix A. Obviously; I made separate measurements for the noun and verb phrase versions of the stimuli. There has been a lot of work on f_0 affecting parsing off-line (Bolinger: 1958; 1965; Hirschberg and Ward: 1991; Birch and Clifton: 1995; Slujiter et al: 1995; Schafer: 1997; Schafer et al: 1996; 2000; Maynell: 2000). I expected there to be significant differences between the f_0 of the first and second word based on which word the speaker intended to stress. Therefore, I also measured the timing of the peak accent in the first and second words, again, in both the noun and verb phrase versions. The timing of the peak accent was measured from the end of the word pair fragment in all cases. I also measured the f_0 of the peak accent of the first word, the f_0 of peak accent of the second word, and the difference in f_0 between the first and second word. The latter was calculated to find out whether a particular fragment had falling or rising pitch. I expected that falling pitch would be associated with the noun phrase interpretation and rising pitch would be associated with the verb phrase interpretation. There are some indications that loudness and amplitude do contribute to parsing (Streeter: 1978; Slujiter et al: 1995; 1996; Faulkner and Rosen: 1996; Kochanski et al: submitted). However, these are less certain than the studies on f_0 guiding parsing. Despite this, I also measured the peak RMS amplitude of the first word, the peak RMS amplitude of the second word, and the difference in peak RMS amplitude between the first and second word. I expected that if amplitude were used by the speaker to signal contrastive stress on one of the two words, difference in amplitude should always be positive in the noun phrase reading, and negative in the verb phrase reading. All these values are given in Table A.1 in Appendix A.

Figures 2.1 and 2.2 plot the duration of the first and second word in the noun and verb phrase versions, respectively. The duration of the first word was almost always greater in the noun phrase than it was in the verb phrase versions.

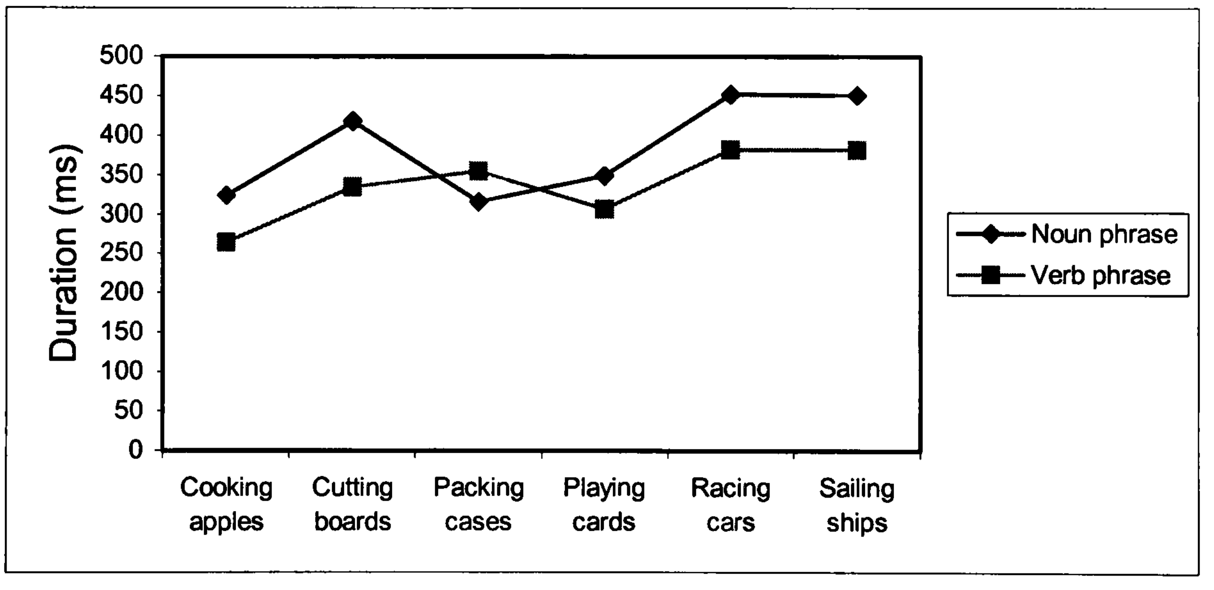


Figure 2.1 Duration of word 1 of the noun vs. verb phrase versions of the ambiguous pairs

Conversely, the duration of the second word was almost always greater in the verb phrase than in the noun phrase versions of the prosodically contrastive word pairs. Only in the recording of *Packing cases* was this reversed. However, in this case, the speaker marked the prosodic contrast between the two alternatives of Packing cases with very strong differences in amplitude between the two words in the noun and verb phrase forms (see Figure 2.5).

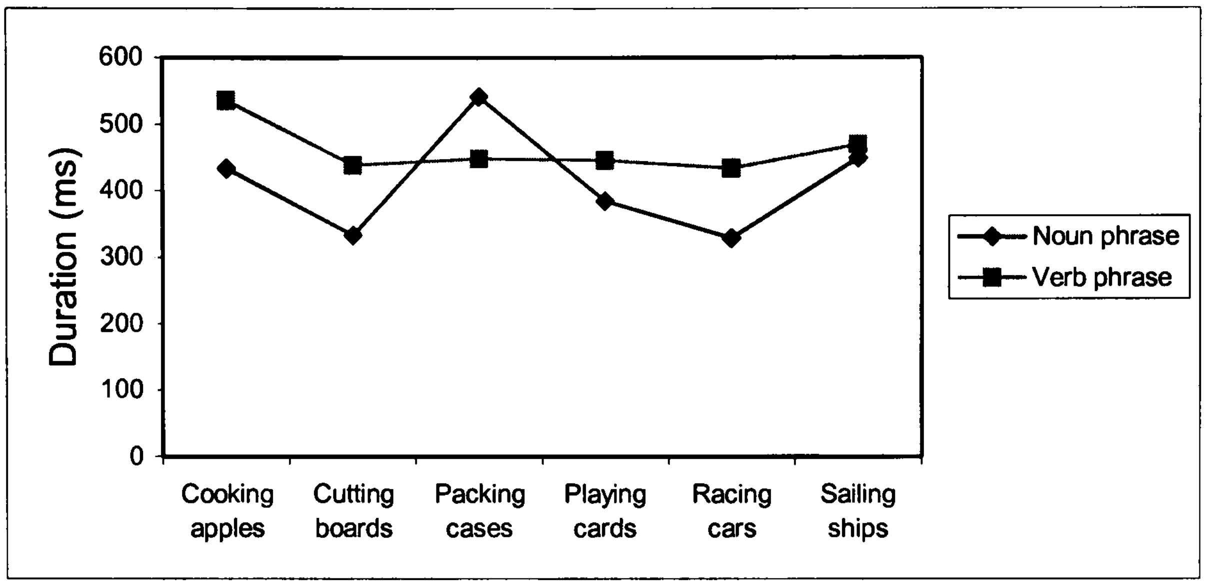


Figure 2.2 Duration of word 2 in the noun vs. verb phrase versions of the ambiguous pairs

I also measured the duration of the vowels in the noun and verb phrase versions of the word pairs. The duration of the first vowel was also almost always greater in the noun phrase than it was in the verb phrase reading. Similar patterns were found in comparing the durations of the three other vowels. The duration of the vowel common to the first word of all the word pairs (in *ing*) was almost always greater in the noun phrase condition than it was in the verb phrase condition. Measurement of duration was done manually and was therefore, subject to human error. These values can vary quite substantially with individual measurement. They are only meant to give the reader an idea of the difference between the first and second word in the two alternatives. There is definitely a trend towards differences in the duration of the different segments in the noun and verb phrase versions of the word pairs. This trend was consistent with the expectation that the first word would be stressed in the noun phrase and the second word would be stressed in the verb phrase reading of the prosodically contrastive word pairs.

Figures 2.3, 2.4, and 2.5 plot the peak RMS amplitude of the first and second word and of the difference in amplitude between the first and second word respectively, in both the noun and verb phrase condition. Figure 2.3 shows that the peak amplitude of the first word in the noun phrase was always greater than the peak amplitude of the first word in the verb phrase. Conversely, the amplitude of the second word in the verb phrase was always greater than the amplitude of the second word in the noun phrase, as we can see from Figure 2.4. This is consistent with the earlier observation that the first word was stressed in the noun phrase condition and the second word was stressed in the verb phrase condition. It is also consistent with suggestions that differences in amplitude can mark differences in stress (Slujiter et al: 1995; 1996; Beckman: 1996; Streeter: 1978; Kochanski et al: submitted). Figure 2.5 plots the difference between the peak amplitude of the first and the second word in both the noun and verb phrase condition for all stimuli. This difference in peak amplitude was always positive in the noun phrase reading – indicating that the first word had consistently greater amplitude than the second word. It was expected that this difference would always be negative in the verb phrase reading of the word pairs. However, two of the verb phrase versions of the stimuli had greater amplitude in the first than in the second word. This goes against the expectation that the second word would always receive greater stress in the verb phrase reading. However, in one of the stimuli – *Racing cars*, the difference in pitch between the two words in the verb phrase reading might make up for this positive difference in amplitude (see Figure 2.9). As for the positive difference in both amplitude and pitch in the verb phrase reading of *Cooking apples*, it is possible that the verb phrase reading might also be cued by a sustained high pitch or amplitude on the second word. This is contrasted with the drop in pitch and amplitude that is consistently displayed in the second word of the noun phrase reading. Therefore, it might not be necessary for the second word of the verb phrase to receive main stress as long as it is sufficiently marked.

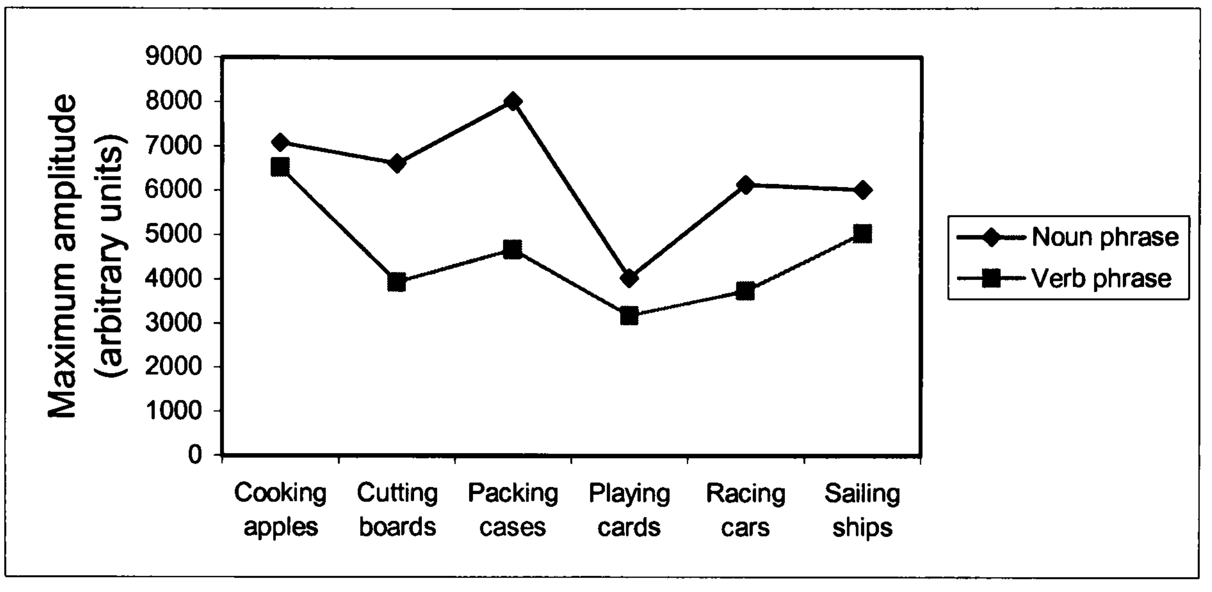


Figure 2.3 Peak amplitude of word 1 in the noun vs. verb phrase versions of the ambiguous word pairs

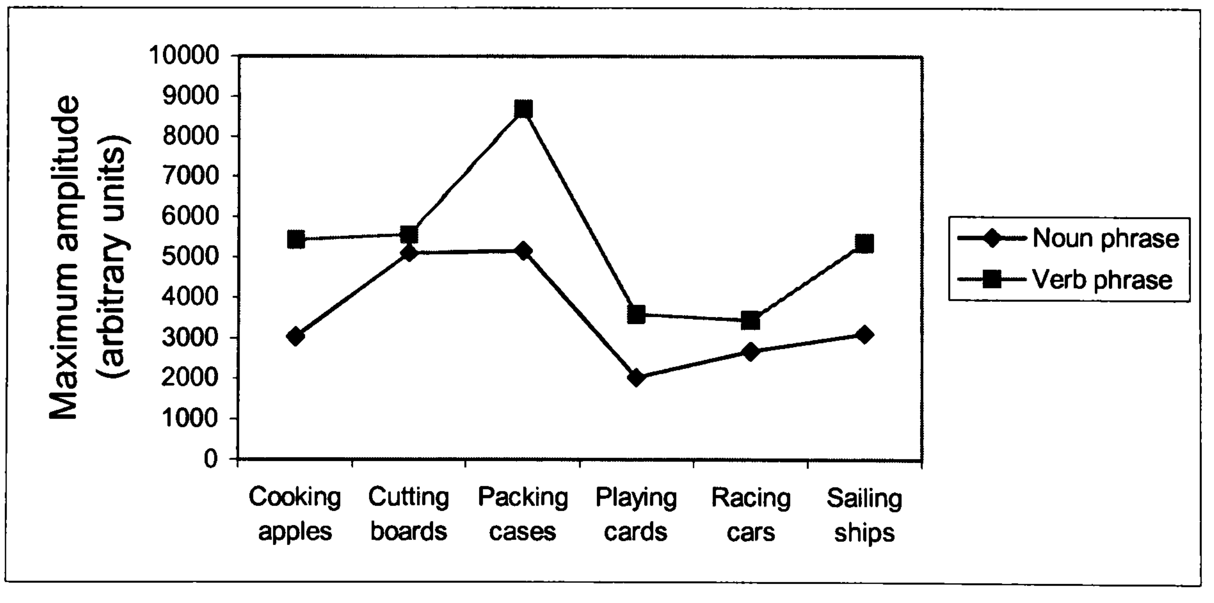


Figure 2.4 Peak amplitude of word 2 in the noun vs. verb phrase versions of the ambiguous word pairs

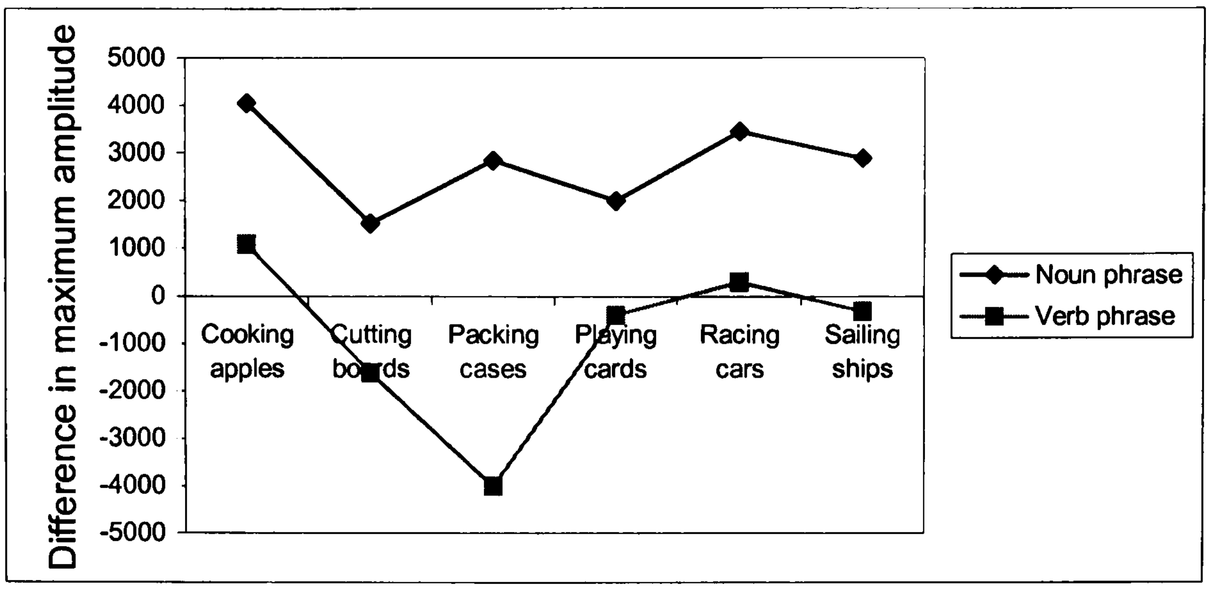


Figure 2.5 Difference in peak amplitude between word 1 vs. word 2 in the noun and verb phrase versions of the ambiguous word pairs

The peak accent of the word pair seemed to be considerably earlier in the noun phrase reading of four of the six word pairs than it was in the verb phrase reading. This is in keeping with the expectation that the first word would be stressed in the noun phrase reading and the second would be stressed in the verb phrase reading. Two of the stimuli did not follow the pattern of the others. However, comparison of Figure 2.9 with Figure 2.5 will show that noun and verb phrase reading of *Packing cases* was marked by a strong contrast in amplitude. The first word had greater amplitude in the noun phrase reading and the second had greater amplitude in the verb phrase reading (see Figure 2.5). Further, the noun and verb phrase reading of *Cooking apples* was marked by a strong contrast in duration, with the first word having greater duration in the noun phrase reading and the second having greater duration in the verb phrase reading. The difference in the timing of the peak accent in the word pair will become clearer when I discuss the differences in pitch across the word pairs.

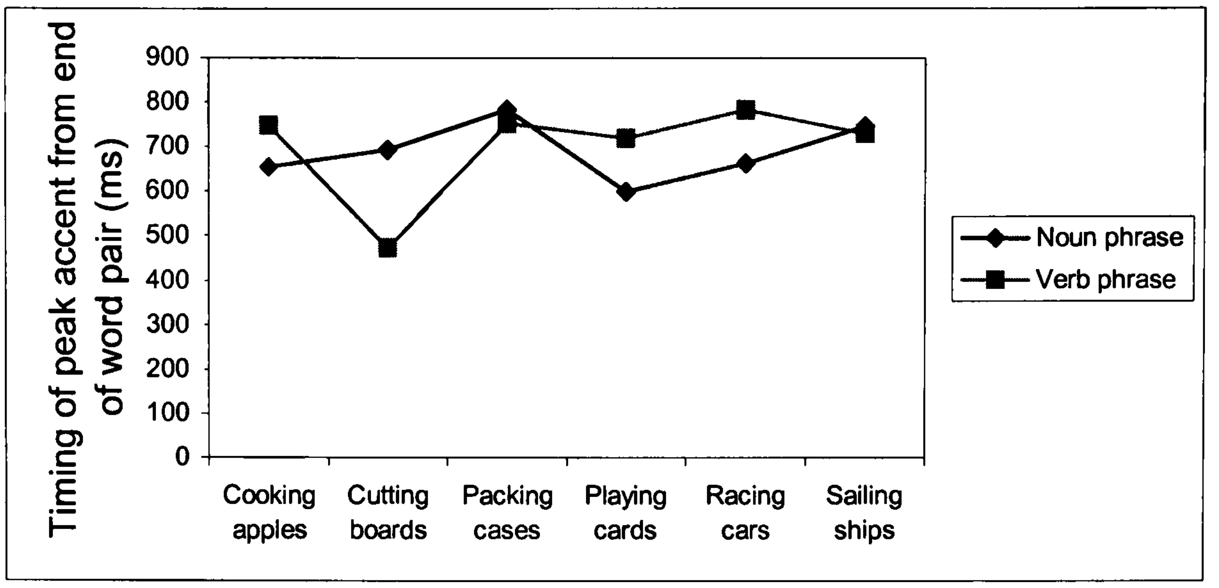


Figure 2.6 Differences in the timing of the peak accent on the word pair in the noun vs. verb phrase versions of the stimuli

Figures 2.7 and 2.8 plot the peak f_0 of the first and second words in the noun and verb phrase versions. Figure 2.9 plots the difference in f_0 between the first and second word. This difference was expected to be negative in the verb phrase versions and positive in the noun phrase versions – if f_0 is a reliable marker of prosodic contrast in these stimuli. Figure 2.7 shows that f_0 of the first word was always greater in the noun phrase reading than in the verb phrase reading. Conversely, f_0 of the second word was always greater in the verb phrase reading than in the noun phrase reading. Therefore, the first word had consistently higher pitch in the noun phrase reading and second word had higher pitch in the verb phrase reading. Figure 2.9 plots the difference in f_0 between the first and second word in the noun and verb phrase versions. This shows that this difference was always greater in the noun phrase than in the verb phrase reading, which also implies that the first word had higher pitch than the second word in the noun phrase reading. All these figures are consistent with the expectation that that the prosodic contrast between the word pairs would be marked by differences in duration, amplitude and f_0. However, two of the stimuli had greater f_0 in the first word than in the second word in the verb phrase versions of the word pairs – *Cooking apples* and *Packing cases*. This goes against the expectation that the nuclear accent of the word pair would be consistently on the second word in the verb phrase reading. However, Figure 2.6 shows that the small positive difference in the case of the verb phrase reading of *Packing cases* can be outweighed by the large negative difference in amplitude. Again, I am unsure as to what to make of the prosodic contour of *Cooking apples*. The prosodic contrast between the two versions *could* also be marked by a sustained high pitch and amplitude on the second word in the verb phrase reading. The prosodic contrast in this case could also be marked by the fairly consistent difference in duration that was found between the first word of the noun and verb phrase reading. The first word had greater duration in the noun phrase while the second word had greater duration in the verb phrase reading.

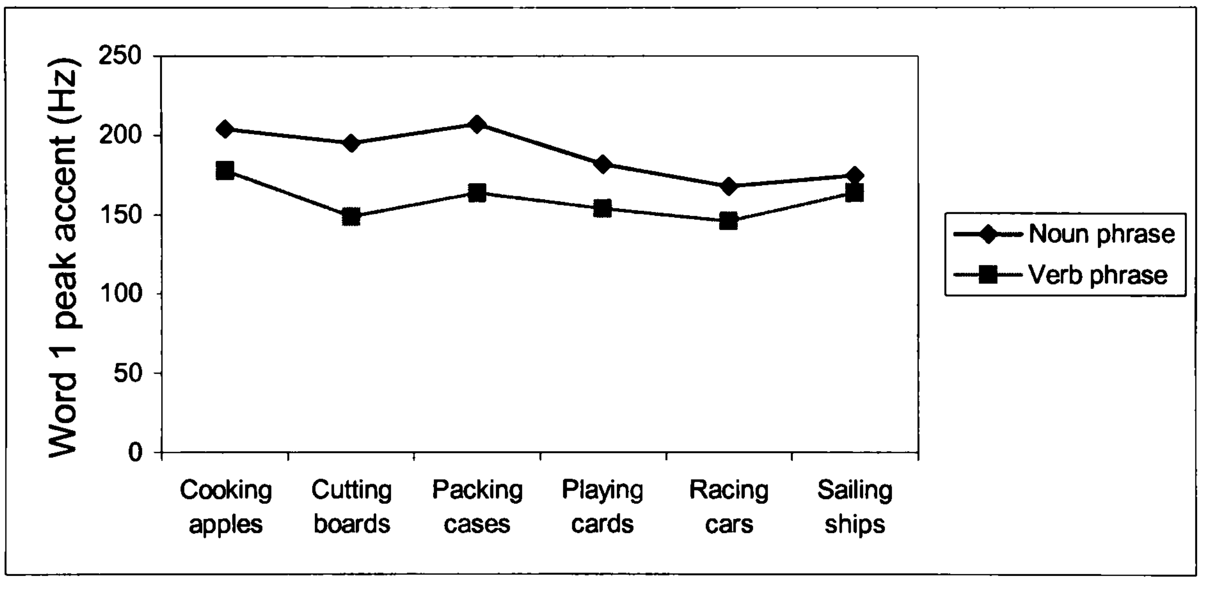


Figure 2.7 Peak word 1 f_0 in the noun vs. verb phrase reading of the ambiguous word pairs

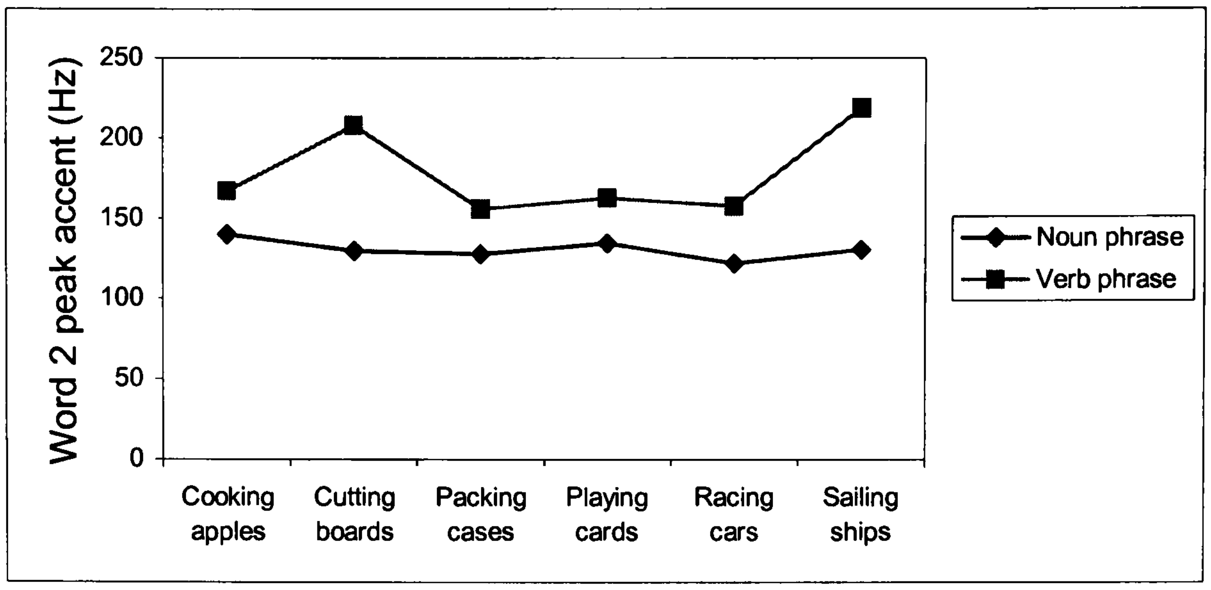


Figure 2.8 Peak word 2 f_0 in the noun vs. verb phrase versions of the ambiguous word pairs

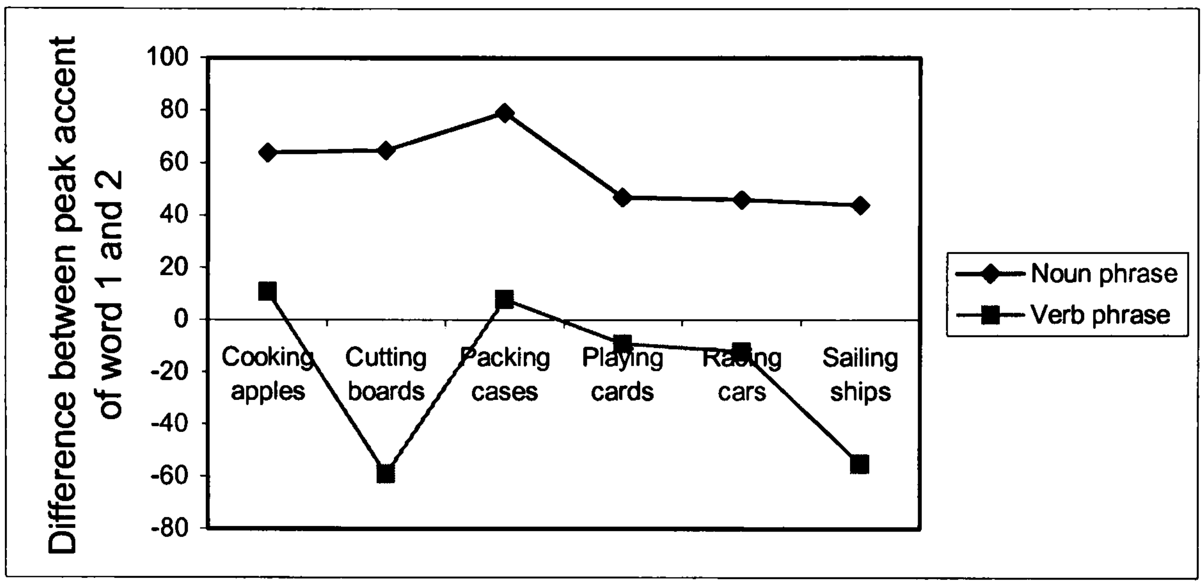


Figure 2.9 Difference in f_0 between word 1 and word 2 in the noun vs. verb phrase versions of the ambiguous word pairs

I also measured the formant frequencies of the middle of the vowels of the words in the noun and verb phrase forms. These values are presented in Table A.3 in Appendix A. While there are differences between the formant frequencies of the vowels in the noun and verb phrase versions, Experiment 1 does not concentrate on the influence of spectral characteristics on parsing. I shall discuss this in Experiment 3.

From the data reported in this section, it is quite clear that the expectation raised at the beginning of this section has been supported. The speaker consistently stressed the first word in the noun phrase form of the ambiguous word pairs and the second word in the verb phrase form. The speaker did not always use the same acoustic feature to mark the contrast between the noun and verb phrase alternations of the prosodically contrastive word pairs. However, there are suggestions that duration, f_0, and amplitude interact to mark contrastive emphasis. This brings up a troubling issue of the variation in the correlation of prosodic to syntactic structure: How is the listener to correlate the two if the same syntactic difference is marked by a difference in amplitude in one case and by a difference in duration in the other? I argue that this kind of variation is not as much of a problem as it is made out to be. The data indicates that speakers mark contrast with not just one but several prosodic features simultaneously. If one of the prosodic features is not by itself contrastive between two ambiguous syntactic alternatives, others might contribute to marking the same contrast between the ambiguous alternatives. This provides the listener with a range of cues to guide parsing of the structure of the ambiguous stimulus. Speech is reliably cooperative, although this cooperation might be marred by variation. It remains to be seen, however, whether listeners can equally reliably disambiguate the stimuli, given some variation in the acoustic characteristics of the word pairs. There is little doubt that listeners can use these prosodic cues to distinguish the two forms. However, what is in doubt is whether listeners can incorporate these cues early enough to guide initial disambiguation.

I also measured the prosodic characteristics of the prosodically non-contrastive control stimuli. These values are presented in Table A.2 in Appendix A. A more detailed description of the prosodic differences between the noun and verb phrase versions of the prosodically non-contrastive stimuli is presented in Appendix B. There did not appear to be any consistent differences between the prosodic characteristics of the noun and verb phrase versions of the prosodically non-contrastive stimuli.

There were syntactic differences between the word pairs. In the example below, *Packing* plays the role of an attribute adjective defining the noun in the noun phrase reading (e.g. *cases*). However, the verb phrase reading is a non-finite clause masquerading as a noun phrase and is the subject of the sentence it is a part of. Verb phrases that are sentential subjects are typically huge PRO subject chunks. The structures of the sentences from which the noun phrase and verb phrase versions of the ambiguous words are taken from are presented in the Figures 2.10 and 2.11 below.

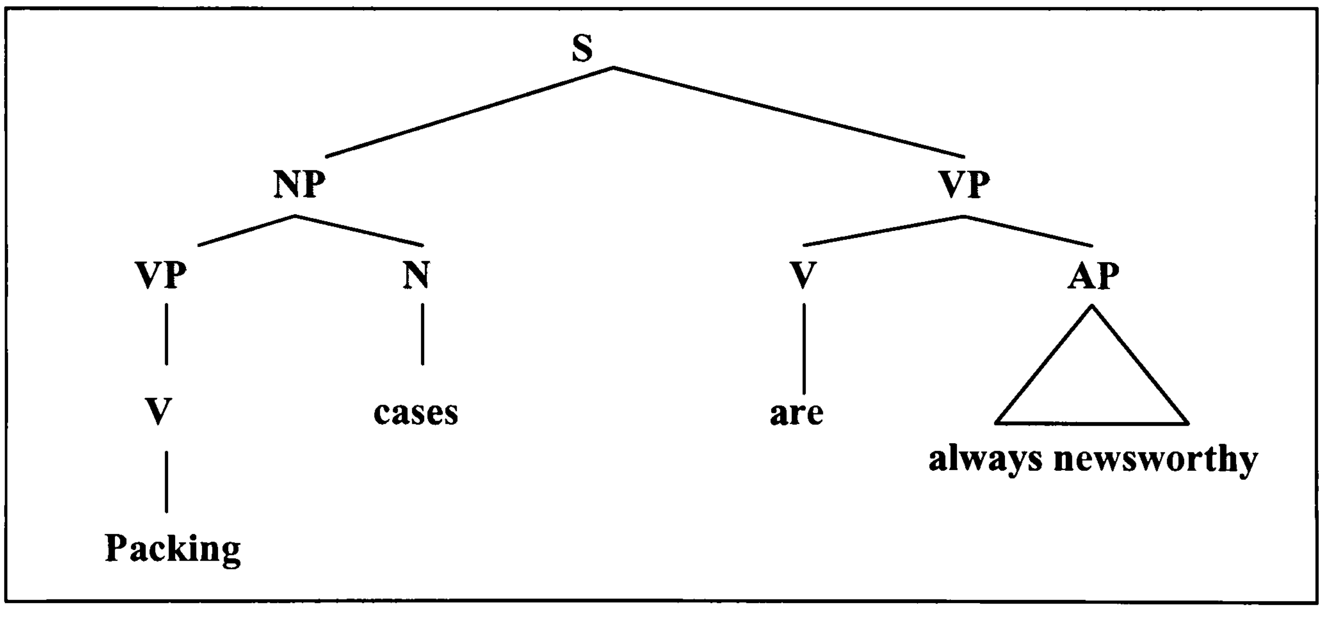


Figure 2.10 Syntactic structure of the original sentence from which the noun phrase reading was clipped

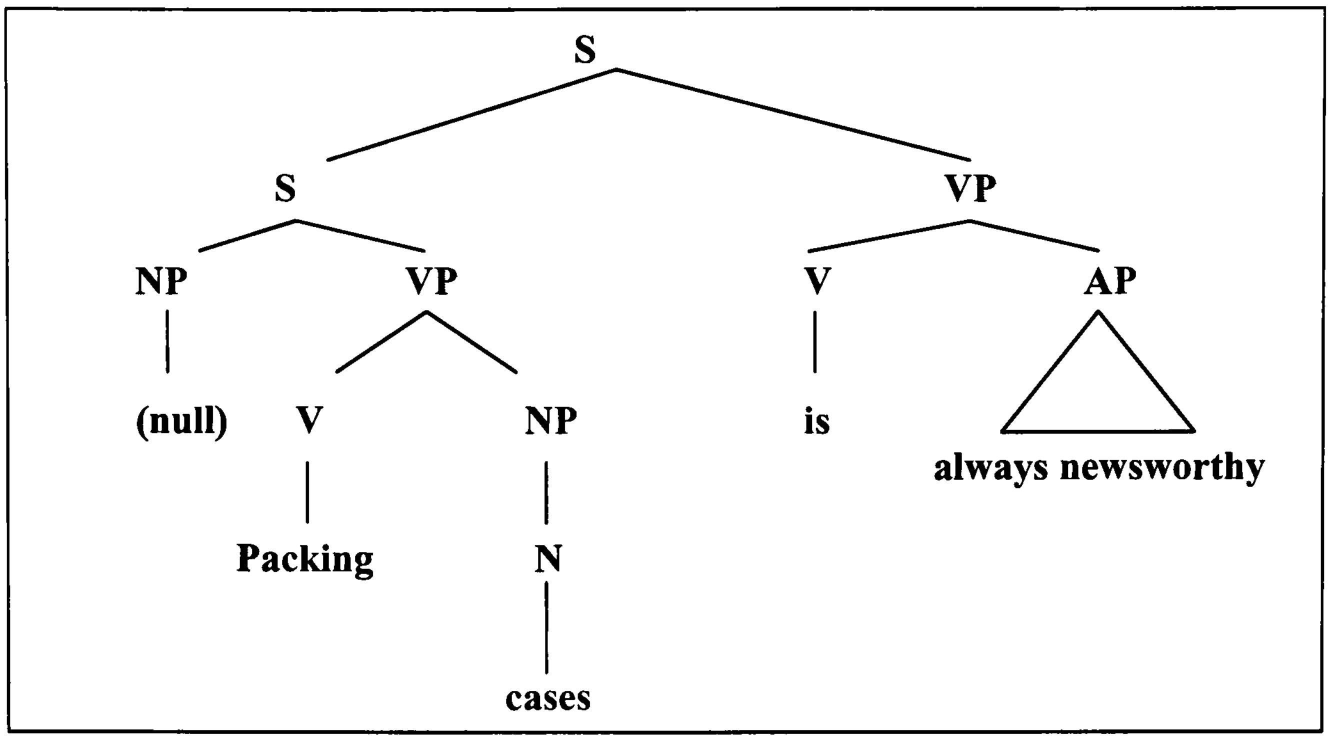


Figure 2.11 Syntactic structure of the original sentence from which the verb phrase reading was taken.

Figures 2.10 and 2.11 show the syntactic differences between the two forms. Carston (1989) argues that a parser following the principles of late closure and minimal attachment would prefer the noun phrase structure. The principle of late closure argues that the clause currently being processed is kept open for as long as possible. The word pair in the verb phrase reading closes the non-finite clause, violating the principle of late closure. The auxiliary verb cannot be added within the non-finite clause. Conversely, the noun phrase structure keeps the finite clause open to the addition of the auxiliary verb, in keeping with the principle of late closure. Moreover, the addition of this S node and the empty category NP subject in the verb phrase reading violates the principle of minimal attachment. Therefore, proponents of a syntax-first serial processing model would suggest that the noun phrase structure would always be suggested first. Experiment 1 tests whether the noun phrase structure is preferred over the verb phrase structure in order to challenge this view.

Conversely, according to a weak interactive parsing model, a syntactically autonomous parser might compute both structures. Both structures would have to be simultaneously checked for compatibility with the structures assigned by the prosodic contrast. Thus, both noun and verb phrase parses should be accessible at the same time. Swinney (1979) argues that multiple meanings of an ambiguous lexical item are simultaneously active during processing. If this extends beyond lexical access to syntactic structure assignment, then both parses should also be active at the same time. This mechanism of syntactic processing is termed 'weak interaction'. Indeed, Marslen-Wilson et al argued that the results of their experiment (Marslen-Wilson et al: 1992) could be explained by proposing a syntactically autonomous parser that creates all plausible parses of an ambiguous stimulus. Prosodic and pragmatic information is brought in subsequently to choose a preferred interpretation only after the parser has created all possible structures. Experiment 1 also tests whether both parses of the stimuli are simultaneously active in order to address this suggestion of weak interactive parsing.

2.3. PROCEDURE

I presented subjects with the centroid tokens of the noun and verb phrase versions of the prosodically contrastive and non-contrastive stimuli. The tokens were presented through a set of headphones. As in the experiment described in Marslen-Wilson et al (1992), a visual probe appeared on a screen in front of the subjects as soon as they finished listening to each word pair. The probe was always either 'is'/'are'. The probe 'are' is an appropriate continuation of the noun phrase versions of the word pairs, and the probe 'is' is an appropriate continuation of the verb phrase versions. However, only in the case of the prosodically contrastive stimuli was there adequate prosodic information to differentiate between the noun and verb phrase versions of the word pairs. There were few consistent prosodic differences between the noun and verb phrase versions of the prosodically non-contrastive words pairs.

Each subject heard sixteen repeats of each of the chosen fragments – both prosodically contrastive and non-contrastive. The fragments were presented in one of two random orders. Subjects were randomly assigned to one of two groups with a random order of presentation of the stimuli. In half of the presentations, the visual probes were consistent with the sentences that the fragments had been taken from. Therefore, fragments that were biased towards the noun phrase interpretation were followed by 'are' and fragments biased towards the verb phrase interpretation were followed by 'is'. For ease of exposition, I will refer to probes that were consistent with the sentences the fragments had been taken from as *appropriate probes*. In the other half of the presentations, the visual probes conflicted with the sentences the fragments had been taken from. Therefore, fragments biased towards the noun phrase interpretation were now followed by 'is'. Conversely, fragments biased towards the verb phrase interpretation were now followed by 'are'. I will refer to probes that conflict with the sentences that the fragments had been taken from as *inappropriate probes*.

The prosodically non-contrastive stimuli were also followed by appropriate probes in half of the presentations and inappropriate probes in the other half.

However, there were few consistent prosodic differences between the two versions of the non-contrastive word pairs for the probes to be inappropriate to either interpretation. Therefore, both visual probes were consistent with the interpretation consistent with the prosody of the stimuli. The probes were appropriate or inappropriate depending on whether the word pairs had been clipped from sentences where the word pairs were followed by 'is' or 'are' probes.

Subjects were seated in a sound proof room and provided with headphones over which the fragments were played. Subjects were given the following instructions by the experimenter before starting the experiment:

> *You are going to hear a number of sentence fragments over the headphones. At the end of the fragment, a single word will appear on the screen in front of you. This word will either be 'is' or 'are'. It will be a possible continuation of what you have heard up to that point. I would like you to listen to the sentence fragments normally, and when the word appears on the screen in front of you, you should press the button on the handset in front of you that matches the word on the screen as quickly as possible. In half of the presentations, the word that follows is a good continuation of what you have just heard. In the other half the word that follows is a bad continuation of what you have heard. The experiment is designed so that it does not record any material over a second, so you have a second to respond to what you hear. A second is a very long time and has been used in most experiments of this kind so far as the cut off*

point. I would like you to take care that you press the buttons in front of you with only one finger and that you do not have more than one finger or thumb near the buttons.

It is important that you concentrate on understanding the meaning of what you hear and the word on the screen in relation to the meaning of the fragment. After you have responded to the word, the screen goes blank for five seconds, during which time I would like you to mark on the paper in front of you whether the word was a good or bad continuation of what you heard.

During a set of pilot tests [not documented here, but reported in Mani (MPhil, 2003)], subjects were asked to make a decision of whether the word was a *good, not so good, not so bad* or *bad* continuation of what they heard. However, it was found that subjects were confused by the number of options. Therefore, the appropriateness-rating task was changed to a *good / bad* decision.

Subjects were paid to participate in the experiment, which lasted around forty minutes per subject. All subjects were native speakers of British English from Southern England and were between 21 to 25 years of age. There were fifteen subjects in all – eight female and seven male – and each subject heard 384 fragments. All subjects were right-handed, in order to control for handedness in the button-pressing task. As soon as each subject arrived, he or she was randomly assigned to one of two different groups. The groups differed in the order of presentation of the material to control for any effects of repetition of stimuli. The groups also differed in the labels of the response buttons that matched with the probe verb to control for left/ right button biases. In the first group, subjects pressed a red button on the left to respond to the 'is' probe and a black button on the right to respond to the 'are' probe. This was swapped around in the second group to control for any possible colour bias. Subjects were put through a practice run of 16 fragments to familiarise them with the task and to check they had understood the task. The fragments tested in the practice run were not repeated in the main experiment. An appropriateness-rating task was included so that subjects did not concentrate only on pressing the right button. This was to focus subjects' attention on the visual probe as a possible continuation of the auditory stimulus. The time taken by the subjects to respond to the probes in each of the conditions was recorded.

2.4. HYPOTHESIS

Since the word pairs were taken from sentences biased towards one of the interpretations of the word pairs, the prosodic characteristics of the stimuli should be consistent with the intended interpretation of the stimuli. Experiment 1 tests whether subjects can use the prosodic contrast between the noun and verb phrase versions of the contrastive word pairs to determine the correct parse of the stimuli. The visual probes are consistent with one of the two interpretations of the stimulus ('are' is consistent with the noun phrase reading and 'is' is consistent with the verb phrase reading). If subjects are guided by prosody, then their responses to the probe is expected to be delayed if the parse consistent with the probe is inconsistent with the parse indicated by the prosody of the fragment. Steinhauer et al (1999) argues from

ERP data that mismatch between prosody and syntax necessitates syntactic and lexical reconfirmation. This should cause a delay in subjects' responses to an inappropriate probe. In contrast, if the parse consistent with the probe is consistent with the parse indicated by the prosody, then subjects' responses to the probe should be facilitated. Therefore, I predict that response times to appropriate probes following the prosodically contrastive word pairs should be shorter than response times to inappropriate probes.

However, the prosodically non-contrastive word pairs do not provide subjects with adequate prosodic information to guide disambiguation. If subjects are using solely prosodic information, then they should not be able to disambiguate the word pairs. Therefore, their response times to appropriate and inappropriate probes following the control word pairs should be similar. This would indicate that subjects were unable to determine the correct parse of the stimuli. The parse preferred by subjects would be the parse confirmed by the probes, and this preference would be established only upon presentation of the probe.

Conversely, if subjects were using a "syntax-first" approach, they might observe the principles of late closure and minimal attachment and would consistently prefer the noun phrase reading of the word pairs (Carston: 1989). Consequently, they should take less time to respond to probes confirming the noun phrase interpretation, i.e. 'are'. Subjects should show a preference for the noun phrase interpretation of the word pairs, irrespective of whether the stimuli are prosodically contrastive. The "syntax-first" hypothesis would hold that subjects do not consider the prosodic characteristics of the stimuli during initial processing. However, an arbitrary syntax-first approach would provide the wrong parse of half the stimuli. Nevertheless, it is possible that the absence of adequate prosodic information stimuli might force subjects towards a noun phrase parse of the control stimuli. This would mean that subjects would show faster responses to 'are' than to 'is' following only the prosodically non-contrastive stimuli.

Alternatively, a syntactically autonomous parser might construct all plausible parses of the input using *solely* syntactic information. Prosodic information would be subsequently incorporated to choose a preferred interpretation of the ambiguous stimulus. Similar to proposals made in some lexical access models (Swinney: 1979; Binder and Morris: 1995; Binder and Rayner: 1998), a weak interactive parsing mechanism would compute multiple syntactic interpretations on-line. Both noun and verb phrase interpretations of the word pairs would be simultaneously active. Subjects should not have a preference for either form of the word pair. They should assign structure to the word pair according to the continuation the probe confirms. Therefore, subjects should take the same amount of time to respond to both 'are' and 'is' probes, irrespective of whether the stimuli are prosodically contrastive.

Therefore, analysis of subjects' responses will pay special attention to whether the results might also be explained by a "syntax-first" parsing mechanism. Specifically, I will test whether subjects prefer the noun phrase structure of the stimuli or whether both structures are simultaneously active. Experiment 1 explores the interaction between prosody and syntax in disambiguation. I propose that the prosodic differences between the noun and verb phrase versions of the prosodically contrastive stimuli are sufficient to guide parsing. Furthermore, these prosodic preferences should establish themselves before the application of syntactic parsing preferences. It seems unnecessary to inflexibly follow the principles of late closure and minimal attachment, when they would predict an inaccurate parse of half the stimuli. In the case of the prosodically non-contrastive word pairs, it is possible that the absence of prosodic preferences might force subjects to determine a parse using solely syntactic information. In the next section, I present the results of Experiment 1, and in Section 2.5, I argue that subjects do not seem to follow these syntactic principles in parsing the word pairs tested here even in the absence of adequate prosodic information.

2.5. RESULTS

18 out of 5760 responses were discounted because of incorrect button pressing (pressing the ‘are’ button for an ‘is’ cue or vice-versa). 64 responses later than 1500 milliseconds were excluded in order to ensure that subjects’ responses were immediate and not off-line. The mean response times to appropriate and inappropriate probes following the prosodically contrastive and non-contrastive stimuli are presented in Table 2.2. Figure 2.12 plots the difference in mean response times to probes that were consistent with the sentences the fragments had been clipped from (appropriate) and probes that conflicted with these sentences (inappropriate) (see Table 2.2).

Stimuli	Mean response time to appropriate probes (ms)	Mean response time to inappropriate probes (ms)
With prosodic contrast	551	573
No prosodic contrast	560	564

Table 2.2 Mean response times to appropriate and inappropriate probes following prosodically contrastive vs. non-contrastive stimuli (ms)

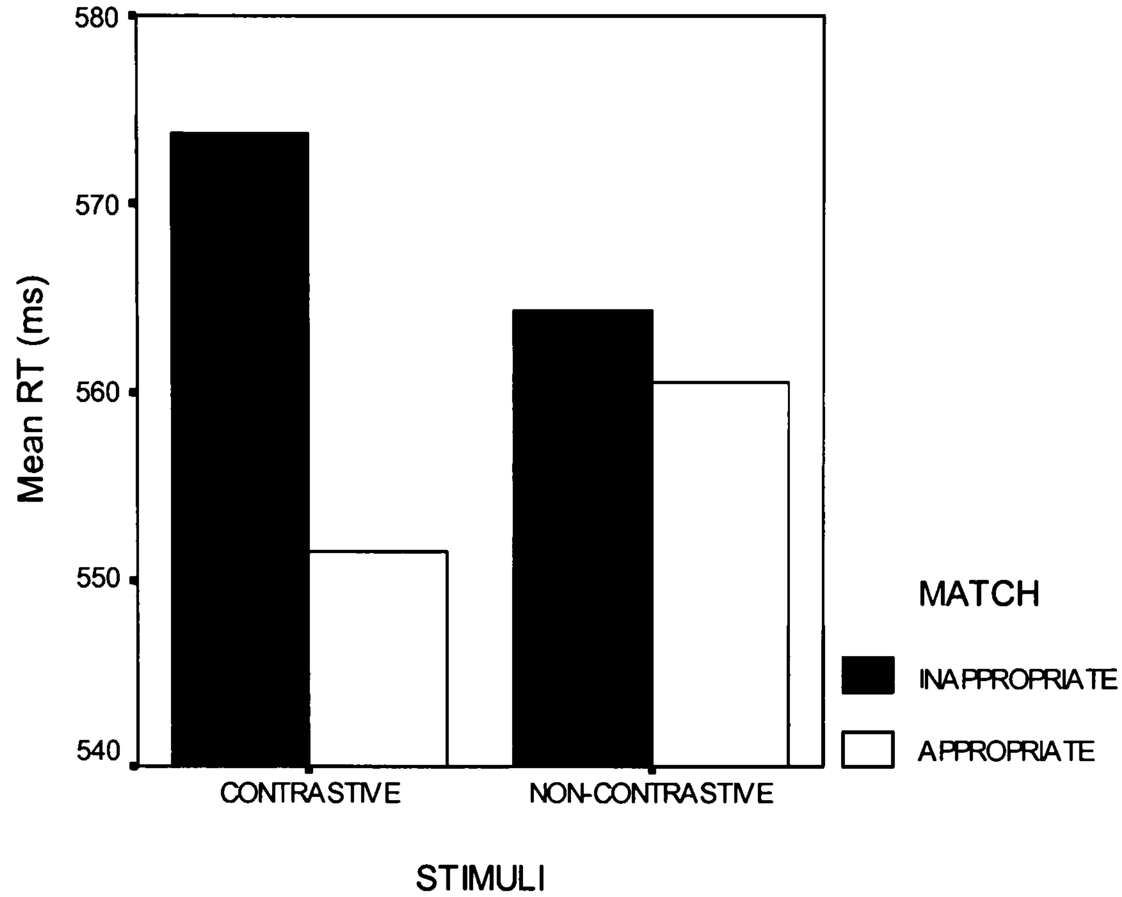


Figure 2.12 Mean response times to appropriate and inappropriate probes following the prosodically contrastive vs. non-contrastive stimuli

The mean response times to appropriate and inappropriate probes following the prosodically non-contrastive word pairs are remarkably similar. However, the mean response times to appropriate and inappropriate probes following the prosodically contrastive stimuli are significantly different. The response times were analysed to see if the results were normally distributed or whether there were significant deviations from the normal distribution. Normal distribution is a required condition for the appropriate use of parametric tests. As is usual with response time data, the responses were not normally distributed around the mean. The Kolmogorov-Smirnov test for normality proved significant ($p < .05$). Therefore, the results of parametric tests such as t-tests or ANOVAs would have to be treated with caution. However, non-parametric tests do not rely on the distribution being normal. Therefore, I conducted only non-parametric tests in this thesis. A non-parametric test analysis of the data revealed that response times did vary significantly according to whether the probes were consistent with the sentences that the fragments had been clipped from ($z = -3.726$; $p < .001$). The tests suggest that there is a less than 0.1% probability that the difference between the means was due to chance. As Table 2.4 shows, subjects took less time to respond to probes that were consistent with the sentences the fragments had been taken from than when they conflicted with these sentences, when the two versions of the sentences were prosodically contrastive.

For the prosodically non-contrastive control stimuli, the mean response times to probes consistent with the sentences the fragments had been taken from was similar to the mean response times to probes conflicting with these sentences. A Mann Whitney test of the responses to the prosodically non-contrastive word pairs revealed that the response times did not vary significantly according to the appropriateness of the probe to the sentences the fragments had been taken from ($z = -.752$; $p > .1$).

I also conducted a separate test to check if responses varied according to whether subjects preferred noun or verb phrase parses of the input. I found that subjects' responses to 'is' probes were faster than subjects' responses to 'are' probes.

(Mean RT (are) = 568 ms; Mean RT (is) = 556 ms; $z = -2.837$; $p < .005$). I also tested whether subjects' preference for 'is' probes was maintained within stimulus-type. I found that subjects' responses to 'is' and 'are' probes following the prosodically contrastive stimuli were similar (Mean RT (are) = 561 ms; Mean RT (is) = 564 ms; $z = -.512$; $p > .5$). However, subjects did seem to prefer verb phrase interpretations of the prosodically non-contrastive stimuli (Mean RT (are) = 576 ms; Mean RT (is) = 548 ms; $z = -4.579$; $p < .001$).

Item	Mean response times appropriate probes (ms)	Mean response times inappropriate probes (ms)	p
Cooking apples	519	563	**p < .005**
Cutting boards	541	551	p >.1
Packing cases	543	551	p > .1
Playing cards	525	540	p > .1
Racing cars	524	560	**p < .01**
Sailing ships	511	525	p < .1

Table 2.3 Mean response times to appropriate and inappropriate probes following prosodically contrastive stimuli per item.

Subject	Mean response times Appropriate probes (ms)	Mean response times Inappropriate probes (ms)	p
1	457	500	**p < .01**
2	778	734	p > .05
3	530	596	**p < .01**
4	685	726	**p < .05**
5	420	426	p > .05
6	502	486	p > .05
7	425	430	p > .05
8	427	466	**p < .05**
9	469	492	p > .05
10	513	556	**p < .05**
11	475	459	p > .05
12	561	592	p > .05
13	612	636	p > .05
14	594	615	p > .05
15	840	909	**p < .01**

Table 2.4 Mean responses for appropriate and inappropriate probes to prosodically contrastive stimuli per subject

I also conducted within-item (Table 2.3) and within-subject (Table 2.4) analyses to check whether the pattern of responses that was obtained for all the data was replicated per item and per subject. Subjects showed significantly different mean response times to appropriate and inappropriate probes following only two of the prosodically contrastive stimuli. Further, only six out of fifteen subjects showed significantly different response times to appropriate and inappropriate probes. This does not mean that the other subjects were not able to use prosodic information to guide parsing. For all except three subjects and for all items, responses to appropriate probes following the prosodically contrastive stimuli were faster than responses to inappropriate probes – in keeping with the pattern of overall responses. I suspect that the number of data points obtained per subject and per item was not adequate to establish true significance. This suggestion is supported by the fact that almost all the subjects showed shorter mean response times to appropriate than to inappropriate probes. I also tested to see whether there was any effect of order on subjects' responses. I found no significant correlation between subjects' responses to the probes and the order of presentation of the stimuli ($z = -.190$; N (5678); $p > .05$).

An appropriateness-rating task was included to monitor subjects' off-line decisions about the appropriateness of the probe as a continuation of the auditory stimuli. This task was also included to focus subjects' attention on the visual probe as a continuation of the auditory stimuli. The results of the appropriateness rating task are provided in Table 2.5.

Evaluation	With prosodic contrast		No prosodic contrast	
	Appropriate probes	Inappropriate probes	Appropriate probes	Inappropriate probes
Good continuation	1057	406	814	773
Bad continuation	284	924	543	549

Table 2.5 Number of good and bad evaluations awarded to appropriate and inappropriate probes following the stimuli with vs. without prosodic contrast

Significantly more *good* evaluations were given to probes which were appropriate continuations than to probes that were inappropriate continuations of the prosodically contrastive stimuli ($t\ (2669) = 28.664$; $p \ll .05$). Furthermore, a much greater number of *bad* evaluations were given to probes that were inappropriate continuations of the prosodically contrastive stimuli than to appropriate continuations. As for the prosodically non-contrastive stimuli, a similar number of *good* evaluations were given to appropriate and inappropriate probes ($t\ (2677) = .419$; $p > .5$). A similar number of *bad* evaluations were given to appropriate and inappropriate probes.

2.6. DISCUSSION

In this experiment, I tested the early incorporation of prosody in parsing. Specifically, I predicted that the prosodic differences between the noun and verb phrase versions of the prosodically contrastive word pairs would guide subjects towards the correct parses of the stimuli. The ambiguous fragments retained the prosodic contour of the word pairs in the sentences that they had been edited from. Therefore, if subjects use prosody to guide parsing, they would prefer parses that were consistent with the sentences the fragments had been clipped from to parses that conflicted with these sentences. Subjects' response times did vary according to whether the probes were consistent with the parse indicated by the prosody of the fragments and this result was restricted to responses to the probes following only the prosodically contrastive word pairs. Response times were shorter when the probe was consistent with the sentences the fragments had been taken from than when they conflicted with these sentences, i.e. subjects' response times were shorter when the parse confirmed by the probe was consistent with the parse indicated by the prosody of the fragment than when the probe conflicted with this parse. Only prosodic information could have guided subjects towards the accurate parse of the stimuli. This indicates that subjects are able to use prosodic information to construct a parse of the fragment. The difference in response times to appropriate and inappropriate probes can be explained as follows. When subjects' parses conflict with the interpretation confirmed by the probes, subjects reconsider their parse of the ambiguous stimuli, which results in increased response times to inappropriate probes.

The stimuli also present subjects with lexical, and therefore, syntactic information to guide parsing. Carston (1989) argues that the noun phrase reading of the word pairs is consistent with the principles of late closure and minimal attachment. If subjects' parsing were controlled by a syntax-first approach, responses to probes confirming the noun phrase parse should be faster than responses to probes confirming the verb phrase parse. However, the tests showed that subjects' responses to 'are' were not faster than responses to 'is' following the prosodically contrastive fragments, providing evidence against the mandatory application of minimal attachment parsing. Perhaps the difference between the results obtained in Experiment 1 and Carston's (1989) reanalysis of Marlsen-Wilson and Tyler (1987) were due to Experiment 1 testing subjects' responses at the initial stage of parsing. In addition, testing the interaction between prosody and syntactic structure is certain to have caused a difference between Marslen-Wilson and Tyler's results and the results of Experiment 1 presented here. Carlson et al (2001) had argued that prosodic information was contained within the syntactic module, while thematic information was accessed only after initial parsing: Perhaps the presence of prosodic cues guiding parsing biased subjects away from a minimal attachment parse of the stimuli.

If a syntactically autonomous parser computed both noun and verb phrase structures of the input, subjects would not show preference for either continuation of the auditory input. This would indicate that the parser had created multiple interpretations on-line. Preference for one interpretation would then be decided by the continuation supported by the probe. This is because there might not have been adequate syntactic information to suggest a preferred parse to the parser. Therefore, if subjects were following this parsing mechanism, their response times to both 'is' and 'are' probes following both noun and verb phrase forms of the stimuli would be comparable. However, this was not the case: Subjects' responses to appropriate probes were faster than their responses to inappropriate probes. This suggests that subjects had constructed a preferred parse of the stimuli that was in keeping with the prosody of the word pairs even at the point of realisation of ambiguity. They had not computed both possible parses of the ambiguous stimuli simultaneously. This casts doubts upon suggestions of a syntactically autonomous parser guiding initial parsing. The results of Experiment 1 cannot be explained by a syntax-first approach enforcing the principles of late closure and minimal attachment or by a weak interactive parsing mechanism. Subjects' response times were not in keeping with the predictions of either of these two models.

For the prosodically non-contrastive stimuli, I predicted that the absence of prosodic information differentiating the two alternatives would make it impossible for subjects to disambiguate them. The results of analysis of the responses to the prosodically non-contrastive stimuli confirm the hypothesis that prosodic cues can guide initial parsing. Subjects took the same amount of time to respond to the probe following the control stimuli irrespective of whether the probe was consistent with the sentences the fragments had been taken from. This suggests that subjects were not able to construct the correct parse of the fragments presented to them . This was to be expected since the alternative versions were prosodically similar. Consequently, they did not provide the listener with sufficient prosodic cues with which to disambiguate the fragments. The stimuli were not disambiguated syntactically either. Subjects did not prefer noun phrase parses of the stimuli over the verb phrase parses. Furthermore, although subjects' responses to appropriate and inappropriate probes were similar, the results I discuss next indicate that both parses of the stimuli were not simultaneously available. The only difference between the prosodically contrastive and non-contrastive stimuli was that the latter did not have prosodic differences between the noun and verb phrase alternatives. Therefore, subjects' inability to predict an accurate parse of the stimuli suggests that their parsing of the prosodically contrastive stimuli was prosodically informed.

Contrary to the minimal attachment and late closure hypothesis, subjects preferred verb phrase parses of the stimuli. Separate tests within stimulus type established that subjects did not arbitrarily prefer verb phrase parses of the prosodically contrastive stimuli. However, they did prefer verb phrase parses of the prosodically non-contrastive stimuli. In recording the prosodically non-contrastive stimuli, the speaker placed the emphasis on the second word of the fragment in five out of six cases, as indicated below:

(2.3) Frying eggs are....

The same pattern of emphasis was used by the speaker in producing the verb phrase interpretation of the prosodically contrastive stimuli. This suggests that subjects might have been biased by the verb phrase versions of the prosodically contrastive word pairs towards verb phrase parses of the control stimuli. This also provides a simple explanation for why subjects preferred verb phrase parses of the stimuli in general. Subjects' overwhelming preference for verb phrase parses of the prosodically non-contrastive stimuli would have affected the significance of difference between responses to 'is' and 'are', since this result was obtained using responses to 'is' and 'are' probes following all the stimuli. Importantly, subjects do not seem to be guided by late closure or minimal attachment preferences even in parsing the control stimuli. I had expected that the absence of adequately contrastive prosodic information would guide subjects towards following late closure and minimal attachment; i.e. towards favouring the noun phrase interpretation. Conversely, subjects seem to give priority to the available, albeit limited prosodic information presented to them. This argues for an especially strong effect of prosody on parsing. There might be other circumstances where parsing is guided by a syntactically motivated parsing principle: I do not contest this claim. However, the present results challenge any suggestion that prosodic information cannot guide initial parsing. Experiment 1 establishes that subjects were able to differentiate between the noun and verb phrase form of ambiguous word pairs, where there is a sufficient prosodic contrast between the two forms.

The order of presentation of the stimuli is crucial in response time tasks of the sort described here. Repeated exposure to the stimuli could increase subjects' sensitivity to the prosodic contrast. Therefore, response times to inappropriate probes following the prosodically contrastive stimuli could increase substantially as the experiment progresses. Conversely, order could adversely affect prosodic sensitivity as subjects get used to the acoustic characteristics of the stimulus. They might then display reduced reaction times to inappropriate probes. However, the order of presentation of stimuli was not found to significantly affect the pattern of responses in Experiment 1. I did not find a significant correlation between the order of presentation of the word pairs and subjects' response times. This was to be expected since the stimuli had been randomised in two different orders and subjects randomly assigned to one of two groups to pre-empt any effects of order of presentation of stimuli.

Experiment 1 establishes that subjects can disambiguate the prosodically contrastive word pairs. This argues for the early incorporation of prosody during the analysis of ambiguous word pairs. At this point, it is natural to ask what acoustic characteristics subjects use to differentiate between the noun phrase and verb phrase interpretations. I have already outlined the acoustic characteristics that make up the contrast between the two forms in Section 2.1. There were consistent differences in pitch, amplitude, and duration that together mark the prosodic contrast between the noun and verb phrase versions of the word pairs. In most cases, all of these cues point towards the first word being stressed in the noun phrase reading and the second word being stressed in the verb phrase reading. However, in some cases, this contrast was not marked by all the prosodic cues simultaneously. This need not indicate that the specific acoustic cue does not contribute to the difference between the noun and verb phrase interpretation. Some of the verb phrase versions were marked by a sustained high pitch and amplitude over the second word. This could be compared to the sharp drop in pitch and amplitude in the noun phrase reading. Therefore, these acoustic characteristics were influential in guiding parsing, especially since subjects' preferred parses that were consistent with the prosody of the stimuli.

However, some of the word pairs were more likely to occur as noun phrases, while the others were more likely to appear as verb phrases in speech. Corpus analysis of the data from the British National Corpus suggested that there was indeed a lexical bias towards one of the interpretations in the stimuli. I shall address the possibility of lexical information influencing parsing in detail in Chapter 5. As far as the conclusions reached by this experiment are concerned, subjects' parses were consistent with the parses suggested by the prosody of the ambiguous stimuli. This was found irrespective of any effect of lexical bias. While there might be a simultaneous effect of lexical bias on parsing, this does not detract from the finding that early incorporation of prosody is possible during on-line parsing.

Further evidence for the likelihood of prosodic information guiding parsing in this experiment is provided by the results of the appropriateness-rating task. In their evaluations of the probes as appropriate continuations of the prosodically contrastive stimuli, subjects provided twice as many good evaluations for appropriate probes than for inappropriate probes and three times as many bad evaluations for inappropriate probes than for appropriate probes. From these evaluations, I conclude that subjects' parses of the prosodically contrastive stimuli were consistent with the parse intended by the speaker's prosody. However, subjects did not suggest correct parses of the prosodically non-contrastive stimuli in both on-line and off-line tests. Subjects gave an almost equal number of 'good' evaluations to appropriate and inappropriate probes, and similarly, gave as many 'bad' evaluations to appropriate and inappropriate probes. This was to be expected, as neither prosodic nor syntactic cues could have guided subjects' parsing of the prosodically non-contrastive stimuli. Subject's off-line parsing was consistent with their on-line parses of the different stimuli.

2.7. CONCLUSION

The results of this experiment provide evidence for prosodic information guiding parsing. These results are stronger than those reported in previous work. In Chapter 1, I proposed four requirements of any experiment claiming to test for the early incorporation of prosodic information in parsing. Experiment 1 therefore employed an on-line task, testing subjects' parses of sentence-initial word pairs at the point of recognition of ambiguity. Experiment 1 tested subjects' initial parsing of ambiguous input before the presentation of the main verb of the sentence. Indeed, Experiment 1 tested subjects' predictions of the number of the main verb. I also included a subsidiary appropriateness-rating task in order to focus subjects' attention on the cohesiveness of the auditory and visual stimuli. Despite all these constraints, I conclude from the positive results of this experiment that subjects use prosodic information to *initiate* parsing. Prosodic information does not merely support a 'post-sentence interpretive judgement process' (Clifton et al: 2002: 106).

I found no evidence of syntactic preferences guiding parsing. The results of Experiment 1 could not be explained using a weak interactive parsing mechanism. I tested subjects' parsing at the very initial stage of recognition of ambiguity and still found an effect of prosody. Therefore, it cannot be argued that a syntactically autonomous parser created all possible parses of the input. This does not mean that listeners do not use either of these syntactic parsing preferences during processing. Experiment 1 does not provide evidence that listeners cannot parse ambiguous stimuli using either a "syntax-first" or a multiple structure strategy. However, the fact that subjects do not seem to be using these strategies to parse the stimuli presented to them does contradict the mandatory application of these approaches. There is a possibility that lexical information might guide parsing of the stimuli. However, since I have been able to report a strong effect of prosody on parsing, the presence of a simultaneous effect of lexical information on parsing does not contradict the conclusion that prosodic information can guide parsing. In response to the first question posed at the beginning of this thesis, I argue that listeners use prosodic cues to differentiate between noun and verb phrase forms of ambiguous word pairs.

Moreover, listeners are able to do this on-line. This provides strong evidence in favour of the early incorporation of prosody in parsing.

Listeners use prosodic cues to parse the stimulus, irrespective of whether the same acoustic characteristics were consistently used to mark the prosodic contrast. In fact, I found that in some cases the acoustic characteristics did not mark the prosodic contrast consistently. Despite this variation, listeners were still able to use prosodic information to choose a preferred interpretation of the ambiguous word pairs. There were indications that the amplitude, f_0, and duration of the word pairs seem to contribute to reliably marking this contrast, albeit somewhat variably. However, more evidence is required to corroborate suggestions of these three acoustic characteristics influencing parsing. Therefore, in the next experiment, I test the contribution of amplitude, f_0, and duration in guiding parsing of the word pairs.

CHAPTER THREE
THE EFFECT OF CONFLICTING PROSODY ON PARSING: EXPERIMENT TWO

3.1. INTRODUCTION

The previous experiment showed that prosodic information influences on-line parsing of syntactically ambiguous stimuli. Experiment 1 tested for prosodically motivated disambiguation at the point of realisation of syntactic ambiguity using an on-line response time task. Syntactic information would have predicted a noun phrase structure of all the stimuli. Lexical bias would predict one of the interpretations of the stimuli all the time, i.e. the interpretation in keeping with the bias of the word pair. I argue that only the prosodic content of the stimuli could guide subjects towards the intended interpretations of the word pairs. Therefore, subjects' preference for accurate parses of the stimuli provides strong evidence for the early incorporation of prosodic information in parsing.

Experiment 2 aims to test this conclusion further. The stimuli used in the second experiment were re-synthesised versions of the original word pairs, in which the timing, amplitude, and f_0 of each of the noun and verb phrase versions of word pairs were swapped with the timing, amplitude, and f_0 contour of the alternative versions. The newly cross-synthesised stimuli retain the segmental and spectral information of the original word pairs. However, the main prosodic characteristics of the stimuli – timing, amplitude, and f_0 – have been swapped over. If prosodic information guides parsing, then swapping the prosodic characteristics should change subjects preferred parses of the stimuli. Subjects should now display a preference for 'is' following the noun phrase versions and 'are' following the verb phrase versions. Experiment 2 addresses the question: *Can prosodic information alone guide the on-line parsing of ambiguous stimuli*?

There is an important reason for testing subjects' parses of stimuli with the "wrong" prosody. Prior research points in the direction of vowel context 'colouring' the spectral characteristics of the fricative, while not changing perception of the fricative (Mann and Ripp: 1980). I had argued that only the prosodic content of the stimuli could guide subjects' parsing towards the intended interpretation in Experiment 1. However, this ignores the contribution of segmental and spectral cues to parsing. The word pairs were edited from sentences where they were followed by one of the probes: the noun phrases followed by 'are' and verb phrases followed by 'is'. There is a possibility that the last segment of the word pairs [s/z] might have been conspicuously co-articulated with the initial vowel of the verb to follow ('is'/'are'), giving subjects an additional cue to the appropriate continuation of the fragments (Shadle and Scully: 1995). Therefore, subjects' accurate disambiguation of the word pairs in Experiment 1 could have been caused by co-articulation and not prosodic information. Subjects' response times to probes that were consistent with the intended interpretation would be shorter than subjects' response times to probes that were not consistent with this interpretation. Schepman (1997) makes a similar observation of the results obtained by Marslen-Wilson et al (1992). Therefore, Experiment 2 tests whether subjects were being guided by spectral information in guiding parsing. Since the stimuli have been cross-synthesised, they retain the spectral characteristics of the non-cross-synthesised word pairs. Experiment 2 tests whether subjects' prefer parses that are consistent with the cross-synthesised prosodic characteristics to parses consistent with the spectral characteristics of the stimuli.

3.2. THE INFLUENCE OF DURATION, AMPLITUDE AND SPECTRAL INFORMATION ON PARSING

Earlier work by Hirschberg and Ward (1992) provides evidence of some influence of spectral information in guiding interpretation, albeit not disambiguation. They explored the importance of f_0, amplitude, duration, and spectral information in the perception of *incredulity* and *uncertainty* in phrases like 'Eleven in the morning'. In a production study, they found that both *uncertainty* and *incredulity* were characterised by an L* + H L H % contour. This means that the sentence is produced with a nuclear accent on *Eleven* (L* + H), and a phrase accent (L) and boundary tone (H %) after *morning*. They argue that this implies a lack of speaker commitment in the content of the message. Although both interpretations were represented by the same contour, they did find consistent prosodic differences between the two interpretations. Tokens of *incredulity* were shorter and louder with a larger f_0 than tokens of *uncertainty*. Therefore, listeners could use duration, amplitude, and f_0 to decide whether the tone of the stimuli was incredulous or uncertain. In order to assess the contribution of these features to listeners' ability to differentiate the tone of the two readings, the authors created a set of 16 sentences with contrasting f_0, amplitude, and duration. The stimuli were prepared by swapping the f_0, amplitude, and duration from the *uncertainty* reading onto the *incredulity* reading and vice versa. Each of the 16 sentences had one or more of the features of the alternative reading while retaining the other prosodic characteristics of the base reading. This gave a number of different combinations of the four features being tested in their research. Subjects were played the different sentences and asked to make either an '*incredulity*' or '*uncertainty*' judgement. They were allowed to listen to the stimuli as often as they wanted before making a decision. However, they were not allowed to change an already made decision. It was found that the maximum number of *incredulity* response times was made when the f_0 and spectral characteristics corresponded to the *incredulity* reading. Spectral characteristics seemed to be less influential than f_0. The authors conclude that f_0 is the most influential in guiding parsing with spectral characteristics coming a close second. However, there seemed to be no influence of duration and amplitude on parsing.

The hypothesis of Experiment 2 of this thesis is similar to the Hirschberg and Ward study. While simultaneously verifying the result of Experiment 1, Experiment 2 also tests the relative contribution of spectral characteristics, and other prosodic characteristics, specifically, amplitude, duration, and f_0 to parsing. Hirschberg and Ward's results provide evidence of f_0 and spectral information influencing interpretation. However, they employed an off-line test of listeners' interpretation of the tone of the speaker as either *incredulous* or *uncertain*. Off-line tests do not adequately test the incorporation of different kinds of information in parsing. Furthermore, Hirschberg and Ward do not test the contribution of f_0, duration, and amplitude to guiding *parsing*. They test the contribution of these prosodic characteristics to subjects' interpretation of the tone of the stimuli. Therefore, we require an on-line test of the possibility of spectral information guiding parsing away from the parses indicated by prosodic information. Similarly, we cannot conclude against the possibility of duration and amplitude guiding parsing using the results of an off-line test. Experiment 2 fills this gap in research by presenting an on-line test of the relative contribution of prosodic and spectral information to parsing.

Previous research has shown that there are significant differences in the duration of identical lexemes in differing syntactic contexts of an utterance, especially depending on whether the lexemes precede a syntactic boundary or not (Lehiste: 1973; Warren: 1985; Klatt: 1975; Cooper et al: 1978; Schepman: 1997; Gee and Grosjean: 1983). Identical lexical items tend to be longer before a syntactic boundary than when they occurred within a syntactic constituent. For example, Lehiste (1973) found significant differences in duration of segments and pauses (or periods of laryngealisation) between the constituents of alternative parses of sentences like.

(3.1) The old men and women stayed at home

She found that the duration of the word *men* increased when it preceded a syntactic boundary than when it did not: in the two interpretations of (3.1), the duration of *men* was longer when the adjective *old* applied only to *men* than when it applied to *men and women*. However, Lehiste also found that the difference between the two interpretations was more marked when speakers were told to indicate the difference in meaning by different ways of producing the sentence. Furthermore, in an off-line perception test, she found that listeners were better at disambiguating productions of the latter than productions that were more natural. The possibility of duration guiding parsing is obviously contingent on these differences being present in natural speech. In addition, Price et al (1991) argue that differences in duration were only found to correlate with minor syntactic boundaries. Major syntactic boundaries were more frequently and consistently marked by differences in f_0. This contradicts earlier findings by Lehiste (1973) reporting significant differences in the duration of segments in contrasting syntactic context. Furthermore, most of the previous work on duration and syntactic ambiguity were production studies (Warren: 1985; Klatt: 1975; Cooper et al: 1978; Sorenson et al: 1978). Therefore, they did not actually test the effect of duration on parsing (apart from Schpeman: 1997). However, they conclude in favour of the production of consistent differences between the duration of identical lexical segments in contrastive syntactic situations.

Gee and Grosjean (1983) presented subjects with sentence fragments that had been taken from sentences where they concluded the sentence or where they were followed by more words. They found that subjects could predict whether the fragments were sentence final or whether they were followed by more words. Importantly, Gee and Grosjean found that subjects' predictions were correlated with differences in the duration of the segments in sentence-final or sentence-medial position. However, the finding that subjects' predictions were correlated with differences in duration does not provide strong evidence that these differences in duration were influential in guiding subjects' parsing. It merely supports suggestions that duration could influence parsing.

Schpeman (1997) also presents results of off-line and on-line tests of prosody guiding parsing. The fourth experiment of her doctoral thesis concludes that prosody can guide parsing. Later experiments presented in the same thesis contradict this result, with the last experiment of nine concluding in favour of syntactic parsing preferences guiding parsing. However, neither her off-line nor her on-line tests specifically addressed the possibility of duration guiding parsing by manipulating the duration of the stimuli.

Similarly, there is some evidence against the possibility of amplitude guiding parsing. This is unsurprising given the loss in perceived loudness that can easily be caused by the speaker turning their head away mid-articulation or walking away. This cannot be equated with a reduction in f_0 or duration that is not as easily affected by external circumstances. The argument against the perceptual salience of differences in amplitude is corroborated by prior research on production data (Warren: 1985). Warren found no consistent difference in amplitude between alternative interpretations of syntactically ambiguous stimuli. Perception experiments have also not concluded in favour of a significant influence of differences in amplitude on parsing. Streeter (1978) tested the perception of the sentence 'A plus E times O'. Speakers were either asked to read out one of two formulae given below:

(3.2) (A + E) * O

(3.3) A + (E * O)

The f_0, amplitude, and duration of the different readings were then crossed to produce different combinations of stimuli. These re-synthesised sentences were then presented to subjects. The preparation of the stimuli was similar to the methods used later by Hirschberg and Ward (1992). However, contrary to Hirschberg and Ward, Streeter found that differences in duration were strongly influential in reversing subjects' interpretations. f_0 was also moderately influential in similarly reversing subjects' preferences for one of the interpretations. However, amplitude did not have any effect on parsing, with differences in amplitude yielding no changes in the perception of the boundary. The results of Hirschberg and Ward (1992) and Streeter (1978) present similar evidence against the possibility of differences in amplitude guiding parsing.

However, recent evidence argues that differences in amplitude and intensity are perceptually salient, at least as far as the perception of stress is concerned (Beckman: 1996; Slujiter et al: 1995; 1996; Kochanski: submitted). The first two studies tested the effect of varying amplitude on the perception of stress. The latter analyses the cues used by speakers to mark prominence in speech. Beckman (1996) tested the factors affecting the perception of stress by English and Japanese subjects. Specifically, she tested the perceptual salience of f_0, amplitude, and duration. However, amplitude was calculated as a measure of power over the duration of the vocalic nucleus. This integrates durational and amplitude cues, rather than using independent measures of amplitude. Beckman terms the former a measure of 'total amplitude'. She reports that English subjects were significantly affected by differences in total amplitude in the perception of stress.

Conversely, Sluijter et al (1995, 1996) argue that simple changes in overall intensity do not affect the perception of stress. Rather, they argue that only a rise in intensity within certain specified bands of frequency has an effect on the perception of stress. They argue that rises in intensity in lower frequency bands are poor indicators of stress. In a production study, consistent rises in intensity were found in higher frequency bands whenever the speaker intended to place stress on an item. The authors also conducted a perception test to check whether listeners were consistently able to perceive these differences in intensity at higher frequency bands. They tested whether listeners perceived changes in intensity on the initial or final syllable of words like *nana* as changes in stress. The stimuli were prepared by varying the intensity by three, six, or nine dB in bands over 500 Hz in either the initial or final syllable of nonsense words. Subjects were presented with sentences like 'Wil je nana zeggen?' [Will you *nana* say?]. They were asked to indicate whether the main stress on the word was on the initial or final syllable of *nana*. They found that listeners were able to use differences in intensity in the higher frequency bands to detect the placement of main stress on the word.

Kochanski et al (submitted) analysed the correlation between speakers' use of five acoustic measures (loudness, duration, aperiodicity, spectral slope and f_0) and perception of prominence by classifiers trained to pick out prominent syllables in speech. Speech stimuli were taken from corpus data of read and spontaneous productions. The classifiers based on the different acoustic measures were then tested on a new set of speech data in their marking of prominent syllables. They noted the rate of errors made by classifiers in picking out prominent syllables and compared the performance of the classifiers based on different acoustic measures. This provides an estimate of speakers' use of the different acoustic measures to mark prominence. They found that the classifiers based on loudness performed at least 50% better than all the others did. The classifier based on f_0 produced unimpressive results. They also found that the classifiers based on loudness and duration performed better when the window of the sample included the neighbouring syllables of the prominent syllable. This indicates that prominence depends on the difference in loudness between syllables and their neighbours.

While the results reported by Beckman (1996); Sluijter et al (1995; 1996) and Kochanski et al (submitted) are promising for research on the perceptual salience of amplitude, intensity and loudness respectively, none of them test the contribution of amplitude to parsing. The results reported by Streeter (1978), on the other hand, provide evidence against the perceptual salience of amplitude in parsing. This is corroborated by the results of Hirschberg and Ward (1992). However, these were off-line tests of the effect of differences in amplitude on parsing. It is possible that an on-line test of parsing will be able to provide stronger evidence of the possibility of amplitude guiding disambiguation.

As the summary of literature presented above will indicate, not much research has tested the effect of manipulated or cross-synthesised prosody on parsing. Most experiments have tended towards testing the effect of cross-spliced prosodic information on parsing. Cross splicing involves swapping a fragment from one sentence into another. In experiments testing the effect of prosody on parsing, the fragments that are swapped are usually lexically identical, although they could show small differences in lexical pronunciation. This makes the prosodic content of an utterance conflict with the semantic and syntactic content of the base utterance. For instance, Kjelgaard and Speer (1999) cross spliced the ambiguous fragment 'When Roger leaves the house' from the late closure parse in (3.4) into the sentence in (3.5):

(3.4) When Roger leaves the house, it's dark (Late Closure)

(3.5) When Roger leaves, the house is dark (Early Closure)

Subjects were presented with sentences with co-operating (normal), baseline, and conflicting prosody (cross-spliced). It was expected that subjects would consider the prosodic content of the baseline stimuli appropriate to either interpretation of the sentences, since they were normal sentences with reduced prosodic cues. They studied speeded phonosyntactic grammaticality judgements, end-of-sentence comprehension, and cross-modal naming. The first task asked listeners to rate sentences played to them as *okay* or as an *error*. They measured subjects' judgements of the sentences as well as the amount of time they took to decide. Subjects took the least time to respond to the task when played the sentences with co-operating prosody. Sentences with conflicting prosody were judged erroneous. They also found that early closure sentences with conflicting prosody took longer in all tasks. The end-of-sentence decision tasks and the grammaticality judgement tasks are subject to the familiar objection that they do not test for the ability of prosody to guide initial parsing. They are more pertinent to prosodic reanalysis *after* initial syntactic analysis.

Kjelgaard and Speer also conducted an on-line test, similar to Marslen-Wilson et al (1992). They measured the time taken by subjects to respond to 'it's' and 'is' probes following different versions of 'When Roger leaves the house'. In table 3.1, I present transcriptions of the acoustic stimuli used in their on-line experiment. The probes that followed the sentences are also presented in the table.

Condition	Auditory fragment	Probe
Cooperating prosody		
(a) Early closure	((When Roger **H*** leaves **L-)$_{PPh}$ L %)$_{IPh}$** ((the house	is
(b) Late closure	((When Roger **H*** leaves the house **L-)$_{PPh}$ L%)$_{IPh}$**	it's
Baseline prosody		
(c) Early closure	((When Roger **L+H*** leaves **L-) $_{PPh}$** (the house	is
(d) Late closure	((When Roger **L+H*** leaves the house **L-) $_{PPh}$**	it's
Conflicting prosody		
(e) Early closure	((When Roger **H*** leaves the house **L-)$_{PPh}$ L%)$_{IPh}$**	is
(f) Late closure	((When Roger **H*** leaves **L-)$_{PPh}$ L %)$_{IPh}$** ((the house	it's

Table 3.1 Transcriptions of auditory fragments and visual probes for the cross-modal experiment from Kjelgaard and Speer (1999)

In the on-line cross-modal task, subjects took the same amount of time to respond to probes confirming early ('is') and late closure ('it's') interpretations following the fragments with co-operating prosody (Table 3.1 (a) and (b)). Furthermore, response times to the probes following the fragments with conflicting prosody (Table 3.1 (e) and (f)) were longer than response times to the probes following the fragments with co-operating prosody (Table 3.1 (a) and (b)). They also found that there was an advantage for fragments with conflicting late closure prosody: response times were significantly faster following late closure fragments with conflicting prosody (Table 3.1 (f)) than early closure sentences with conflicting prosody (Table 3.1.(e)).

However, despite claiming to test for an effect of conflicting prosody on parsing, Kjelgaard and Speer (1999) do little more than repeat Marslen-Wilson et al (1992). While the prosody of the stimuli conflicts with the interpretation confirmed by the probes, Kjelgaard and Speer do not test the effect of prosody conflicting with the spectral and segmental content of the stimuli. In addition, they find an unexplained advantage for fragments with conflicting late closure prosody over fragments with conflicting early closure prosody. If initial parsing did take place using prosodic information, then there should be no advantage for late closure sentences. The only effect should have been delayed response times to sentences with 'conflicting' prosody – as found in Marslen-Wilson et al (1992). Kjelgaard and Speer propose that the IPh boundary was disturbed for the early closure conflicting prosody condition because of the way the stimuli were cross spliced: cross splicing might have removed important phonetic information of the IPh from the end of the early closure sentences (see Table 3.1 (e)). Since the IPh occurs earlier in the late closure sentences (Table 3.1 (f)), cross-splicing preserved the phonetic information required for parsing.

Therefore, late closure sentences provided subjects with more information for parsing, giving these sentences an advantage. However, the auditory stimuli presented to subjects in the 'co-operating' prosody sentences were also truncated at the same point. Therefore, there should have been an advantage for early closure 'co-operating' prosody sentences. No such effect was observed and Kjelgaard and Speer do not offer any explanation for this. It is possible that a late closure advantage is found in parsing stimuli in the difficult processing conditions caused by the conflicting prosodic content of the stimuli. Initial parsing might have taken place using solely prosodic information. When these parses conflict with the interpretations confirmed by the probe, late closure structures might be suggested to aid disambiguation.

Experiment 1 of this thesis also tested subjects' response times to probes that conflicted with the interpretation appropriate to the prosody of the stimuli. No advantage for late closure fragments was found. Similar results in favour of prosodic information influencing parsing were reported by other experiments using cross-spliced stimuli (Shapiro and Nagel: 1995; Nagel et al: 1994; Speer et al: 1996). These authors argued that conflicting prosody delayed parsing in off-line and on-line tasks. However, in a number of cases, it is impossible to distinguish the effects of acoustic disruptions produced by cross splicing from effects of prosody. Secondly, the experiments reported by Kjelagaard and Speer (1999), and indeed, Shapiro and Nagel (1995); Nagel et al (1995; 1995) and Speer et al (1996) did not test whether conflicting prosody leads to an alternative parse of the stimuli. They all tested whether conflicting prosody delays parsing. They tested subjects' response times when the prosodic content of the stimuli conflicted with the interpretation confirmed by the probe. Consequently, the results of previous cross-splicing studies could also be explained by a parsing mechanism where prosodic information is given an inhibitory role – where prosodic information can inhibit the construction of a parse suggested by syntactic information, but cannot suggest a probable parse of the stimuli.

EXPERIMENT TWO

In Experiment 2, I test subjects' parsing of stimuli with cross-synthesised, rather than cross-spliced prosody. Cross-synthesised stimuli have the f_0, timing and amplitude of one of the versions of the word pair swapped with the f_0, timing and amplitude of the alternative version. This retains the spectral and segmental content of the stimuli, while swapping three main prosodic features of one of the versions of the word pair with the other. Experiment 1 found that subjects were being guided by prosodic information in on-line parsing. Therefore, if prosodic information were guiding parsing, swapping the prosodic content of the syntactically contrastive stimuli should lead subjects to alternative parses of the word pairs. This would provide conclusive evidence of the early incorporation of prosodic information in parsing. I test whether conflicting prosodic information can *guide* initial parsing away from the original interpretation of the stimuli.

Experiment 2 also tests the relative contribution of spectral, segmental, and prosodic information in guiding parsing. There were differences in the spectral characteristics of the stimuli. Table A.3 in Appendix A presents the spectral characteristics of the different vowels of the word pairs in the noun phrase versions of the stimuli. Table A.4 in Appendix A presents the spectral characteristics of the different vowels in the verb phrase versions. There are also differences between the formants of the vowels. These differences are quite large and consistent in some cases. However, it is uncertain whether these differences were significant enough to guide parsing. Therefore, Experiment 2 tests whether the spectral differences between the word pairs can guide parsing. Hirschberg and Ward (1992) argued for a moderately influential role for spectral information in parsing using an off-line task. Experiment 2 tests a similar hypothesis to Hirschberg and Ward using an on-line task. I test whether subjects' parses are in keeping with the spectral and segmental characteristics of the stimuli or with the prosodic characteristics of the stimuli. In addition, there was a possibility that the final fricative of the word pairs was co-articulated with the initial vowel of the auxiliary verbs following the fragments in the original sentences. This might provide subjects with an additional cue towards the intended interpretation of the word pairs. Therefore, subjects' parsing in Experiment 1 might not have been caused by prosodic information, but by the cues provided by co-articulation. Again, there were differences between the spectral characteristics of the word-final fricatives in the noun and verb phrase version. These measures are in Table A.5 in Appendix A. However, these differences were not always consistently on the same formant of the segments. Therefore, it is not possible to suggest whether these differences were sufficient to guide parsing. Experiment 2 tests whether the effects explained by the early incorporation of prosody in parsing in Experiment 1 could be explained by co-articulation between the initial vowel of the probe and the final fricative of the word pairs.

3.3. STIMULI

As in Experiment 1, I recorded twelve pairs of syntactically ambiguous word pairs – 24 stimuli altogether. These 12 pairs of sentences were each repeated five times in the recording. Of the twelve, six were pairs of sentences whose first two words were prosodically minimally contrasting as well as syntactically distinct (Experimental stimuli). The other six pairs were of sentences whose accentual alternations were prosodically non-distinct: there were no prosodic differences between the noun and verb phrase versions of the stimuli (Control stimuli: e.g. *Frying eggs*).

The speaker was a Southern British male, of the same age and social background as the speaker in Experiment 1. As in Experiment 1, recordings were made to recordable audio CD (CD-R) and were subsequently transferred to the computer with a sampling rate of 16 kHz at a resolution of 16 bits. The sentences were then edited so that each experimental fragment consisted of an adjective and a noun (the noun phrase), or a verb and a noun (the verb phrase). LPC analysis was performed using a window of 160 samples, which overlapped every 80 samples – a temporal overlap every 5 ms – to derive 22 filter parameters, of which 16 were reflection coefficients. The centroid of the five recorded tokens of each of the items was calculated using the method in Experiment 1.

Each of the centroid stimuli was then cross-synthesised using an in-house program. The second speaker's productions cross-synthesised better than the speaker in Experiment 1. Cross-synthesis swapped the timing, f_0, and amplitude of each of the word pairs with the acoustic characteristics of the alternative reading of the word pair. First, the word pairs were time-warped, so that the durations of all parts of the noun phrase forms of the word pairs were either lengthened or shortened to the durations of the corresponding verb phrase versions. Similarly, the durations of the verb phrase versions were time-warped to fit the durations of the original noun phrase versions of the word pairs. Time-warping repeats or deletes the required number of frames from one of the versions of the word pair to match the number of frames in the other version of the word pair. Therefore, the timing of the two versions is now the same. Time-warping only manipulates the duration of the stimuli. The time-warped stimuli were then re-synthesised with the f_0 and amplitude contour of the alternative reading of the word pair. The re-synthesised stimuli thus had the timing, f_0 and amplitude contours of the alternative interpretation of the word pair.

The measures of duration, amplitude, f_0 of the word pairs are presented in Table A.1 in Appendix A. The measures of the duration, amplitude and f_0 of the cross-synthesised word pairs are presented in Table A.6 in Appendix A. In addition, Figures 3.1 and 3.2 present the f_0 of the noun and verb phrase version of *Racing cars*. Figures 3.3 and 3.4 present the f_0 of the noun and verb phrase version of the cross-synthesised *Racing cars*. The peak accent of the original noun phrase reading of the word pair falls on the first word. The peak is around 170 Hz and is near the end of the first syllable (Figure 3.1).

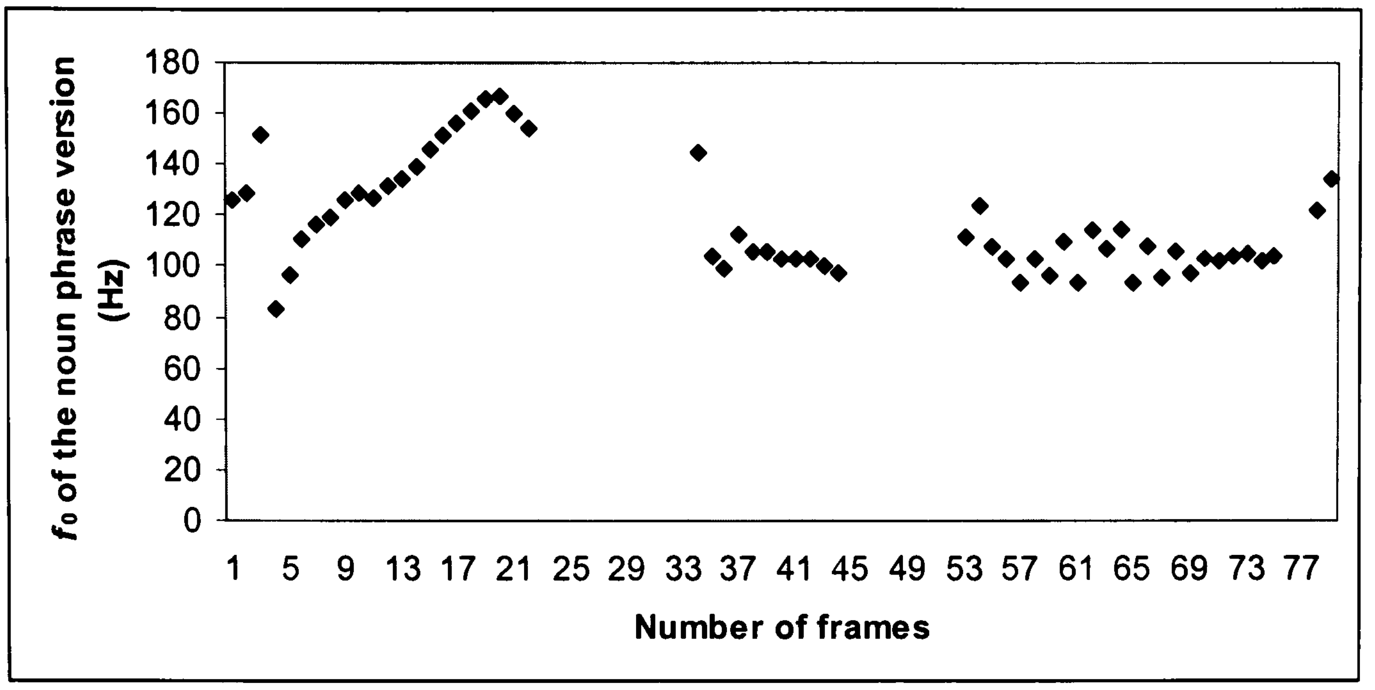


Figure 3.1 f_0 of the noun phrase version of *Racing cars*

Conversely, the peak accent of the verb phrase version falls on the second word. Again, the peak is around 160 Hz on the first syllable (Figure 3.2).

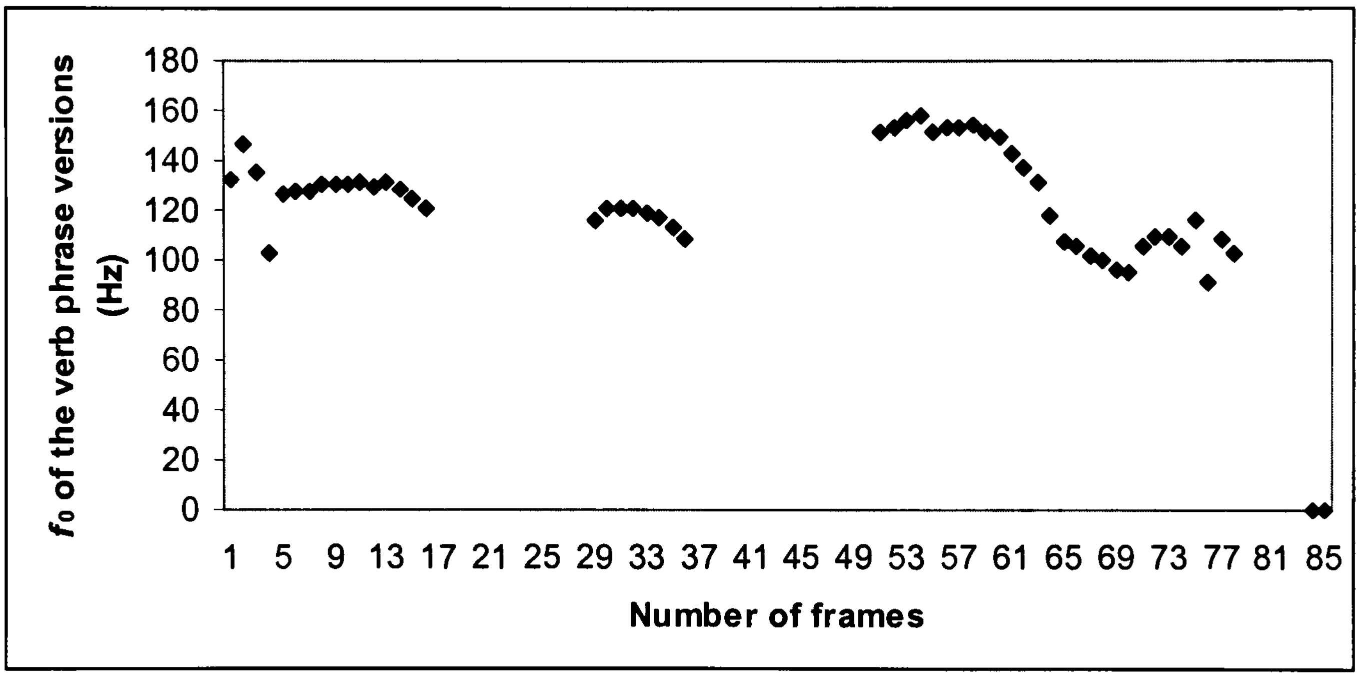


Figure 3.2 f_0 of the verb phrase version of *Racing cars*

The pitch contour of the original noun phrase reading has an initial steep rise followed by a fall, with the pitch levelling out over the rest of the word pair. Conversely, the pitch of the cross-synthesised version of the noun phrase is level throughout the first word. The peak accent is on the first syllable of the second word (Figure 3.3).

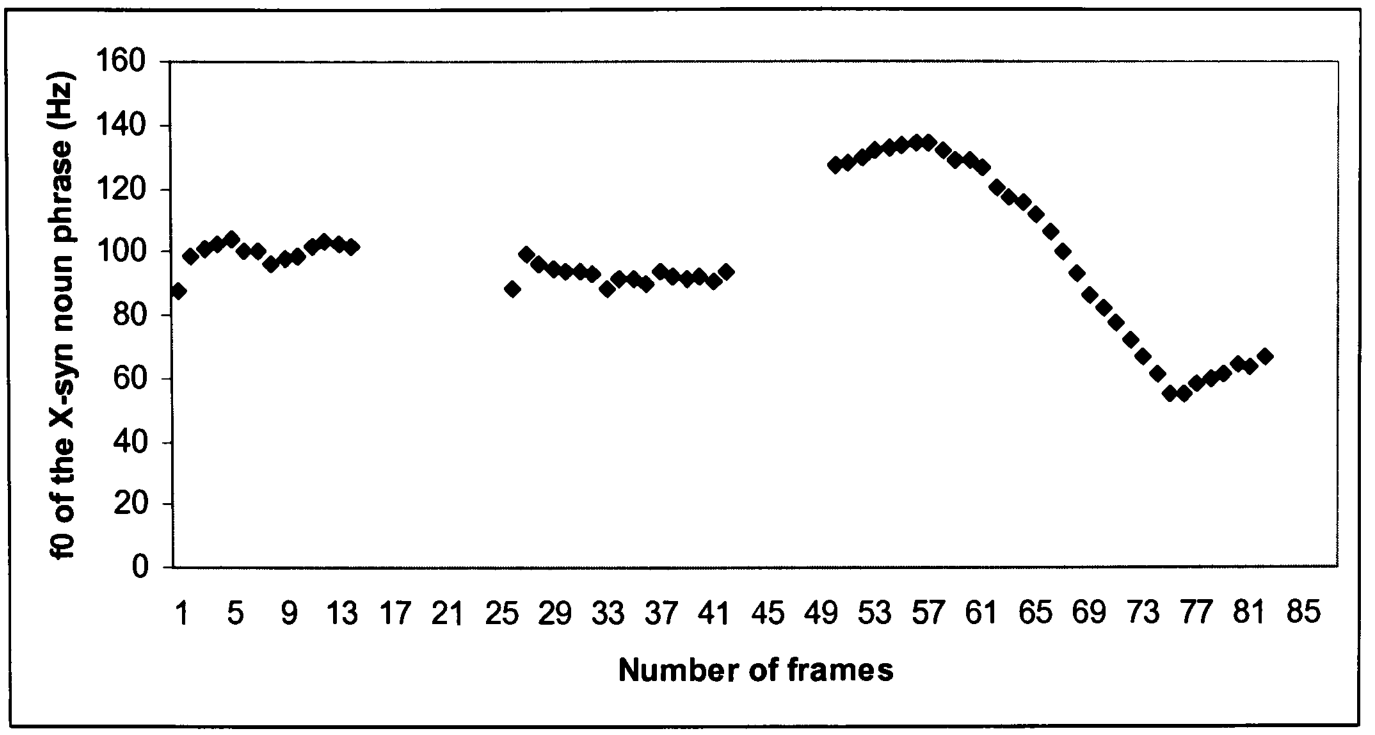


Figure 3.3 f_0 of the cross-synthesised noun phrase version of *Racing cars*

This is very similar to the contour of the original verb phrase version (Figure 3.2). Conversely, the cross-synthesised verb phrase version follows the contour of the original noun phrase version. The peak accent is at the beginning of the first syllable of the first word (Figure 3.4). This is followed by a fall through the rest of the word and a lower peak on the final syllable of the first word.

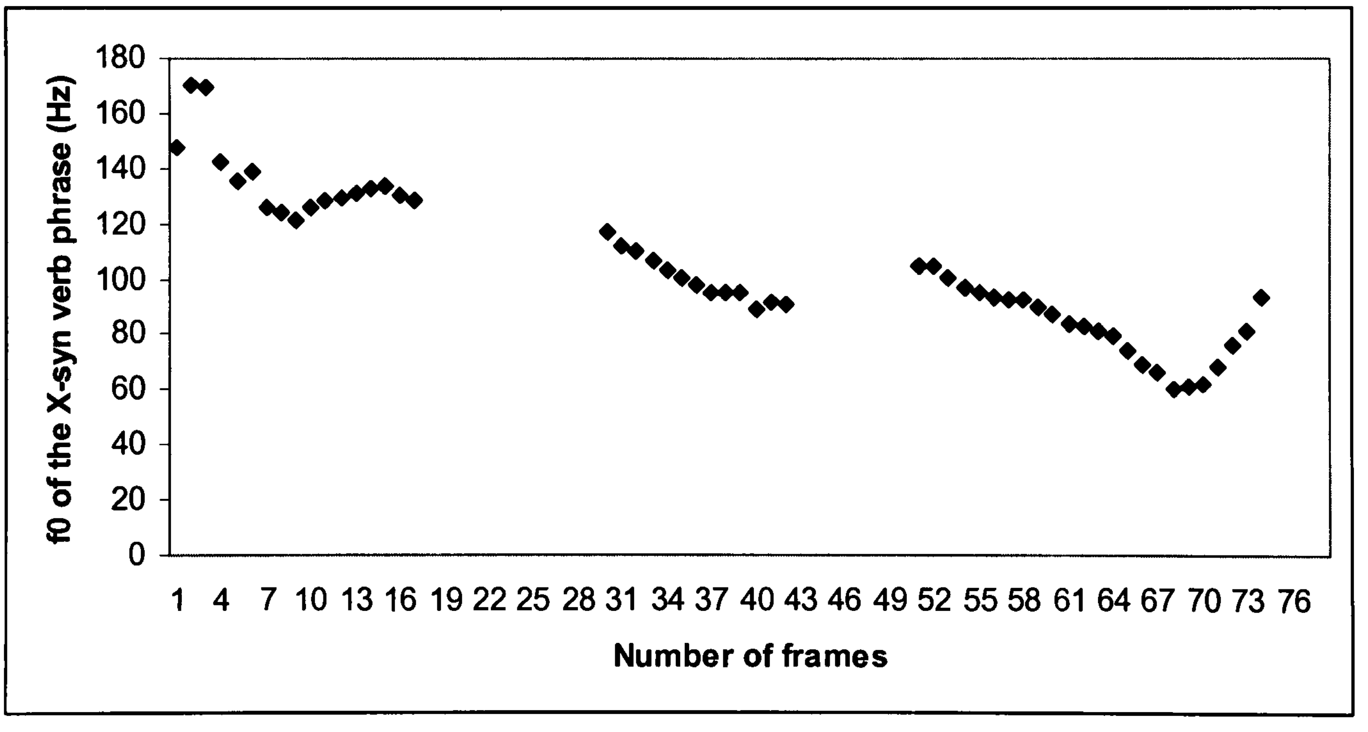


Figure 3.4 f_0 of the cross-synthesised verb phrase version of *Racing cars*

The peak RMS amplitude of the original noun phrase version is on the first word, with very low amplitude on the rest of the word pair (Figure 3.5).

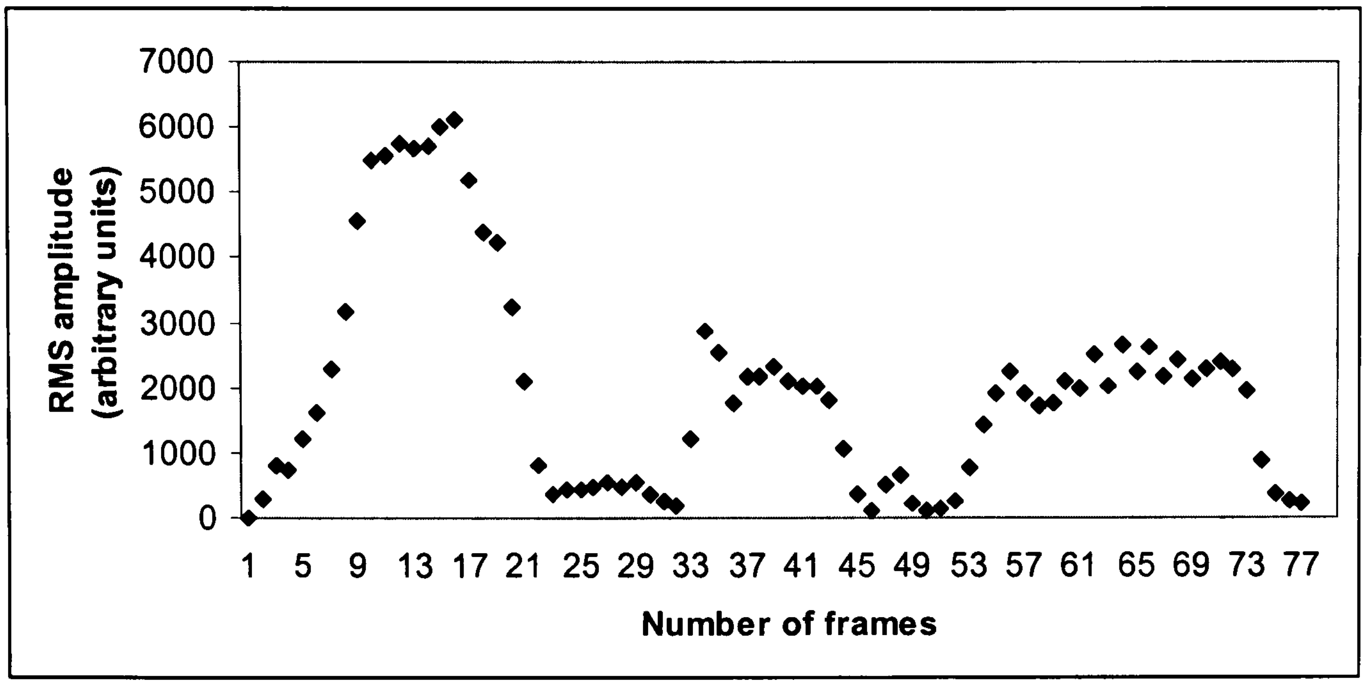


Figure 3.5 RMS amplitude of the noun phrase version of *Racing cars*

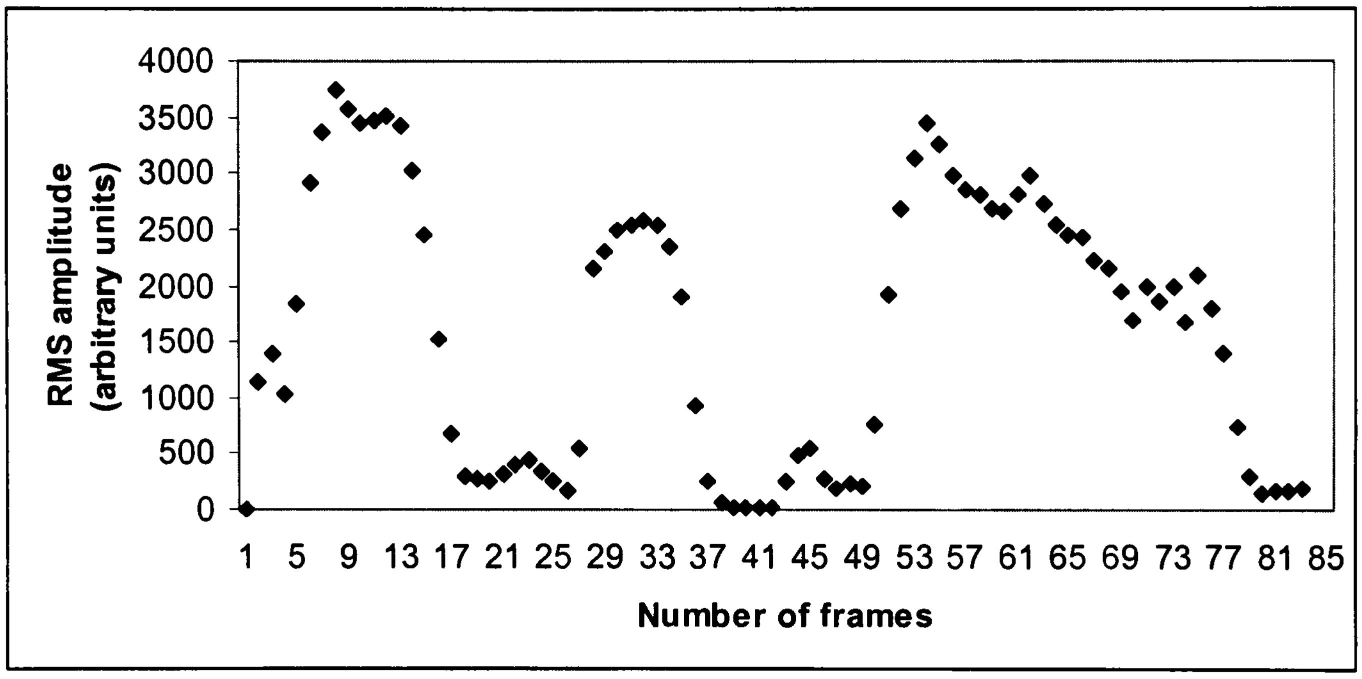


Figure 3.6 RMS amplitude of the verb phrase version of *Racing cars*

The peak RMS amplitude of the cross-synthesised noun phrase is on the second syllable of the first word with a similar peak on the first syllable of the second word (Figure 3.7). These peaks are consonantal bursts produced by the release of the stop at each of these points. The latter is similar, though not identical to the amplitude contour of the original verb phrase version (Figure 3.6).

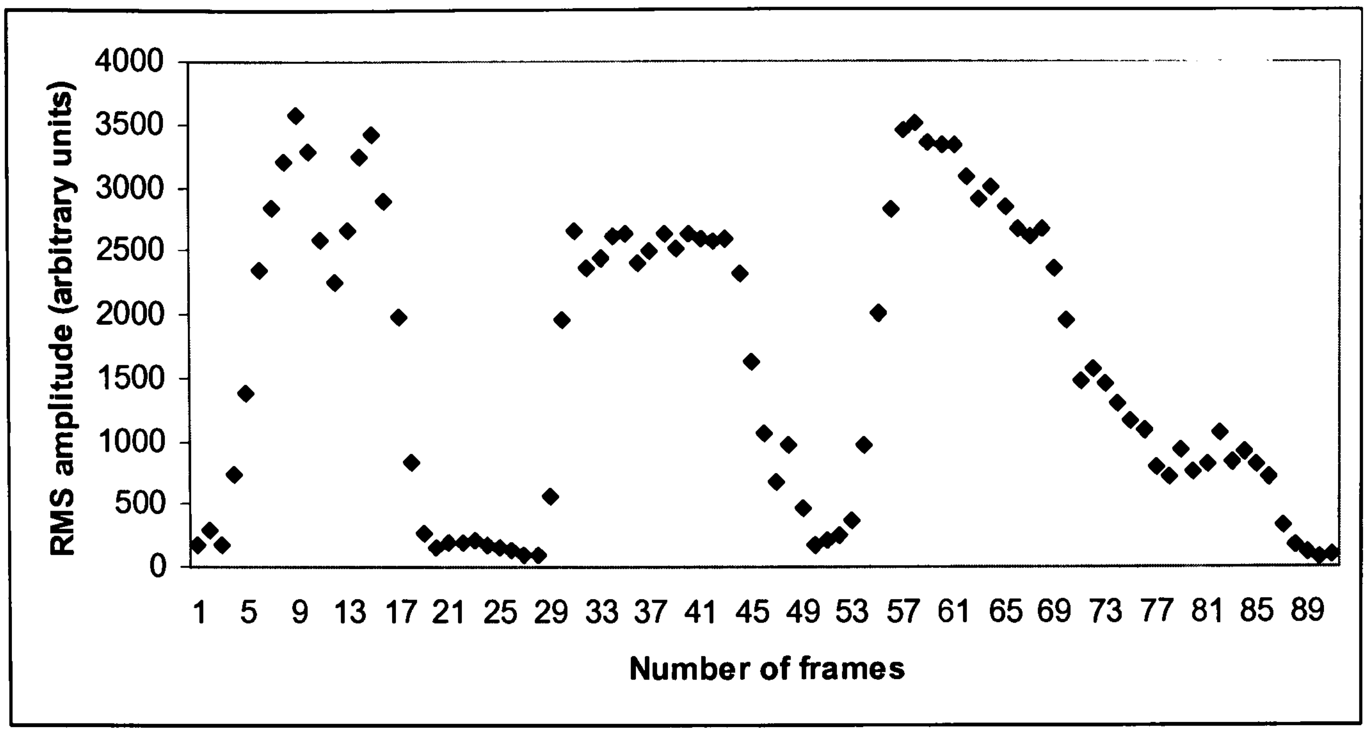


Figure 3.7 RMS amplitude of the cross-synthesised noun phrase version of *Racing cars*

Both the first and second word have high amplitude in the original verb phrase version. However, the cross-synthesised verb phrase peaks on the first word with a secondary peak on the second syllable of the first word (Figure 3.8).

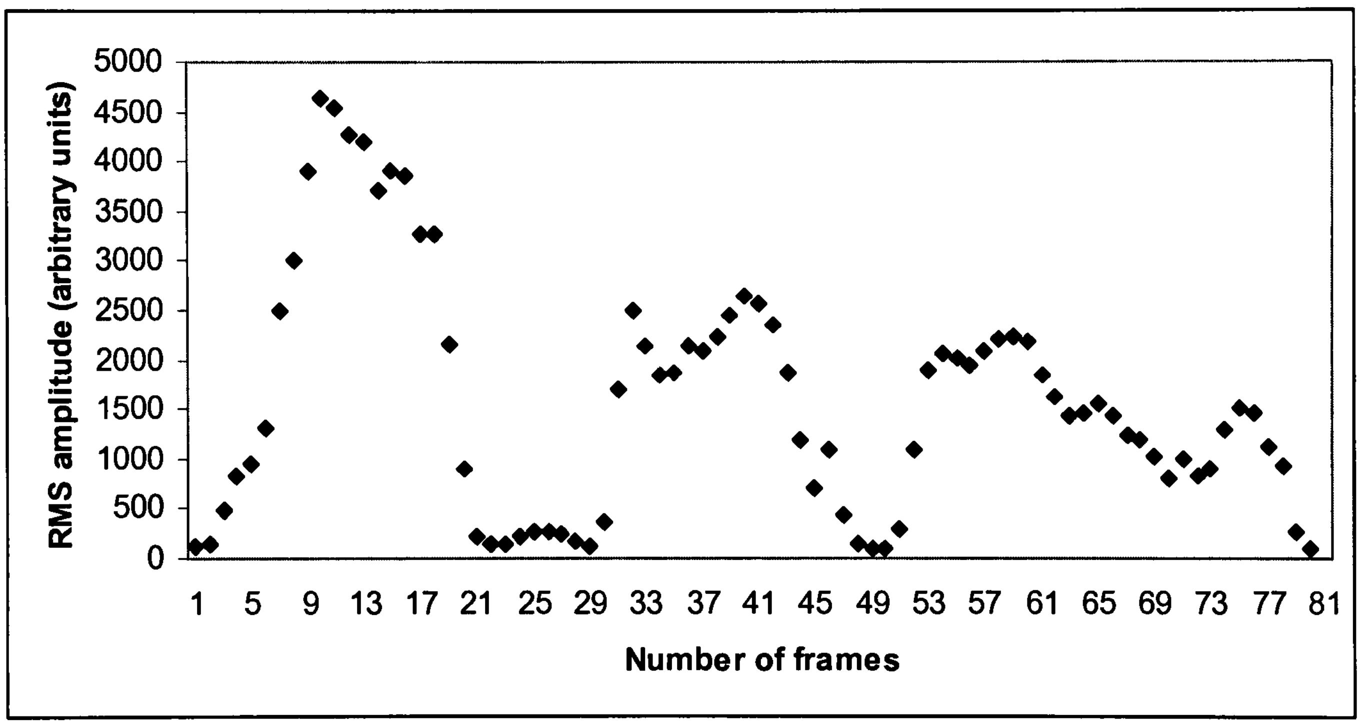


Figure 3.8 RMS amplitude of the cross-synthesised verb phrase version of *Racing cars*

This contour is extremely similar to that of the non-cross-synthesised noun phrase (Figure 3.5). While the figures do not show clearly the differences in duration of the segments between, Appendix A presents the duration, peak amplitude and peak f_0 of the original and cross-synthesised versions of the stimuli.

3.4. PROCEDURE

Each subject heard 384 stimuli whose prosody had been cross-synthesised. Subjects were presented with an equal number of noun and verb phrase versions of the prosodically contrastive and prosodically non-contrastive stimuli. The 384 fragments were randomised in two different orders and individually presented to subjects randomly assigned to one of two different groups.

As in Experiment 1 and 2, subjects were seated in a sound proof room. The fragments were played to them over headphones. As soon as each fragment finished, a word appeared on a screen in front of the subjects. Half of the stimuli were followed by visual probes confirming the interpretations consistent with the prosody of the original stimuli ('is' to follow noun phrase versions and 'are' to follow verb phrase versions – appropriate probes). The other half of the auditory stimuli were followed by visual probes conflicting with the interpretations consistent with the prosody of the original stimuli ('is' to follow verb phrase versions and 'are' to follow noun phrase versions – inappropriate probes). Subjects were given the same instructions as in Experiment 1 and told to press the button on the button set that corresponds to the word they saw on the screen in front of them. However, they were told that the stimuli had been manipulated and sounded a bit tinny and metallic. As in the first two experiments of this thesis, each presentation of the auditory stimuli was followed by an appropriateness-rating task.

Subjects were paid to participate in the experiment. All subjects were native speakers of English from Southern England. Of the twelve subjects who took part in Experiment 3, six were female and six were male. All subjects were aged between 22 to 29 years and were right-handed. Each subject heard 16 repeats of the noun and verb phrase versions of each fragment in a random order. The stimuli presented to the different groups varied in the order of presentation of the fragments. The buttons that subjects pressed in response to the probes also varied between groups. Subjects in one group pressed a red button to the left to respond to 'is' probes and a black button to the right to respond to 'are' probes. The red and black buttons corresponded to the alternative probe in the other group. Subjects were also run through a practise set of fragments to familiarise themselves with the requirement of the experimental task. The fragments in the practice test were different from the fragments used in the main experimental task. The time taken by subjects to respond to the probes by pressing the button was recorded.

3.5. HYPOTHESIS

The stimuli presented to subjects in this experiment were cross synthesised so that the timing, f_0, and amplitude of the noun phrases were swapped with the timing, f_0 and amplitude of the verb phrases, and vice versa. Therefore, the stimuli had the prosodic content of the alternative interpretations of the word pairs, for these prosodic parameters at least. They retained the spectral (i.e. mainly segmental) characteristics of the original word pairs. I hypothesise that if prosodic information were guiding parsing, subjects would construct noun phrase parses of the cross-synthesised verb phrases, and verb phrase parses of the cross-synthesised noun phrases. Therefore, subjects' response times to 'is' probes following the cross-synthesised noun phrases should be shorter than their response times to 'is' following the cross-synthesised verb phrases, despite the fact that 'is' is an appropriate continuation of the original verb phrases and an inappropriate continuation of the original noun phrases. Similarly, subjects' response times to 'are' probes following the cross-synthesised verb phrases should be shorter than their response times to 'are' probes following the cross-synthesised noun phrases, despite the fact that 'are' is an appropriate continuation of the original noun phrases and an inappropriate continuation of the original verb phrases. In summary, if subjects were guided by prosody, their response times to appropriate probes (probes confirming the interpretation intended by the cross-synthesised prosody) would be shorter than their response times to inappropriate probes (probes confirming the interpretation intended by the original stimuli). This would provide strong evidence that subjects were using prosodic information to parse the stimuli presented to them in both Experiments 1 and 2.

However, cross-synthesising and time-warping should not have an effect on the prosodically non-contrastive stimuli. Since there were few consistent differences in the prosody of the two words between the noun and verb phrase versions, the time-warped and cross-synthesised stimuli should display prosodic characteristics similar to the original prosodically non-contrastive stimuli. Therefore, subjects would not be able to differentiate the noun and verb phrase versions of these stimuli. Their response times to appropriate and inappropriate probes following the cross-synthesised prosodically non-contrastive stimuli would be similar, i.e. response times to 'is' following the noun phrases or verb phrases would be similar to response times to 'are'.

All of the cross-synthesised stimuli sounded somewhat tinny and robotic. This might degrade the quality of prosodic information and encourage subjects to resort to a syntax-first approach to disambiguate the stimuli. Consequently, subjects would be unable to use prosodic information to parse the stimuli presented to them. A parser following the principles of late closure and minimal attachment would prefer the noun phrase structure of the stimuli. Therefore, subjects' response times to probes confirming the noun phrase structure 'are' would be shorter than response times to probes confirming the verb phrase version 'is'. Subjects should not find any significant difference between the prosodically contrastive and non-contrastive stimuli if prosodic information is not incorporated early.

Alternatively, the degraded quality of prosodic information might force subjects towards a weakly interactive parsing mechanism. A parser following weak interactive parsing would predict that initial syntactic processing would construct both parses of the input. Therefore, both noun and verb phrase structures should be simultaneously active. Subjects would show no preference for either of the probes. Both probes would be consistent with one of the parses that have been constructed, and parsing would continue with the interpretation confirmed by the probe. Again, if early incorporation of prosodic information does not take place, subjects would not find any significant difference between the prosodically contrastive and non-contrastive word pairs. Therefore, response times to both probes following either the prosodically contrastive or the non-contrastive stimuli should be comparable.

I had suggested that there was the possibility of co-articulation between the initial vowel of the probes and the final fricative of the word pairs in the original sentences from which the stimuli were edited. If subjects were being guided by co-articulation, they would predict parses in keeping with the original interpretation of the word pairs. Therefore, subjects' response times to probes that followed the word pairs in the original sentences that the stimuli had been taken from would be shorter than response times to the alternative probe. Therefore, their response times to inappropriate probes (probes consistent with the interpretation intended by the original word pairs) would be shorter than their response times to appropriate probes (probes consistent with the interpretation intended by the cross-synthesised word pairs).

Experiment 1 found that subjects were not using a weak interactive approach or favouring a noun phrase parse of the stimuli. On the contrary, subjects showed every indication of parsing the stimuli in keeping with the interpretation intended by the prosody of the word pairs. Furthermore, even their parses of the prosodically non-contrastive stimuli also seemed to be guided by prosodic information. Therefore, I do not expect subjects to follow a syntax-first approach to parsing in Experiment 2.

3.6. RESULTS

A total of 92 responses were omitted because of incorrect responses (pressing the 'is' button for an 'are' probe or vice versa) or because of excessive delay in pressing the response buttons: any responses over 1500 msec were omitted. The mean response times of the remaining measures are given in Table 3.2.

	Mean response time to appropriate probes (ms)	Mean response time to inappropriate probes (ms)
With prosodic contrast	614	622
No prosodic contrast	615	617

Table 3.2 Mean response times to appropriate and inappropriate probes following prosodically contrastive vs. non-contrastive cross-synthesised stimuli (ms)

Subjects had faster mean response time to appropriate probes than to inappropriate probes following the prosodically contrastive stimuli – a difference of 8 ms. Subjects also had faster mean response time to appropriate than to inappropriate probes following the prosodically non-contrastive stimuli – a difference of 2 ms. The difference in response times was then analysed to find out if the data was normally distributed. As is usual with response time data, the response times were not normally distributed. The Kolmogorov-Smirnov test for normality proved significant ($p < .050$). Therefore, the results of any t-tests or ANOVAs would have to be treated with caution.

Therefore, I only performed nonparametric tests on the data. The results of a Mann Whitney test indicated that subjects' mean response times to appropriate and inappropriate probes following the prosodically contrastive stimuli were significantly different ($z = - 1.885$; $p < .05$). Conversely, subjects' response times to appropriate and inappropriate probes following the prosodically non-contrastive stimuli were not significantly different ($z = - 0.016$; $p > .5$).

I also conducted a separate test to check whether subjects preferred noun phrase parses of the input. I found that response times to 'is' probes were faster than response times to 'are' probes. This difference was found to be statistically significant ($z = -9.802$; $p < .001$). This indicates that subjects seemed to prefer verb phrase parses of the stimuli to noun phrase parses. The test was performed using response times to both probes following all the stimuli, irrespective of whether the probes were appropriate or inappropriate continuations of the stimuli. This was in order to ensure an effect of syntactic preference irrespective of prosodic interference. Tests on response times to appropriate probes following the prosodically contrastive stimuli also found significant difference between mean response times to 'is' and 'are' (RT ('is') = 603; RT ('are') = 633; $z = -3.815$; $p < .001$). I also tested response times to appropriate and inappropriate probes following the prosodically non-contrastive stimuli. I found significant differences between response times to 'is' and 'are' probes (RT ('is') = 580; RT ('are') = 653; $z = -10.004$; $p < .001$).

I performed a within-subject analysis to check whether each subject had the same pattern of response times to appropriate or inappropriate probes. I also performed a within-item analysis to see whether this pattern was also found for response times to individual items.

Item	Mean response times inappropriate probes (ms)	Mean response times appropriate probes (ms)	Statistical significance
Cooking apples	604	653	**p << .05**
Cutting boards	630	630	p >.I
Packing cases	620	590	p > .1
Playing cards	646	603	**p < .05**
Racing cars	605	630	p > .1
Sailing ships	628	582	**p < .05**

Table 3.3 Mean response times to appropriate and inappropriate probes following prosodically contrastive stimuli per item.

Subject	Mean response times inappropriate probes (ms)	Mean response times appropriate probes (ms)	Statistical significance
1	581	576	$p > .05$
2	526	537	$p > .05$
3	775	750	$p > .05$
4	639	522	**$p < .001$**
5	609	641	$p > .05$
6	691	683	$p > .05$
7	550	557	$p > .05$
8	632	632	$p > .05$
9	632	628	$p > .05$
10	542	569	$p > .05$
11	634	674	$p > .05$
12	659	614	**$p < .05$**

Table 3.4 Mean response times for appropriate and inappropriate probes to prosodically contrastive stimuli per subject.

Only three of the items showed significant differences between response times to appropriate and inappropriate probes. Therefore, I cannot comment on the persistence of the effect of prosody within-item. Mean response time to inappropriate probes was actually shorter than mean response time to appropriate probes following two of the items. Subjects' response times were actually faster when the probe confirmed the interpretation that conflicted with the cross-synthesised prosody. Taken together, the results of the item analysis are insufficient to argue for a within-item effect of prosody on parsing. However, this might be due to the small size of the data-set being taken into consideration.

Almost half the subjects showed faster mean of response times to inappropriate than to appropriate probes. Taking the cross-synthesis into consideration, subjects showed faster response times when the probes conflicted with the parse intended by the prosodic content of the stimuli, then when they confirmed this interpretation. This finding contradicts the results of the first and second experiments. However, as in the case of the within-item analysis, this effect could merely be due to a small number of data points being taken into consideration. The within-subject analysis only considers around 180 tokens. This is not enough data to acquire a true picture of the significance of difference between response times to appropriate and inappropriate probes. Interestingly, two subjects showed significantly faster response times to probes that confirmed the interpretation intended by the cross-synthesised stimuli than when they conflicted with this parse. Again, this finding supports the early incorporation of prosody in parsing.

I performed Pearson correlation tests between subjects' response times and the order of presentation of repetitions of the different stimuli and found there to be no significant correlation ($z = -0.007$; $p > .5$). I could not find significant correlations between response times and the subject factor ($z = .001$; $p > .5$) or between response times and group ($z = -0.015$; $p > .5$). This was to be expected since the experiment controlled for effects of order, subject, and group by randomly assigning subjects to one of two groups, which differed in the order of presentation of the stimuli.

Figures 3.9 plots the difference in response times to appropriate and inappropriate probes following the prosodically contrastive and non-contrastive stimuli. Mean response times to probes confirming the interpretation intended by the cross-synthesised prosody was lesser than the mean of response times to probes conflicting with this interpretation. Furthermore, the interpretation intended by the cross-synthesised prosody conflicted with the interpretation intended by the original word pairs. Therefore, mean response time to probes confirming the parses intended by the original word pairs was longer than the mean response time to probes conflicting with this interpretation. Conversely, the mean response times to appropriate and inappropriate probes following the prosodically non-contrastive word pairs are similar.

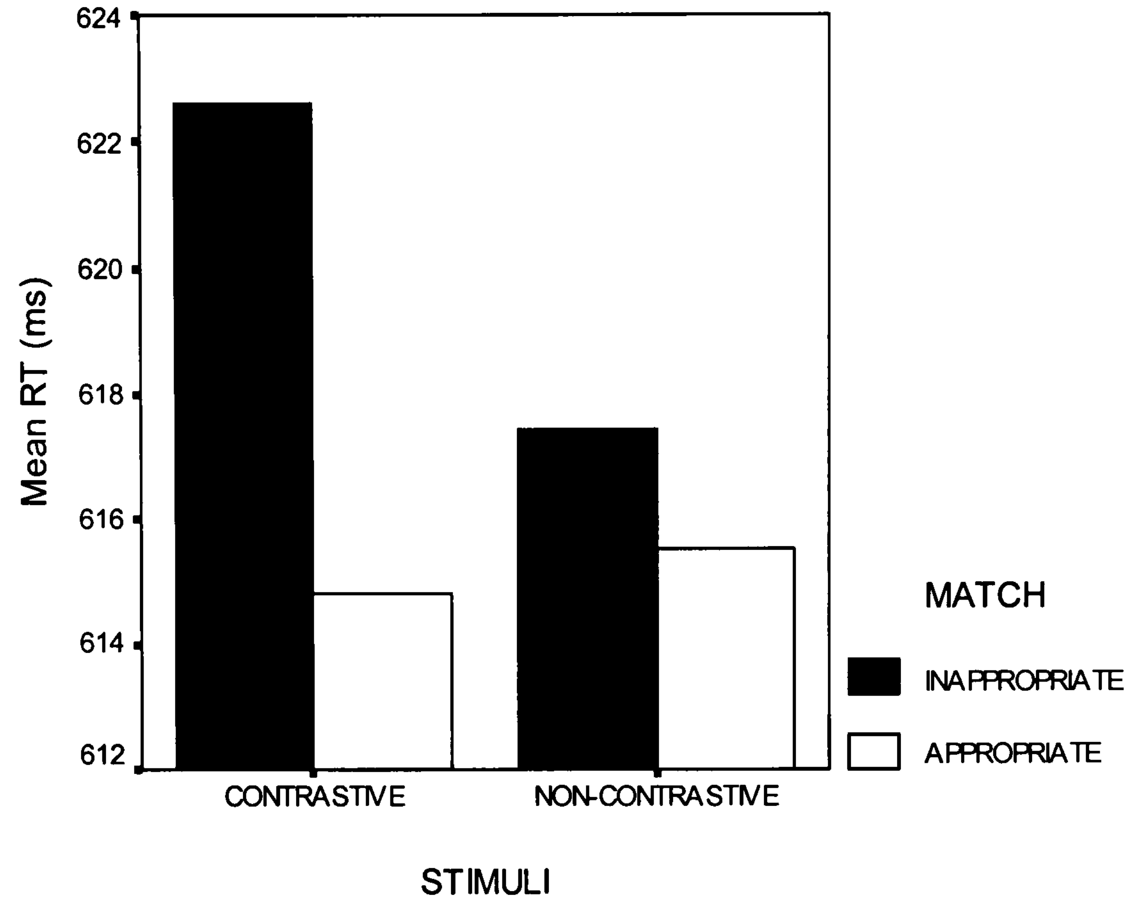


Fig. 3.9 Mean response times to appropriate and inappropriate probes following stimuli with vs. without prosodic contrast (ms)

I also conducted a number of correlation tests between subjects' response times to appropriate probes following the prosodically contrastive stimuli and the main acoustic characteristics that were tested in this experiment – amplitude, pitch and duration. There were two separate data sets for the correlations. One consisted of response times to appropriate probes following the cross-synthesised prosodically contrastive noun phrases, the other, verb phrases. The probes were inappropriate to the interpretation intended of the original stimuli. Therefore, they were expected to be appropriate to the interpretation intended of the cross-synthesised stimuli. The correlations that were significant are reported in the tables below. Table C.1 in Appendix C presents the correlations between subjects' response times to appropriate probes following the prosodically contrastive cross-synthesised verb phrases. Table

C.2 in Appendix C presents the correlations between subjects' response times to appropriate probes following the prosodically contrastive cross-synthesised noun phrases.

In the appropriateness-rating task, subjects were given five seconds at the end of each presentation to provide an evaluation of the visual probe as a suitable continuation of the auditory stimuli. This checked subjects' off-line parsing of the ambiguous stimuli. This task was also included to emphasise the cohesiveness of the auditory and visual stimuli.

	With prosodic contrast		No prosodic contrast	
Evaluation	Inappropriate	Appropriate	Inappropriate	Appropriate
Good	361	956	761	740
Bad	785	191	403	390

Table 3.5: Number of good and bad evaluations awarded to appropriate and inappropriate probes following the stimuli with vs. without prosodic contrast

Subjects gave significantly more good evaluations of appropriate probes than inappropriate probes following the stimuli with prosodic contrast (t (2290 = 29.524; p << .05). Subjects did not give significantly more good or bad evaluations to appropriate and inappropriate probes following the stimuli with no prosodic contrast (t (2291) = .026; p > .5). Some of the stimuli received no evaluation and were excluded from the count. However, they did not make a serious difference to the count, as only a few stimuli received no evaluation.

I also checked to see whether subjects showed an off-line preference for verb phrase parses of the stimuli. Table 3.6 presents the evaluations of 'is' and 'are' probes as appropriate continuations for all the stimuli.

Evaluation	'is'	'are'
Good	1304	1275
Bad	823	820

Table 3.6 Evaluations of noun and verb phrase parses of the stimuli.

365 stimuli did not receive any evaluation at all. Subjects gave almost equal number of good and bad evaluations to 'is' and 'are' probes following the stimuli.

3.7. DISCUSSION

There were two main reasons why I conducted Experiment 2. The first was to check for the relative contribution of prosodic and spectral information in guiding parsing. The second was to provide further evidence for on-line parsing using prosodic information. Experiment 2 provides strong evidence in favour of the early incorporation of prosodic information in parsing. The cross-synthesis and time warping swapped the prosodic characteristics of each of the versions of the word pairs with that of the alternative reading. Therefore, I had predicted that if prosodic information were incorporated early, subjects should predict noun phrase parses of the cross-synthesised verb phrases and verb phrase parses of the cross-synthesised noun phrases.

The results were in keeping with these predictions. Subjects did take longer to respond to probes that conflicted with the cross-synthesised prosody than probes that were consistent with the cross-synthesised prosody. Consequently, subjects took longer to respond to probes that confirmed the interpretation intended by the original word pairs than probes that conflicted with this interpretation. This suggests that subjects had parsed the cross-synthesised verb phrases as noun phrases, and the cross-synthesised noun phrases as verb phrases. Only prosodic information could have guided subjects towards a noun phrase parse of the cross-synthesised verb phrases and a verb phrase parse of the cross-synthesised noun phrases. Furthermore, parsing was tested at the point of recognition of ambiguity, before the completion of the clause that the word pairs were taken from and before the presentation of the main verb of the sentences the word pairs were taken from. Experiment 2 provides strong evidence in favour of the early incorporation of prosodic information in parsing.

There were no consistent differences between the noun and verb phrase versions of the prosodically non-contrastive stimuli. If the results reported for the experimental stimuli were not caused by the early incorporation of prosodic information, response times to the prosodically non-contrastive stimuli should be similar to response times to the prosodically contrastive stimuli. This was not so. Subjects' preference for the parses consistent with the prosody of the cross-synthesised stimuli could only have been caused by the presence of contrastive prosodic information. This was absent in the case of the control stimuli. Hence, subjects were not able to predict the parses intended of the cross-synthesised stimuli. Again, this provides evidence in favour of the early incorporation of prosodic information in parsing.

Experiment 2 also tested the relative contribution of spectral information to the parsing of ambiguous stimuli, since the spectral content of the stimuli was not affected by cross-synthesis. The results of Experiment 2 establish that these differences are not sufficient to guide parsing in the presence of conflicting prosodic information. Subjects did not parse the stimuli in keeping with the parse suggested by the spectral content of the stimuli. They did not prefer parses that confirmed the interpretation intended by the original word pairs. However, the difference between mean response times to appropriate and inappropriate probes was not as large as the difference between means observed in Experiment 1. This might have been caused by the presence of conflicting prosodic vs. spectral information. As far as the aim of this experiment is concerned, I conclude that prosodic information of f_0, duration, and amplitude can guide parsing of the ambiguous stimuli tested here. Moreover, this influence is registered on-line, immediately upon receipt of the ambiguous stimuli.

Subjects did not prefer noun phrase parses of the stimuli, indicating that they did not use syntactic information to guide parsing. Subjects also showed a preference for one interpretation of the word pairs, indicating that they did not create all plausible parses of the word pairs. As in Experiment 1, subjects showed a preference for verb phrase parses of the prosodically non-contrastive stimuli, indicating that they did not use syntactic information or a weak interactive parsing mechanism to disambiguate the non-contrastive stimuli either. Only the prosodic content of the stimuli could have guided subjects towards their preferred parses of the stimuli. Since cross-synthesis swaps only the f_0, amplitude and timing of the word pairs, Experiment 2 provides evidence of an influential role for these three features in parsing.

This conclusion is supported by the results of the correlations. I performed the correlations using two data sets. The first included response times to appropriate probes following the prosodically contrastive noun phrases. The second included response times to appropriate probes following the prosodically contrastive verb phrases. Appropriateness of the probe was decided with reference to the interpretation intended of the cross-synthesised stimuli. I found a number of significant correlations between duration, f_0, and amplitude and subjects' response times. The results suggest that response times were correlated with the duration, amplitude, and f_0 of the word pairs. These results provide further evidence that subjects' parsing was affected by the different prosodic characteristics.

However, there is a caveat to the claims I make here. While this does not provide evidence against the early incorporation of prosodic information in parsing, it does raise additional issues that need to be addressed. Irrespective of the influence of prosodic information on parsing the stimuli, there also seemed to be a preference for verb phrase parses of the stimuli: response times to 'is' were significantly shorter than response times to 'are' following the prosodically contrastive and non-contrastive stimuli. There was a similar preference for verb phrase parses of the prosodically non-contrastive stimuli in Experiment 1. I suggested that this was because of the prosodic contour of the control stimuli resembling the verb phrase versions of the prosodically contrastive stimuli. However, there is a difference between the results of Experiment 1 and 2 in this regard. There was a preference for verb phrase parses of the prosodically contrastive stimuli as well in Experiment 2. I suggest that this is also because of the similarity in the prosodic contour of the control stimuli and verb phrase stimuli. This similarity might have increased subjects' sensitivity to the prosodic contour of the verb phrase stimuli, therefore decreasing their reaction times to probes confirming the verb phrase interpretation. While I do not have empirical support for this possibility, it does provide a suitable explanation for subjects' faster responses to probes confirming the verb phrase interpretation of the stimuli. Interestingly, I did not find the late closure advantage reported by Kjelgaard and Speer (1999) for conflicting prosody sentences.

Subjects' off-line evaluations of the appropriateness of the probes were also in favour of the incorporation of prosodic information in parsing. Subjects provided 956 *good* evaluations for probes that were consistent with the cross-synthesised prosody of the prosodically contrastive stimuli. Conversely, they provided 361 *good* evaluations for probes that were not consistent with the prosody of the prosodically contrastive stimuli. Subjects provided more *bad* evaluations for probes that were not consistent with the prosody of the cross-synthesised prosodically contrastive stimuli. More importantly, a similar dissociation was not found in subjects' evaluations of the probes following the prosodically non-contrastive stimuli. They provided equal number of *good* and *bad* evaluations for appropriate and inappropriate probes. This provides a strong argument in favour of prosodic information guiding parsing off-line as well. Subjects' evaluations did not suggest that they preferred verb phrase parses of the stimuli. They gave an equal number of *good* and *bad* evaluations to both 'is' and 'are' probes. Their on-line preference for verb phrase parses seems to have disappeared during off-line analysis.

3.8. CONCLUSION

Experiment 2 tested whether cross-synthesising the prosodic characteristics of the word pairs would lead subjects to construct different parses of the prosodically contrastive stimuli. Following cross-synthesis, subjects preferred verb phrase parses of the original noun phrase stimuli and noun phrase parses of the original verb phrase stimuli. Only prosodic information could have guided subjects towards alternative parses of the stimuli. The spectral and segmental content of the stimuli would have guided subjects towards the originally intended parses of the word pairs. The syntactic content of the stimuli would have guided subjects towards the noun phrase parses. A weak interactive parsing mechanism would have predicted both parses of the stimuli. Therefore, subjects' preference for the parses indicated by the cross-synthesised prosody provides strong evidence in favour of the early incorporation of prosody in parsing.

Cross-synthesis consisted of swapping the duration, f_0, and amplitude of the noun and verb phrase interpretation of the stimuli. Subjects' preference for the parses indicated by the cross-synthesised prosody provides evidence in favour of duration, f_0, and amplitude guiding parsing. This influence was registered on-line. Previous experiments on swapping the prosodic contour of ambiguous stimuli have used cross splicing to test a similar hypothesis. Experiment 2 avoids the criticism of cross-splicing experiments that I have made in Section 3.2. I did not merely test whether prosody conflicting with the original reading delays subjects' parsing. I tested whether conflicting prosody could guide subjects' parsing towards the alternative parse of ambiguous stimuli. This provides much stronger evidence in favour of prosody guiding parsing than the results of prior experiments using cross-spliced data. However, while subjects' parsing was guided by prosodic and not spectral information, the difference in response times to appropriate and inappropriate probes following the prosodically contrastive stimuli was reduced from Experiment 1. It is possible that the presence of conflicting spectral information delayed subjects' response times to appropriate probes (probes that would have been inappropriate to the original versions of these stimuli).

Furthermore, as with Experiment 1, prosodic information was incorporated on-line, during initial parsing. Subjects' parsing was tested before the presentation of the main verb of the sentences, and before the completion of the clauses the word pairs were taken from. This provides very strong evidence for the early incorporation of prosodic information in parsing. Experiment 2 concludes that f_0, amplitude, and duration can guide parsing. However, it does not disentangle the contribution of each of these prosodic characteristics to parsing. Therefore, in Experiment 3, I test the contribution of f_0 in guiding parsing. There were differences in f_0 between the two words in the noun and verb phrase versions of the word pairs. I test whether flattening the f_0 affects subjects' disambiguation of the word pairs: i.e. in Experiment 3, I present subjects with stimuli at a monotone to test the contribution of f_0 to parsing.

CHAPTER FOUR

THE ROLE OF f_0 IN PARSING AMBIGUOUS SENTENCES: EXPERIMENT THREE

4.1. INTRODUCTION

Table A.1 in Appendix A presents the measures of the different acoustic characteristics of the noun and verb phrase versions of the word pairs. There were differences between the f_0 of the noun and the verb phrase versions of the word pairs. Similarly, there were also differences in the timing and amplitude between the noun and verb phrase versions of the word pairs. Experiment 1 established that prosodic information could guide subjects towards the intended parse of the stimuli. Experiment 2 swapped the f_0, timing, and amplitude of the noun and verb phrase versions of the word pairs. Subjects then preferred noun phrase parses of the cross-synthesised verb phrases and verb phrase parses of the cross-synthesised noun phrases. Therefore, f_0, timing, and amplitude were influential in guiding parsing. This conclusion was supported by the results of correlation tests that I conducted. I found that subjects' response times were strongly correlated with the f_0, amplitude and timing of the stimuli. This provides additional evidence in favour of the early incorporation of prosodic information in on-line parsing.

Having concluded that f_0, timing, and amplitude can guide parsing, the next experiment in this thesis attempts to disentangle the relative contribution of one of the characteristics to parsing. The question addressed in Experiment 3 is: what is the contribution of f_0 to guiding parsing. Experiment 3 tests whether flattening the f_0 of the word pairs (i.e. setting the f_0 of all the stimuli to a constant value) attenuates subjects' ability to differentiate between the noun and verb phrase versions of the word pairs.

Liberman and Prince (1977) write that it is a 'commonplace observation' that there are differences in the stress patterns of lexical compounds and phrasal collocations. For instance, lexical compounds like 'keelhaul' and 'stress-shift' have primary stress on the first word, with falling pitch over the second. Conversely, phrasal collocations like 'red cows' and Sam left have primary stress on the second word. Consistent with Liberman and Prince's study, I found that the noun and verb phrase versions of the prosodically contrastive stimuli differed in the placement of stress on the words of the word pairs, e.g.

(4.1) Cútting bòards are

(4.2) Cùtting bóards is

The noun phrase versions of the prosodically contrastive word pairs display a similar pattern of prominence with falling pitch from the first to the second word. The peak accent of the word pair was consistently on the first word. In most cases, there was a nominal rise in f_0 towards the end of the second word. The rise was insignificant compared to the peak at the beginning of the first word.

Conversely, the verb phrase versions of the prosodically contrastive word pairs were more varied in their acoustic characteristics. The verb phrase versions were usually produced with multiple peaks of prominence. This was different from the noun phrase versions, all of which exhibited single peaks of prominence. Therefore, three of the verb phrase versions – *Packing cases*, *Playing cards*, *Racing cars* – had high f_0 on both the first and second words. In four cases, the peak of the second word in the verb phrase versions was nominally higher than the peak of the first word – *Cutting boards*, *Playing cards*, *Racing cars*, *Sailing ships*. However, not all of the verb phrase versions were produced with the peak accent on the second word. The verb phrase versions of two of the word pairs were produced with the peak accent on the first word and sustained high f_0 on the second word e.g. *Cooking apples* and *Packing cases*. The verb phrase reading of *Cooking apples* was produced with one peak at the beginning of the first word and another at the beginning of the second word. Conversely, the noun phrase versions of both *Cooking apples* and *Packing cases* were produced with a single peak at the beginning of the first word and falling f_0 through the rest of the word pair. Therefore, irrespective of the speaker producing the verb phrase reading in two ways, there were differences between the f_0 of the noun and verb phrase versions of all the word pairs. Experiment 3 tests whether these differences in f_0 alone can guide subjects towards noun or verb phrase parses of the prosodically contrastive stimuli. The peak f_0 and timing of the peak of the first and second word of the noun and verb phrase versions of all the word pairs are presented in Table 4.1. The stimuli in bold are all prosodically contrastive. The remaining stimuli are prosodically non-contrastive.

Stimuli	NP/ VP	Word 1		Word 2		Word pair	
		Timing of peak (ms)	Peak f_0 (Hz)	Timing of peak (ms)	Peak f_0 (Hz)	Timing of word pair peak accent (ms)	Peak on word
Cooking apples	NP	655	204	405	140	655	1
Cooking apples	VP	747	178	427	167	747	1
Cutting boards	NP	694	195	314	130	694	1
Cutting boards	VP	473	149	423	208	423	2
Packing cases	NP	783	207	443	128	783	1
Packing cases	VP	753	164	303	156	753	1
Playing cards	NP	601	182	331	135	601	1
Playing cards	VP	721	154	361	163	361	2
Racing cars	NP	665	168	325	122	665	1
Racing cars	VP	784	146	264	158	264	2
Sailing ships	NP	748	175	238	131	748	1
Sailing ships	VP	732	164	342	219	342	2
Breaking glasses	NP	825	142	345	168	345	2
Breaking glasses	VP	679	155	339	170	339	2
Burning trees	NP	740	138	360	197	360	2
Burning trees	VP	669	136	249	175	249	2
Flying kites	NP	447	247	307	164	447	1
Flying kites	VP	665	177	315	187	315	2
Melting glaciers	NP	907	144	417	175	417	2
Melting glaciers	VP	873	150	403	175	403	2
Ringing bells	NP	691	154	341	147	691	1
Ringing bells	VP	679	144	359	141	679	1
Visiting relatives	NP	962	166	442	150	962	1
Visiting relatives	VP	1047	170	477	151	1047	1

Table 4.1 Timing and measure of the peak f_0 of the words of the word pair

There were no consistent differences between the f_0 of the noun and verb phrase versions of the prosodically non-contrastive word pairs. All except one of the word pairs showed similar f_0 contours for both the noun and verb phrase versions. Most of the word pairs were produced with the nuclear accent on the second word of the word pair. The f_0 of the noun and verb phrase versions of these word pairs were extremely similar with the peak accent on the second word of the word pair. The peak accent in the noun and verb phrase versions of one of the word pairs – *Ringing bells* – was on the first word. However, the f_0 contour of the noun and verb phrase versions of this word pair were again extremely similar. In addition, one of the word pairs was produced with differences between the f_0 of the noun and verb phrase versions –

Visiting relatives. The main stress in both the noun and verb phrase reading was on the first word. However, the accent on the first word in noun phrase reading was higher than the accent on the first word on the verb phrase reading.

There were also differences in the timing and amplitude of the noun and verb phrase versions of the word pairs. However, the difference between the noun and verb phrase versions was signalled by contrastive amplitude in some cases – i.e. the amplitude of the first word was greater than the amplitude of the second word in the noun phrase reading and vice versa in the verb phrase reading (e.g. *Packing cases*). In other cases, the difference between the noun and verb phrase versions was similarly signalled by differences in f_0 (e.g. *Cutting boards*) or timing (e.g. *Cooking apples*). Therefore, f_0 was only one of the prosodic cues marking the contrast between the ambiguous alternatives. This has prompted arguments against the early incorporation of prosodic information in parsing. How is a listener to use prosodic information to guide parsing if the same syntactic contrast is marked by differences in timing in some cases and by differences in f_0 in others?

Marslen-Wilson et al (1992) argue that syntactic information is not as variable and elusive as prosodic information, except in cases of ambiguity. Therefore, subjects might be forced towards employing a syntax-first approach in parsing because of the "overtness" of syntactic cues in the face of more elusive prosodic cues. Admittedly, there were differences in prosodic realisation of the different verb phrase versions of the different word pairs. Subjects had to interpret either sustained high f_0 on both words of the word pair or a peak accent on the second word of the word pair as prosodically cuing in the verb phrase reading. However, the wealth of prosodic cues marking the same syntactic contrast strengthens the argument in favour of early incorporation of prosody in parsing. If the difference between the timing of the noun and verb phrase reading is not adequately contrastive, then subjects may rely on the other prosodic characteristics of the stimuli – for instance, f_0, to guide parsing. Experiment 3 tests this suggestion by presenting subjects with flattened stimuli. I test whether subjects continue to accurately parse the noun and verb phrase versions of the flattened stimuli. Does the absence of only one of the prosodic characteristics of the stimuli make them syntactically and prosodically ambiguous?

A considerable amount of research tests the effect of f_0 on structural assignment. However, research to date has not yet provided strong evidence of an on-line effect of prosody on parsing. Consequently, there is no prior research reporting an on-line effect of f_0 on on-line parsing. Most models of speech processing present tentative suggestions of the possibility of prosody influencing on-line structural assignment (Frazier and Fodor: 1978; Carston et al: 2001). For instance, although Frazier and Fodor argue that prosodic information is a part of the general structural principles guiding parsing, there is a considerable gap in outlining how prosodic information guides parsing. Strong evidence of f_0 influencing parsing might clarify the architecture of models of initial prosodic processing.

Finally, evidence of the flattening of f_0 affecting accurate parsing would provide additional support for the early incorporation of prosody in parsing. Subjects' preference for the interpretations consistent with the prosody of the stimuli in Experiments 1 and 2 provides evidence of prosody guiding parsing. The word pairs presented to subjects in Experiment 3 are identical to those used in Experiment 1 except that the f_0 of the word pairs was flattened. Therefore, the absence of a similar preference for the intended interpretation would provide additional evidence in favour of prosody guiding parsing. In this chapter, first, I review the prior research testing for an influence of f_0 on parsing. I focus on outlining the problems facing suggestions of f_0 guiding parsing. Then, I present the details of Experiment 3, which tests the relative contribution of f_0 to parsing.

4.2. THE PROBLEM OF PROSODIC INCONSISTENCY: PRODUCTION

Shattuck-Hafnagel and Turk (1996) define prosody as the combination of acoustic characteristics – such as f_0, timing, amplitude, spectral lift, segmental reduction, and their articulatory correlates – defined by relation to the higher level structures like intermediate phrase boundaries, intonational phrase boundaries and phrasal stress patterns. The influence of acoustic characteristics such as timing, f_0, amplitude and some higher-level structures such as prosodic breaks between syntactic elements on parsing have received considerable attention. In Chapter 3, I reviewed the work testing the production of consistent differences in timing between alternative interpretations of syntactically ambiguous stimuli. In this chapter, I concentrate on the problems in production studies testing the production of consistent differences between the f_0 of similar segments of ambiguous stimuli.

There is considerable debate on the prosodic characteristics of syntactically ambiguous alternatives, especially with regard to the f_0 of the alternatives. For instance, Albritton et al (1996) argue that naïve speakers do not usually produce sufficiently contrastive prosodic cues between ambiguous syntactic alternatives. They report that most production studies employ trained speakers who are aware of the potential ambiguity of the stimuli. The sentences tested in Albritton et al were taken from Price et al (1991). The sentences were spoken by trained and naïve speakers. In addition, some of the trained speakers were told to produce the sentences in such a way as to signal the intended interpretation of the stimuli. An independent 'rater' then judged whether the sentences were produced with adequate prosodic information to guide disambiguation. Trained speakers only produced sufficiently contrastive prosodic differences between ambiguous sentences when told to signal the intended interpretation using prosody. In these cases, trained speakers produced around 80% of the sentences with appropriate prosody. However, even in these productions, only differences in timing were consistently correlated with differences in meaning. There were significant differences between the f_0 of the alternative versions in only a few cases. When trained speakers were not instructed to produce the sentences with contrastive prosody, the sentences were as indistinct as those of the untrained speakers. This could explain the differences between the results of the production studies reviewed in Chapter 3. Moreover, this raises a number of problems for studies arguing in favour of early incorporation of prosody. If speakers rarely produce adequate prosodic information to guide disambiguation, then situations that necessitate the early incorporation of prosody are few. This might also mean that there is little need to test the early incorporation of prosody, since speakers might not actually produce adequate prosodic information in conversation. However, the added complications of using only one independent rater need to be kept in mind. The results of one listener's off-line decisions are not sufficiently strong evidence of differences between the f_0 of productions by trained and naïve speakers.

Furthermore, Blasko and Hall (1998) present contradictory results using an on-line perception study to argue against significant differences between the prosody of productions of trained and naïve speakers. First, a production study explored the differences between productions of trained and naïve speakers. They compared the prosodic characteristics of (4.3) and (4.4) when produced by a trained speaker to productions of three naïve speakers.

(4.3) As you know Mike, kids around the office can be trouble

(4.4) As you know, Mike kids around the office an awful lot

The experimenters were confident that the naïve speakers were unaware of the nature of the experiment. Therefore, speakers were neither aware of any potential ambiguity nor were they instructed to produce adequate prosodic cues to differentiate the two interpretations. However, Blasko and Hall do not clarify why they thought the speakers were unaware of potential ambiguity and the purpose of the task. For instance, did they ensure that each speaker had only one variation of the temporarily ambiguous sentences? Blasko and Hall report that both naïve and trained speakers produced consistent differences between the prosodic characteristics of the ambiguous portions of the two sentences. They found that (4.4) was produced with a larger f_0 rise between *Mike* and *kids* than (4.3). They found no significant differences between the amplitude of the ambiguous portions in the two sentences. They also found that there were significant differences in the pauses across syntactic boundaries in (4.3) and (4.4). (4.3) had a greater pause between *Mike* and *kids*, while (4.4) had a greater pause between *know* and *Mike*. Furthermore, they found that the prosodic cues produced by the naïve speakers and trained speaker were similar.

Blasko and Hall then conducted an off-line perception study to test the effect of f_0 on parsing. They reduced the differences in timing of the pauses between words in both the sentences. They also flattened the stimuli in order to eliminate differences between the f_0 of the sentences. Subjects were played repetitions of both manipulated and normal stimuli. Subjects were asked to indicate whether they thought the initial phrase was *long* (included *Mike*) or *short* (did not include *Mike*). They found that subjects' interpretations of the normal sentences were consistent with the interpretation intended by the speaker. However, subjects' interpretations were significantly biased towards the short reading (4.4) when played the flattened sentences. The authors argue that subjects' inability to infer the intended interpretation indicates that parsing was affected by flattening the stimuli. However, the short version was produced with a larger f_0 rise. Therefore, flattening the f_0 should lead subjects away from the short version. The prosodic characteristics of the long version would be more similar to the prosodic characteristics of the flattened stimuli. The long version would also be preferred if the parser were following the principle of late closure. Consequently, it is unclear why subjects were biased towards this interpretation of the stimuli.

Blasko and Hall also performed an on-line experiment testing for an effect of prosodic information on parsing. Their on-line task suffered from many of the failings of experiments mentioned in Section 1.2.3. However, contrary to Albritton et al (1996), Blasko and Hall (1998) present evidence of naïve and trained speakers producing similar prosodic cues. Furthermore, both naïve and trained speakers in Blasko and Hall's study reliably produced sufficiently contrastive prosodic information without explicitly being made aware of potential ambiguity.

However, additional problems are raised by studies reporting differences between the f_0 of productions of spontaneous and read speech. Previous studies reported finding a decline in f_0 from the beginning to the end of the utterance when speakers were asked to read out sentences (Maeda: 1976; Cooper and Sorenson: 1981). Conversely, similar declines in f_0 were not found when speakers were producing spontaneous speech (Umeda: 1982; Cooper et al: 1985). If the f_0 of spontaneous speech is significantly different from the f_0 of read speech, this might complicate the results of earlier production studies reporting differences between the f_0 of ambiguous alternatives. This would also complicate suggestions of these f_0 differences guiding parsing. Noticeably most of the studies reviewed in this thesis analysed productions where speakers were asked to read out individual sentences. The experiments in this thesis also tested the parsing of read speech. However, Anderson and Cooper (1986) found that speakers consistently produced similar falls in f_0 even in spontaneous speech. Spontaneous speech samples were produced by asking speakers to describe a picture. Matching read speech samples were produced by asking speakers to read out target sentences printed on cards. They found no significant effects of speech mode. They conclude that the decline in f_0 towards the end of a sentence is not only a characteristic of read speech. Speakers reliably produce similar prosodic cues in both spontaneous and read speech. However, speakers in Anderson and Cooper were asked to produce single sentences. It is possible that the prosodic characteristics of stimuli are affected by the length of the stimuli. This would explain the difference between the results reported by Anderson and Cooper (1986) and prior studies.

In addition to the variability in the f_0 of productions of naïve and trained speakers producing read and spontaneous speech, there is equal variability in the prosodic characteristics used to signal clause closure. For instance, Cooper and Sorenson (1977) measured the f_0 peaks and falls in sentences such as

(4.5) Anthony was surprised and Raymond was upset

(4.6) Anthony was surprised Andrea became upset

They measured the f_0 of /raɪzd / and /reɪ/ in both sentences. Although they found that f_0 was higher in sentences such as (4.5) than in (4.6), these differences were not statistically significant. I have already reviewed the work of Warren (1985), which reports significant differences between the f_0 of the ambiguous fragments of sentences such as:

(4.7) Before the king rides his horse it takes ages to groom.

(4.8) Before the king rides his horse is groomed for him.

(4.9) The actor learnt the text amused his cast.

(4.10) The actor learnt the text and knew his role.

As I have commented in an earlier chapter, Warren found significant differences between the timing of similar segments in (4.7) and (4.8). Warren also concluded that differences between the f_0 of the two versions could be expected near major syntactic boundaries. The fall in f_0 on *rides* was smaller in (4.8) than in (4.7) where it was the end of the opening clause. Conversely, the fall on *horse* was greater in (4.8) where it was the end of the longer clause than in (4.7) where it is the beginning of the second clause.

However, consider the sentences in (4.9) and (4.10). Prior research (O'Shaugnessy: 1979; Maeda: 1976; Cooper and Sorenson: 1981) found that the ends of the phrases were usually accompanied by falls in f_0, while the beginnings of phrases were accompanied by rises in f_0 (O'Shaugnessy: 1979). Warren reports a much greater fall in f_0 over *learnt* in (4.9) than in (4.10). However, *learnt* is not the end of the clause in either (4.9) or (410). If anything, it is close to the *beginning* of the subordinate clause. Following, O'Shaugnessy, there should be a sharp rise in f_0 on *learnt* in (4.9) since it is near the beginning of the subordinate clause. Warren does not report finding any sharp rises in f_0 at this point. In addition, Warren did not find significant differences between the f_0 of sentences such as (4.11) and (4.12).

(4.11) He climbed the peak with snow on top

(4.12) He climbed the peak with Pete and Dave

Warren argues that differences between the f_0 of the ambiguous stimuli are found more reliably in situations of clause closure ambiguity than PP attachment ambiguity. Furthermore, Warren argues that the differences in f_0 were less marked than the differences in timing between the productions of the ambiguous stimuli.

Schepman (1997) also reports not finding consistent differences between the f_0 of alternative interpretations of ambiguous fragments such as (4.13) to (4.15) below. The sentences are ambiguous in attachment:

(4.13) The lawyer greeted the powerful barrister and the old judge was walking to the courtroom.

(4.14) The lawyer greeted the powerful barrister and the old judge who was walking to the courtroom.

(4.15) The lawyer greeted the powerful barrister and the old judge who were walking to the courtroom.

She measured the f_0 peaks and valleys on *barrister* and *judge* in all the sentences. She also measured change in f_0 on *barrister* and *judge.* She did not find significant differences between the f_0 of the alternative versions of the similar words. There were consistent differences in the placement of the peak accent in (4.14) and (4.15). Furthermore, there was a greater drop in pitch after the peak in (4.15) than in (4.14). However, this difference was only significant by items and not by speakers. There were no consistent differences between the placement of the peaks in (4.13) and in (4.14) and (4.15). Schepman concluded that although there was a trend towards the peaks being different, there was too much inter-subject variability to propose consistent differences between the f_0 of similar words in the alternative versions.

In addition, Beckman (1996) argues that often the same prosodic cues are used to mark both intonational boundaries and differences in emphasis. She discusses this problem with respect to the prosodic differences between a vocative reading (4.16) and an object reading (4.17) of *Manny* in the sentences below:

(4.16) Mary will marry, Manny.

(4.17) Mary will marry Manny.

From a number of production studies, Beckman and her colleagues concluded that *marry* was emphasized by a nuclear accent in both (4.16) and (4.17). There were no consistent differences in the timing of either *marry* or *Manny* tokens between the two sentences. Beckman argues that listeners would not be able to disambiguate the two structures in (4.16) and (4.17) using prosodic information because of the similarity in the prosodic contour of both sentences. However, the parse in (4.17) does not violate the principle of late closure, since *Manny* is a part of the main clause. Therefore, Beckman argues that listeners would always prefer the reading in (4.17), advocating a syntax-first approach to parsing.

The problems with proposing that f_0 can influence parsing are evident when we analyse the different ways of producing (4.18):

(4.18) We know you have great credentials.

We're looking for someone with just such credentials.

(4.18) could be produced with a high tone on *just* and another accent and boundary tone on *credentials* (We're looking for someone with just $_{H*}$ such $_{!H*}$ credentials $_{!H*\ L-L\%}$). However, it could also be produced without the accent and boundary tone on *credentials* (We're looking for someone with just $_{H*}$ such $_{!H*}$ credentials) (Pierrehumbert and Hirscheberg: 1990). Both these contours would indicate that the *credentials* referred to in the second utterance co-reference with the *credentials* presented in the first utterance. Conversely, listeners might also interpret the second accent on *credentials* to indicate that the speaker is referring to a different set of credentials. Therefore, there might not be any significant differences between the f_0 of alternative interpretations of ambiguous utterances. In addition, there might be significant and consistent differences between the f_0 of two productions of the same utterance used to indicate similar meanings. This variability in the prosodic characteristics of speech is a considerable obstacle to proposing that f_0 can influence parsing on-line. How is a listener to use f_0 to guide parsing when there seems to be little consistency to the correlations between f_0 and syntactic structure?

I present a final instance of this inconsistency before outlining the details of Experiment 3. Some of the research mentioned above argues that the presence of a peak accent on an item indicates the presence of a syntactic boundary. However, Schafer et al found that the most prominent word in a phrase was not always the final word in the intermediate phrase to carry an accent. Therefore, parsing cannot automatically assume that the final pitch accented word in a phrase is the focus of the sentence. They recorded pairs of sentences that were syntactically ambiguous yet prosodically distinct such as:

(4.19) I asked $^{L* H-L\%}$ the pretty ^{L*} little ^{L*}girl $^{L* L-L\%}$ WHO'S ^{H*} cold $^{L-L\%}$

(4.20) I asked $^{L* H-L\%}$ the pretty ^{L*} little ^{L*} girl $^{L* L-}$ who's ^{L*} COLD $^{H* L-L\%}$

The sentences were ambiguous between the subordinate clause acting as either an embedded question or a temporal adjunct. Typically, the embedded question versions were produced with a high pitch accent on the interrogative constituent (4.19). Conversely, the temporal adjunct versions were produced with a peak accent on the final word of the string (4.20). Schafer et al found that pitch accents did influence parsing. Subjects were played different versions of the sentences and were asked to paraphrase what they heard. In one experiment, subjects were asked to paraphrase the sentences in their own words. In another experiment, subjects were asked to choose one of two paraphrases of the sentences. It was found that the presence of a pitch accent on the interrogative item biased listeners to an embedded question reading of the sentence. The absence of a peak accent on the interrogative items biased listeners towards the temporal adjunct reading.

However, sometimes the sentences were also produced with down-stepped high or low accents on words after the most prominently accented words in the phrase. Therefore, there was a possibility that the peak accent did not coincide with the syntactic boundary. Consequently, Schafer et al conclude that resolution of syntactic ambiguity would have to consider the kinds of pitch accents on key words in the sentence. This has two main implications for any theory suggesting that can f_0 guide parsing. Firstly, as Schafer et al conclude, it indicates that listeners cannot conclude that phrase accents always demarcate syntactic boundaries. In cases where the most prominent accent (H*) was followed by a pitch accent (L*), parsing would have to consider the H* accent as the focus and the L* as marking the syntactic boundary. Secondly, Schafer et al conclude that 'the resolution of syntactic ambiguity is influenced by the kinds of pitch accents on key words in the sentence – for example, L* versus H* – and not by the location of the most prominent accented word (Schafer et al: 2001: 93). However, the accents on words are not inherently low or high by themselves. The peak f_0 of an utterance is the peak accent only if f_0 over the segment is greater than the f_0 of the other segments in the utterance. Therefore, the f_0 of all the prior segments in the utterance would need to be analysed before the H* accent can be located. It could be argued that relative prominence is decided within intermediate prosodic phrases, so that parsing compares the H* and L* accents within the intermediate phrase. Even so, this would mean that all the f_0 peaks within an intermediate phrase have to be processed before the nuclear accent is located. This would indicate that f_0 could not influence on-line parsing of ambiguous stimuli until the closure of the intermediate phrase or sentence. For instance, Frazier and Fodor (1978) propose an alternative parsing mechanism that breaks speech stimuli into manageable chunks. The size and location of these splits is decided by the prosodic characteristics of the boundaries that intersect the stimuli. The boundaries themselves would be decided on the information already presented to the parser. The f_0 of a particular segment can only influence parsing relative to the f_0 of other segments in the utterance. Relative prominence might be decided within these chunks and not necessarily compared to the pitch contour of the entire stimulus. Therefore, prosodic parsing could be retained as a plausible online option. This parsing mechanism would then maintain that speech is organised into chunks of information. The pitch contour of this 'chunk' is analysed to locate the nuclear and secondary accents. This information is used to parse the input accurately.

However, there are problems with this account. The first is an inherent problem with the 'sausage packing mechanism' – Frazier and Fodor's term for this parsing model. The model was suggested to account for late closure parsing preferences in English. Using sentences like (4.21) they argue that late closure parsing preferences are caused by the high attachment site (*Tom said*) being within a prosodic package. Conversely, early closure parsing preferences are caused by the high attachment site (*Tom said*) being found between packages. Therefore, the parser would always parse the stimulus as a late closure parse when the high attachment site was within a prosodic package.

(4.21) Tom said Bill left yesterday

Therefore, (4.21) is open to two interpretations where either Bill left '*yesterday*', or the interpretation where Tom said yesterday that Bill left. However, Warner (1980) found that subjects parsed sentences such as (4.21) as late closure sentences even when both potential attachment sites were within the same prosodic package. In other words, even where there was no prosodic boundary between *yesterday* and *Tom said*, subjects preferred the adjunct modifying the lower clause *Bill left* more often *Tom said*.

Therefore, a later paper by Fodor (1998) argued that attachment would also be decided by a desire to break the utterance into equal sized chunks. If (4.21) were parsed with high attachment of 'yesterday', then this would split the utterance in two unequal chunks of *Tom said Bill left* and *yesterday*. However, low attachment of 'yesterday' would require the text to be parsed as merely *Bill left* and *yesterday*. They argue that parsing would ignore the weight of *Tom said*, as the restriction of equality in size refers only to sister constituents and not to the entire phrase. This explains subjects' preference for low attachment of the adjunct.

However, even the sausage packing mechanism cannot adequately address all the issues of f_0 guiding parsing on-line. Restricting the debate to the possibility of the inclusion of f_0 during initial parsing, this solution cannot also account for the parsing of f_0 in sentences such as in

(4.22) Did Bill say that, or did Mary?

The sausage parser only considers the prosodic characteristics within chunks before continuing to parse the rest of the sentence. However, a greater accent on *Mary* than on *Bill* might be used to cue the added interpretation that it mattered more if it was Mary who said x than if it was *Bill* who said x. Since the two nouns would be separated by their being in separate chunks, would the parser be able to recognise the implications of a greater accent on *Mary* than on *Bill*? Irrespective of whether a greater accent on *Mary* does cue in the added implication suggested above, this example raises an interesting question. If f_0 can guide parsing, then can it guide parsing when the accents are in two prosodic chunks or separated by an intermediate phrase boundary?

It is possible that parsing using f_0 is incremental. In this case, the parser would compare the accents on the words already presented to it and use this information to guide parsing. Therefore, the parser would not need a cut-off point where it decides to stop and look back to compare the accents in the sentence to find the nuclear accent. Neither is there a need to suggest that differences in pitch only influence parsing during re-analysis. However, evidence of the early incorporation of f_0 in parsing is required to substantiate claims of incremental parsing using f_0.

There is considerable debate on whether speakers reliably produce consistent differences between the f_0 of syntactically ambiguous alternatives. The arguments take many directions, discussed above. Firstly, there is evidence that only trained speakers who are explicitly aware of potential ambiguity produce consistent prosodic differences between the ambiguous alternatives of the stimuli. Secondly, the results of prior production studies do not consider possible differences between productions of spontaneous and read speech. Earlier studies had claimed that the f_0 of spontaneous speech is different from the f_0 of read speech. Thirdly, I have also reviewed research supporting arguments that two ambiguous alternatives can often be produced with the same prosodic characteristics. Conversely, there is also evidence that the same interpretation of an ambiguous alternative can be marked by different prosodic contours (Pierrehumbert and Hirschberg: 1990). The research reviewed in this section raises a few problems for arguments in favour of f_0 influencing on-line parsing. Firstly, does the argument for f_0 influencing parsing wither if speakers do not reliably produce consistent differences between the f_0 of ambiguous alternatives? Secondly, is it possible for f_0 to influence on-line parsing in an incremental manner, continually comparing the accent on the newly introduced item to the accents on the parts of the sentence that have already been processed? Experiment 3 considers possible solutions to these questions by testing whether f_0 can guide parsing on-line. Strong evidence of f_0 guiding parsing on-line would support arguments of incremental parsing using prosody. This would also support arguments of prosody guiding parsing despite inconsistency in the prosodic characteristics of test stimuli. Similar inconsistency is found in the production of segmental information in speech. However, cue trading research has provided a plausible solution to the issues raised by inconsistency in the segmental layer (Luke and Piconi: 1986). I employ a similar cue-trading argument to propose that f_0 can guide parsing despite variability in the prosodic characteristics of speech.

EXPERIMENT THREE

Having established that prosodic information can guide parsing on-line, the next question to ask is what these prosodic cues are. Experiment 2 found that swapping the f_0, timing, and amplitude of the noun and verb phrase versions of the stimuli reversed subjects' parses of the stimuli. This provides strong evidence of f_0, timing, and amplitude guiding parsing, without further insight into the relative contribution of these features to guiding parsing. Therefore, Experiment 3 tests the parsing of stimuli whose f_0 has been flattened. If flattening the f_0 affects subjects' parsing, this provides strong evidence of f_0 guiding parsing in Experiment 1 and 2. In Chapter 3, I argued that there was similar uncertainty about differences in timing and amplitude between ambiguous syntactic alternatives being produced consistently. There was more evidence of f_0 influencing parsing than there was of an influence of timing and amplitude on parsing (Hirschberg and Ward: 1992; Price et al: 1991;

Streeter: 1978; Slujiter et al: 1995; 1996). Therefore, I test the contribution of f_0 alone to guiding parsing on-line.

In addition, this will also examine the consequences of removing one of the major prosodic cues contributing to marking the syntactic contrast between the noun and verb phrase versions. Flattening the f_0 does not flatten the amplitude nor remove the differences in timing between the noun and verb phrase versions. Therefore, subjects might still use amplitude and timing to differentiate the stimuli. This explores the flexibility of prosodic parsing, testing whether subjects can cope with less than complete prosodic information. The uncertainty about the consistent production of f_0 differences between ambiguous syntactic alternatives makes testing the flexibility of subjects' use of prosodic information in parsing all the more important. This uncertainty affects arguments proposing the early incorporation of prosodic information in guiding parsing.

4.3. STIMULI

The stimuli used in Experiment 3 were modified versions of the auditory stimuli presented to subjects in Experiment 1. The prosodically flattened stimuli were obtained using the PRAAT speech editor (Boersma & Weenink: 2000). The f_0 contour of the original stimuli was extracted and flattened to a frequency of 115 Hz. This flattened f_0 contour was then re-synthesised with the original signal using PSOLA re-synthesis. All the word pairs were flattened for the sake of uniformity, despite the similarity in the f_0 of the noun and verb phrase versions of the prosodically non-contrastive word pairs. All the stimuli had a monotone. They did not exhibit any of the differences between the f_0 of the noun and verb phrase versions of the stimuli. However, they retained the amplitude, timing, and spectral characteristics of the original stimuli.

4.4. PROCEDURE

Each subject heard 384 stimuli whose f_0 had been flattened. Subjects were presented with an equal number of noun and verb phrase versions of the prosodically contrastive and prosodically non-contrastive stimuli. The 384 fragments were randomised in two different orders and individually presented to subjects randomly assigned to one of two different groups.

As in Experiment 1 and 2, subjects were seated in a sound proof room. The fragments were played to them over headphones. As soon as each fragment finished, a word appeared on a screen in front of the subjects. Half of the stimuli were followed by visual probes confirming the interpretations consistent with the prosody of the stimuli ('is' to follow verb phrase versions and 'are' to follow noun phrase versions – appropriate probes). The other half of the auditory stimuli were followed by visual probes conflicting with the interpretations consistent with prosody of the stimuli ('is' to follow noun phrase versions and 'are' to follow verb phrase versions – inappropriate probes). Subjects were given the same instructions as in Experiments 1 and 2 and told to press the button on the button set that corresponds to the word they saw on the screen in front of them. However, they were told that the stimuli had been manipulated and sounded a bit tinny and metallic. As in the first two experiments of this thesis, each presentation of the auditory stimuli was followed by an appropriateness-rating task.

Subjects were paid to participate in the experiment. All subjects were native speakers of English from Southern England. Of the twelve subjects who took part in Experiment 3, six were female and six were male. All subjects were aged between 22 to 29 years and were right-handed. Each subject heard 16 repeats of the noun and verb phrase versions of each fragment in a random order. The stimuli presented to the different groups varied in the order of presentation of the fragments. The buttons that subjects pressed in response to the probes also varied between groups. Subjects in one group pressed a red button to the left to respond to 'is' probes and a black button to the right to respond to 'are' probes. The red and black buttons corresponded to the alternative probe in the other group. Subjects were also run through a practise set of fragments to familiarise themselves with the requirement of the experimental task. The fragments in the practice test were different from the fragments used in the main experimental task. The time taken by subjects to respond to the probes by pressing the button was recorded.

4.5. HYPOTHESIS

In this experiment, I present subjects with stimuli whose f_0 had been flattened. There were consistent differences between the f_0 of the noun and verb phrase versions of the prosodically contrastive word pairs presented to subjects in Experiments 1 and 2. There was also evidence that subjects might have been guided by f_0 in parsing the stimuli. Consequently, subjects' parsing of the stimuli in the third experiment might be affected by flattening the f_0 of the stimuli. I predict that subjects would no longer be able to disambiguate the prosodically contrastive stimuli presented to them. Therefore, their response times to probes confirming the interpretation consistent with the prosody of the original stimuli would be similar to their response times to probes conflicting with this interpretation. In other words, I predict that subjects' response times to appropriate probes (probes that confirm the interpretation intended by the original stimuli) would be similar to subjects' response times to inappropriate probes (probes that conflict the interpretation intended by the original stimuli).

However, the f_0 of the stimuli was not the only prosodic cue differentiating the noun and verb phrase versions of the word pairs. There were still differences between the amplitude of the noun and verb phrase versions of the word pairs. Similarly, there were differences between the timing of the noun and verb phrase versions of the word pairs. Therefore, subjects' parsing *might* be guided by differences in amplitude and timing between the noun and verb phrase versions of the stimuli. If so, subjects would still be able to predict the intended parses of the word pairs. Their response times to probes that confirm this interpretation (appropriate probes) would be shorter than their response times to probes that conflicted with this interpretation (inappropriate probes). Such a result would indicate that subjects were able to disambiguate the word pairs despite flattening the f_0 of the stimuli.

Subjects were not able to disambiguate the prosodically non-contrastive stimuli in Experiments 1 and 2. Therefore, I do not expect that subjects will be able to disambiguate the flattened prosodically non-contrastive stimuli presented to them. Their response times to probes that confirm the intended interpretation of the stimuli (appropriate probes) would be similar to their response times to probes that conflict with this interpretation (inappropriate probes). The differences in amplitude and timing might guide subjects' disambiguation of the prosodically contrastive stimuli in this experiment. However, in Chapter 2, I found no consistent differences in amplitude or timing of the noun and verb phrase versions of the prosodically non-contrastive stimuli. Therefore, amplitude and timing would not guide subjects towards the intended interpretations of the prosodically non-contrastive stimuli. Subjects' response times to appropriate and inappropriate probes should be similar.

4.6. RESULTS

As in Experiments 1 and 2, only correct responses under 1500 milliseconds were analysed. 4277 responses from the original 4560 responses were included. The mean response times to appropriate and inappropriate probes following prosodically contrastive and prosodically non-contrastive stimuli are presented in Table 4.2.

Stimuli	Mean response time to appropriate probes (ms)	Mean response time to inappropriate probes (ms)
With prosodic contrast	751	772
No prosodic contrast	747	764

Table 4.2: Mean response times to appropriate and inappropriate probes following prosodically contrastive vs. non-contrastive stimuli (ms)

The mean response time to appropriate probes was less than the mean response time to inappropriate probes following both the prosodically contrastive and non-contrastive stimuli. The response time data were then analysed to see if the response times were normally distributed or not. A Kolmogorov-Smirnov test established that the data were not normally distributed round the mean ($p < .001$). Consequently, the results of parametric tests would have to be treated with caution. Therefore, I conducted non-parametric tests to ensure that any conclusions were not contingent on normally distributed data.

I conducted two different kinds of non-parametric tests. First, I conducted a Mann Whitney test – a non-parametric equivalent of the independent samples test. Next, I conducted a Wilcoxson Signed ranks test, which is the non-parametric equivalent of the paired samples t-test. The Mann Whitney test indicated that there was a possibility that the difference between response times to appropriate and inappropriate probes was marginally significant ($z = -1.609$; $p > .05$). This value was at the borderline of significance. Standard expectations of significance would not consider the mean of response times significantly different (Kinnear and Gray: 2000: 150). In addition, I found that the mean response time to appropriate probes following the prosodically non-contrastive stimuli was significantly different from the mean response time to inappropriate probes ($z = -1.673$; $p < .05$). However, this effect was restricted to the results of the Mann Whitney tests.

Finally, I also conducted Wilcoxson Signed ranks tests on the significance of the difference between the two conditions. The Wilcoxson Signed ranks test was performed on the median of reaction times to appropriate and inappropriate continuations of the individual stimuli. The results suggested that the difference between response times to appropriate and inappropriate probes following the prosodically contrastive stimuli were significant ($z = -1.883$; $p < .05$). Analyses of the response times following prosodically non-contrastive stimuli suggested that response times to appropriate and inappropriate probes were not significantly different (z = –

1.334; $p < .1$). However, the Wilcoxson signed ranks test is dependent on the distribution shape of the populations and can be biased by outliers. Therefore, the results of the Wilcoxson tests have also to be treated with caution.

I also performed within-item and within-subject analyses. These figures are reported in the tables below.

Subject	Mean response times (ms) Appropriate probes	Mean response times (ms) Inappropriate probes	Statistical significance
1	956	993	$p > .1$
2	1020	1067	**$p < .1$**
3	827	862	$p > .1$
4	523	545	$p > .1$
5	947	960	$p > .1$
6	493	529	**$p < .1$**
7	683	690	$p > .1$
8	681	711	$p > .1$
9	1147	1121	$p > .1$
10	605	620	$p > .1$
11	564	559	$p > .1$
12	764	791	$p > .1$

Table 4.3: Mean response times for appropriate and inappropriate probes to prosodically contrastive stimuli per subject

Item	Mean response times (ms) Appropriate probes	Mean response times (ms) Inappropriate probes	Statistical significance
Cooking apples	732	755	$p > .1$
Cutting boards	771	778	$p > .1$
Packing cases	769	780	$p > .1$
Playing cards	756	740	$p > .1$
Racing cars	737	807	**$p < .05$**
Sailing ships	743	772	$p > .1$

Table 4.4: Mean response times to appropriate and inappropriate probes following prosodically contrastive stimuli per item.

Two subjects showed slower response times to inappropriate than to appropriate probes following the flattened versions of the prosodically contrastive stimuli (Subjects 9 and 11). This indicates that the other subjects might have been able to differentiate between the noun and verb phrase versions of the stimuli.

However, none of these differences was significant. This does not entail that subjects were not able to differentiate between the ambiguous stimuli, since the size of the data set might have been too small for realistic expectations of significance. The results of the within-subject analysis in Experiment 3 do not contribute to or detract from the conclusions I present here. The mean response time to appropriate probes was less than the mean response time to inappropriate probes following all except one item – *Playing cards*. In addition, this difference was significant in the case of one of the items – *Racing cars*. This was to be expected since there remained consistent differences in the timing and amplitude of the words between the noun and verb phrase versions of the word pair.

I also performed Pearson correlations to test whether subjects' response times were affected by the order of presentation of the stimuli. The stimuli had been randomly distributed and subjects were randomly assigned to one of two groups with different orders of presentation of the stimuli. Therefore, the likelihood of finding a confounding effect of order was minimal. I found that there was no effect of order on subjects' response times ($z = -0.057$; $p > .05$).

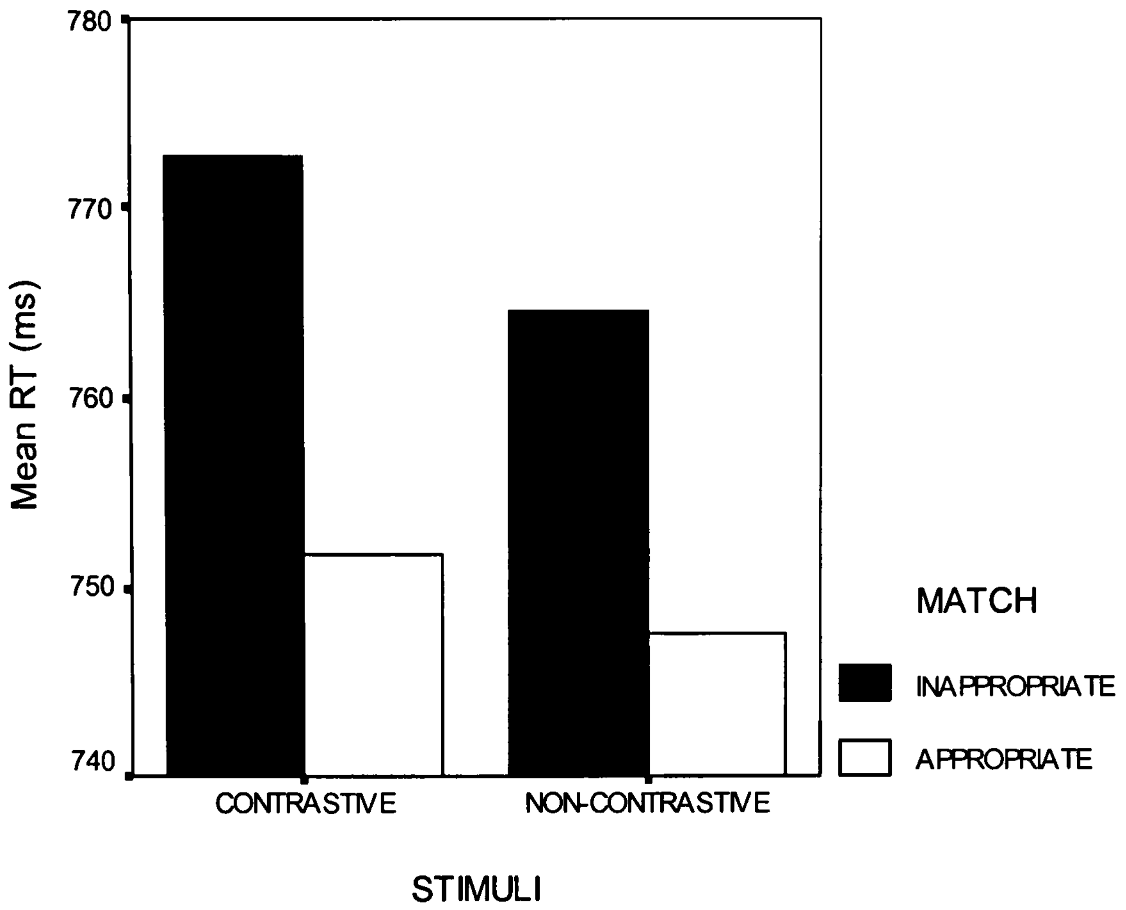


Figure 4.1 Mean response times to appropriate and inappropriate probes following the stimuli with and without prosodic contrast.

Figure 4.1 plots the mean of the response times to appropriate and inappropriate probes following the prosodically contrastive and the prosodically non-contrastive stimuli. There appears to be a large difference between response times to appropriate and inappropriate probes following the contrastive and non-contrastive stimuli. However, further observations are constrained by the results of the statistical tests.

I also performed Pearson correlation tests on the significance of the correlation between subjects' response times and the amplitude and timing of the stimuli. Response times were not significantly correlated with any of the acoustic measures tested in Experiments 1 and 2 (all p's > .1).

I also tested whether subjects' response times to probes that confirmed the noun phrase reading were faster than response times to probes that confirmed the verb phrase reading. I found that subjects' response times to 'is' and 'are' probes following all the stimuli were similar. ('is' (rt) = 756 ms; 'are' (rt) = 761 ms; $z = -.929$; $p > .05$). The lack of difference between response times to 'is' and 'are' probes was maintained in independent tests of response times to only the prosodically contrastive ('is' (rt) = 766 ms; 'are' (rt) = 758 ms; $z = -.511$; $p > .05$)) and non-contrastive stimuli ('is' (rt) = 746 ms; 'are' (rt) = 765 ms; $z = -1.8$; $p > .05$).

Finally, I also conducted an appropriateness-rating task. This was an off-line test of subjects' parses of the stimuli. It also focused subjects' attention on the cohesiveness of the visual and auditory stimuli. The results of this task are given in Table 4.5, based on the number of *good* and *bad* evaluations given to cases where the probes were good or bad continuations of the auditory stimuli. The implications of these results are discussed in the next section. Subjects gave significantly more good evaluations of appropriate probes than inappropriate probes following the stimuli with prosodic contrast (t (2282 = 11.765; $p << .05$). Subjects did not give significantly more good or bad evaluations to appropriate and inappropriate probes following the stimuli with no prosodic contrast (t (2298) = 0.166; $p > .5$).

Evaluation	With prosodic contrast		No prosodic contrast	
	Appropriate probes	Inappropriate probes	Appropriate probes	Inappropriate probes
Good	827	559	682	665
Bad	316	582	480	473

Table 4.5 Number of good and bad evaluations awarded to appropriate and inappropriate probes following the stimuli with vs. without prosodic contrast

4.7. DISCUSSION

Experiment 1 established that subjects could differentiate between noun and verb phrase versions of the prosodically contrastive stimuli presented to them. Furthermore, parsing was guided by the differences between the prosodic characteristics of the noun and verb phrase versions of the word pairs. Parsing was guided by differences in the f_0, amplitude, and timing of the word pairs (Experiment 2). This prompted the question being tested by Experiment 3: What would happen to subjects' ability to differentiate between the noun and verb phrase versions of the word pairs if the f_0 of the stimuli was flattened? The results of Experiment 3 establish that subjects' parsing was affected by flattening the f_0 of the stimuli. Subjects were not able to differentiate between the noun and verb phrase versions of *some* of the word pairs because the stimuli had been flattened.

The mean of response times to appropriate probes was less than the mean response time to inappropriate probes. However, there was a less than a one in thousand probability that the difference was due to chance in Experiment 1. This indicated that almost all response times to appropriate probes were faster than response times to inappropriate probes. This was reduced to close to a one-in-ten probability that difference between response times to appropriate and inappropriate probes was due to chance in Experiment 3. This indicated that some response times to appropriate probes were faster than response times to inappropriate probes. However, this was not so in other cases. The only difference between the stimuli presented to subjects in Experiments 1 and 3 was that the stimuli in Experiment 3 had been flattened. The strong possibility that the difference between mean response time to appropriate and inappropriate probes was due to chance in Experiment 3 suggests that subjects' parsing was affected by the flattening of f_0. This indicates that f_0 was influential in guiding parsing of the ambiguous alternatives of the word pairs in Experiments 1 and 2.

This conclusion is not affected by the possibility that the difference between mean response times was not due to chance – given the one-in-ten probability that they could be. Subjects might have been able to disambiguate some of the word pairs presented to them. However, flattening the f_0 made it difficult for subjects to disambiguate *all* the prosodically contrastive word pairs presented to them. The discussion of the acoustic characteristics of the stimuli in Chapter 3 shows that differences in f_0 were not the only prosodic cues marking the contrast between the noun and verb phrase versions. There were also consistent differences in the amplitude and timing of the noun and verb phrase versions. Therefore, subjects might have been guided by these differences in amplitude and timing in disambiguating *some* of the stimuli presented to them.

Furthermore, there were large differences between the f_0 of the noun and verb phrase versions of some of the stimuli – especially *Cutting boards*, *Sailing ships* and *Playing cards*. It is possible that flattening the f_0 affected subjects' parsing of these stimuli adversely. Conversely, the differences between the f_0 of the noun and verb phrase versions of other word pairs were not as large, e.g., *Cooking apples*, *Racing cars*. It is possible that flattening the f_0 did not affect subjects' parsing of the latter stimuli.

This was a possibility that became apparent during the running of the experiment. Subjects' feedback suggested that some pairs had proved easier to distinguish than others. When questioned about which pairs stood out as being sufficiently dissimilar, subjects sometimes reported the pairs *Cooking apples* and *Packing cases*. There were large differences between the amplitude of the noun and verb phrase versions of *Packing cases*. In fact, as is clear from Figure 2.5, the difference between the amplitude of the noun and verb phrase reading of *Packing cases* was the greatest of all the stimuli-pairs. I did not find similar consistent differences between the timing of the noun and verb phrase reading of *Packing cases*. There were also differences between the amplitude of the noun and verb phrase versions of *Cooking apples*. However, this difference was not as large as the difference between the noun and verb phrase versions of *Packing cases*. In fact, the amplitude of the first word in the verb phrase reading of *Cooking apples* was greater than the amplitude of the second word, although there was sustained high amplitude on the second word of the verb phrase reading. Contrarily, the noun phrase reading displays a drop in amplitude over the second word (see Figure 2.3 and 2.4). In addition, there were consistent differences between timing of the noun and verb phrase versions of *Cooking apples*. The timing of the first word in the noun phrase reading was greater than the timing of the first word in the verb phrase reading. Conversely, the timing of the second word in the noun phrase reading was less than the timing of the second word in the verb phrase reading. These differences in timing and amplitude might have made the noun and verb phrase versions of these word pairs sufficiently dissimilar. It is possible that subjects' used these prosodic differences to disambiguate these word pairs.

There were no significant correlations between subjects' response times and the remaining prosodic characteristics of the stimuli. Correlation tests considered subjects' response times to appropriate probes following the prosodically contrastive stimuli alone. Subjects' response times did not vary significantly with changes in the amplitude or timing of the stimuli. This might suggest that subjects' response times were not affected by differences in amplitude and timing. The hypothesis behind my use of the correlation tests is that the stimuli with greater differences between the amplitude of the noun and verb phrase versions might guide subjects towards the intended interpretation more quickly. Therefore, subjects' response times to probes confirming this interpretation might be faster with greater differences in amplitude between the two alternative interpretations. It is possible that such an effect might either not occur or not be recognised by a correlation test. The absence of significant correlations does not provide strong evidence against amplitude or timing influencing parsing. However, the absence of significant correlations does suggest that there might be other factors guiding subjects towards the intended interpretation of the stimuli. I consider this next in discussing subjects' response times to appropriate and inappropriate probes following the prosodically non-contrastive stimuli.

The mean response time to appropriate probes was less than the mean response time to inappropriate probes following the prosodically non-contrastive stimuli. However, there was a close to one-in-ten probability that this difference was due to chance. Interpretation of the results may vary depending on independent statistical expectations of significance. This might indicate that subjects were not able to differentiate between the noun and verb phrase versions of the flattened prosodically non-contrastive stimuli. This was expected since there were no consistent differences between the amplitude and timing of the noun and verb phrase versions to guide parsing. Subjects in Experiments 1 and 2 were unable to disambiguate the prosodically non-contrastive stimuli. Consequently, I did not anticipate that subjects would be able to disambiguate the flattened stimuli, since they presented subjects with even less prosodic information.

The possibility of subjects disambiguating the noun and verb phrase versions of the prosodically contrastive stimuli was a reasonable argument: there were still consistent differences in amplitude and timing that subjects could have used to disambiguate the stimuli. However, this was not the case for the prosodically non-contrastive stimuli. There were no consistent differences between the amplitude and timing of the prosodically non-contrastive stimuli. Again, this points to the existence of other factors guiding subjects towards the intended interpretation of the stimuli.

The word pairs were edited from sentences where they were followed by verbs confirming the interpretation intended of the stimuli. The noun phrase versions were followed by 'are' and verb phrase versions were followed by 'is' in the original sentences. There is a possibility that the final fricative of the word pairs might have been co-articulated with the initial vowel of the verb to follow. Therefore, the spectral characteristics of the final fricatives would have been different in the noun and verb phrase versions of the word pairs. This effect would be common for the prosodically contrastive and non-contrastive stimuli. In Chapter 3, I report that there were differences between the spectral characteristics of all the segments in the noun and verb phrase versions. Subjects' ability to differentiate between the noun and verb phrase versions of the stimuli might have been guided by the spectral differences between the alternative forms. However, Experiment 2 had tested this possibility: I found that subjects' interpretations of the stimuli were not in keeping with the interpretations confirmed by the spectral characteristics of the word pairs. On the other hand, Hirschberg and Ward argue that spectral characteristics were not as influential as f_0 in guiding parsing. It is possible that in the absence of f_0 information, the spectral characteristics cue the intended interpretation of the prosodically contrastive and non-contrastive stimuli. This would account for spectral information not influencing parsing in Experiment 2, while playing a major role in Experiment 3. I did find a stronger possibility that the difference between response times to appropriate and inappropriate probes was due to chance in Experiment 2. This might have been because of conflicting spectral and prosodic information. Similarly, subjects' ability to disambiguate the prosodically non-contrastive stimuli in Experiment 3 could also be explained by attributing an influential role in parsing to spectral information. This might also account for the lack of significant correlations between subjects' response times to appropriate probes following the prosodically contrastive stimuli and differences in amplitude and timing. Hirschberg and Ward had argued that amplitude and timing do not guide parsing. It is possible that spectral information, and not amplitude and timing of the stimuli, were guiding parsing of the prosodically contrastive and non-contrastive stimuli in Experiment 3. However, further research into these effects is required to provide strong evidence of an influential role for spectral information in guiding parsing on-line, especially given the results of Experiment 2 of this thesis.

Subjects also gave more 'good' evaluations to appropriate probes than to inappropriate probes following the prosodically contrastive stimuli and almost twice as many 'bad' evaluations to inappropriate probes than to appropriate probes following the prosodically contrastive stimuli. The results of the appropriateness-rating task add support to the suggestion that subjects were able to differentiate between the noun and verb phrase forms of some of the stimuli. The results of the off-line test show that subjects were able to disambiguate the prosodically contrastive stimuli presented to them. The results of the on-line tests show that subjects were able to disambiguate *some*, but not a majority of the stimuli presented to them. Therefore, subjects' parsing was affected by flattening the f_0 of the stimuli. In addition, subjects' off-line preferences indicated that they had not been able to differentiate the noun and verb phrase versions of the prosodically non-contrastive stimuli. Subjects gave almost the same number of 'good' evaluations to both appropriate and inappropriate probes. Similarly, they gave an equal number of 'bad' evaluations to both appropriate and inappropriate probes. However, the results of the on-line tests indicated that subjects might have been able to disambiguate some of the prosodically non-contrastive stimuli.

Importantly, subjects did not follow the principles of late closure or minimal attachment in initial parsing. They did not prefer noun phrase parses of the stimuli, even in the absence of adequate prosodic information. Carston (1989) argues that the noun phrase reading of the stimuli is syntactically preferable. It is possible that the syntactic content of the stimuli is not adequate to trigger late closure or minimal attachment parsing preferences. There was also no indication of subjects' preferring verb phrase parses of the stimuli as in Experiment 1 and 2. This does not mean that f_0 was guiding parsing towards verb phrase parses. It does support a connection between prosody and subjects' preference for verb phrase parses of the stimuli. Subjects probably used the spectral content of the stimuli to disambiguate any word pairs they could.

There is a difference between subjects' response times in Experiment 3 and the other experiments reported earlier. Subjects' response times were considerably later than in Experiments 1 and 2. The mean response time was around 750 milliseconds in Experiment 3. It is also possible that subjects' response times were delayed because of the quality of the stimuli. The stimuli sounded tinny and unnatural. This might have delayed parsing, with subjects still getting used to the quality of the stimuli. Therefore, it is also conceivable that subjects' parsing was not affected by flattening the f_0, as much as subjects' response times were affected by the quality of the stimuli. In other words, accurate parsing was not affected by the absence of f_0 – f_0 might not be essential to guiding parsing. However, the stimuli played to subjects in Experiment 2 sounded equally tinny. Subjects' parsing was not affected by the quality of the cross-synthesised stimuli. I do not believe that they were affected by the quality of the flattened stimuli, although the flattened stimuli did sound more unnatural because of the monotone. Admittedly, this raises a possible objection for future research to consider.

At the beginning of this chapter, I raised the issue that previous research found that speakers do not reliably produce consistent differences between the f_0 of ambiguous alternatives. Does this mean that f_0 cannot influence parsing? This question has been countered by the results of Experiment 3: flattening the f_0 of the stimuli does affect subjects' disambiguation of the word pairs. Having established that

f_0 does guide parsing, it remains to be addressed whether the absence of consistent prosodic cues in speech necessitate a lesser role for prosody in parsing.

Earlier cue trading research has demonstrated that listeners are flexible in their use of acoustic cues in perception. Speech perception is marred by variability in the acoustic characteristics of the percept, with different acoustic cues signalling the same phonetic contrast. In prior cue trading research, the stimulus is manipulated to diminish the effectiveness of one cue in signalling a phonetic contrast. It was found that listeners could still perceive speech using the other cues signalling the same contrast (Luce and Pisoni: 1986). Hermann et al (2003) present a similar argument for the processing of syntactic contrasts using prosodic information. They test the online perception of f_0 contrasts by recording subjects' ERP measures when presented with normal and flattened speech. Hermann et al suggested that the absence of adequate prosodic information would lead to decreased activity in the right hemisphere when subjects were presented with prosodically flattened speech. Conversely, they found that there was increased activity in the right hemisphere when subjects were presented with prosodically flattened stimuli. Hermann et al argue that this increased activity might have been caused by increased processing load. They propose a prosodic correlate of the Gestalt phenomenon, arguing that the brain compensates for the lack of adequate information by filling in the missing prosodic information. However, the f_0 of a stimulus does not compose the entire prosodic component of speech. For instance, the contrast between the noun and verb phrase versions of the word pairs presented to subjects in these experiments was also marked by differences in amplitude and timing. There were also spectral differences between the two versions of the word pair, though these are sometimes excluded from purist accounts of prosody. Flattening the f_0 does not remove these other prosodic cues. The brain still has to process the differences in amplitude and timing. While processing the other prosodic cues (such as amplitude and timing) might account for continued activity in the right hemisphere, it does not explain the *increased* activity reported by Hermann et al. The explanation provided by the Gestalt phenomenon could provide a solution to the problems of inconsistency in the production of sufficiently contrastive prosodic information. If the brain does actually fill in missing prosodic information, then the fact that speakers do not always produce sufficiently contrastive prosodic information does not contradict initial prosodic processing. However, there is not much evidence apart from the argument suggested by Hermann et al in favour of a processing correlate of the Gestalt phenomenon. In addition, Hermann et al do not provide details of the exact mechanism by which this Gestalt phenomenon would apply during on-line speech processing.

One of the ways to situate a Gestalt version of prosodic processing is by assuming a prosodic correlate of the Fuzzy Logic model of speech perception (Massaro and Cohen: 1983). Arguments against the perception of speech segments using acoustic information have drawn strength from the supposed lack of invariance of the acoustic content of a stimulus. However, proponents of FLMP models of speech perception argue that acoustic information can guide phonemic perception. They argue that perception occurs by matching the features of the physical stimulus to a prototype. As long as a required number of features are sufficiently similar, the stimulus is recognised as the prototypical segment. Similarly, Gestalt models could argue that prosodic processing takes place by matching the features of the physical stimulus to a prosodic prototype. Perception could take place by comparing the features of the prototype to that of the new stimulus. Once the stimulus is recognised as the prototype, the missing information is filled in by the brain. In the case of flattened stimuli, recognition takes place using the other prosodic characteristics that have been provided.

Using the results of Experiment 3, I argue that when the content of one of these cues is affected, listeners use other prosodic cues to guide parsing. It is uncertain whether listeners use the spectral content of the stimuli or the amplitude and timing of the stimuli to guide parsing. The analysis of response times following the prosodically non-contrastive stimuli suggests that spectral information was guiding parsing of these word pairs. Therefore, it is conceivable that listeners were using spectral information to guide parsing of the prosodically contrastive stimuli. However, the flexibility suggested by cue trading entails that there is no surety that listeners were using only spectral information and not amplitude and timing to guide parsing of the prosodically contrastive stimuli. Currently, I do not have evidence that either the Gestalt model or FLMP models can describe the mechanism of on-line processing of prosodic information. However, these theories present a possible explanation for the flexibility exhibited by subjects in Experiment 3. They also present a solution to the problems of inconsistency in the production of sufficient prosodic cues in natural spontaneous speech.

In addition, the possibility that speakers might not produce consistent differences between the f_0 of the ambiguous alternatives only raises problems for suggestions that parsing cannot take place without f_0. I approach this problem from another angle. I tested whether speakers were able to use prosodic information to guide parsing. Experiments 1 and 2 established that prosodic information could guide parsing. The results of Experiment 2 could only be explained by suggesting that prosodic information was guiding parsing. This has many implications for the arguments reported in this section. The speaker obviously produced sufficiently contrastive prosodic information to guide parsing. It could be argued that the speaker only produced sufficient prosodic cues because he was aware of potential ambiguity. Furthermore, the stimuli were produced by asking the speaker to read out sentences presented to him. These factors might have improved the consistency with which contrastive prosodic cues were produced. However, listeners were able to use these prosodic cues to guide parsing. Listeners were able to associate a particular contour with the noun phrase reading and an alternative prosody with the verb phrase reading. They did not learn this association during the experiment. There were no effects of order on subjects' parsing of the stimuli. This implies that listeners are already familiar with the prosody of the noun and verb phrase versions. They must have been exposed to it before the experiment. This can only imply that speakers do produce these cues in spontaneous speech outside of laboratory situations. Finally, earlier experiments have provided evidence for prosodic information influencing parsing. Due to reasons discussed in Chapter 1, none of these experiments provides strong evidence for prosodic information guiding initial analysis of ambiguous stimuli. However, the lack of invariance of prosodic cues does not stand in the way of prosodic information influencing parsing during reanalysis. Consequently, this lack of invariance may not be an obstacle to the possibility of prosodic information *guiding* parsing. Indeed, despite the variability in the production of consistent differences between the f_0 of the noun and verb phrase versions of the stimuli, subjects were able to use f_0 to guide parsing on-line.

4.8. CONCLUSION

The key concern of Experiment 3 was to test whether flattening the f_0 of the stimuli affected subjects' ability to disambiguate the noun and verb phrase versions of the word pairs. Subjects were still able to disambiguate some of the noun and verb phrase versions of both the prosodically contrastive and non-contrastive word pairs. However, there was a one-in-ten probability that the difference between response times to the appropriate and inappropriate probes was due to chance. The only difference between the stimuli presented to subjects in Experiments 1 and 3 was that the stimuli had been flattened in Experiment 3. Therefore, subjects' inability to accurately parse *all* of the stimuli suggests that parsing was affected by flattening the f_0 of the word pairs. Furthermore, this provides evidence that f_0 was influencing on-line parsing in Experiments 1 and 2. Experiment 3 establishes an influential role for f_0 in parsing. Moreover, since the experiment was an on-line test of parsing, I conclude that f_0 is accessed during the initial stage of parsing. This suggests that continual monitoring of the accents in a stimulus is a real possibility with the parser comparing the accent of a newly presented item to the accents on the already processed parts of the stimuli. Therefore, there is little need to propose either a parsing-by-chunks mechanism or necessitate off-line parsing using f_0. However, further evidence of this is required before concluding that parsing using f_0 follows continual monitoring.

In Chapter 2, I detail the differences between the prosodic characteristics of the noun and verb phrase versions of the prosodically contrastive stimuli. There were differences in the f_0, timing, amplitude, and spectral characteristics of the noun and verb phrase versions. However, in some cases, there were differences only between the f_0 of the noun and verb phrase versions of the word pairs. In other cases, there were differences between, for instance, the amplitude and timing of the noun and verb phrase versions of the word pairs. This presents obstacles to suggesting that prosodic information can influence parsing on-line, since there were few consistent prosodic differences between the noun and verb phrase versions of the stimuli. However, the results of Experiment 3 provide evidence that listeners can cope with this variability in the production of prosodic information. Listeners are flexible in their use of prosodic information in parsing. Therefore, despite the absence of f_0, subjects were able to use either the spectral or the remaining prosodic content of the stimulus to guide parsing of *some* of the stimuli.

The possibility that subjects were still able to disambiguate some of the flattened stimuli raises the question whether there are other cues being presented to subjects in the word pair stimuli – most noticeably, spectral and lexical information. Chapter 3 looked at the role of spectral information in guiding parsing in detail. Given the results of Experiment 3, evidence suggests an influential role for spectral information in the absence of. This is in keeping with the results of Hirschberg and Ward (1992). However, the influence of amplitude and duration needs to be disentangled from the influence of spectral information. In addition, the possible influence of lexical information needs to be explored. Therefore, in Chapter 5, I analyse whether subjects were being guided by lexical preferences in parsing. Chapter 6 of this thesis takes this issue further by looking at whether subjects continue to use prosodic information to guide parsing in the absence of any lexical information.

CHAPTER FIVE
THE ROLE OF LEXICAL INFORMATION IN INITIAL PROSODIC PROCESSING

5.1. INTRODUCTION

Experiments 1 and 2 provide evidence of the early incorporation of prosodic information in parsing. Subjects' ability to make accurate parses of the stimuli presented to them could not have been guided by the syntactic or spectral content of the stimuli. However, there are two caveats to this claim that I address in this chapter. Firstly, there is the possibility that prosodic processing is dependent on the lexical content of the stimuli, i.e. prosodic information cannot guide parsing when the lexical content of the stimuli is removed. Therefore, parsing using prosody must be an interactive process combining the prosodic and lexical resources of the word pairs. In order to verify this claim, Experiment 4 investigates the autonomy of prosodic on-line processing by testing subjects' parsing of delexicalised stimuli. Delexicalised stimuli sound like hearing someone speaking with a hand over his or her mouth or someone speaking in another room with the door closed. The lexical content of the stimuli is obscured, leaving the purely prosodic content of the stimuli intact. Neither syntactic, nor spectral, nor lexical information could guide subjects towards the intended parses of the delexicalised stimuli. Therefore, subjects' ability to disambiguate such stimuli would confirm independent prosodic processing of ambiguous stimuli, even in the absence of lexical information

Fodor (1983) proposes the existence of a syntactically autonomous module geared towards processing only the syntactic content of speech. Fodor argues that the efficiency of the module is increased by the fact that it can process only syntactic information. Fodor argues that the other information provided by the stimulus would subsequently be parsed by central systems, which are not restricted to processing only the syntactic content of the stimulus. The syntactic module would ensure that processing of syntactic information would be earlier than the processing of the other information provided by the signal. This proposal gives syntactic information priority in initial parsing. Fodor argues that the other information provided by the signal cannot *independently* guide parsing. Consequently, he argues that nothing apart from syntactic information can be parsed independently by informationally encapsulated modules. Suspecting otherwise, I test the possibility of autonomous prosodic processing – I test whether prosodic information can independently guiding parsing.

The second caveat to the conclusions reached by Experiments 1, 2, and 3 concerns the possibility of an influence of lexical bias on parsing. Current literature argues that the likelihood of the interpretations of ambiguous stimuli in speech influences the ease with which either alternative is parsed (Binder and Morris: 1995; Binder and Rayner: 1998; Carpenter and Daneman: 1981; Rayner, Pacht and Duffy: 1994; Tabossi et al: 1987). Specifically, the dominant interpretations are easily available, while the less frequent interpretations are accessed with more difficulty. Therefore, it is possible that the likelihood of the noun or verb phrase interpretations influences subjects' parsing in Experiments 1, 2 and 3 – i.e., if the noun phrase interpretation was more likely, subjects would find it easier to parse the word pair as a noun phrase. This does not detract from the results of the first three experiments of this thesis. Lexical bias would not lead subjects towards the intended parses of the stimuli in half of the presentations. Therefore, subjects' ability to make accurate parses of the stimuli in the first three experiments was guided by prosodic information. In this chapter, I consider the possibility of an additional influence on the stimuli competing with the prosodic effects already established.

Next, I shall review arguments against the perceptual salience of prosodic information in the absence of lexical information. I also raise the possibility of prosodic bootstrapping of syntactic structure to provide evidence for the perceptual salience of purely prosodic information. Then, I review some of the prior work testing the parsing of delexicalised speech and I argue that such research does not provide answers to the issues raised in this introduction. Finally, I take another look at the data obtained in Experiments 1, 2, and 3 to test whether there is an influence of lexical bias on parsing that is coincident with the already established influence of prosodic information.

5.2. THE PERCEPTUAL SALIENCE OF PURELY PROSODIC 'SPEECH'

There is considerable debate about the possibility of prosodic cues influencing parsing in the absence of lexical information. For instance, Toepel and Alter (2002) argue that purely prosodic information cannot guide parsing. They studied ERPs when subjects were listening to delexicalised stimuli to see whether they could find the closure positive shift reported by Steinhauer et al (1999) using normal speech (see Chapter 1). Toepel and Alter (2002) studied ERPs when subjects were listening to delexicalised and normal versions of sentences such as (5.1) and (5.2), which had previously been tested by Steinhauer et al (1999):

(5.1) Peter verspricht Anna zu arbeiten ...und das Buro zu putzen

Peter promises Anna to work ... and to clean up the office

(5.2) Peter verspricht Anna zu entlasten ...und das Buro zu putzen

Peter promises to support Anna... and to clean up the office.

The differences in syntactic structure between the two sentences are presented in the figures below, from Steinhauer (2003).

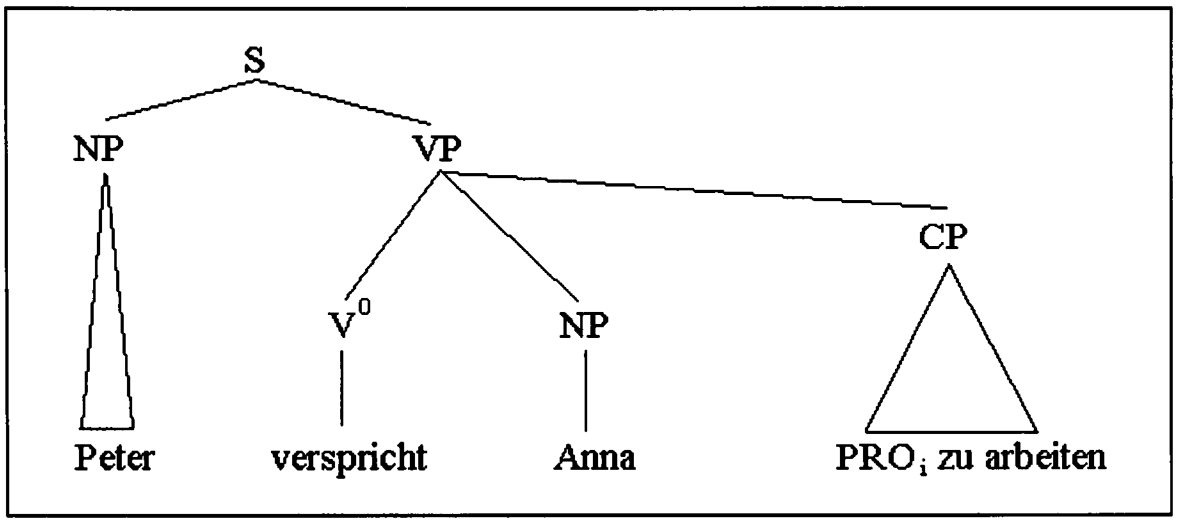


Figure 5.1 Proposed syntactic structure of part of (5.1)

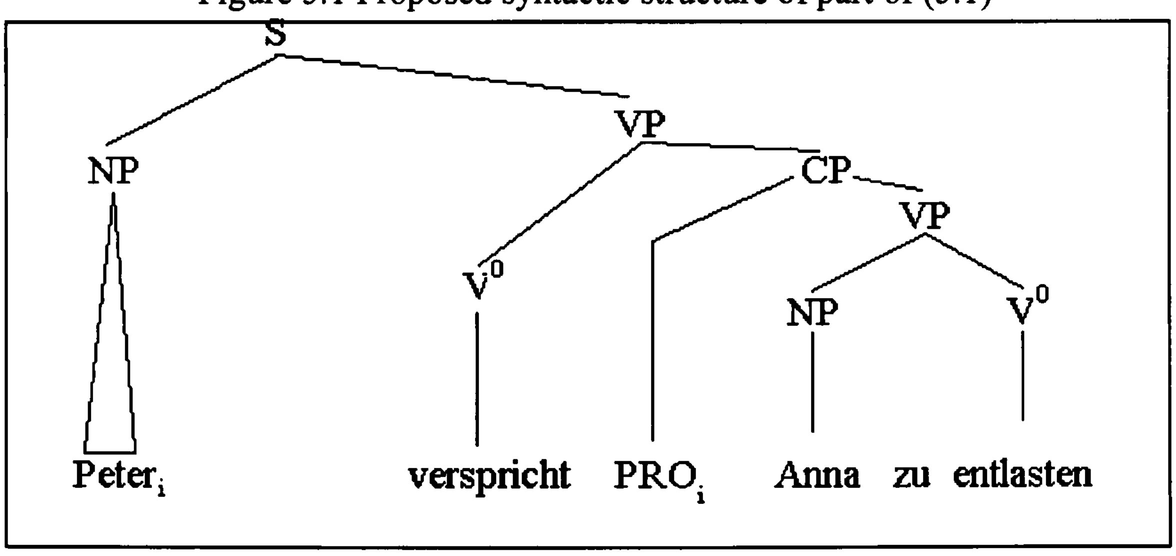


Figure 5.2 Proposed syntactic structure of part of (5.2)

The syntactic and lexical content of the stimuli are ambiguous until the second noun phrase. In (5.1), the second verb *arbeiten* is intransitive. *Anna* in (5.2) is the object of the first verb *verspricht*. However, the second verb in (5.2) is transitive. *Anna* is not the object of the second verb. A parser applying the principle of minimal attachment would predict a direct object parse of (5.2), with *Anna* as the object of the first verb *verspricht*. The parser would be forced to backtrack upon receiving the second verb, *entlasten* and would have to reparse *Anna* as the object of the second verb. Steinhauer et al (1999) argue that there are prosodic differences between the two structures that can cue the deeper embedding of *Anna* in (5.2). (5.2) is produced with signals to an additional boundary after the first verb. In addition, the sentence accent shifts from the second verb in (5.1) to *Anna* in (5.2). As I mentioned earlier, Steinhauer et al (1999) reported finding a closure positive shift after the first verb in (5.2). They argue that this was caused by the additional processing required by the presence of the extra boundary after the first verb in (5.2).

In order to test this hypothesis further, Toepel and Alter presented subjects with delexicalised and normal versions of the sentences. If prosody were causing the closure positive shift, then it should be found even when subjects were listening to delexicalised stimuli. Subjects were played entire sentences and later asked questions related to one of the interpretations of the sentences. Using normal sentences, they found that a closure positive shift was present only after the second verb in (5.1), while it was present after the first and second verbs in (5.2). When presented with delexicalised sentences, they found a large negativity at the beginning of processing of the delexicalised stimuli. Toepel and Alter argue that this was because of the increased load on processing caused by insufficient information being presented. However, they did not find a closure positive shift after the first verb when subjects were presented with delexicalised versions of (5.2). Toepel and Alter argue that the absence of a closure positive shift was caused by the difficulty in forming associations between intonational phrase boundaries and delexicalised speech. They suggest that the IPh in delexicalised stimuli cannot be used to parse speech because the stimuli lack semantic content. Toepel and Alter's conclusions were corroborated by the fact that subjects were not able to choose the right interpretation of the stimuli when presented with delexicalised speech. They also found that subjects were not able to compare the acoustic characteristics of the two sentences. Toepel and Alter argue that their results present evidence of the lack of perceptual salience of prosodic cues in the absence of lexical information. This suggests that prosodic processing of ambiguous stimuli cannot take place in the absence of lexical information.

However, Steinhauer (2003) presents a second ERP experiment with results contradictory to Toepel and Alter (2002). Steinhauer (2003) played subjects delexicalised versions of both sentences. The auditory stimuli were followed by the sentences being presented word-by-word on a screen. Subjects were asked to read the sentences silently and then replicate the prosody of the auditory stimulus while interpretation. Subjects were presented with both appropriate and inappropriate written sentences following the auditory stimuli. Therefore, subjects were presented with the delexicalised version of the sentence in (5.1) followed by the written presentation of (5.1) in half of the presentations and written presentations of (5.2) in the other half. They found a closure positive shift at the boundary after *verspricht* when listening to delexicalised versions of (5.2), *before* version the word-by-word presentations of the sentences. They found a similar closure positive shift after *verspricht* during word-by-word presentations of the sentences when subjects had been presented with delexicalised versions of (5.2).

The results reported by Steinhauer (2003) provide evidence for the processing of prosodic information in the absence of lexical information. The discrepancy between the results reported by Toepel and Alter and Steinhauer (2003) might be due to the difference in quality of the signals. The boundaries and the placement of accents in Steinhauer (2003) might have been exaggerated to compensate for the difficulty of the task. Irrespective of this possibility, subjects seem to be able to process prosodic information in the absence of lexical content of the stimuli. However, this does not provide evidence that subjects can use purely prosodic information to guide parsing of syntactic structure. Neither Toepel and Alter (2002) nor Steinhauer (2003) test the prosodic processing of delexicalised speech. Firstly, the tasks were focussed towards specification of brain activation during processing. Although they test subjects' ERPs at different points in the sentence, they do not actually test subjects' *parsing* at these different points. The tests lacked the appropriate controls required to argue against prosodic processing of delexicalised speech. Parsing was tested off-line, after subjects were presented with delexicalised versions of *full* sentences. Consequently, they do not test subjects' parsing at the points of syntactic ambiguity, where sufficiently contrastive prosodic information is presumably available to guide parsing. A suitable test would be to test subjects' parsing after the presentation of the first verb in (5.1) and (5.2), where adequate prosodic information is available to guide parsing. Therefore, Experiment 4 of this thesis tests subjects' parsing of delexicalised speech.

5.3. THE PERCEPTUAL STRENGTH OF PROSODIC BOUNDARIES

The last quarter century has produced a lot of research using delexicalisation methods. For instance, Lehiste and Wang (1976) tested subjects' perception of sentence or paragraph boundaries using delexicalised speech. Subjects were played both delexicalised and normal sentences and asked to mark the sentence and paragraph boundaries they perceived. Lehiste and Wang found that subjects predicted more boundaries in delexicalised speech than in normal speech. Lehiste and Wang argue that these boundaries were predicted using prosodic cues such as pause length, low f_0, and laryngealisation. Lehiste and Wang argue that fewer boundaries were predicted in normal speech because subjects often waited for additional syntactic and lexical information before predicting a boundary in normal speech. They argue that this provides evidence that lexical and syntactic cues override prosodic cues suggesting sentence closure. Lehiste and Wang conclude that purely prosodic information cannot guide accurate generation of syntactic structure.

Conversely, Kreiman (1982) argues that Lehiste and Wang's method of signal manipulation causes a loss of the signal's speech-like characteristics and can distract the subjects' attention with its tinny quality. Kreiman found improved boundary detection rate in the delexicalised versions compared to Lehiste and Wang's results. She argues that the positive results found in her experiment could be attributed to the enhanced quality of the modified stimulus. Kreiman also found no difference in boundary detection between the normal and modified stimulus. She contradicts Lehiste and Wang's argument that subjects were waiting for the next word before they decided whether to follow the interpretation consistent with the prosodic cues.

Kreiman argues that normal sentences provide subjects with adequate semantic and prosodic cues to clause closure *before* the boundary. Therefore, additional semantic information should only decrease boundary prediction times. Indeed, Kreiman found that subjects were faster at predicting boundaries in normal speech than to delexicalised speech. However, Kreiman found a mismatch between the boundaries perceived in normal and delexicalised speech. She argues that this was because of the imperfect alignment of prosodic and lexical information in normal speech. She argues that subjects would respond either to purely lexical information or to a mismatch between syntactic and prosodic information. This would cause either very fast or very slow response times to boundaries in the normal speech condition. However, she does not provide adequate information as to the number or type of response times that were high or low outliers from the mean. Kreiman's results were also supported by research by De Pijper and Sanderman (1994). They argue that listeners' ability to perceive prosodic boundaries in delexicalised speech was affected by differences in pause duration, reset and discontinuity. Similar results were reported by Swerts and Geluykens (1994) studying the effect of speech melody in marking discourse boundaries in delexicalised speech. These results provide evidence of the perceptual salience of prosodic cues in parsing delexicalised stimuli.

However, Schepman (1997) provides evidence against the possibility of subjects using purely prosodic information to guide parsing. As reported in Chapter 1, a production study found that speakers produced differences in duration and f_0 between alternative versions of ambiguous stimuli. Further, some on-line tests similar to those of Marslen-Wilson et al provided evidence of the early incorporation of prosodic information in parsing. However, Schepman argues that it was possible that the difference in subjects' response times could have been caused by their preference for one kind of prosody over another. She tests whether it was possible for subjects to respond faster after one kind of prosody than the other. Subjects were presented with delexicalised versions of ambiguous fragments in (5.3) and (5.4):

(5.3) The lawyer greeted the powerful barrister and the wise judge who was walking to the courtroom

(5.4) The lawyer greeted the powerful barrister and the wise judge who were walking to the courtroom

The sentences were followed by *was* and *were* probes. Schepman found that subjects' response times to *was* and *were* probes following the delexicalised stimuli were not significantly different. She concluded that this indicates that subjects did not arbitrarily prefer one melody to another. She argues that this provides evidence against the possibility that subjects responded faster to one type of prosody rather than to another. However, Schepman's results also provide evidence against initial processing using only prosodic information. Irrespective of whether subjects preferred one type of prosody to another, subjects did not seem to have made correct parses of the delexicalised stimuli. However, Schepman tests subjects' parsing at a point much after the presentation of adequate prosodic information to guide parsing. In a production study, Schepman found that only two speakers produced consistent prosodic cues to differentiate between the two interpretations at the point after *judge*. Conversely, speakers consistently differentiated between the two interpretations using pause duration after *barrister*. It is possible that the prosodic cues surrounding *barrister* more consistently cue the difference between the two interpretations and might make a more reliable testing point. For instance, a greater pause between *barrister* and *and the wise judge* would provide the first indication of the sentence complement interpretation of (5.3). An accent on *barrister* would also cue in the sentence complement interpretation. Therefore, it would be more appropriate to test parsing after *barrister* rather than after *judge*. The delexicalised versions of the two sentences seem too long to be assimilated during initial processing. Furthermore, parsing was tested four words past the point where sufficiently contrastive prosodic information was produced, presenting an unfair test of the effect of delexicalisation on parsing.

Typical of research in sentence processing, most of the studies contradict earlier conclusions for or against the possibility of solely prosodic processing. In addition, most of the research has focussed on the perception of sentence or discourse structure in speech and not on clause- or phrase-internal boundaries. The most convincing evidence against processing using solely prosodic information comes from Schepman's research. However, none of these studies adequately tested the possibility of initial analysis using solely prosodic information. The questions raised by the contradictions between Toepel and Alter (2002) and Steinhauer (2002) remain. Do prosodic cues retain their perceptual salience in the absence of lexical information? Furthermore, can listeners use these prosodic cues to guide initial parsing in the absence of lexical information? In the next section, I present evidence reported by Soderstrom et al (2003) arguing for the perceptual salience of purely prosodic information in infant speech processing. It remains to be seen, however, whether this ability persists in adulthood.

5.4. PROSODIC BOOTSTRAPPING: A DEVELOPMENTAL TREND?

In this section, I concentrate on the research of Soderstrom et al (2003) because it focuses on infants' disambiguation of noun and verb phrases using only prosodic information. In addition, it addresses the question of the importance of semantic content in the prosodic processing of speech signals. Soderstrom et al (2003) tested the ability of infants to use prosodic cues to differentiate between similar lexical items. Soderstrom et al (2003) conclude that infants as young as 6 months old are able to recognise phrase-internal syntactic units using prosodic information. Infants in one group were familiarised with stimulus (5.5). Infants in another group were familiarised with stimulus (5.6).

(5.5) new watches for men / gnu watches for men

(5.6) people by the hole / people buy the whole

All infants were then tested with passages (5.7) and (5.8). The prosody of the phrase *new watches for men* in (5.7) and *people by the hole* in (5.8) indicated that there were no major syntactic boundaries between the words. The prosody of the phrase *gnu watches for men* in (5.8) and *people buy the whole* in (5.7) indicated the presence of a syntactic boundary between the words:

(5.7) At the discount store, **new watches for men** are simple and stylish. In fact, some **people # buy the whole** supply of them

(5.8) In the field, the old frightened **gnu # watches for men** and women seeking trophies. Today, **people by the hole** seem scary.

If infants familiarised with (5.5) preferred listening to the passage in (5.7) over (5.8), this would indicate a preference for prosodically well-formed passages. A similar conclusion could be drawn if infants familiarised with (5.6) preferred listening to the passage in (5.8) over (5.7). They found that infants familiarised with the sentence in (5.5) preferred to listen to the passage in (5.7) rather than (5.8). The same effect was not found for the infants familiarised with (5.6). Infants familiarised with sequence in (5.5) were able to use prosodic information to differentiate between alternative productions of ambiguous items. The authors argue that the absence of a similar effect with the other group of infants might have been caused by the lack of adequate prosodic information marking the phrase boundary in (5.6) when compared to (5.5). Soderstrom et al also tested infants' perception of verb phrases such as (5.9) and (5.10):

(5.9) design telephones

(5.10) promise surprises

All infants preferred listening to passages that maintained the prosodic well-formedness of the stimuli. This provides evidence that infants were sensitive to prosodic boundaries within clauses. Infants were able to use prosodic cues to disambiguate the two versions. The authors call this process prosodic bootstrapping, referring to the way infants segment speech using prosodic information.

Importantly, infants were not familiar with the meaning of the words presented to them in the experiments. However, they were still able to use prosodic information to segment speech. The semantic content of the stimuli does not seem an essential requirement for the prosodic bootstrapping of speech. Toepel and Alter (2002) argued that adults would not be able to parse delexicalised speech, because they lack semantic content. However, the results presented by Soderstrom et al (2003) and other experiments suggest that semantic content is not an essential requirement for the prosodic bootstrapping of speech (Jusczyk et al: 1992). Admittedly, there are a number of obstacles to using the results of Soderstrom et al to argue against Toepel and Alter's conclusion. Firstly, although infants were not aware of the semantic content of the stimuli, they were still aware of the segmental content of the phrases. Secondly, Soderstom et al tested infant perception. It is possible that the ability to use purely prosodic information to guide parsing is a developmental trend. This ability could disappear once infants have acquired their language to a greater extent. However, the results indicate that prosodic information is used to structure speech in the absence of semantic content. It would be interesting to verify the continued presence of this ability in adulthood. There is little evidence of adults using purely prosodic information to guide parsing in the absence of lexical information. Consequently, Experiment 4 tests the processing of solely prosodic information to assess whether the abilities demonstrated by infants is a developmental trend that wanes through the acquisition of language. Research arguing that infants can use purely prosodic information to guide segmentation necessitates further research on adults. Before I present the details of Experiment 4, I test whether the early incorporation of lexical information takes place, as there is little need to test whether prosodic information guides parsing independently otherwise. I test this by studying the influence of lexical bias on parsing. I also test the autonomy of prosodic processing by analysing the interaction between lexical bias and prosodic information in parsing the stimuli in the earlier experiments of this thesis.

5.5. LEXICAL BIASES IN PARSING

Boland and Blodgett (2001) found that lexical bias could influence subjects' processing of ambiguous stimuli. They measured subjects' eye movements to test whether the noun and verb forms of homographs such as *duck* were more easily parsed when they were the dominantly occurring form. Subjects were presented with four different short stories into which target sentences such as (5.13) and (5.14) containing the ambiguous homographs were embedded. Target sentences were immediately preceded in the story by context sentences such as (5.11) or (5.12) that biased the reader towards one interpretation of the homograph. The context and target sentences were fully crossed so that noun or verb form biasing context sentences preceded both noun and verb form target sentences. The stories were taken from an earlier off-line study, which also reported an effect of lexical bias on parsing (Boland: 1997)

Context sentences:

(5.11) Noun context: As they walked around, Kate looked at all of Jimmy's pets.

(5.12) Verb context: As they walked around, Kate watched everything Jimmy did.

Target sentences:

(5.13) Noun target: She saw his duck and chickens near the barn.

(5.14) Verb target: She saw him duck and stumble near the barn.

They calculated the lexical bias of the ambiguous homographs using measures of the log of noun and verb frequency obtained from the Francis and Kucera norms on the Brown corpus. In order to obtain a measure of lexical bias, overall frequency was calculated as the log of noun frequency minus the log of verb frequency. Overall frequency measures suggested that the homographs had a mean noun usage of 51 per million and a mean verb usage of 18 per million. However, using a measure of contingent frequency, Boland (1997) argues that the homographs were more frequent as the verb form. Contingent frequency measures of lexical bias were calculated from data obtained in a sentence completion task, where native speakers of English were asked to complete sentences unbiased towards either interpretation to see whether they preferred the noun or verb forms.

Boland and Blodgett (2001) tested whether subjects' parsing of the ambiguous homographs was influenced by lexical bias using measures of contingent and overall frequency. They found that the amount of time subjects spent looking at the ambiguous string was influenced by the likelihood of occurrence of the form of the homograph. Subjects spent longer looking at the ambiguous homograph if the context sentences were biased towards the interpretation that conflicted with the lexical bias of the homograph. This was only true if lexical bias was calculated using overall frequency measures and not contingent frequency measures. Boland and Blodgett (2001) argue that overall frequency measures might be a better predictor of lexical bias effects than contingent frequency measures. The effect was found using initial fixation data. Consequently, they argue that lexical bias affects the initial processing of ambiguous stimuli.

However, the stimuli were visually presented to the subjects. Consequently, Boland and Blodgett did not test the interaction between prosody and the lexicon in parsing. There is a possibility that prosodic and lexical information may simultaneously influence parsing of the ambiguous word pairs. A parser following lexical bias would guide parsing towards the more likely interpretation of the stimuli. The first experiment of this thesis established that prosody guides parsing towards the intended parse of the stimuli. Therefore, in some cases, the parses confirmed by the lexical bias would conflict with the parses proposed by prosodically motivated parsing. Therefore, besides testing for an influence of lexical bias on parsing, I also test the interaction between prosodic and lexical information in parsing.

A relative frequency test indicated that the noun phrase interpretation was more likely than the verb phrase interpretation. Conversely, there was a preference for the verb phrase form of the prosodically non-contrastive stimuli. The measures of frequency of occurrence of the ambiguous forms of the word pairs were taken from the British National Corpus (http://www.natcorp.ox.ac.uk/). I calculated the log of the number of times the word pairs occurred as the noun or verb phrase forms in the corpus. Relative frequency is the difference between the log of the frequency of occurrence of the noun phrase form and the log of the frequency of occurrence of the verb phrase form. A positive value of relative frequency indicates that the noun phrase form is dominant and a negative value indicates that the verb phrase form is dominant.

Stimuli	Log NP frequency I	Log VP frequency II	Relative frequency (I – II)
Cooking apples	- 6.676024	- 7.977054	1.30103
Cutting boards	-7.278084	- 7.977054	0.69897
Packing cases	- 6.596843	- 7.977054	1.38021
Playing cards	- 6.323842	- 6.170874	- 0.15296
Racing cars	- 6.485693	- 7.499933	1.0142
Sailing ships	- 6.170874	- 7.977054	1.8061
Breaking glasses	- 7.977054	- 7.676024	- 0.30102
Burning trees	- 7.977054	- 7.676024	- 0.30102
Flying kites	- 7.977054	- 7.374994	- 0.60205
Melting glaciers	- 7.374994	- 7.977054	0.60205
Ringing bells	- 7.499933	- 7.374994	- 0.12493
Visiting relatives	- 7.374994	- 6.721782	- 0.65321

Table 5.1 Log of the frequency of occurrence of the word pairs as the noun or verb phrase forms.

The noun phrase form was the dominant interpretation of all of the prosodically contrastive word pairs, apart from *playing cards*. Conversely, the verb phrase form was the dominant interpretation of all of the prosodically non-contrastive word pairs, apart from *melting glaciers*. Quite a few of the tokens have the minimum log value (- 7.977054). This meant that there were no tokens of this version of the word pair in the sample set. Zero probabilities complicate analysis of the data set. Therefore, in order to assign a non-zero value to the zero-probability, we added one to all frequency counts. This maintains the relative frequency of the alternative versions of the word pairs.

I tested whether subjects' parsing of the stimuli was influenced by the likelihood of occurrence of the word pair as the noun or verb phrase version: I tested whether subjects preferred probes that confirmed the parses favoured by the lexical bias of the stimuli to probes that conflict with these parses. This investigated an effect of lexical information on parsing independent of an influence of prosodic information. Consequently, I considered subjects' response times following the prosodically contrastive and non-contrastive stimuli together.

I also test the interaction between prosodic and lexical information in initial parsing. The parses consistent with the prosody of the prosodically contrastive stimuli were contradictory to the parses favoured by the lexical bias of the stimuli in half of the presentations. Five of the prosodically contrastive word pairs were more likely to occur as noun phrase versions than as verb phrase versions. Therefore, the parse favoured by the lexical bias of these word pairs conflicted with the parses consistent with the prosody of every occurrence of the verb phrase versions of the stimuli. The other word pair was more likely to occur as the verb phrase form. Therefore, the parse favoured by the lexical bias of this word pair conflicted with the parse consistent with the prosody of the noun phrase versions. Conversely, in the other half of the presentations, the parses consistent with the prosody of the prosodically contrastive stimuli were the same as the parses favoured by the lexical bias of the stimuli. I test whether subjects' response times to probes were shorter when the lexical bias favoured the parses consistent with the prosody than when they conflicted with them.

5.5.1. Hypothesis

If subjects were affected by the lexical bias of the stimuli, they would take less time to respond to probes confirming the noun phrase version ('are') than to probes confirming the verb phrase version ('is') following items more likely to occur as the noun phrase forms. Conversely, they would take less time to respond to probes confirming the verb phrase versions ('is') than probes confirming the noun phrase version ('are') following items more likely to occur as the verb phrase forms. This tests for an effect of lexical bias independent of prosodic effects on processing. Therefore, I incorporated response times following all the stimuli and irrespective of whether the probes that followed confirmed the parses consistent with the prosody or not.

Assuming there is no interaction between lexical and prosodic information, the influence of lexical bias would not be affected by swapping the f_0, duration, and amplitude of the word pairs. Therefore, subjects would continue to respond faster to probes confirming the noun phrase versions ('are') than probes confirming the verb phrase versions ('is') following items biased towards the noun phrase version. Conversely, subjects would respond faster to probes confirming the verb phrase versions ('is') than probes confirming the noun phrase versions ('are') following items biased towards the verb phrase versions. Similar results are expected following the flattened stimuli. If there were no interaction between lexical and prosodic information, subjects should not be influenced by the flattening of the f_0 of the stimuli.

Next, I tested whether the effect of lexical bias is independent of the effect of prosodic information on parsing. I only consider response times following prosodically contrastive stimuli, since only they provide adequate prosodic information to guide parsing. Therefore, only response times to appropriate probes, i.e. probes that confirm the interpretation consistent with the prosody of the stimuli were considered. The parses consistent with the prosody of the stimuli contradicted with the parses favoured by the lexical bias of the stimuli in half of the presentations. In the other half of the presentations, the parses consistent with the prosody of the stimuli were the parses favoured by the lexical bias of the stimuli. If there were a significant interaction between prosody and lexical information, subjects' response times to appropriate probes would be shorter when the probes confirmed the parses favoured by the lexical bias of the stimuli than when they conflicted with these parses.

The cross-synthesised stimuli merely swap the prosodic content of the noun and verb phrase versions. Consequently, the number of stimuli whose prosody was consistent with the noun or verb phrase version remained the same in Experiments 1 and 2. Again, in half of the presentations, the prosodic and lexical bias of the stimuli conflicted with each other. In the other half of the presentations, the prosodic and lexical bias of the stimuli would support the same parse. I only consider response times to probes confirming the interpretation consistent with the cross-synthesised prosody of the stimuli, i.e. appropriate probes. Subjects' response times to appropriate probes would be shorter to probes confirming the parses favoured by the lexical bias of the stimuli than when they conflicted with these parses, if there were any interaction.

Flattening the f_0 of the stimuli might affect any interaction between prosody and lexical information during initial processing. The stimuli now provide less prosodic information to guide parsing. Subjects' ability to disambiguate the word pairs was attenuated in Experiment 3. Therefore, the conflict between parses favoured by the prosody and the lexical bias of the stimuli might not be as pronounced compared to when the stimuli provide adequate prosodic information. Again, I only consider response times to probes confirming the parses consistent with the prosody of the stimuli. Response times to appropriate probes confirming the parses favoured by the lexical bias would still be shorter than response times to probes conflicting with this parse. However, because of the reduced prosodic content, the difference between response times might be less than in Experiments 1 and 2. This would provide strong evidence of an interaction between prosodic and lexical information in on-line parsing.

5.5.2. Results

5.5.2.1. Experiment 1

The mean response times to stimuli biased towards the noun and verb phrase forms are given in Table 5.2. Non-parametric tests conducted confirmed that response times to 'are' were significantly shorter than response times to 'is' when the stimuli were biased towards the noun phrase form ($z = -2.483$; $p < .01$). A similar effect was found analysing subjects' response times to 'is' probes. I found that subjects' response times to 'is' were shorter when the items were biased towards the verb phrase versions ($z = -2.794$; $p < .01$). Again, the mean response times are given in Table 5.2.

Stimuli	Word pairs with NP bias	Word pairs with VP bias
Response times to 'are' (ms)	558	578
Response times to 'is' (ms)	565	547

Table 5.2 Experiment 1: Response times to the visual probes as a measure of lexical bias

Correlation tests indicated that subjects' response times to 'are' probes following the stimuli with a noun phrase lexical bias were significantly correlated with the strength of the lexical bias towards the noun phrase form ($r = -.050$; $p = .05$). The negative correlation indicates that subjects' response times to 'are' probes decreased as the strength of the lexical bias towards the noun phrase form increased. Again, response times to 'is' probes following the stimuli with a noun phrase lexical bias were significantly correlated with the strength of the lexical bias towards the noun phrase form ($r = .074$; $p < .05$). The positive correlation indicates that subjects' response times to the 'is' probes increased as the strength of the lexical bias towards the noun phrase forms increased.

I tested whether response times to appropriate probes following stimuli whose lexical and prosodic content conflicted with each other were significantly different following stimuli where they did not conflict with each other. The mean response time to appropriate probes was 551 ms when the probes confirmed the parses favoured by lexical bias. The mean response time to appropriate probes was 552 ms when the probes conflicted with the parses favoured by lexical bias. This difference was not significant ($z = -0.187$; $p > .1$).

5.5.2.2. Experiment Two

Non-parametric tests confirmed that subjects' response times to 'are' were significantly shorter when the stimuli were biased towards the noun phrase form ($z = -3.209$, $p < .001$). The mean response times are given in Table 5.3. A similar effect was found analysing subjects' response times to 'is' probes. I found that subjects' response times to 'is' were shorter when the items were biased towards the verb phrase versions ($z = -4.149$; $p < .001$). Again, the mean response times are given in Table 5.3.

Stimuli	Word pairs with NP bias	Word pairs with VP bias
Response times to 'are' (ms)	627	660
Response times to 'is' (ms)	606	578

Table 5.3 Experiment 2: Response times to the visual probes as a measure of lexical bias

There was a noticeable difference between response times to 'is' and 'are' in Experiments 1 and 2. Response times to 'are' were shorter than response times to 'is' following items with a noun phrase bias in Experiment 1. However, in Experiment 2, response times to 'is' were shorter than response times to 'are' following items with a noun phrase bias. In addition, correlation tests indicated that subjects' response times to 'are' probes following the stimuli with a noun phrase lexical bias were not significantly correlated with the strength of the lexical bias towards the noun phrase form ($r = -.021$; $p > .5$). Similarly, response times to 'is' probes following the stimuli with a noun phrase lexical bias were not significantly correlated with the strength of the lexical bias towards the noun phrase form ($r = - .030$; $p > .5$).

Response times to appropriate probes following stimuli whose lexical and prosodic content conflicted with each other were not significantly different following stimuli whose lexical and prosodic content did not conflict with each other ($z = -0.273$; $p > .1$). The mean response time to appropriate probes when the probes confirmed the parses favoured by lexical bias was 627 ms. The mean response time to appropriate probes when the probes conflicted with the parses favoured by lexical bias was 618 ms.

5.5.2.3. Experiment Three

I performed similar tests on the data from Experiment 3. I found that subjects' response times to 'are' were not significantly shorter when the stimuli were lexically biased towards the noun phrase form than when they were biased towards the verb phrase form ($z = -1.232$; $p > .1$). The mean response times are given in Table 5.4. Subjects' response times to 'is' probes were shorter when the stimuli were lexically biased towards the verb phrase form ($z = -.2.837$; $p = .005$). The mean response times are given in Table 5.4.

Stimuli	Word pairs with NP bias	Word pairs with VP bias
Response times to 'are' (ms)	754	769
Response times to 'is' (ms)	772	739

Table 5.4 Experiment 3: Response times to the visual probes as a measure of lexical bias

There was no significant correlation between the strength of the lexical bias towards the noun phrase and subjects' response times to 'is' probes following the stimuli with a noun phrase lexical bias ($r = .051$; $p > .1$). There was also no significant correlation between the strength of the lexical bias towards the noun phrase and subjects' response times to 'are' probes ($r = -.022$, $p > .1$).

I also tested for an interaction between prosody and lexical bias in on-line parsing. However, since there appeared to be less evidence of an influence of lexical bias, I did not anticipate finding an interaction between prosodic and lexical information. I tested whether response times to appropriate probes following stimuli whose lexical and prosodic content conflicted with each other were significantly different from response times following stimuli where they did not conflict with each other. The mean response time to appropriate probes was 748 ms when the probes confirmed the parses favoured by lexical bias. The mean response time to appropriate probes was 754 ms when the probes conflicted with the parses favoured by lexical bias. Although there was an 8 ms difference between response times, this difference was not significant ($z = -0.294$; $p > .1$).

5.5.4. Discussion

The statistical tests confirm that lexical information can influence on-line parsing. In Experiment 1, I found that subjects' response times to probes confirming the noun phrase interpretation ('are') were shorter when the words pairs were lexically biased towards the noun phrase version than when they were biased towards the verb phrase version. Conversely, subjects' response times to probes confirming the verb phrase version ('is') were shorter when the words pairs were biased towards the verb phrase version than when they were biased towards the noun phrase version. The results confirm that subjects' parsing of the word pair as either a noun or a verb phrase version was influenced by the likelihood of occurrence of this version in speech: if the noun phrase version was the more likely interpretation of the word pair, then subjects were more likely to parse the word pair as a noun phrase. Similarly, if the verb phrase version was the more likely interpretation, then subjects preferred verb phrase parses of the stimuli. Correlation tests confirmed this conclusion. Subjects' reaction times to 'are' probes decreased as the strength of lexical bias towards the noun phrase form increased. Conversely, subjects' response times to probes confirming the verb phrase version increased as the strength of lexical bias towards the noun phrase interpretation increased.

Earlier research described in this thesis also argued for the early incorporation of prosodic information in parsing. Therefore, there are two different influences working on the stimuli around the same time. I tested whether there was an interaction between these two influences. Does one take priority over the other so that the parse consistent with one of these components is de facto preferred to the other? On the other hand, do they work independent of each other in as much as the parse that is finally presented could just as easily be based on lexical or prosodic preferences due to the effectiveness of that particular influence for processing at the time? This could be stimulus, context, or listener dependent. It could also vary at different times with one influence being preferred to the other at the first presentation of the input, and the other having greater influence during repeated presentations. Obviously, any theory of the processing of speech would have to test for the interaction between lexical and prosodic information during on-line parsing.

There was minimal interaction between prosodic and lexical information during parsing. Subjects' response times to probes confirming the parses consistent with prosody were not delayed by the probes conflicting with the parses favoured by the lexical bias of the stimuli. Neither were response times to probes confirming the parses consistent with prosody speeded by the probes confirming the parses favoured by the lexical bias of the stimuli. There is little indication that that parsing is influenced by interaction between prosodic and lexical processing on-line. Therefore, the two influences seem to be simultaneously yet independently working on the stimuli. The results are more in keeping with suggestions that the final parse could just as easily be decided on lexical or prosodic information.

Swapping the f_0, duration, and amplitude of the stimuli did not affect the influence of lexical bias on parsing. Subjects' response times to probes confirming the noun phrase version ('are') were still shorter when the word pairs were more likely to occur as noun phrases than verb phrases. Conversely, subjects' response times to probes confirming the verb phrase version ('is') were shorter when the word pairs were more likely to occur as verb phrases than as noun phrases. The influence of lexical bias on parsing persists in Experiment 2. However, I did find that response times to 'are' were slower than response times to 'is' following items with a noun phrase bias, despite the fact that lexical bias should have decreased response times to probes confirming the parse consistent with lexical bias of the stimuli. This might be because subjects responded faster to 'is' probes overall in Experiment 2. The similarity in the prosody of the control stimuli and the verb phrase versions might have increased subjects' sensitivity to the verb phrase prosody, and consequently, decreased their reaction times to 'is' probes. There seem to be a number of influences working simultaneously on the input, prosodic, spectral, and lexical. However, subjects' response times to probes confirming the parse consistent with the prosody were not significantly different when these probes confirmed or conflicted with the parse favoured by lexical bias. There was no indication of interaction between prosodic and lexical information in parsing.

Furthermore, even assuming there was an interaction between prosodic and lexical information on parsing, prosodically flattening the stimuli might amplify the influence of lexical bias. However, the influence of lexical bias was not as strong in the parsing of the flattened stimuli as in Experiment 1. Subjects' response times to probes confirming the noun phrase version ('are') were not influenced by whether the word pairs were biased towards the noun or verb phrase version. Conversely, subjects' response times to probes confirming the verb phrase version ('is') were significantly shorter when the word pairs were biased towards the verb phrase version rather than the noun phrase version. In the last chapter, I reported a possibility that response times to appropriate and inappropriate probes following the flattened stimuli were significantly different. Subjects' continued ability to differentiate the stimuli might have been guided by spectral differences between the noun and verb phrase versions. These spectral differences centre on the cues provided by co-articulation between the word final fricative and the initial vowel of the probe word following the word pairs in the unspliced sentences. These segmental cues might be more pervasive than lexical cues and override the lexical bias effect in parsing. It is possible that the absence of prosodic cues amplifies the effect of segmental cues causing lexically guided parsing to take a secondary role. This does not mean that there is interaction between prosodic and lexical information. Contrarily, prosodically flattening the stimuli did not make subjects focus on the influence of lexical bias. The absence of strong effects of lexical bias on parsing the flattened stimuli only provides further evidence against the likelihood of interaction between prosodic and lexical information.

I found a significant difference between response times to 'is' probes following items with a noun and verb phrase bias in Experiment 3, despite not finding a similarly significant difference between response times to 'are' probes. This difference in response times might have been caused by an association between 'is' probes and items with a verb phrase bias. As we can see from Table 5.1, almost all of the items with a verb phrase bias were prosodically non-contrastive. I argue that prosodically flattening the prosodically non-contrastive stimuli would not remove any essential cues towards disambiguation, as prosodic cues were not used to disambiguate the stimuli in any case. Since prosodic flattening is not expected to have an effect on the parsing of these stimuli, it is possible that the segmental cues are not as effectively activated in the parsing of these prosodically non-contrastive stimuli, leading to a greater effect of lexical bias towards the verb phrase form. This might have led to a greater distinction between items with a noun and a verb phrase lexical bias. Again, there is little evidence in favour of this explanation. It is clear that the relationship between segmental cues and lexical bias appears to be more complicated than anticipated.

Flattening the stimuli did not increase the influence of lexical bias on parsing. This provides additional evidence that there is no interaction between prosodic and lexical influences in parsing. If lexical and prosodic processing were interactive, then reducing the prosodic cues of a stimulus might increase the influence of lexical bias on parsing. However, flattening the stimuli seemed to attenuate the influence of lexical bias on parsing as well. These results could also be used to argue greater interaction between prosodic and lexical information in parsing. Affecting the prosodic content of the stimulus seems to have an effect on the influence of lexical bias on parsing. Conversely, while it is possible that prosodic processing cannot take place in the absence of lexical information, should lexical processing be similarly affected by the absence of prosodic information? If there is an effect of lexical bias at all, the direction of bias cannot be decided by the prosodic content of the stimulus. Consequently, I maintain that manipulating the prosodic content of the stimuli does not influence the effect of lexical bias on parsing. There are other factors influencing the parsing of the flattened stimuli that attenuate the simultaneous influence of lexical bias. Within the limits of my research, there are suggestions that these factors centre on the segmental content of the stimuli. However, further research is required before conclusive arguments can be made on the parsing of prosodically flattened stimuli.

5.6. CONCLUSION

There is a definite effect of lexical bias on initial processing in Experiments 1 and 2. Lexical bias and prosodic information affect parsing independently of each other, with lexical bias leading parsing in some cases and prosody in other cases. The statistical tests reported in this chapter provide evidence of the early and simultaneous incorporation of lexical and prosodic information. They also support suggestions that lexical processing is independent of any influence of prosodic information. I test the consequences of these findings on serial and parallel models of processing in Chapter 7.

However, the contradictions between the studies reported in this chapter raise doubts as to the perceptual salience of prosodic information in the absence of lexical information. The hypothesis tested in Experiment 4 pushes the notion of prosodic bootstrapping to a further level. Gleitman and Wanner (1982) argue that infants could use the prosodic cues in speech to parse speech before they have acquired the required lexical vocabulary to guide parsing. Specifically, this proposal incorporates three main suggestions:

> First, that syntax is reliably correlated with acoustic properties; second, that infants are sensitive to these properties; and third, that infants use these acoustic cues in their processing of speech (Soderstrom et al: 2003: 250)

Research by Soderstrom et al provides strong evidence of infants using solely prosodic information to guide parsing. This does not provide evidence of a similar ability during adult parsing. It is possible that this is merely a developmental trend used to acquire the syntactic structure of a language, which wanes when the lexical and syntactic repertoire is completed.

Toepel and Alter (2002), for instance, argue that prosodic boundaries lose meaning in the absence of lexical information. However, Steinhauer (2003) presents contradictory evidence for purely prosodic processing. Similar contradictions abound in results of tests on subjects' ability to perceive prosodic boundaries in delexicalised speech. Research by Kreiman (1982), De Pijper and Sanderman (1994) and Swerts and Geluykens (1994) argues in favour of prosodic processing of delexicalised speech. Conversely, research by Lehiste (1976) and Schepman (1997) argues against attributing perceptive significance to solely prosodic information.

In addition, most of the prior research in the field has focussed on the perception of sentence, paragraph, and discourse boundaries. Research on the perception of phrase-internal structure is vital to discussions of models of speech processing. The experiments in this thesis were the first to test the parsing of phrase internal syntactic ambiguity on-line. The parsing of phrases has been completely ignored in the literature on ambiguity resolution. Phrase internal parsing represents a vital component of any parsing mechanism. It is the initial stage of processing in an incremental parsing situation. Alternatively, it is the final stage of processing in a top-down parser. Either way, parsing is incomplete without the resolution of phrase internal syntactic ambiguity.

Therefore, Experiment 4 tests the possibility of initial phrase-internal parsing using solely prosodic information. Experiment 1 provided evidence against the possibility of syntactic parsing preferences guiding parsing. Experiment 2 provided evidence against the possibility of spectral information guiding parsing. However, this chapter presents evidence for the early incorporation of lexical information on parsing. Consequently, it is possible that the early incorporation of prosody is dependent on the lexical content of the stimuli. . Delexicalised stimuli do not retain the spectral, segmental, syntactic, or lexical content of the original signal. Therefore, Experiment 4 tests whether subjects still show a preference for the intended interpretation of the ambiguous word pairs in the absence of any lexical, syntactic, and spectral information.

CHAPTER SIX
THE EARLY INCORPORATION OF LEXICAL INFORMATION IN PARSING: EXPERIMENT FOUR

6.1. INTRODUCTION

Toepel and Alter (2002) argue that prosodic information loses perceptual salience in the absence of lexical information – prosodic processing cannot take place when the lexical content of the stimuli is removed. This indicates that parsing using prosody either interacts with lexical processing or is subsequent to lexical processing. Conversely, research by Steinhauer (2003) argues that prosodic information retains perceptual salience even in the absence of lexical information. However, neither Steinhauer (2003) nor Toepel and Alter (2002) actually test *parsing* using solely prosodic information. While it is possible that prosodic processing is contingent upon lexical access, any conclusions regarding the perceptual salience of prosodic information in the absence of lexical information using their data are premature. Therefore, Experiment 4 of this thesis tests the parsing of delexicalised stimuli – i.e. stimuli devoid of the lexical content, retaining only the prosodic melody of the original signals. I test whether prosodic information can guide parsing in the absence of lexical information. If subjects still prefer the intended parses of the delexicalised stimuli, this would provide evidence of prosodic processing in the absence of lexical information. This would also resolve doubts about the perceptual salience of purely prosodic information.

In addition, the results of Soderstrom et al (2003) provide evidence that infants as young as six months old segment speech using prosodic information, despite not having access to the semantic content of the stimuli. It is possible that these abilities persist into adulthood – adult listeners might parse ambiguous stimuli based on solely prosodic information. Experiment 4 pushes the notion of prosodic bootstrapping of syntactic structure further. Can listeners' assign syntactic structure to the ambiguous stimuli using solely prosodic information, in the absence of lexical, syntactic, and spectral information?

The word pairs were the same as in the previous three experiments. However, the signals were delexicalised so that the listener was not provided with the lexical content of the stimulus. Delexicalised speech sounds like someone speaking in the next room. The listener is provided with the melody of the original signal, but nothing of the semantic component. Most of the prior research testing the perceptual salience of purely prosodic information tests the parsing of sentence or paragraph boundaries. Although a number of delexicalisation methods were proposed in the last quarter of a century, none of them tested phrase-internal disambiguation using solely prosodic information. Testing shorter stimuli provides a fair test of the ability of solely prosodic information to guide parsing. This ensures that Experiment 4 tests parsing at the point of recognition of ambiguity and at the point of presentation of adequate prosodic information to guide parsing.

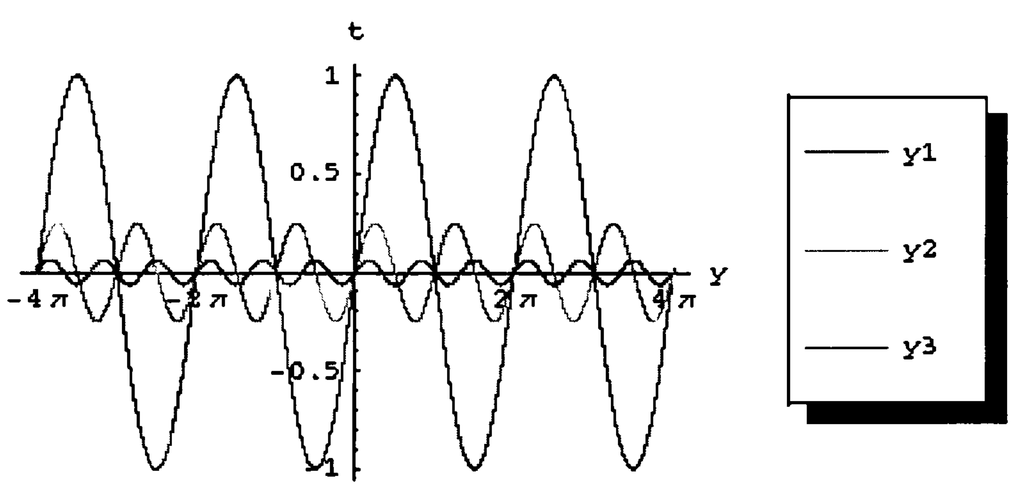
In the next section, I present the delexicalisation method used and the reasons for choosing this method. I then present the details of Experiment 4 and the results of the experiment. Finally, I present the conclusions of Experiment 4– prosodic information can guide parsing in the absence of lexical, segmental, and syntactic information. This provides evidence, contrary to Fodor (1983), against mandatory initial syntactic processing.

6.2. DELEXICALISATION: THE PURR PROGRAM

One of the earliest and simplest methods of delexicalisation utilises the common signal manipulation technique of low-pass filtering. In these cases, the filter is automatically set to zero at the voiceless segments. This ensures that all the voiceless segments are turned into pauses. The other (voiced) segments are all filtered at a cut off frequency just above f_0. The signal produced provides information about the f_0 and the temporal structure of the original utterance.

Lehiste and Wang (1976) propose a similar technique where the sign of every alternate sample of the signal is reversed. This maintains the prosodic characteristics of the signal, while removing the segmental content. Kreiman (1982) suggests an alternative method using a combination of low pass filtering and high pass filtering of a spectrally inverted input. There are a number of variations of this twin filter, but Kreimann (1982) suggests that the values set on her filters provide the best output – "a cross between a conversation heard through a thick wall and one in a foreign language" (1982: 166). The signal to be delexicalised is high pass filtered at 600 Hz and then spectrally inverted at a sampling frequency of 8000 Hz. The output from the inversion routine is low pass filtered at 2000 Hz and then combined with a low pass filtered version of the original signal, filtered at 200 Hz. Most of the experiments I reviewed in Chapter 5 used variations of the Kreiman procedure, sometimes changing the values of the filters. However, the ERP studies (Toepel and Alter: 2002; Steinhauer: 2003) used variations of the delexicalisation procedure suggested in Sonntag and Portele (1998) – the PURR program. Experiment 4 of this thesis uses a combination of the PRAAT delexicalisation program and a variation of the PURR program, provided by Wagner (2004).

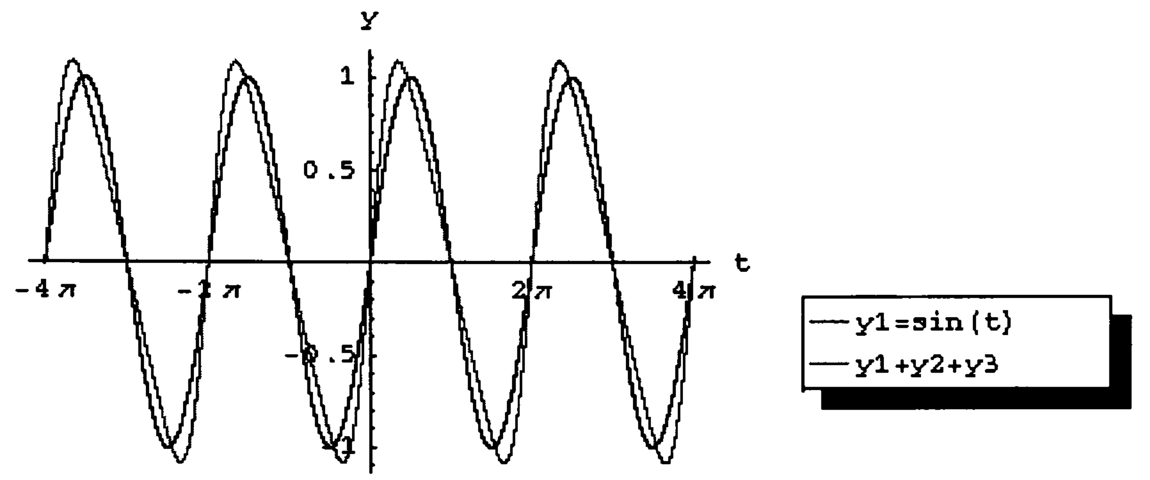
PURR re-synthesises the recorded data in order to convey only information of intensity, f_0, and temporal structure. The program first determines the pitchmarks of the stimulus to be delexicalised. The cut-off frequency for the aperiodic segments is set to zero, while the rest of the segments are low pass filtered with a frequency below the third harmonic. The pitchmarks are then replaced by three superimposed sine-waves. The first of these is a wave of the same amplitude and frequency of the stimulus. The second is a wave with double the frequency and $1/4^{th}$ the amplitude of the original signal. The third is a wave with triple the frequency and $1/16^{th}$ the amplitude of the stimulus.



where y1 is the original utterance, y2 is the sinus with double the frequency and a fourth of the amplitude and y3 is the sinus with triple the frequency and one sixteenth of the amplitude of the original.

Figure 6.1.Superposition of the three different waveforms in the delexicalisation procedure

The resultant signal is not a sine wave, though it is periodic with the original utterance frequency and resembles a sawtooth wave. Since we have added three different waves which all go through a number of cycles in the time given by the frequency of the original utterance, the pattern repeats with the original frequency. As we can see from the resultant signal (Figure 6.2), the wave is slightly skewed from the original wave form with an amplitude and intensity very slightly different from the original.



where y1 is the original signal and the resultant y1+y2+y3 is the delexicalised signal in the form of a sawtooth wave slightly different from the first.

Fig 6.2 Resultant delexicalised stimulus using the PURR program

In the signal thus obtained, aperiodic voiceless segments are turned into pauses. The original rhythmic structure of the signal is maintained by a string of voiced segments and pauses. This sequence also represents the temporal structure of the utterance and sustains the appearance of speech. The stimulus obtained in this manner sounds muffled, almost as if someone were speaking with their hand over their mouth.

Sonntag and Portele (1998) conducted a number of tests to determine the perceptual salience of PURR delexicalised stimuli. These tests focus on the perception of intra-phrasal boundaries and accents and are in keeping with the aims of Experiment 4. The PURR program was tested alongside five other possible delexicalisation strategies. The five programs included Kreiman's technique and simple low pass filtering. The first of the other three procedures involved filling the pitchmarks with an excitation signal proposed in CCITT (1989). Alternatively, the pitchmarks could be replaced with a glottal flow model as recommended by Fant, Lijlencrants and Lin (1985). The final alternative involved replacing the pitchmarks with a simple sawtooth signal.

Sonntag and Portele delexicalised a number of sentences using the six different methods described above. Each of the delexicalised stimuli was played to subjects in five different tests of perceptual salience. In the first test, subjects were asked to count the number of syllables they perceived in what was played to them. In the second, subjects could view a cursor moving down an oscillogram that marked the boundaries of the different syllables while simultaneously listening to the stimulus. Subjects were asked to pick out the most prominent of the syllables marked by the cursor. The third experiment consisted of a phrase-boundary detection test. Subjects were asked to mark the two prosodic boundaries in the sentence on a visually presented oscillogram annotated with all the possible boundaries between syllables. The pauses between potential boundaries were manually removed. The fourth tested subjects' ability to accurately decide the modality of the delexicalised stimuli presented to them. The three choices of modality were interrogative, continual pitch movement or terminal. The fifth test was repeated after each of the four tests above. Subjects were asked to mark the stimulus they found the easiest and the most pleasant to work with. Subjects' results on the different tests on stimuli prepared by the different methods are presented in the table below. Results are presented in percentage of correct answers provided by the subjects.

	CCITT	Low pass filtering	Kreiman spectral inversion	Fant, Liljencrants and Lin	Saw tooth signal	PURR
Syllable recognition	58%	50%	58%	54%	69%	59%
Phrase accent	77%	70%	72%	80%	72%	68%
Phrase boundary	74%	69%	58%	74%	71%	79%
Phrase modality	84%	85%	82%	81%	85%	85%

Table 6.1 Percentage of correct answers to the four tests using the differently delexicalised speech stimuli

Sonntag and Portele decided to eliminate the spectral inversion method based on its poor performance on the phrase-boundary detection test. All the other techniques had similar results on all the tests. Subjects indicated that PURR stimuli were the easiest to listen to. Therefore, I decided to use the PURR stimulus based on PURR's consistently good performance in the five other tests of signal quality. Almost all of Sonntag and Portele's subjects showed an above chance ability to pick out the syllable and phrase boundaries in the stimuli presented to them. Despite the lack of lexical content, subjects could still extract prosodic and syntactic structure from the stimuli. Experiment 4 tests this conclusion using an on-line perceptual test.

The stimuli in Experiment 4 were delexicalised using a revised version of the PURR delexicalisation program proposed by Wagner (2004). It follows the old PURR approach; while at the same time compensating for what Wagner considers PURR's instability towards pitch calculation errors. Wagner provides a compromise between the old PURR program and PRAAT delexicalisation procedures. Wagner argues that the PRAAT delexicalisation procedure leaves out the amplitude contour of the original signal, while superimposing a quasi-vowel quality. Her re-working of the PURR-PRAAT script chooses to not introduce imitated vowels, by filtering a formant near the second harmonic. The intensity and pitch contour of the signal to be delexicalised are first extracted. The mean pitch is calculated. Information of the resulting pitch contour is used to posit glottal pulses where the original sound contains energy, while retaining the formants of the original signals. The time-stamped intensity contour of the signal is then multiplied by this glottal pulse chain. This is equivalent to the original sound being multiplied by a linear interpolation of the intensity tier. Finally, a formant is introduced around the second harmonic. This value is determined by calculating the value of double pitch – which is twice the value of the mean pitch calculated above. As with the PURR stimuli, the amplitude and frequency contour of the original signals were retained, while the segmental content of the stimuli was completely distorted. The word pairs in Experiment 1 were all delexicalised using the newly re-worked PURR/PRAAT script.

6.3. PROCEDURE

Each subject in Experiment 4 now heard twenty repeats of each of the noun and verb phrase forms of all the stimuli – a total of 480 word pair fragments. This was to adequately test subjects' perception of the delexicalised stimuli. This provided a stricter test of subjects' parsing than Experiments 1, 2, and 3. As in Experiment 1, the 480 stimuli were presented to subjects in a random order. There were two random orders of presentations of the stimuli in order to control for any possible effects of order of presentation. In half of the presentations, the delexicalised stimuli were followed by probes consistent with the prosody of the stimuli – noun phrases followed by 'are' and verb phrases followed by 'is'. I refer to these probes as 'appropriate probes'. In the other half of the presentations, the visual probe that followed the stimuli was not consistent with the prosody of the stimuli – noun phrases followed by 'is' and verb phrases followed by 'are'. I refer to these probes as 'inappropriate probes'. Subjects were seated in a sound-proof booth and given the same instructions as in Experiments 1 and 2. They were also informed that the stimuli had been manipulated so that they sounded muffled. They were told that the stimuli would sound like hearing someone speaking in the next room with the door closed. I decided to exclude the appropriateness-rating task from the experiment. I felt it was inappropriate to ask subjects whether they thought the word was a good or bad continuation of the stimuli in the absence of lexical information. This might have indicated the purpose of the experiment to the subjects, since prosodic information was all that remained for subjects to base their evaluations on. This also made the experimental run shorter, allowing the intended increase in the number of repetitions subjects heard of each of the word pairs.

Subjects were paid to participate in the experiment. All subjects were right handed native speakers of Southern British English. All subjects were aged between 22 to 29 years. Six subjects were male and six were female. Subjects were randomly assigned to one of two groups. The groups differed in the order of presentation of the stimuli and the buttons that corresponded to 'is' and 'are' probes. This was in order to control for any effects of the order of presentation of the stimuli. In addition, Group A had a red 'is' button to the left and a black 'are' button to the right. Group B had a red 'are' button to the left and a black 'is' button to the right. Subjects were put through a practice run to familiarise themselves with the task and to check whether they had understood the task properly. The experiment lasted a little over 35 minutes per subject. The time taken by subjects to respond to the probes was noted and analysed by an in-house response time program.

6.4. HYPOTHESIS

The stimuli presented to subjects in Experiment 4 retained none of the lexical content of the original signals. If prosodic information cannot guide parsing in the absence of lexical information, subjects would not be able to disambiguate the noun and verb phrase readings of the delexicalised stimuli. Therefore, their response times to either of the probes should take around the same time, irrespective of whether the probe was consistent with the prosody of the stimuli (appropriate) or not (inappropriate). The probe 'are' is an appropriate continuation of the noun phrase forms of the original word pairs, and the probe 'is' is an appropriate continuation of the verb phrase forms of the original word pairs. Despite this, subjects' mean response times to 'are' following the noun and verb phrase versions would be similar. Similarly, subjects' mean response times to 'is' following the verb and noun phrase versions would be similar.

Subjects would not be able to differentiate between the noun and verb phrase forms of the prosodically non-contrastive stimuli either. The delexicalised versions of the prosodically non-contrastive stimuli do not present adequate syntactic, prosodic, segmental, or lexical information to guide parsing. Therefore, any lexical biases that might have influenced subjects' parsing in the earlier experiments in this thesis would have been removed by the delexicalisation procedure. Therefore, subjects' response times to probes consistent with the prosody of the stimuli (appropriate probes) would be similar to response times to probes conflicting with the prosody (inappropriate probes).

However, Steinhauer (2003) reports finding a closure positive shift even when subjects were played delexicalised stimuli. This presents evidence against the argument that prosodic information loses meaning in the absence of lexical information. In addition, the perceptual studies reviewed in Sections 5.2 and 5.3 suggest that subjects can detect sentence, paragraph, and discourse boundaries in delexicalised speech. Therefore, subjects might be able to differentiate between the noun and verb phrase readings of the delexicalised prosodically contrastive stimuli presented to them. Subjects' response times to probes consistent with the prosody of the stimuli (appropriate) would be faster than response times to probes inconsistent with the prosody (inappropriate). Therefore, subjects' response times to 'are' following the delexicalised forms of the prosodically contrastive noun phrase stimuli would be faster than subjects' response times to 'is'. Conversely, subjects' response times to 'is' following the delexicalised forms of the prosodically contrastive verb phrases would be faster than subjects' response times to 'are'. This would indicate that prosody was guiding subjects towards the intended interpretations of the stimuli even in the absence of lexical information.

There are inadequate and inconsistent prosodic differences between the alternative versions of the prosodically non-contrastive stimuli. Therefore, subjects would be unable to disambiguate these stimuli even if the prosodic content were perceptually salient in the absence of lexical information. Therefore, subjects' response times to probes consistent with the prosody of the stimuli (appropriate) would be similar to subjects' response times to probes conflicting with the prosody (inappropriate). This would indicate that subjects could not disambiguate the stimuli. In addition, this would provide evidence of prosodic information guiding parsing of the prosodically contrastive word pairs. The only difference between the experimental and control stimuli was that there were consistent prosodic differences between the noun and verb phrase readings of the former, while the alternative control stimuli were prosodically similar. Therefore, subjects' inability to disambiguate the control stimuli would provide evidence that prosodic information was guiding parsing of the experimental stimuli.

6.5. RESULTS

17 out of 5760 response times were discounted because of incorrect button pressing or response times greater than 1500 milliseconds. The mean response times to appropriate and inappropriate probes following the prosodically contrastive and non-contrastive stimuli is presented in the table below and plotted in Figure 6.3.

Stimuli	Mean response time to appropriate probes (ms)	Mean response time to inappropriate probes (ms)
With prosodic contrast	435	443
No prosodic contrast	438	441

Table 6.2: Mean reaction times to appropriate and inappropriate probes following prosodically contrastive vs. non-contrastive stimuli (ms)

The response times were then analysed to see if the distribution was normal. As is usual in a database of this size, the response times were not normally distributed according to a Kolmogorov-Smirnov test for normality ($p < .05$). Hence, any t-tests or ANOVAs conducted would have to be invalid.

Figure 6.3 plots the means of response times to appropriate and inappropriate probes following the prosodically contrastive and the prosodically non-contrastive stimuli. As is obvious from the graph, response times to appropriate probes were significantly faster than response times to inappropriate probes following stimuli with prosodic contrast. Conversely, there was little difference between response times to appropriate and inappropriate probes following stimuli with no prosodic contrast between the noun and verb phrase reading. This graph is similar to the graphs in Experiments 1 and 2. However, it must be remembered that the scales of the graphs are considerably different. Therefore, interpretation of these graphs is considered in conjunction with the results of the statistical tests reported.

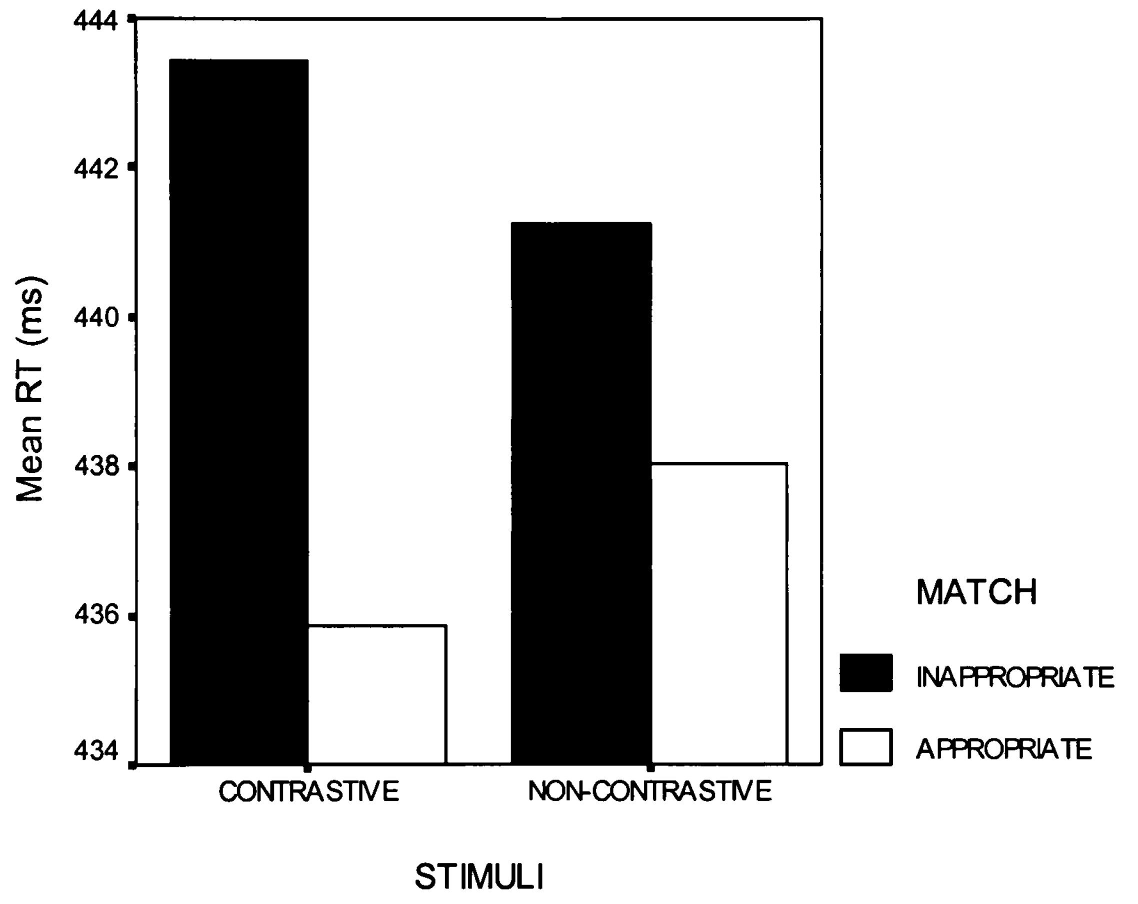


Figure 6.3 Mean response times to appropriate and inappropriate probes following stimuli with vs. without prosodic contrast

A Mann-Whitney test showed that the mean response times to appropriate and inappropriate probes were significantly different ($z = -2.003$; $p = .05$), although the difference between means is not large (Table 6.2). The mean response time to appropriate probes was significantly lower than the mean response time to inappropriate probes.

As we can see from Table 6.2, subjects' mean response times to appropriate and inappropriate probes following the prosodically non-contrastive stimuli were quite similar. Again, the response times were not normally distributed. A Kolmogorov-Smirnov test for normality was significant ($p < .05$). A Mann-Whitney test revealed that the mean response times to appropriate and inappropriate probes following the prosodically non-contrastive stimuli were not significantly different ($z = -1.130$, $p > .1$).

The results of within-subject analyses are reported in Table 6.4. This was to find out whether the pattern of response times obtained cross subject was maintained within individual subjects' response times. Table 6.3 presents the results of an item analysis to check whether the pattern of results discussed above was maintained within item.

Item	Mean response times Appropriate probes (ms)	Mean response times Inappropriate probes (ms)	Results
Cooking apples	441	429	p > .1
Cutting boards	432	448	p > .1
Packing cases	439	447	p > .1
Playing cards	434	445	p > .1
Racing cars	431	444	p > .1
Sailing ships	427	428	p > .1

Table 6.3 Mean response times for appropriate and inappropriate probes to prosodically contrastive stimuli per item

Subject	Mean response times Appropriate probes (ms)	Mean response times Inappropriate probes (ms)	Statistical significance
1	499	515	$p > .05$
2	411	441	**$p < .05$**
3	426	408	$p > .05$
4	405	409	$p > .05$
5	442	486	**$p < .05$**
6	462	460	$p > .05$
7	393	399	$p > .05$
8	585	548	**$p < .05$**
9	370	406	**$p < .01$**
10	444	437	$p > .05$
11	408	437	**$p < .05$**
12	378	369	$p > .05$

Table 6.4 Mean response times for appropriate and inappropriate probes to prosodically contrastive stimuli per subject

Seven subjects showed faster responses to appropriate probes than to inappropriate probes following the prosodically contrastive stimuli. However, only four of these differences in response times were significant. As argued in earlier chapters, it is quite possible that there was not enough response time data per subject to find true significance. It is quite encouraging that despite considering only 196 responses per subject, seven subjects showed faster responses to appropriate probes than to inappropriate probes following prosodically contrastive stimuli. Again, despite using only 196 responses per subject, four subjects showed significantly faster responses to appropriate probes than to inappropriate probes following prosodically contrastive stimuli.

The item analysis was also conducted using only subjects' response times following the prosodically contrastive stimuli. Subjects showed lower response times to appropriate probes than to inappropriate probes for almost all items. However, none of these differences was significant. Again, since within-item analysis only considers 196 response times per item, it is possible that this is not enough to show true significance of difference.

I also performed a few tests to compare with the results of the earlier experiments of this thesis. Despite the absence of syntactic information, I tested whether subjects arbitrarily preferred one of the probes as an appropriate continuation of the auditory stimuli. A Mann-Whitney test between subjects' response times to 'is' and 'are' probes following all the stimuli indicated that subjects continued to prefer 'is' probes to 'are' probes (Mean RT ('is') = 431 ms; Mean RT ('are') = 447ms; $z = -5.430$; $p < .001$). This preference for 'is' probes was maintained in testing response times following only the prosodically contrastive (Mean RT ('is') = 431 ms; Mean RT ('are') = 447 ms; $z = -3.865$; $p < .001$) and only the prosodically non-contrastive stimuli (Mean RT ('is') = 432 ms; Mean RT ('are') = 447 ms; $z = -3.809$; $p < .001$).

I also tested to see whether there was an influence of lexical bias on parsing the stimuli. This was to verify that the pattern of results reported in Chapter 5 was caused by an influence of lexical bias. The results of the Mann-Whitney test confirmed that subjects' response times to 'are' were not significantly faster when the stimuli were biased towards the noun phrase form than when the stimuli were biased towards the verb phrase form ($z = -.139$, $p >. 8$). Similarly, subjects' response times to 'is' were not significantly faster when the items were biased towards the verb phrase versions than when the items were biased towards the noun phrase version ($z = -1.559$; $p > .1$).

I also performed correlation tests between subjects' response times and the different prosodic characteristics that were consistently different between the noun and verb phrase versions of the stimuli. I only considered subjects' response times to appropriate probes following the prosodically contrastive stimuli for these correlations. If subjects were influenced by the prosodic characteristics of the stimuli, then it is reasonable to suggest that their response times might be faster or slower; depending on variation in the strength of the acoustic characteristic. The correlations that were significant are listed in Table 6.5.

	Peak RMS amplitude of word 1	Peak RMS amplitude of word 2	Difference in amplitude between word 1 and word 2	Difference in peak f_0 between word 1 and word 2
Correlation significance	z = .052; p < .05	z = - .053, p < .05	z = .057; p < .05	z = .053, p < .05

Table 6.5 Significant correlations between response times to appropriate probes following the prosodically contrastive stimuli and contrastive prosodic characteristics.

6.6. DISCUSSION

The mean of subjects' response times to appropriate probes following the prosodically contrastive stimuli was less than the mean of subjects' response times to inappropriate probes. There was a 95% probability that the difference between mean response times to appropriate and inappropriate probes was not due to chance. This indicates that subjects were able to make accurate parses of the delexicalised prosodically contrastive stimuli. The only difference that remains between the alternative readings of the word pairs is the prosodic contrast between the two forms. Therefore, subjects' ability to disambiguate the stimuli was based on these prosodic differences. In other words, the prosodic contrast between the ambiguous alternatives could guide initial parsing independent of segmental, lexical, and syntactic information. This provides evidence against the arguments proposed in Toepel and Alter (2002), Meyer et al (2002), Schepman (1997) and Lehiste and Wang (1976). If prosodic information can guide parsing in the absence of lexical information, removing the semantic content of a message does not remove the salience of prosodic information.

This conclusion is supported by the results of analysis of the prosodically non-contrastive stimuli in this experiment. The means of response times to appropriate and inappropriate probes following the prosodically non-contrastive stimuli were similar. Subjects' response times to probes consistent with the interpretation intended of the original stimuli were not faster or slower than response times to probes conflicting with this interpretation. Non-parametric tests suggested that there was a 43% chance that the difference between mean response times to appropriate and inappropriate probes was due to chance. This was because of the absence of consistent prosodic differences between the noun and verb phrase readings. The only difference between the prosodically contrastive and non-contrastive stimuli was the presence of consistent prosodic differences between the noun and verb phrase readings of the former. Subjects' inability to make accurate parses of the prosodically non-contrastive stimuli provides evidence that subjects' parsing of the prosodically contrastive stimuli was guided by prosodic information *even in the absence of lexical information.*

I also conducted within-subject and within-item analyses to find out whether this pattern of response times following the prosodically contrastive stimuli was maintained within individual subjects' response times and within items. As argued in the previous section, it is possible that there was not enough response time data per subject for there to be significant differences within each subjects' response times. The item analysis was also conducted using only subjects' response times following the prosodically contrastive stimuli. Subjects showed faster response times to appropriate probes than to inappropriate probes within almost all items. However, none of these differences was significant. Irrespective of the lack of significance of some of the within-subject and within-item analysis, the overall results of Experiment 4 provide strong evidence that prosodic information can guide parsing even in the absence of lexical information.

Subjects' could make accurate parses of the prosodically contrastive stimuli using only prosodic information, in the absence of lexical information. However, subjects' preferred verb phrase parses of the stimuli as in Experiments 1, 2, and 3. As in the case of the last three experiments, this preference for verb phrase parses of the stimuli might have been motivated by the prosodic contour of the prosodically non-contrastive stimuli. The prosody of these control stimuli was similar to the prosody of the verb phrase parses. Therefore, subjects were exposed to a greater number of stimuli with the prosody of the verb phrase reading. This might increase their sensitivity to the prosody of the verb phrase versions, thereby decreasing their reaction times to probes consistent with the verb phrase interpretation of the stimuli. This provides a reasonable explanation for their shorter response times to 'is' than to 'are' probes in all four experiments. Carston (1989) argues that the noun phrase parse of the stimuli is syntactically simpler. Consequently, if subjects were following syntactic parsing principles such as late closure or minimal attachment, they would prefer noun phrase parses of the stimuli. This preference for verb phrase parses cannot be based on segmental information either, since Experiment 4 does not provide subjects with any segmental information. The influence of lexical bias would not consistently lead subjects towards the verb phrase parses of the stimuli. In Experiment 4, there was no possibility of this preference for verb phrase parses being motivated by anything other than prosodic information, since subjects were presented with delexicalised stimuli.

I also tested to see whether there was an influence of lexical bias on parsing. I had used the results of similar tests to conclude in favour of an influence of lexical bias on parsing in Experiments 1 and 2. Therefore, the absence of a similar preference for the parses favoured by lexical bias when listening to delexicalised stimuli would confirm that subjects' parsing was influenced by lexical bias in Experiments 1 and 2. Subjects no longer preferred parses favoured by the lexical bias of the stimuli. This confirms the influence of lexical bias on parsing in Experiments 1 and 2, an influence that is possibly simultaneous with the incorporation of prosodic information in parsing.

Finally, I discuss the results of the correlation tests conducted between subjects' response times and the acoustic characteristics that contributed to the prosodic differences between noun and verb phrase readings of the prosodically contrastive stimuli. The correlations were conducted using response times to appropriate probes following the prosodically contrastive stimuli. The hypothesis for the correlation tests was that variation in the prosodic characteristics of the stimuli might increase or decrease subjects' response times to probes consistent with the interpretation intended by the prosody. As is clear from Table 6.5, there was a strong correlation between the peak RMS amplitude of the first and the peak of the second word and subjects' response times. There was a similarly strong correlation between subjects' response times and the difference in amplitude between the two words. Similarly, subjects' response times were also correlated with the difference in fundamental frequency between the two words. Amplitude and f_0 seem to play a major role in influencing subjects' parsing. This was already indicated by Experiment 2. There were consistent differences in the amplitude and f_0 between the alternative versions of the prosodically contrastive stimuli. The results of the correlations do not prove that prosodic information can guide parsing, but they provide additional evidence in support of the conclusions of Experiment 4.

6.7. CONCLUSION

In the conclusion of Chapter 5, I argue that there were still a number of unanswered questions regarding solely prosodic processing of speech. Firstly, Toepel and Alter (2002) argue that prosodic information lacks perceptual salience in the absence of lexical information. Similar results were reported by Meyer et al (2002), Schpeman (1997) and Lehiste (1982). However, experiments by Steinhauer (2003) report findings contradictory to Toepel and Alter (2002). He found the expected closure positive shift even when subjects were listening to delexicalised speech.

However, there are reasons for the contradictions in results between Experiment 4 and other experiments reviewed in the previous chapter. Firstly, neither Steinhauer nor Toepel and Alter tested subjects' parsing, while Experiment 4 of this thesis focussed on testing subjects' *parsing* of delexicalised stimuli. This difference in focus might be the reason for the difference in results. The presence of the closure positive shift might suggest prosodic processing, but there is no evidence that the absence of the closure positive shift is irrevocably linked to the absence of prosodic processing. Schepman's (1997) experiment did test on-line parsing of delexicalised stimuli. However, her experiments did not test subjects' parsing at the point of presentation of adequately contrastive prosodic information. Therefore, irrespective of the contradictions between the results reported here and by previous research, there is strong evidence of the possibility of prosodic information guiding parsing in the absence of lexical information. In other words, prosodic information does not lose salience in the absence of lexical information.

Secondly, there were consistent prosodic differences between the noun and verb phrase readings of the stimuli presented to subjects in the experiments in this thesis to guide parsing. It is possible that Toepel and Alter's stimuli did not present subjects with adequate prosodic information to distinguish between the two readings. This would explain the contradictions in the results reported by Toepel and Alter (2002) and Steinhauer (2003). Consequently, there is a caveat to the conclusions presented here that shapes the discussion of models of speech processing in Chapter 7. Experiment 4 provides evidence that prosodic information can independently guide resolution of ambiguous stimuli in the presence of sufficiently contrastive prosodic information. There is always the possibility that the prosodic content of daily speech may not be adequate to guide assignment of syntactic structure. However, it has to be remembered that the subjects in the experiments in this thesis were not trained to perceive the prosodic contrast between the alternative versions of the word pairs. In other words, they were not told that greater stress on the first word should be used to cue in the noun phrase reading. Any preferences they indicated were guided by their prior experience with the melody of their native language. This indicates that subjects had been exposed to similarly contrastive stimuli in the past. At the very least, they were aware that it was natural for the first word to be stressed in the noun phrase versions and the second to be stressed in the verb phrase version. This suggests that these prosodic contrasts do exist in everyday speech. They can be used in the service of the comprehension of everyday speech in an immediate and on-line manner, even in the absence of adequate syntactic, segmental, and lexical information.

The results of Experiment 4 also support suggestions that the results reported by Soderstrom et al (2003) were not evidence of a developmental trend. Infants' ability to use prosodic information to segment speech is not limited to the period before the acquisition of the syntactic structure of their language. Adults display a strong preference for parses intended by the prosody of the stimuli even in the absence of lexical information. Since the infants were also presented with segmental information, this ability of adults to use *only* prosodic information to guide parsing pushes the notion of prosodic bootstrapping of syntactic structure further than most prior studies. However, it is a definite possibility that infants' ability to segment speech using prosodic cues draws upon different resources than adults' ability to parse delexicalised speech.

Adults' parsing using delexicalised stimuli was a within-phrase parsing preference. Experiment 1 established that subjects could use prosodic cues in the initial processing of phrase-internal ambiguity. Experiment 4 takes a step further in testing the parsing of delexicalised stimuli. Prosodic information can independently guide the parsing of phrase-internal ambiguity and is not restricted to disambiguating between-sentence and intermediate phrase boundaries. Conversely, testing the parsing of within-phrase ambiguity was necessary for the results reported by Experiment 4. I test parsing immediately after the presentation of adequately contrastive information. The stimuli I test were not too long and did not add an unnecessary load on processing. Experiment 4 presents a reasonable test of the ability of prosodic information to guide parsing.

The results of Experiment 4 establish that prosodic information can guide initial parsing independent of lexical, syntactic, and spectral information. Disambiguation using purely prosodic information can and does take place. Removing the lexical content of the stimuli did not affect subjects' ability to differentiate between alternative forms of the ambiguous stimuli. Consequently, this provides evidence that prosodic initial processing is not affected by the absence of lexical bias. These results are contrary to what would be expected by both serial and parallel models of processing. Serial processing models would predict that the absence of syntactic information would prevent disambiguation or further parsing. Parallel processing models would expect some interaction between prosodic and lexical biases in parsing. I examine the consequences of the results of Experiment 4 in Chapter 7. I argue that despite the suggested lack of interaction between prosodic and lexical information in parsing, the results of Experiment 4 are more easily explained by interactive models of processing.

CHAPTER SEVEN
DISCUSSION AND CONCLUSIONS

7.1. INTRODUCTION

This chapter begins with a recapitulation of the results of the four experiments in this thesis, focussing on the findings that shape this conclusion. Then I provide a more detailed description of previously suggested interactive and serial models of processing. I argue that the results of the four experiments in this thesis support parallel processing or interactive activation models. Conversely, the results are not compatible with the stress on syntactic autonomy usually associated with serial processing models. Finally, I analyse the role of prosodic information in parsing, given the results of the four experiments in this thesis. First, I recapitulate the results of the four experiments in this thesis.

Production data

- The speaker consistently emphasised the first word of the noun phrase version and the second word of the verb phrase version of the prosodically contrastive word pairs.
- Conversely, the speaker emphasised the second word of the noun and verb phrase versions of the prosodically non-contrastive word pairs, more often that not.
- There is evidence that f_0, timing, and amplitude interact to cue contrastive emphasis.

Experiment 1

- Subjects' response times were significantly lower when the probes were consistent with the prosody of the word pairs than when they conflicted with the prosody. Subjects were using prosody to provide an initial parse of the prosodically contrastive word pairs.
- Subjects did not prefer noun phrase parses of the word pairs, despite this interpretation being consistent with the principles of minimal attachment and late closure. The verb phrase reading violated the principles of minimal attachment and late closure (Carston: 1989).
- Subjects did not construct all plausible parses of the stimuli. They showed a strong preference for one of the interpretations of the prosodically contrastive stimuli.
- Subjects did prefer verb phrase parses of the stimuli – this also provides evidence of prosody influencing on-line parsing.
- Although there might be a number of variations in the productions of the prosodic characteristics of these word pairs in normal situations, listeners could use these prosodic cues to guide parsing in Experiment 1. This suggests that the prosodic characteristics of the stimuli were not different from listeners' expectations.

Experiment 2

- Subjects preferred verb phrase parses of the cross-synthesised noun phrases and noun phrase parses of the cross-synthesised verb phrases. This result was restricted to response times to probes following the prosodically contrastive stimuli.

- Subjects did not prefer noun phrase parses of the stimuli – as would be predicted by a processing system following the principles of minimal attachment or late closure.
- Subjects did not predict all plausible parses of the stimuli.
- There was a continued simultaneous preference for verb phrase parses of the stimuli.
- Prosodic information of f_0, timing, and amplitude does guide initial analysis of the word pairs.

Experiment 3

- Subjects could not disambiguate *some* of the prosodically contrastive word pairs when the stimuli were flattened.
- Subjects might have been guided by amplitude, timing, or spectral information in disambiguating *some* of the prosodically contrastive word pairs.
- Subjects could disambiguate *some* of the prosodically non-contrastive word pairs, possibly guided by spectral information.
- Subjects did not prefer noun phrase parses of the stimuli.
- Subjects did not predict all plausible parses of the stimuli. They still showed a slight preference for the parses indicated by amplitude, duration, and spectral information, despite reduced prosodic input.
- Listeners are flexible in their use of prosodic information to guide parsing. When the required prosodic cues are present, subjects could use them to guide parsing. When the required prosodic cues are not present, subjects use other information to guide parsing.

The influence of lexical bias

- Most of the prosodically contrastive word pairs were more likely to occur as noun phrases, while most of the prosodically non-contrastive word pairs were more likely to occur as verb phrases.
- Subjects' parsing of the word pair as the noun or verb phrase reading was influenced by the likelihood of occurrence of this reading in speech.
- The parse finally decided was based on either prosodic or lexical bias. There was little indication of an interaction between the two – subjects did not take longer to respond to probes confirming the prosodically appropriate parse when the lexical bias of the word pair conflicted with this interpretation than when it confirmed it.
- Prosodically flattening the stimuli did not encourage subjects to pay more attention to the influence of lexical bias. Cross-synthesising the stimulus also did not encourage subjects to pay more attention to the influence of lexical bias.

Experiment 4

- Subjects could disambiguate the noun and verb phrase versions of the prosodically contrastive word pairs even when the stimuli were delexicalised.
- Removing the lexical content of the stimuli did not affect the perceptual salience of prosodic information.
- Subjects continued to prefer verb phrase parses of the stimuli.
- Subjects did not show a preference for noun phrase parses or predict all plausible parses of the stimuli.

- Subjects were no longer influenced by the likelihood of occurrence of one of the alternative readings – there was no influence of lexical bias
- The influence of lexical bias appears to be independent of the influence of prosodic information.
- ***Prosodic information can initiate on-line parsing in the absence of lexical, segmental, and syntactic information.***

These were the main findings of the four experiments reported in this thesis. I now analyse whether these results would be predicted by serial and parallel processing models that have been proposed so far.

7.2. SERIAL PROCESSING

Serial processing models propose that linguistic input is analysed through a series of processes dedicated to decoding the syntactic, semantic, prosodic, and pragmatic content of the stimulus. There are numerous variations of this central claim. However, most maintain that the initial step of the series is the processing of only the syntactic content of the stimulus. The purpose of this initial syntactic processing is to produce a single syntactic interpretation of the ambiguous input. The proposed theories differ on whether initial syntactic analysis takes into consideration only a few syntactic principles, such as Minimal Attachment and Late Closure (Frazier et al: 1983; Frazier: 1985) or all of them (Crain and Fodor: 1985). However, most models invoke the notion of modularity (Fodor: 1983) in order to ensure solely syntactic initial processing of the signal. Fodor proposed that processing involves the transformation of the input into a form that is more accessible to higher-level processing such as the fixation of belief or the incorporation of the representation into the encyclopaedic memory. This transformation of the perceptual stimulus is performed by informationally encapsulated input systems called *modules*. Specifically, the modular stage of speech processing involves the transformation of the linguistic signal into a representation of the syntactic structure of the input. Fodor argues that this initial analysis of the input does not consider the semantic content of the signal:

> ...semantic information is never used to predict syntactic structure, but a line of analysis on which the parser is engaged can be aborted whenever it produces structures that resist contextual integration...all that the context analyser is allowed to say to the parser is either 'yes' (continue with the present line of analysis) or 'no' (try something else, I can't fit what you're giving me to the context). What the context analyser is prohibited from doing is telling the parser which line of analysis it should try next – i.e., semantic information can't be used predictively to guide the parse (Fodor: 1983: 134-5)

Only once the syntactic structure of the input has been constructed is the parse verified by the semantic content of the signal. This implies that the meaning of an utterance cannot be accessed until the entire sentence has been parsed. There are obviously a number of variations of models of serial processing. However, three strains of thought dominate discussions of serial processing. Firstly and rather obviously, processing is serial: i.e., perception and recognition takes place through a series of different analyses. This is undisputed in all the different models of serial processing. Secondly, initial processing uses only the syntactic content of the stimulus (First analysis models/ Syntax-first models: Frazier et al: 1983; Frazier: 1985). Thirdly, initial processing is modular. However, evidence against initial syntactic processing does not provide evidence against serial processing. Similarly, evidence against the modularity of initial processing does not provide evidence against serial processing – although the links between the two are stronger than between syntactic and serial processing. The serial nature of processing is the only distinctive claim of serial processing models. The other two claims are more open to debate in discussions of serial processing. A lot of this debate concerns the role of prosodic information in processing. This will, obviously, form the central concern of my discussion of speech processing models.

Chapter 1 outlined a few main proposals of serial processing models arguing that prosodic information is a part of the 'general structural principles' guiding parsing (Frazier: 1987; Carlson et al: 2001). In addition, I outlined the contradictions inherent in proposing that prosodic information can influence on-line parsing. Frazier and colleagues argue that the parser cannot provide an initial parse without using all the information in the syntactic module (Frazier: 1990). This syntactic module also contains information about the prosodic content of the message. However, their work resounds with uncertainty of the possibility and the mechanism of early incorporation of prosodic information in parsing (Carlson, Clifton and Frazier: 2001; Carlson: 2001). These suggestions leave open the possibility of initial analysis using solely syntactic cues with prosodic cues coming into play later, as I have argued in Section 1.2.1.

This position is complicated further by considering Frazier's outline of initial processing using this syntactic-prosodic module:

> At each point in the ongoing analysis of a sentence, a decision will not be made within any module without checking that the decision is consistent with all within module information (Frazier: 1990: 415)

Since the module includes both prosodic and syntactic information, the parse chosen by the module would have to be consistent with both the syntactic and prosodic content of the stimulus. However, this standpoint is ambiguous. I have already described the principles guiding initial autonomous syntactic parsing according to Frazier (1989). These were the principles of Minimal Attachment and Late Closure. According to Carston (1989), a parser following the principles of minimal attachment and late closure would always prefer the noun phrase parse of the stimuli tested in the experiments in this thesis. However, the prosodic content of the stimulus would guide the parser towards a verb phrase parse in half of the presentations. Therefore, the syntactic and prosodic content of the stimuli would make contradictory parses of the stimuli of the verb phrase stimuli if adhering to the principles of minimal attachment and late closure. Consequently, it would be impossible for the parse constructed to be simultaneously consistent with the prosodic and syntactic content of the verb phrase versions of the stimuli. If the preferred parse has to be consistent with all the information within the module, then the parser would not be able to predict a parse of the stimulus of the verb phrase stimuli. If it proposed an initial parse using syntactic information, then this would conflict with the parse confirmed by the prosodic content of the stimuli. Conversely, if it proposed the parse indicated by the prosody of the stimuli, then this parse would conflict with the syntactic preferences of the syntactic module. This standpoint entails a conflict within the syntactic-prosodic module. In the next half of this section, I analyse a number of ways in which this conflict can be resolved.

On the one hand, we could propose that syntactic information has priority within the syntactic-prosodic module. An initial parse is created using solely syntactic information. If the parse constructed within the module is inconsistent with prosodic information, then the parser produces an alternative parse of the stimulus. In fact, this seems similar to serial processing as proposed by Carlson et al:

> The simple structure hypothesis predicted that the simplest syntactic structure would be the preferred analysis, and this was clearly the dominant factor in the auditory processing of potentially gapping sentences as well. Other factors like prosodic parallelism only partially influenced the choice between structures. (Carlson: 2001: 19).

Prosodic information is incorporated only after initial syntactic processing within the module. However, this removes the motivation for suggesting that prosodic information is incorporated within the syntactic module. In addition, the experiments in this thesis show that syntactic information does not always have priority over prosodic information in parsing. At the very least, this is true if we consider that the syntactic principles guiding parsing are restricted to the principles of minimal attachment and late closure. Subjects' response times to probes consistent with verb phrase parses of the stimuli were actually faster than subjects' response times to noun phrase parses – the preferred parse following the principles of minimal attachment and late closure (Carston: 1989). Therefore, the results of the experiments in this thesis provide evidence against suggestions of priority always being given to these syntactic parsing principles in parsing. Subjects' initial parses were consistent with the parses suggested by the prosody of the stimuli in the first and second experiments in this thesis.

However, would the results in this thesis be compatible with serial processing models if Minimal Attachment and Late Closure were not the only principles guiding syntactic processing? Frazier and Rayner (1987) argue that category ambiguous word pairs, such as those tested in the experiments in this thesis, are parsed using a delay strategy – upon parsing the first word, the parser realises that this is a case of category ambiguity and waits to receive further disambiguating information. They measured subjects' reading times with sentences such as (7.1), (7.2), (7.3) and (7.4):

(7.1) I know that the desert trains young people to be especially tough.

(7.2) I know that the desert trains are especially tough on young people.

(7.3) I know that this desert trains young people to be especially tough.

(7.4) I know that these desert trains are especially tough on young people.

Frazier and Rayner argue that the default interpretation of *dessert trains* would be that of a noun and verb (similar to Carston (1989)). Therefore, they expected longer reading times to (7.2) than for (7.1), (7.3), or (7.4). However, subjects had shorter reading times in the ambiguous region of (7.1) and (7.2) than in (7.3) and (7.4). This was explained by arguing that subjects were waiting for further information in (7.1) and (7.2), due to realisation of category ambiguity. Therefore, Frazier and Rayner (1987) would predict a different parsing mechanism for category ambiguous words that would not operate on a minimal attachment parsing preference. However, MacDonald (1993) and Warren (1985) argue that the determiners preceding the ambiguous words in (7.3) and (7.4) warrant the longer reading times in the ambiguous region of these sentences. Both determiners indicate that the *desert* is given information. Therefore, subjects' longer reading times in the ambiguous region of (7.3) and (7.4) are an indication that subjects are merely confused by the status of givenness implied by the determiner. MacDonald (1993) supports this observation by claiming that subjects' reading times in the ambiguous region were longer even when the phrase is made unambiguous (I know that the/these deserted trains...). There is little evidence for a delay strategy of parsing if the evidence provided by Frazier and Rayner (1987) is called into question. Conversely, if the parser were to delay decisions regarding category ambiguous stimuli till further information is received, this might give prosodic information priority in parsing. However, further evidence of a delay strategy in parsing needs to be acquired before this can be proposed as a viable parsing alternative.

It could also be argued that prosodic input always has priority within the module. The results of the experiments in this thesis are consistent with this claim. However, Carlson et al (2001) disagree with this stipulation. They claim that factors such as prosodic parallelism only partially affected parsing. The results reported by Carlson et al (2001) and Schepman (1997) contradict the findings of the four experiments presented here. It is possible that the contradiction in results was caused by the fact that Carlson et al employed off-line tests of parsing. However, if prosodic information has priority, then subjects should have been guided towards the intended parse of the stimuli early in processing. Therefore, it should make no difference whether parsing is tested on-line or off-line. Both tests should show positive results in favour of prosody guiding parsing. However, Carlson et al (2001) reports finding only a marginal effect of prosody, while the results in this experiment found evidence of prosody guiding parsing. Therefore, proposing that prosodic information has priority within-module also does not seem a consistently viable option. A suitable model of processing has not only to be consistent with the results of the experiments reported in this thesis, but also with prior experiments establishing an influence of syntactic preferences guiding parsing.

Another way of resolving this issue is to propose that serial processing does not actually take place. At the very least, there is no evidence of mandatory initial syntactic parsing of ambiguous stimuli, given that this initial parsing must always follow the principles of Minimal Attachment and Late Closure. Admittedly, there is considerable evidence of the early incorporation of syntactic information in processing in the earlier literature referred to in Chapter 1. Similarly, the research reported in this thesis provides evidence of the early incorporation of prosodic information in processing. It also provides evidence of the early incorporation of lexical information in processing. These facts taken together provide evidence that is more in keeping with interactive processing models that do not allow default priority to any component. Therefore, in the next section, I analyse whether it is more logical to ignore revisions of serial processing models in favour of a parallel processing explanation of the results in this thesis.

7.3. PARALLEL PROCESSING

Parallel processing models hold that processing of the syntactic, semantic, and prosodic content of the stimuli can take place simultaneously. They do not presuppose inflexible priority for either syntactic or prosodic or pragmatic input during processing. Therefore, the integration of pragmatic and syntactic meaning need not take place after the construction of an initial syntactic structure. According to the strong version of parallel processing, semantics or syntax (or presumably, prosody) can *direct* the course of action the parser is to follow (Tyler and Marslen-Wilson 1977; Crain 1980). On the other hand, according to weak interactive processing, the syntactic parser proposes all plausible structures of an ambiguous input. Semantics (and presumably, prosody) cannot *direct* the course of action the parser is to follow. The semantic processor can prevent the syntactic processor from continuing on a line of analysis that is inconsistent with the semantic characteristics of the input.

Differences in the timing of interaction between the semantic and the syntactic processor affect the strength of the interaction. For instance, Altmann and Steedman (1988) argue that if this influence were on a word-by-word basis, there would be few differences between the strong and weak interactive parser – semantics could influence parsing on a word-by-word basis. Therefore, Altmann and Steedman propose a 'fine-grained' weakly interactive parser, where semantics has the ability to either agree or disagree with a current syntactic analysis on a nearly word-by-word basis.

The strong version of parallel processing models was proposed in the early works of Marslen-Wilson and Tyler, Crain and Steedman (1988), and Pynte and Prieur (1996) among others. However, Marslen-Wilson et al (1992) later modified their proposed model to highlight the effect of prosody on the construction of syntactic structures. They had found that prosodic information could be used to resolve ambiguities in the assignment of syntactic structures. However, they could not be sure of the nature of the interface between the syntactic and prosodic information within the processing network. They also suggested a 'weaker account of interaction': the processor constructs all possible structures of the input using purely syntactic information. All the structures thus computed are *then* tested for compatibility with the prosodic and pragmatic content of the message. This presupposes that syntactic information retains some prioritised access to the stimuli.

Parallel processing models can be explained in terms of interactive activation models of sentence processing (McClelland, St John, and Taraban: 1980; McClelland: 1987; McClelland and Kawamoto: 1986; St John and McClelland: 1992 among others). In this thesis, I will concentrate on the characteristics of interactive activation models described by McClelland (1987) and St John and McClelland (1992).

A parallel distributed processing network is divided into a number of levels, each level consisting of units of alternative representations of the different characteristics of the input. McClelland (1987) proposes that the organisation of the levels in a network could be as in Figure 7.1.

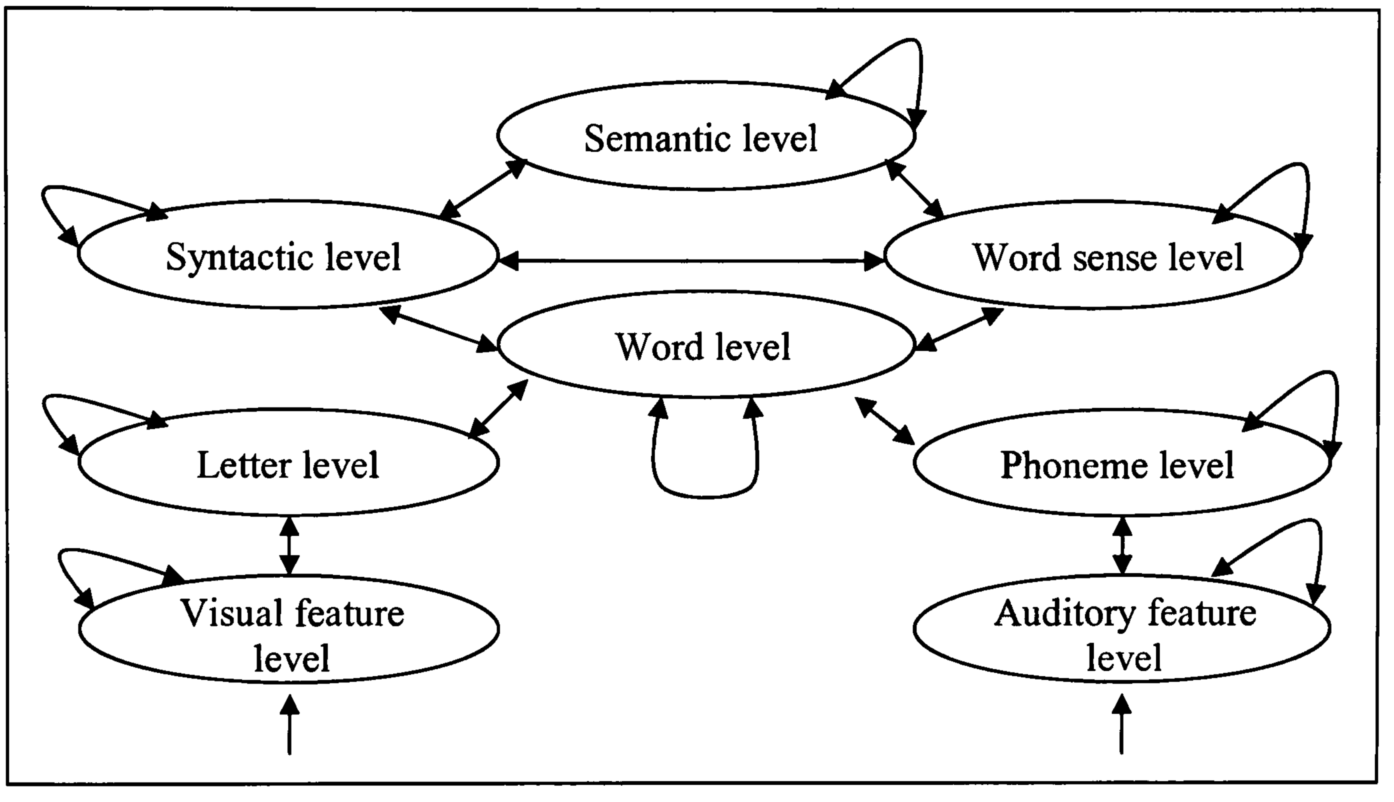


Figure 7.1 A possible representation of the levels involved in an interactive activation network.

Different models of parallel processing would include different levels. For instance, McClelland (1987) proposes that processing speech would include a phonetic level. The representation created by any one level will be influenced by the representations at the other levels of processing. Interaction between and within levels is bi-directional. Connections between levels are restricted to adjacent levels alone. However, each level can influence another level by indirect connections. Connections between levels cannot inhibit the activation of an alternative in a different level. This allows for the plausible alternative to receive support from different sources simultaneously. McClelland terms this a 'best match' strategy that, for instance, will allow a sequence of phonemes to trigger the activation of a word even if all the phonemes do not match the word entirely. The strength of influence between levels is graded, so a particular level can assert a stronger influence for one input and a weaker influence for another. This is decided by strength of the activation for the input within level. The knowledge within levels is encapsulated, so while connections run between levels, no level can access the knowledge that is present in another level. The input a level receives from another level is merely a constraint on the representations that another level can construct. McClelland's model of processing simulates strong interactive processing.

Admittedly, since connections between the lower levels and the higher levels are indirect, there is still a delay between activation at lower and higher levels, for instance, the activation at the phoneme level and activation at the word level. This position has been challenged by empirical data showing that activation at the phoneme level does not always precede activation at the word level (Pitt and Samuel: 1995). In addition, there is evidence of top-down semantic effects on word-recognition (Warren: 1970). While these top-down effects could be explained by McClelland's model, recent models by Gaskell and Marslen-Wilson (1997) provide an outline of distributed models that allow access from the auditory features to both lexical and phonological levels simultaneously. This would remove delays between activation at lower and higher levels.

Interactive activation models provide a non-serial explanation for experiments suggesting that semantic information cannot influence syntactic parsing (Rayner et al: 1983; Ferreira and Clifton: 1986; Frazier: 1987). The influence of semantics can vary depending on plausibility of a semantic representation. Therefore, MacDonald (1993) argues that prior results reporting evidence against interaction between semantics and syntax have tested stimuli that do not provide adequate semantic context to the network. In these cases, the influence of semantic context does not form a strong constraint on syntactic processing. Conversely, experiments providing evidence in favour of semantic influences in on-line parsing have provided adequate semantic context to strengthen the constraint on syntactic parsing. The results of the experiments reported in this thesis could also be explained by arguing that the prosodic content constitutes a strong constraint on syntactic parsing. Conversely, the syntactic content may not provide adequate information to guide parsing towards the minimally attached parse. This notion of probabilistic constraints on syntactic resolution was explored by MacDonald (1993) using sentences such as (7.5) and (7.6):

(7.5) The ruthless dictator fought in the coup was hated throughout the country.

(7.6) The ruthless dictator fought just after dawn was hated throughout the country.

She found that subjects' reading times throughout the sentence were longer in stimuli with poor constraints such as (7.6) than stimuli like (7.5) with stronger semantic constraints. MacDonald (1993) also found that reading times were also sensitive to whether the constraints conflicted or converged. This provides additional evidence for simultaneous interaction between constraints in parsing.

Steedman (1992) proposes a weak interactive parsing model where syntactic cues are given priority to construct all plausible initial parses, based on syntactic parsing preferences. These parses are then checked by the prosodic and semantic cues present in the stimulus. Steedman proposes that this checking for compatibility is carried out in modules specific to the processing of this information. Steedman extends Fodor's conception of syntactic modules to allow the possibility of semantic and prosodic components that "are all formally and computationally autonomous and domain-specific" (1992:495). However, there are no connectionist models of interactive activation that consider the influence of prosodic information on parsing to my knowledge. For instance, Christiansen et al (1998) analyses the effect of lexical stress on the segmentation of speech, but does not look at parsing using prosodic constraints. However, an interactive activation approach could invoke an influence of prosodic constraints in a similar manner to semantic constraints. Consider, for instance, the SRN suggested by Christiansen et al (1998), presented in Figure 7.2.

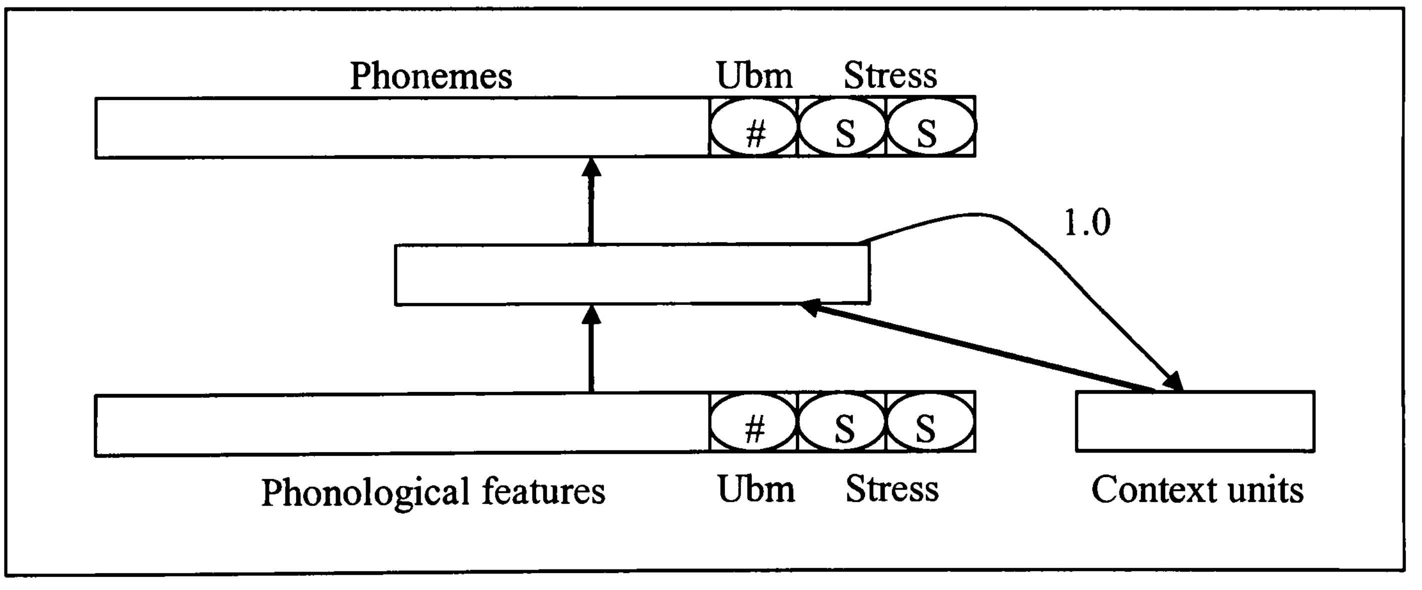


Figure 7.2 SRN used by Christiansen et al (1998)

The input to the model is phonological, based on the activation of 11 phonetic features or an utterance boundary marker (Ubm). The two S units represent lexical stress on each syllable. When both units have the value '0', the syllable is unstressed. When the first unit is on ('1') and the second unit is off ('0'), the syllable has intermediate stress. The syllable has strong stress if the first unit is on ('0') and the second unit is on ('1'). Both stress and phonological units are off when the Ubm is on. The output layer has 36 phonological units, an utterance boundary unit, and stress units. The layer in the middle represents the hidden layer interacting with context units. Christiansen et al (1988) found that, given phonological, utterance boundary and stress information, the network easily learnt to segment speech input. While this model focussed on the segmentation of speech, it would be easy to extend it to the parsing of ambiguous input. Figure 7.3 presents a diagrammatic representation of the role of prosody in disambiguating word pairs that would be in keeping with interactive activationist accounts of parsing.

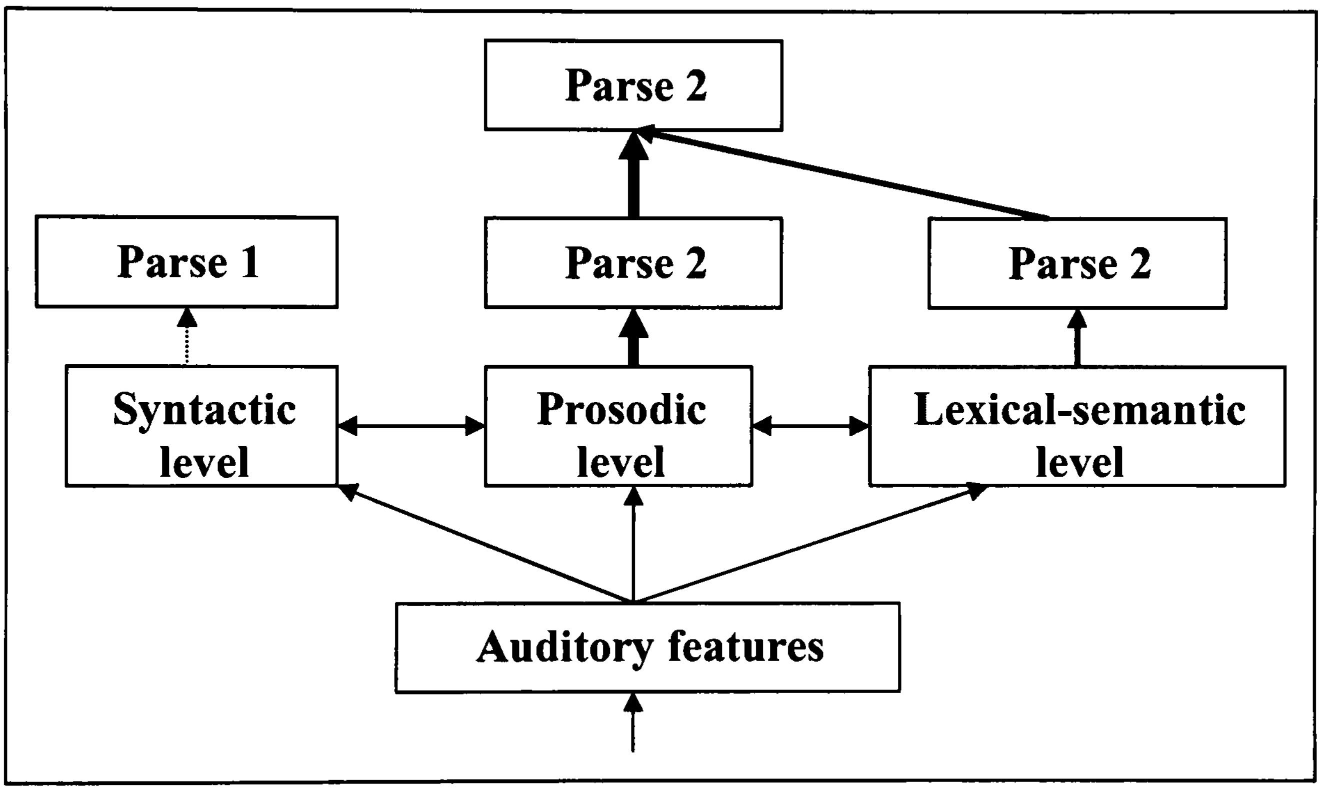


Figure 7.3 Diagrammatic representation of the role of prosody in a parallel processing model

According to Figure 7.3, the auditory features of the input are simultaneously analysed by syntactic, prosodic, and semantic levels. It is possible that there are additional levels between the receipt of the auditory features and analysis by the different levels, e.g. word levels. Once the input is parsed by the syntactic, prosodic, and semantic levels, each level suggests a representation of the input that is compatible with the information in it. 'Parse 1' and 'Parse 2' represent the two alternative interpretations of the input. As with McClelland (1987), there is interaction between the levels, but this is only to increase the activation of a representation in another level. Connections between levels cannot inhibit the activation of a representation in another level. The thickness of the arrows between the parses suggested and the levels indicate the weights attached to these parses (Christiansen et al 1988). For instance, in Figure 7.3, the prosodic constraint on 'Parse 2' has stronger weights than either the syntactic constraint on 'Parse 1' or the semantic constraint on 'Parse 2'. Therefore, 'Parse 2' would finally be preferred by the network.

If we include a prosodic level in the network, this would allow prosodic constraints to operate on syntactic resolution. The experiments in this thesis presented subjects with more than adequate prosodic information and ambiguous syntactic information. Therefore, the prosodic constraints on resolution would have been strong enough for the prosodically preferred parse to be activated. It is possible that the stimuli in other experiments did not provide sufficient prosodic information to enable strong constraints against prosodically inappropriate parses. In these cases, parsing might have been motivated by the presence of strong constraints towards syntactically preferred parses. Similarly, Experiment 2's results could be explained by adequate prosodic information posing a strong constraint on resolution, despite within-level competition between spectral and other purely prosodic influences (amplitude and timing). Experiment 3's results could be explained by arguing that the amplitude, timing and spectral content of the stimuli provide adequate information to pose a strong constraint on resolution in the case of some of the stimuli. The flexibility provided by graded constraints – the fact that these constraints can vary in their strengths depending on the stimuli – allows a suitable analogue to the cue trading data reported in the discussion of Chapter 3. Finally, the results of Experiment 4 would be similarly explained by the absence of adequate syntactic, lexical and segmental information providing the prosodic constraint unrivalled ability to guide parsing.

However, subjects' response times to appropriate probes did not vary irrespective of whether there was conflict between the parses consistent with the lexical bias and prosody of the stimuli. This is contradictory to MacDonald's data reporting longer response times to sentences where constraints conflict than when they converge. This is not an insurmountable problem. It could be argued that in cases of conflict, one of the constraints was always strong enough to disregard the proposal of an alternative parse by the other constraint. Therefore, in some cases, there was adequate prosodic information to rule out any other parses proposed by the semantic level. In other cases, there was adequate prosodic information to rule out any parses proposed by the prosodic level. Connections in an interactive activation network were not inhibitory between levels. Therefore, the proposal of a parse by one level would not inhibit the conflicting parse proposed by a different level. Conversely, the presence of a stronger preference for one parse in one level could override the proposal from another level. The difference between the results reported by MacDonald (1993) and the interaction between lexical bias and prosody in the experiments in this thesis can also be explained. It could be argued that there was a lesser difference between the strengths of conflicting constraints in MacDonald's experiments. This would lead to greater conflict and, consequently, significantly longer reading times. Conversely, it is possible that the conflict between the prosodic and semantic constraints in the experiments in this thesis was not strong enough. The difference between the strengths of the conflicting constraints might have been significant enough to warrant one of them winning out. Admittedly, the lack of interaction between prosody and lexical bias can also be explained by proposing that there is no interaction between the prosodic and lexical-semantic level during parsing. However, ERP data presented in the next section provide considerable evidence for interaction between prosody and syntax (Friederici 2001) and prosody and semantics Asetesano et al 2004) during parsing. Consequently, the results of the experiments in this thesis can be explained by adapting current interactive activation models. Considering the problems outlined in Section 7.2, this thesis concludes that interactive activation accounts might accommodate the results of the experiments in this thesis more easily than serial processing models can. In the next section, I take a final look at the questions raised in this thesis regarding the status of prosodic processing to analyse whether the results of this thesis adequately address these issues.

7.4. PROSODIC PROCESSING

The results of this thesis provide evidence that prosodic information can guide parsing earlier than has been established to date. In Chapter 1, I argued that there was considerable uncertainty regarding the role of prosodic information in parsing. This was mainly due to a number of experiments reporting contradictory results for (Marlsen-Wilson et al: 1992; Grabe et al: 1994; Schepman: 1997; Kjelgaard and Speer: 1999) and to a great extent, against (Watt and Murray: 1996; Warren: 1995) the early incorporation of prosody in parsing. This was probably because most of them (excluding Grabe et al: 1994 and Warren: 1995) tested stimuli well into the ambiguous sentence. I have argued that testing was almost always at a point much after syntactic ambiguity had become apparent. Adequate prosodic information had been provided to guide parsing. However, perhaps due to the delay in testing parsing, a strong effect of prosody could not be found in some experiments. For instance, Schepman (1997) argues that the effect of prosody could be reduced further into the sentence as other information becomes available from the stimuli. Conversely, experiments reporting positive results in favour of prosodic information guiding parsing had to concede that a weaker form of interaction between prosody and syntax could also explain the results of their experiments.

All four experiments in this thesis tested parsing very early in the sentence – earlier than has ever been tested in experiments to date. They provide strong evidence that prosody can influence parsing at the beginning stage of parsing. The results of prior experiments could usually be explained by a weaker form of interaction between syntax and prosody. However, because the experiments in this thesis tested parsing early, the results can only be explained by a parsing mechanism that allows strong interaction between prosody and syntax even in the early stages of parsing. Admittedly, following Altmann and Steedman (1988), it is possible that the results in this thesis could be explained by a fine-grained weakly interactive model. This would ensure that prosody could influence the course of action of syntactic parsing on a word-by-word basis. However, as concluded by Altmann and Steedman, this would leave little difference between strong and weak interactive processing models. In addition, the fourth experiment of this thesis provides evidence that prosody can guide parsing even in the absence of syntactic, lexical, and segmental information. This would imply that prosodic information *guides* parsing independently. This provides evidence against a weak interactive parser, according to which prosodic information can merely inhibit the parses provided by initial syntactic parsing. Therefore, the results in this thesis can only be explained by very strong interaction between prosody and syntax during the initial stages of parsing. This interaction also extends to the early incorporation of lexical bias during parsing. The analysis presented in Chapter 5 showed that subjects' parses were also affected by lexical bias.

The guiding role attributed to prosodic information in processing is supported by results of numerous ERP experiments reporting finding a closure positive shift at prosodic boundaries. This closure positive shift was even found while listening to delexicalised speech, providing strong evidence that the closure positive shift is related to solely prosodic processing. Similar results are reported by Pannekamp et al (2005). They played subjects sentences in which all content words were replaced by meaningless words, sentences in which all function words were replaced by meaningless words, and finally delexicalised sentences. The fact that the closure positive shift was obtained with respect to the intonational boundaries in all cases suggests that this effect was caused by prosodic processing. However, Steinhauer (2003) found that although the CPS was found while listening to normal and delexicalised speech, there were differences between the two closure positive shifts. This supports suggestions of interaction between prosody and syntax and prosody and semantics during processing. Asetesano et al (2004) found that the positive shift found when listening to prosodically incongruous sentences was larger when the sentences were also semantically incongruous than when they were semantically congruous. They use this to argue for interaction between prosody and semantics in parsing. Similarly, using ERPs obtained while subjects listened to sentences with varying lexical-semantic, syntactic, and prosodic information, Friederici (2001) argues that there is interaction between prosody and syntax during parsing. However, Friederici found no evidence of interaction between semantic and syntactic information. Interaction between prosody and syntax and prosody and semantics during parsing can be easily explained by interactive activation models. Conversely, the absence of evidence of interaction between semantics and syntax could be due to inadequate semantic information to sufficiently constrain interaction between the two components.

I conclude that the results of the experiments presented here can be easily explained by current interactive activation models, given the inclusion of a prosodic level in processing. However, in analysing the stimuli presented to subjects in the experiments in this thesis, I found that there were considerable differences in the prosodic realisation of the verb phrase versions of the word pairs. Given this variability, can prosodic information guide parsing? I argue that despite differences in the prosodic realisation of the verb phrase versions, there were adequate contrastive emphasis between the noun and verb phrase versions of the word pairs. Subjects could have used these differences to guide parsing. In addition, as I have argued in Chapter 3, I approach this issue from a different angle. Subjects in all experiments were able to use prosodic information to guide parsing. This was despite variability in the prosodic realisation of the different versions of the word pairs. Therefore, the latter cannot be a valid reason to discount *adequate* prosodic information from guiding parsing. Again, as I argue in Chapter 3, subjects' ability to parse some of the flattened stimuli demonstrates flexibility in the use of prosodic information to guide parsing. This flexibility can be characterised by prototype matching, similar to a Fuzzy Logic model of parsing and can easily be incorporated into an interactive activation model, as I have argued in Section 7.3.

There is a final issue to be raised concerning the role of delay strategies in parsing. There was no evidence that subjects were being guided by syntactic information in the experiments presented here. However, it is possible that the syntactic content of the stimuli was not adequate to trigger Minimal Attachment and Late Closure. This would support Frazier and Rayner's suggestion that serial processing is sensitive to category ambiguities, causing the parser to wait until further information is provided. This would indicate that, in these cases, prosodic information could be used to guide parsing, as has been suggested in some of the literature in serial processing (Carlson et al: 2001). There is insufficient evidence in favour of delay strategies being activated in the case of category ambiguous stimuli. If further evidence of this influence were found, then it might be possible to explain the results in this thesis using a serial processing model. However, this would need to be explored in greater detail once there is strong evidence in favour of delay strategies in parsing.

In conclusion, the results of the experiments in this thesis suggest that parsing using prosody is an interactive process that commences immediately upon receiving ambiguous input. Disambiguation considers the prosodic, syntactic, and semantic-lexical content of the stimuli simultaneously, as would be suggested by interactive-activation models of processing. There is more evidence in favour of strong than weak interactive models of processing, since subjects were able to disambiguate even the delexicalised stimuli at the point of realisation of syntactic ambiguity. This suggests that prosody can guide parsing independently, and does not merely inhibit syntactic processing of prosodically inappropriate parses. The results of this thesis are incompatible with serial models of processing unless further evidence in favour of delay models of processing is reported. At the very least, the results of this thesis provide strong evidence against the mandatory application of syntactic parsing principles like Minimal Attachment and Late Closure in parsing ambiguous sentences. Prosody can guide disambiguation independently of syntactic, lexical, or even segmental information.

REFERENCES

Albritton, D., McKoon, G., & Ratcliff, R. 1996. 'Reliability of prosodic cues for resolving syntactic ambiguity', *Journal of Experimental Psychology: Learning, Memory and Cognition, vol. 22.3*, pp. 714-735.

Altmann, G., & Steedman, M. 1988. 'Interaction with context during human sentence processing', *Cognition, vol. 30*, pp. 191-238.

Anderson, S. W., & Cooper, W. E. 1986. 'Fundamental frequency patterns during spontaneous picture description', *Journal of the Acoustical Society of America*, pp. 1172-1174.

Asetesano, C., Besson, M., & Alter, K. 2004. 'Brain potentials during semantic and prosodic processing in French', *Brain Res Cogn Brain Res, vol. 18*, pp: 172-184.

Beach, C.M. 1991. 'The interpretation of prosodic patterns at points of syntactic structure ambiguity: evidence for cue trading relations', *Journal of Memory and Language, vol. 30*, pp. 644-663.

Beckman, M.E. 1996. 'The Parsing of Prosody', *Language and Cognitive Processes, vol. 11(1/2)*, pp. 17-67.

Behrens, S. 1989. 'Characterising sentence intonation in a right-hemisphere damaged population', *Brain and Language, vol. 37*, pp.181-200.

Berndt, R.S., Salasoo, S., Mitchum, C.C., & Blumstein, S.E. 1988. 'The role of intonation cues in aphasic patients' performance of the grammaticality judgement task', *Brain and Language, vol. 34*, pp. 65-97.

Binder, K.S., & Morris, R.K. 1995. 'Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic', *Journal of Experimental Psychology: Learning, Memory, and Cognition, vol. 21*, pp. 1186-1196.

Binder, K.S., & Rayner, K. 1998. 'Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading', *Psychonomic Bulletin & Review, vol. 5*, pp. 271-276.

Birch, S., & Clifton, C., Jr. 1995. 'Focus, accent, and argument structure: Effects on language comprehension', *Language and Speech, vol. 33*, pp. 365-391.

Birch, S., & Clifton, C., Jr. 2002. 'Effects of varying focus and accenting of adjuncts on the comprehension of utterances', *Journal of Memory and Language, vol. 47*, pp. 571-588.

Blasko, D. G., & Hall, M. D. 1998. 'Influence of prosodic boundaries on comprehension of spoken English sentences', *Perceptual and Motor Skills, vol. 87*, pp. 3-18.

Blumstein, S., & Cooper W. 1974. 'Hemispheric processing of intonation contours', *Cortex, vol. 10*, pp. 146-158.

Bock, K., & Mazella, J.R. 1983. 'Intonational marking of given and new information: Some consequences for comprehension', *Memory and Cognition, vol. 11*, pp. 64-76.

Boersma, P., & Weenink, D. 2000. Praat computer program. On-line, Institute of Phonetic Sciences, University of Amsterdam, http://www.praat.org/

Boland, J.E., & Blodgett, A. 2001. 'Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements', *Journal of Memory and Language, vol. 45*, pp. 391-411.

Boland, J.E. 1997. 'Resolving syntactic category ambiguity in discourse context: Probabilistic and discourse constraints', *Journal of Memory and Language, vol. 36*, pp. 588-615.

Bolinger, D. 1958. 'A theory of pitch accents in English', *Word, vol. 14*, pp. 109-149.

Bolinger, D. 1965. 'Pitch accent and sentence rhythm', In Bolinger, D. (ed) *Forms of English: Accent, Morpheme, Order*, Harvard University Press, Massachusetts.

Carlson, K. 2001. 'The effects of parallelism and prosody on the processing of gapping structures', *Language and Speech, vol. 44*, pp. 1-26.

Carlson, K., Clifton, C., Jr., & Frazier, L. 2001. 'Prosodic boundaries in adjunct attachment', *Journal of Memory and Language, vol. 45*, pp. 58-81.

Carpeneter, P.A., & Daneman, M. 1981. 'Lexical retrieval and error recovery in reading: A model based on eye fixations', *Journal of Verbal Learning and Verbal Behaviour, vol. 20*, pp. 137-160.

Carston, R. 1989. 'Modularity and Linguistic Ambiguity' *University College London Working Papers in Linguistics. vol. 1*, pp. 340-353.

CCITT.1989. *Blue Book 5, Telephone Transmission Quality, Series P Recommendations, IX.* Plenary Assembly, Geneva, Recommendation, P.50, pp. 87 –98.

Cho, T., & Keating, P. 2001. 'Articulatory and acoustic studies of domain-initial strengthening in Korean', *Journal of Phonetics, vol. 28.* p. 155-190.

Chomsky, N. 1959. 'A review of B.F. Skinner's Verbal Behaviour', *Language, vol. 35(1)*, pp. 26-58.

Christiansen, M.H., Allen, J., & Seidenberg, M.S. 1998. 'Learning to segment speech using multiple cues: A connectionist model', In Plunkett, K.R. (ed.) *Language Acquisition and Connectionism*. Psychology Press Ltd. Sussex.

Clifton, C., Jr., & Ferreira, F. 1987. 'Modularity in sentence comprehension', in Garfield , J.L. (ed) *Modularity in Knowledge Representation and Natural Language Understanding,* MIT Press, Massachusetts.

Clifton, C., Jr., Carlson, K., & Frazier, L. 2002. 'Informative prosodic boundaries', *Language and Speech, vol. 45*, pp. 87-114.

Cohen, H., Douaire, J., & Elsabagh, M. 2001. 'The role of prosody in discourse processing', *Brain and Cognition, vol. 45*, pp. 73-82.

Coleman, J. 2003. 'Discovering the acoustic correlates of phonological contrasts', *Journal of Phonetics, vol.31*, pp. 351-372.

Cooper, W.E., & Paccia-Cooper, J. 1980. *Syntax and Speech*, Harvard University Press, Cambridge Massachusetts.

Cooper, W.E., & Sorenson, J.M. 1977. 'Fundamental frequency contours at syntactic boundaries', *Journal of the Acoustical Society of America, vol.62 (3)*, pp. 683-692.

Cooper, W.E. & Sorenson, J.M. 1981. Fundamental Frequency in Sentence Production, Springer.

Cooper, W.E., Eady, S.J., & Muller, P.R. 1985. 'Acoustical aspects of contrastive stress in question-answer contexts', *Journal of the Acoustical Society of America, vol. 77*, p. 2142-2156.

Cooper, W.E., Paccia, J.M., & Lapointe, S.G. 1978. 'Hierarchical coding in speech timing', *Cognitive Psychology, vol. 10*, pp. 154-177.

Crain, S., & Fodor, J.D. 1985. 'How can grammars help parsers?', In D.R. Dowty, L. Kartunnen, & A.M. Zwicky (Eds.), *Natural Language Processing: Psychological, computational, and theoretical perspectives*. New York: Cambridge University Press.

Crain, S., & Steedman, M. 1985. 'On not being led up the garden path', In D.Dowty, L.Karttunen & A.Zwicky (eds.) *Natural language parsing: psychological, computational and theoretical perspectives*. CUP, 320-358.

Cuetos, F., & Mitchell, D.C. 1988. 'Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish', *Cognition, vol. 30*, pp 73-105.

Darkins, A.W., Fromkin, V.A., & Benson, D.F. 1988. 'A characterisation of the prosodic loss in Parkinson's disease', *Brain and Language, vol. 34*, pp. 315-327.

De Pijper, J.R., & Sandermann, A.A. 1994. 'On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues', *Journal of the Acoustical Society of America, vol. 96*, pp. 2037-2047.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. 1996. 'Glottalization of vowel-initial syllables as a function of prosodic structure', *Journal of Phonetics, vol. 24*, pp. 423-444.

Elman, J. L., Bates E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Emmorey, K.D. 1987. 'The neurological substrates for prosodic aspects of speech', *Brain and Language, vol. 30*, pp. 305–329.

Fant, G., Liljencrants, J., & Lin, Q. 1985. 'A four parameter model of vocal flow', *STL-QPSR, vol. 4*, pp. 1-13.

Faulkner, A., & Rosen. S. 1996. 'The contribution of temporally-coded acoustic speech patterns to audio-visual speech perception in normally hearing and profoundly hearing-impaired listeners'. *ABSP-1996*, pp. 261-264.

Ferreira, F., & Clifton, C. 1986. 'The independence of syntactic processing', *Journal of Memory and Language, vol. 25*, pp. 348-368.

Ferreira, F., Anes, M.S., & Horine, M.D. 1996. 'Exploring the use of prosody during language comprehension using the auditory moving window technique', *Journal of Psycholinguistic Research, vol. 25.2*, pp. 273-289.

Fodor, J. D. 1998. 'Parsing to learn', *Journal of Psycholinguistic Research, vol. 27*, pp. 339–74.

Fodor, J.A. 1983. *The Modularity of the Mind: An Essay on Faculty Psychology*. MIT Press: Cambridge Massachusetts.

Ford, M., Bresnan, J., & Kaplan, R. 1982. 'A Competence-based theory of syntactic closure. In Bresnan, J. (ed) *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA.

Forster, K.I. 1974. 'The role of semantic hypotheses in sentence processing', In *Problemes Actuels en Psychologistique*, Colloques Internationaux du C.N.R.S.

Forster, K.I. 1976. 'The autonomy of syntactic processing', Paper presented to the MIT –AT & T Convocation on Communications, Massachusetts.

Francis, W. N., & Kucera, H. 1982. *Frequency analysis of English Usage: Lexicon and Grammar*, Boston: Houghton Mifflin.

Frazier, L., & Fodor, J.D. 1978. 'The Sausage Machine: A new two-stage parsing model', *Cognition, vol. 6*, pp. 1-34.

Frazier, L., & Raynor, K. 1982. 'Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences', *Cognitive Psychology, vol. 14*, pp. 178-210.

Frazier, L., & Rayner, K. 1987. 'Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences', *Journal of Memory and Language, vol. 26*, pp. 505-526.

Frazier, L. 1987. 'Sentence processing', In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Frazier, L. 1979. *On Comprehending sentences: Syntactic Parsing Strategies*, Indiana University Linguistics Club, Bloomington.

Frazier, L. 1985. 'Syntactic complexity.' In Dowty et al (eds), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives.* Cambridge University Press.

Frazier, L. 1989. 'Sentence processing: a tutorial review', In Coltheart, M. (ed) *Attention and Performance XII: The Psychology of Reading*; Erlbaum.

Frazier, L. 1990. 'Exploring the architecture of the language-processing system', In Altmann, G.T.M. (ed) *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*; MIT Press, Massachusetts.

Frazier, L., Clifton, C, Jr., & Randall, J. 1983. 'Filling gaps: Decision principles and structure in sentence comprehension', *Cognition, vol. 13*, pp. 187-222.

Frazier,L. Clifton, C., Jr., & Carlson, K. 2004. 'Don't break, or do – Prosodic boundary preferences', *Lingua, vol. 114.1*, pp. 3-27.

Gaskell, M.G., & Marslen-Wilson, M.W. 1992. 'Integrating form and meaning: A distributed model of speech perception', *Language and Cognitive Processes, vol. 12 (5/6)*, pp. 613-656.

Gee, J.P., & Grosjean, F. 1983. 'Performance structures: a Psycholinguistic and linguistic appraisal', *Cognitive Psychology, vol. 15*, pp. 411-458.

Geluykens, R., & Swerts, M. 1994. 'Prosodic cues to discourse boundaries in experimental dialogues', *Speech communication, vol. 15*, pp. 69-77.

Gilboy, E., Sopena, J.M., Clifton, C., & Frazier, L. 1995. 'Argument structure and association preferences in Spanish and English Complex NPs', *Cognition, vol. 54,* pp. 131-167.

Gleitman, L., & Wanner, E. 1982. 'Language acquisition : The state of the art', In Gleitman, L. & Wanner, E. (ed) *Language acquisition :The state of the art*, New York : Cambridge University Press.

Gocken, J.M., & Fox, R.A. 2001. 'Neurological evidence in support of a specialised phonetic processing module', *Brain and Language, vol.78*, pp. 241-253.

Grabe, E., & Warren, P. 1995. 'Stress Shift: do speakers do it or do listeners use it?' In Connell, B. and Arvaniti, A. (ed), *Papers in Laboratory Phonology IV. Phonology and Phonetic Evidence.* Cambridge: Cambridge University Press.

Grabe, E., Warren, P., & Nolan, F. 1994. 'Resolving category ambiguities: Evidence from stress shift', *Speech Communication, vol. 15,* pp. 101-11.

Gussenhoven, C. 1994. 'Focus and sentence accents in English', In Bosch, P. & Van der Sandt, R (ed), *Focus & Natural Language Processing: Working Papers of the Institute for Logic and Linguistics.* IBM Deutschland Infomations systeme GmbH, Heidelberg.

Herrmann, C.S., Friederici, A.D., Oertel, U., Maess, B., Hahne, A., & Alter, K. 2003. 'The brain generates its own sentence melody: A Gestalt phenomenon in speech perception', *Brain and Language, vol. 85*, pp. 396 – 401.

Hirschberg, J., & Ward, G. 1991. 'Accent and bound anaphora', *Cognitive Linguistics, vol.2*, pp. 101-121.

Hirschberg, J., & Ward,G. 1992. 'The Influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English', *Journal of Phonetics, vol.20*, pp. 241-251.

Holmes, V.M., Kennedy, A., & Murray, W.S. 1987. 'Syntactic structure and the garden path', *Quarterly journal of experimental psychology, vol. 39A*, pp. 277-293.

Isel, F., Alter, K., & Friederici, A.D. 2005. 'Influence of prosodic information on the processing of split particles: ERP evidence from spoken German', *Journal of Cognitive Neuroscience, vol. 17*, pp: 154-167.

Jescheniak, J.D., & Schriefers, H.J. 1998. 'Serial versus cascaded processing in lexical access in language production: Further evidence from the coactivation of near-synonyms', *Journal of Experimental Psychology: Learning, Memory and Cognition,vol. 24.5*, pp. 1256-1274.

Jusczyk, P., Hirsh-Pasek, K., Kemler Nelson, K.G., Kennedy, L.J., Woodward, A., & Piwoz, J. 1992. 'Perception of acoustic correlates of major phrasal units by young infants', *Cognitive Psychology, vol. 2*, pp. 252–29.

Kamide, Y., & Mitchell, D.C. 1997. 'Relative clause attachment: Non-determinism in Japanese parsing', *Journal of Psycholinguistic Research, vol. 26*, pp. 247-254.

Keller, F., & Zechner, K. 1995. 'A Connectionist Model of Lexical and Contextual Influences on Ambiguity Resolution in Human Sentence Processing', In

Proceedings of the 3rd Natural Language Processing Pacific-Rim Symposium, pp.592-597. Seoul.

Kennedy, A., Murray, W.S., Jennings, F., & Reid, C. 1989. 'Parsing complements: Comments on the generality of the principle of minimal attachment', *Language and Cognitive Processes, vol.4*, pp. 51-76.

Kinnear, P.R., & Gray, C.D. 2000. *SPSS for Windows made simple: Release 10.* Psychology Press Ltd.

Kjelgaard, M. M., & Speer, S. R. 1993. 'Prosodic structure, syntactic structure, and lexical preferences in the comprehension of temporary syntactic ambiguity', Presentation at the CUNY Sentence Processing Conference, Amherst, MA.

Kjelgaard, M.M., & Speer, S.R. 1999. 'Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity', *Journal of Memory and Language, vol. 40*, pp. 153-194.

Klatt, C. 1976. 'Linguistic uses of segmental duration in English: Acoustic and perceptual evidence', *Journal of the Acoustical Society of America, vol. 59*, pp. 1208-1221.

Kreiman, J. 1982. 'Perception of sentence and paragraph boundaries in natural conversation', *Journal of Phonetics, vol. 10*, pp. 163-175.

Kutas, M., & Hillyard, S.A. 1980. 'Reading senseless sentences: Brain potentials reflect semantic incongruity', *Science, vol. 207*, pp. 203-204.

Lehiste, I., & Wang, W.S.Y. 1976. 'Perception of sentence boundaries with and without semantic information', In Dressler,W., Pfeiffer,O. (ed) *Phonologica, vol. 19*, pp.277-283.

Lehiste, I. 1972. 'Timing of utterances and linguistic boundaries', *Journal of the Acoustical Society of America, vol. 51*, pp. 2018-2024.

Lehiste, I. 1973. 'Phonetic disambiguation of syntactic ambiguity', *Glossa, vol. 7*, pp. 107-122.

Lehiste, I. 1979. 'Perception of sentence and paragraph boundaries', In Lindblom, B. & Ohman, S. (ed) *Frontiers of Speech Communication Research*, New York: Academic Press.

Lehiste, I., Olive, J.P., & Streeter,L.A. 1976. 'The role of duration in disambiguating syntactically ambiguous sentences', *Journal of the Acoustical Society of America, vol. 60*, pp. 1199-1202.

Liberman, A., Cooper, F., Shankweiller, D., & Studdart-Kennedy, M. 1967. 'The perception of the speech code', *Psychological Review, vol. 74*, pp. 431-461.

Liberman, M., & Prince, A. 1977. 'On stress and linguistic rhythm', *Linguistic Inquiry, vol. 8.2*, pp. 249-336.

Luce, P.A., & Pisoni, D.B. 1986. 'Trading Relations, Acoustic Cue Integration, and Context Effects in Speech Perception', In Schouten, M.E.H. (ed.) *The Psychophysics of Speech Perception*. Dordrecht: M. Nijhoff.

Lujan, M. 1985. 'Stress and binding of pronuns', *Papers of the Twenty-first Regional Meeting*, Chicago: Chicago Linguistic Society, pp. 248-262.

McClelland, J.L. 1987. 'Interaction in Language Processing', In Carlson and Tanenhaus (eds) *Psycholinguistics LI144*, Stanford University Linguistic Institute.

McClelland, J.L., St John, M., & Taraban, R. 1989. 'Sentence comprehension: A parallel distributed processing approach', *Language and Cognitive Processes, vol. 4*, SI 287-335.

MacDonald, M.C. 1993. 'Probabilistic constraints and syntactic ambiguity resolution', *Language and Cognitive Processes, vol. 9 (2).*

Macdonald, M.C. 1993. 'The interaction of lexical and syntactic ambiguity', *Journal of Memory and Language, vol. 32*, pp. 692-715.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. 1994. 'The lexical nature of syntactic ambiguity resolution', *Psychological Review, vol. 101*, pp. 676-703.

Maeda, S. 1976. *A characterization of American English intonation*, Ph.D. thesis, MIT, Cambridge.

Mann, V.A., & Ripp, B.H. 1980. 'Influence of vocalic context on perception of the [sh]-[s] distinction', *Perception and Psychophysics, vol. 28*, pp 213-228.

Marslen-Wilson, W.D., & Tyler, L.K. 1975. 'Processing structure of sentence perception', *Nature, vol. 257*, pp. 784-786.

Marslen-Wilson, W.D., & Tyler, L.K. 1987. 'Against modularity', In Garfield, J.L. (ed) *Modularity in Knowledge Representation and Natural Language Understanding*, Massachusetts: MIT Press.

Marslen-Wilson, W.D. 1973. 'Linguistic structure and speech shadowing at very short latencies', *Nature, vol. 244*, pp. 522-523.

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

Marslen-Wilson, W.D., Tyler, L.K., Warren, P., Grenier, P., & Lee, C.S. 1992. 'Prosodic effects in minimal attachment', *Quarterly Journal of Experimental Psychology, vol. 45A*, pp. 73-87.

Massaro, D.W., & Cohen, M.M. 1983. 'Evaluation and integration of visual and auditory information in speech perception', *Journal of Experimental Psychology: Human Perception and Performance, vol. 8*, pp. 753-771.

Mattingly, I.G., Liberman, A.M., Syrdal, A.K., & Hayes, T. 1971. 'Discrimination in speech and nonspeech modes', *Cognitive Psychology, vol. 2*, pp. 131-157.

Maynell, L. 1999. 'Effect of pitch accent placement on resolving relative clause ambiguity in English', Poster presented at the 12th Annual CUNY Conference on Human Sentence Processing.

Meyer, M, Friederici, A.D., & von Cramon, D.Y. 2000. 'Neurocognition of auditory sentence comprehension: Event related fMRI reveals sensitivity to syntactic violations and task demands', *Cognitive Brain Research, vol. 9.1*, pp. 19-33.

Miller, G. A. 1956. 'The magic number seven, plus or minus two: Some limits of our capacity for processing information', *Psychological Review vol. 63*, pp. 81-93.

Mitchell, D. C. & Brysbaert, M. 1998. 'Challenges to recent theories of cross-linguistic variation in parsing: Evidence from Dutch', In Hillert, D. (ed.) *Sentence Processing: A Cross-linguistic Perspective*. New York: Academic Press.

Mitchell, D.C., Corley, M.M.B., & Garnham, A. 1992. 'Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism', *Journal of Experimental Psychology: Learning, Memory and Cognition, vol. 18*, pp. 69-88.

Monrad-Krohn, G.H. 1947. 'Dysprosody or Altered Melody of Language', *Brain, vol. 70*, pp. 405-415.

Monrad-Krohn, G.H. 1963. 'The third element of speech prosody and its disorders', In Halpern, L (ed.) *Problems of Dynamic Neurology*, pp. 101-117, Jerusalem: Hebrew UP.

Nagel, H., Shapiro, L. & Nawy, R. 1994. 'Prosody and the processing of filler-gap sentences', *Journal of Psycholinguistic Research, vol. 23*, pp. 473-485.

Nagel, H., Shapiro, L., Tuller, B., & Nawy, R. 1995. 'Prosodic influences on the processing of attachment ambiguities', *Journal of Psycholinguistic Research, vol. 25.2*, pp. 319-343.

Nazzi, T., Kemler-Nelson, D. G., Jusczyk, P. W., & Jusczyk. A. 2000. 'Six month olds' detection of clauses embedded in continuous speech: Effects of prosodic wellformedness', *Infancy, vol. 1*, pp. 1-30.

Nespor, M.A. & Vogel, I. 1983. 'Prosodic structure above the word', In Cutler, A. & Ladd, D.R. (ed) *Studies in Sentence Perception*, New York: Wiley.

O'Shaughnessy, D. 1979. 'Linguistic features in fundamental frequency patterns', *Journal of Phonetics, vol.7* , 119-145.

Osterhout, L., & Holcomb, P. J. 1992. 'Event-related brain potentials elicited by syntactic anomaly', *Journal of Memory and Language, vol. 31*, pp. 785-806.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A.D. 2005. 'Prosody-driven sentence processing: an event-related brain potential study', *Journal of Cognitive Neuroscience, vol. 15*, pp: 407-421.

Pell, M., & Baum, S. 1997a. 'The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults', *Brain and Language, vol. 57*, pp. 80-99.

Pell, M., & Baum, S. 1997b. 'Unilateral brain damage and the acoustic cues to prosody: Are prosodic comprehension deficits perceptually based?', *Brain and Language, vol. 57*, pp. 195-214.

Perkins, J.M., Baran, J.A., & Gandour, J. 1996. 'Hemispheric specialisation in processing intonation contours', *Aphasiology, vol. 10.4*, pp. 343-362.

Pierrehumbert, J.B., & Hirschberg, J. 1990. 'The meaning of intonational contours in the interpretation of discourse', In Cohen, J.M.P & Pollack, M. (eds) *Intentions in Communication*. Massachusetts: MIT Press.

Pierrehumbert, J.B. 1980. *The Phonology and Phonetics of English Intonation*. Doctoral dissertation, MIT.

Pijper, J. R. de, & Sanderman, A. A. 1994. 'On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues', *Journal of the Acoustical Society of America, vol. 96*, pp. 2037-2047.

Pitt, M.A., & Samuel, A.G. 1995. 'Lexical and sublexical feedback in auditory word recognition', *Cognitive Psychology, vol. 29*, pp. 149-188.

Price, P.J., Ostendorf, M., Shattuck-Hafnagel, S., & Fong, C. 1991. 'The use of prosody in syntactic disambiguation', *Journal of the Acoustical Society of America, vol. 90*, pp. 2956- 2970.

Pynte, J., & Prieur, B. 1996. 'Prosodic breaks and attachment decisions'. *Language and Cognitive Processes, vol.11.1-2*, pp. 165-191.

Rayner, K., Pacht, J.M., & Duffy, S.A. 1994. 'Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations', *Journal of Memory and Language, vol.33*, pp. 527-544.

Raynor, K., & Frazier, L. 1987. 'Parsing temporarily ambiguous complements', *Quarterly Journal of Experimental Psychology, vol. 39A*, pp. 657-673.

Raynor, K., Carlson,M., & Frazier, L. 1983. 'The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences', *Journal of Verbal Learning and Verbal Behaviour, vol. 22*, pp. 358-274.

Raynor, K., Sereno, S.C., Morris, R.K., Schmauder, A.R., & Clifton, C. 1989. 'Eye-movements and on-line language comprehension processes', *Language and Cognitive Processes, vol. 3/4*, pp. 21-49.

Riesbeck, C., & Schank, R.C. 1978. 'Comprehension by computer: expectation-based analysis of sentences in context', In W. J. M. Levelt and G. B. Flores D'Arcais (eds.), *Studies in the Perception of Language*. Chichester: Wiley.

Ross, E., & Mesulam, M. 1979. 'Dominant language functions of the right hemisphere?: Prosody and emotional gesturing', *Archives of Neurology, vol. 36*, pp. 144-149.

Schafer, A., Carlson, K., Clifton, C.Jr., & Frazier, L. 2000. 'Focus and the interpretation of pitch accent: disambiguating embedded questions', *Language and Speech, vol. 43.1*, pp. 75- 105.

Schafer, A.J. 1997. *Prosodic Parsing: The Role of Prosody in Sentence Comprehension*. University of Massachusetts Doctoral Dissertation. Forthcoming, Amherst.

Schafer, A.J., Carter, J., Clifton, C., Jr & Frazier, L. 1996. 'Focus in relative clause construal', *Language and Cognitive Processes, vol. 11*, pp. 135- 163.

Schaffer, D. 1983. 'The role of intonation as a cue to turn taking in conversation', *Journal of Phonetics vol.11*, pp. 243-257.

Schaffer, D. 1984. 'The role of intonation as a cue to topic management in conversation', *Journal of Phonetics vol. 12*, pp. 327-344.

Schepman, A, & Rodway, P. 2000. 'Prosody and on-line parsing in co-ordination structures', *The Quarterly Journal of Experimental Psychology: A, vol. 53 (2)*, pp. 377-396.

Schepman, A.H.B.C. 1997. *Prosody and On-Line Parsing*. D.Phil Thesis, University of Sussex. Downloading available from: http://www.ling.ed.ac.uk/~astrid/thesisdownload.html

Selkirk, E.O. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press

Shadle, C.H., & Scully, C. 1995. 'An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences', *Journal of Phonetics, vol. 23*, pp. 53-66.

Shapiro,L. & Nagel, H. 1995. 'Lexical properties, prosody, and syntax: implications for normal and disordered language', *Brain and Language, vol. 50*, pp. 240-257.

Shattuck-Hafnagel, S. & Turk, A.E. 1996. 'A prosody tutorial for investigators of auditory sentence processing', *Journal of Psycholinguistic Research, vol. 25.2*, pp. 193-241.

Sluijter, A. 1995. *Phonetic Correlates of Stress and Accent.* HIL dissertations 15, Leiden University.

Sluijter, A.M.C., & van Heuven, V.J. 1996. 'Spectral balance as an acoustic correlate of linguistic stress', *Journal of the Acoustical Society of America, vol. 100 (4),* pp. 2471-2485.

Sluijter, A.M.C., van Heuven, V.J., & Pacilly, J.J.A. 1996. 'Spectral balance as a cue in the perception of linguistic stress', *Journal of the Acoustical Society of America, vol. 101 (1),* pp. 503-513.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P.W. 2003. 'The prosodic bootstrapping of phrases: Evidence from prelinguistic infants', *Journal of Memory and Language, vol. 49*, pp.249-267.

Sonntag, G.P., & Portele, T. 1998a. 'PURR – a method for prosody evaluation and investigation', *Journal of Computer Speech and Language, vol. 12.4*, pp. 437-451.

Sonntag, G.P., & Portele, T. 1998b. 'Comparative evaluation of synthetic prosody with the PURR method', *Proceedings of the ICSLP, Sydney, vol. 5*, pp. 1715-1718.

Speer, S.R., Crowder, R.G., & Thomas, L.M. 1993. 'Prosodic structure and sentence recognition', *Journal of Memory and Language, vol. 32*, pp. 336-358.

Speer, S.R., Kjelgaard, M.M., & Dobroth, K.M. 1996. 'The influence of prosodic structure on the resolution of syntactic closure ambiguities', *Journal of Psycholinguistic Research, vol. 25*, pp. 249-272.

Speer, S.R., Shih, C.L., & Slowiaczek, M. 1989. 'Prosodic structure in language understanding: evidence from Tone Sandhi in Mandarin, *Language and Speech, vol. 32.4*, pp. 337-354.

Speer, S.R., Warren, P., Schafer, A.J., White, S.D., & Kneale, J. 1999. Situational constraints on the prosodic resolution of syntactic ambiguity. *Proceedings of the Fourteenth International Congress of Phonetic Sciences.* 1-7 August 1999. Berkeley, CA. 1301-1304.

Steedman, M. J. 1991. 'Structure and intonation', *Language, vol. 67*, pp. 260-296.

Steinhauer, K. 2003. 'Electrophysiological correlates of prosody and punctuation', *Brain and Language, vol. 86.1*, pp. 142-164.

Steinhauer, K., Alter, K., & Frederici, A.D. 1999. 'Brain potentials indicate immediate use of prosodic cues in natural language processing', *Nature Neuroscience, vol. 2.2*, pp. 191-196.

Stirling, L., & Wales, R. 1996. 'Does prosody support or direct sentence processing?', *Language and Cognitive Processes, vol. 11*, pp. 193-212.

Streeter, L.A. 1978. 'Acoustic determinants of phrase boundary perception', *Journal of the Acoustical Society of America, vol. 64*, pp. 1582-1591.

Swerts, M., & Geluykens, R. 1994. 'Prosody as a marker of information flow in spoken discourse', *Language and Speech, vol. 37.1*, pp. 21-43.

Swinney, S. 1979. 'Lexical access during sentence comprehension: (Re) consideration of context effects,' *Journal of Verbal Learning and Verbal Behaviour, vol. 18*, pp. 645-660.

Tabossi P., Colombo L., & Job R. 1987. 'Accessing lexical ambiguity: Effects of context and dominance', *Journal of Psycholiguistic Research, vol.49*, pp. 161-167.

Tanenhaus, M.K., Delll, G.S., & Carlson, G. 1987. 'Context effects in lexical processing: A connectionist approach to modularity', In Garfield, J.L. (ed.) *Modularity in Knowledge Representation and Natural-Language Understanding*, Cambridge, MA: MIT Press, pp. 83-108.

Terken, J., & Nooteboom, S. 1987. 'Opposite effects of accentuation and deaccentuation on verification latencies for given and new information', *Language and Cognitive Processes, vol. 2*, pp. 145-163.

Toepel, U., & Alter, K. 2002. 'Cerebral strategies in the segmentation and interpretation of speech', *Proceedings of Speech Prosody 2002*, Aix-en-Provence, pp. 71-76.

Tyler, L.K., & Marslen-Wilson, W.D. 1977. 'The on-line effects of semantic context on syntactic processing', *Journal of Verbal Learning and Verbal Behavior, vol. 16*, pp. 683-692.

Tyler, L.K., & Marslen-Wilson, W.D. 1982 'Processing utterances in discourse contexts', *Journal of Semantics, vol. 1*, pp. 297-314.

Umeda, N. 1982. 'Boundary: Perception and acoustic properties and syntactic and statistical determinants', *Speech and Language, vol. 7*, pp. 333-371.

Van Lancker, D., & Fromkin, V.A. 1973. 'Hemispheric specialisation for pitch and 'tone': Evidence from Thai', *Journal of Phonetics, vol. 1*, pp. 101-109.

Wambacq, I.J., & Jerger, J.F. 2004. 'Processing of affective prosody and lexical-semantics in spoken utterances as differentiated by event-related potentials', *Brain Research and Cognitive Brain Research, vol. 20*, pp: 427-437.

Wales, R., & Toner, H. 1979. 'Intonation and ambiguity', In Cooper, W.E. & Walker, E.C. (ed) *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, Erlbaum, Hillsdale.

Wanner, E. 1980. 'The atn and the sausage machine: Which one is baloney', *Cognition, vol. 8.*

Warren, R.M. 1970. 'Restoration of missing speech sounds, *Science, vol. 167*, pp.392-393.

Warren, P. 1985. *The temporal organisation and perception of speech.* Unpublished PhD dissertation, University of Cambridge, Cambridge.

Warren, P., Nolan, F., Grabe, E., & Holst, T. 1995. 'Post-lexical and prosodic phonological processing', *Language and Cognitive Processes, vol. 10*, pp. 411-417.

Watt, S.M., & Murray, W.S. 1996. 'Prosodic form and parsing commitments', *Journal of Psycholinguistic Research, vol. 25.2*, pp. 291- 317.

Whitney, P. 1998. *The Psychology of Language.* Boston: Houghton Mifflin Company.

Wingfield, A., & Klein, J.F. 1971. 'Syntactic Structure and Acoustic Pattern in Speech Perception'. *Perception & Psychophysics, vol. 9*, pp. 23-25.

Zurif, E.B. 1974. 'Auditory Lateralisation: Prosodic and syntactic factors', *Brain and Language, vol. 1*, pp. 391-404.

APPENDIX A

DO NOT FILM