THE EFFECT OF NOISE ON SEGMENTAL AND PROSODIC TIMING IN SPEECH PRODUCTION

by

LOUISE BRADFORD

B.Sc., The University of British Columbia, 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(School of Audiology and Speech Sciences)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1995

© Louise Bradford, 1995
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **Audiology + Speech Sciences**

The University of British Columbia

Vancouver, Canada

Date **Apr. 24, 1995**

DE-6 (2/88)
ABSTRACT

The present investigation examined the effects of noise on prosodic and segmental timing in speech production, focusing on preboundary lengthening, pausing, speaking rate and trends in segmental duration. Preboundary lengthening was examined using normalized duration measurements. Four subjects participated in this study. Each read syntactically ambiguous sentences to a listener under conditions of quiet and multitalker babble. These sentences were then labeled by listeners according to a seven-point scale of perceived boundary strength (or break index). Each break index corresponded to a level of prosodic phrasing proposed in the literature.

In general, speakers were highly individualistic in the ways they altered the segmental components of speech in noise. Three of four subjects decreased their rates of speaking in noise to varying degrees. The greatest changes in segmental durations were observed in vowels and sonorants. Changes in consonant durations were more subject specific.

Both the distribution and durations of pauses were characterized by a large degree of intersubject variability. The number of pauses tended to increase with increasing perceptual boundary strength, and pauses at the larger phrase boundaries were significantly longer than those at lower levels of phrasing. Preboundary lengthening was found to increase rapidly across three lower levels of phrasing, while almost no
differences were observed between the two highest levels of phrasing across both speaking conditions. This suggests that preboundary lengthening is an important perceptual cue to the lower levels of phrasing, which are marked by few other phonetic cues.

The effect of noise on overall timing relationships between prosodic constituents was minimal. The degree of preboundary lengthening remained relatively constant across speaking conditions for all four subjects participating in the experiment. It is proposed that speakers try to maintain prosodic timing relationships in order to preserve speech intelligibility in noise.
Table of Contents

Abstract ii
List of Figures vi
List of Tables vii
Acknowledgements viii
1. Introduction 1
2. Literature Review 5
   2.1 Introduction 5
   2.2 The Prosodic Hierarchy 8
      2.2.1 A Description of the Hierarchy 10
   2.3 Alternative Approaches to Phrasal Phonology 14
      2.3.1 Recursive Structures 14
      2.3.2. Tonal structures 15
   2.4 An Amalgamation of Hierarchies 16
   2.5 The Mapping Between Syntax and Phonology 18
   2.6 Timing Phenomena and Linguistic Structure 18
      2.6.1 Preboundary Lengthening and Pausing 18
      2.6.2 The Relative Importance of Prosodic Cues to Structure 21
      2.6.3 Isochrony 23
      2.6.4 Prosodic Labeling 24
      2.6.5 Phonetic Cues to Prosodic Structure 25
   2.7 Successful Communication 29
      2.7.1 Investigations of Speech in Noise and Related Studies 31
   2.8 Hypotheses 40
3. Methods 42
   3.1 Subjects 42
List of Figures

Figure 1. The metrical organisation of an utterance, with labelled prosodic constituents. 9

Figure 2. Representation of Prosodic Constituency (modified from Beckman & Edwards, 1992). 17

Figure 3. Percentage of change in segmental durations for each subject in noise, with reference to durations observed in the quiet condition. 60

Figure 4. Break Index versus mean normalized duration for (a) vowel nucleus in the final syllable before a boundary, (b) nonvocalic sonorants in coda position; subject 2B1. 70

Figure 5. Break Index versus mean normalized duration for (a) vowel nucleus in the final syllable before a boundary, (b) nonvocalic sonorants in coda position; subject 2A1. 71

Figure 6. Break Index versus mean normalized duration for (a) vowel nucleus in the final syllable before a boundary, (b) nonvocalic sonorants in coda position; subject 1B1. 72

Figure 7. Break Index versus mean normalized duration for (a) vowel nucleus in the final syllable before a boundary, (b) nonvocalic sonorants in coda position; subject 1A2. 73

Figure 8. Frequency plots of pause durations for the quiet and noisy conditions for subjects 2A1 and 1B1. 75

Figure 9. Frequency plots of pause durations for the quiet and noisy conditions for subjects 2B1 and 1A2. 76

Figure 10. Distribution of pauses by break index and speaking condition. (a) 1A2 (n=16 in quiet, n=17 in noise), (b) 1B1 (n=6 in both quiet and noise). 78

Figure 11. Distribution of pauses by break index and speaking condition.
(a) 2B1 (n=13), (b) 2A1 (n=26). 79

Figure 12. Distribution of pause lengths across break indices; (a) 1A2, (b) 1B1. 82

Figure 13. Distribution of pause lengths across break indices; (a) 2B1, (b) 2A1. 83
List of Tables

Table 1. Counterbalancing of materials and speaking conditions (quiet vs. noise) across subjects. 47
Table 2. Average speaking rates in words per minute. 54
Table 3. Average articulation rates in syllables per second. 55
Table 4. The distribution of speech sounds in the stimulus materials. 56
Table 5. Mean durations and standard deviations for the speech sound categories in quiet and noise. 58
  The percentage of change in the mean, with reference to the duration observed in quiet, is also provided. 58
Table 6. Percent change in mean duration of vowels and consonants, with reference to duration observed in quiet. 61
Table 7. Distribution of break indices across speaking conditions. 65
Table 8. Percentage of each type of boundary that is marked with a pause. 77
Acknowledgements

This study could not have been completed without the participation of students and faculty in the School of Audiology and Speech Sciences. I would like to thank all those who gave their time to participate as listeners and/or labelers: John Gilbert, Carolyn Johnson, Sharon Lenz, Michelle McPhee, Carole Nonis, Sheryl Palm and Glynnis Tidball. Thank you also to John Nicol and Glynnis Tidball for providing much needed instruction on the NeXT computer. I would also like to thank Monica Sanchez, in the department of Linguistics, for providing assistance in determining the syntax of the experimental sentences used in this study.

I would like to extend sincere thanks to John Gilbert for his enthusiasm for this project, and for his guidance over the course of its completion. I would also like to thank Kathy Fuller, Andre-Pierre Benguerel and Carolyn Johnson for providing input into the design of the experiment, editorial assistance, and advice regarding the analysis of data.

I would like to thank my family and friends for their patience and support over the course of this study. I would especially like to thank my husband, John Bradford, for countless hours of his own time spent assisting me on this project. Without his love and support, and without his unwavering encouragement and sense of humour, this project would never have been completed.
1. Introduction

Spoken language is highly structured. Speakers tend to group words into phrases, phrases into sentences or single utterances, and utterances into larger chunks of discourse (Wightman & Ostendorf, 1991). The phrasal structure of utterances is hierarchical, with each level of phrasing serving as the domain for a number of different prosodic phenomena and phonological rules. In recent phonological theory, this prosodic hierarchy forms part of the phonological representation of a language. It is related to the syntactic representation but is not completely identical to it (Selkirk, 1978).

Breaks in the flow of speech occur at the boundaries between prosodic phrases. Phonetic cues to these prosodic breaks include durational markers, such as insertion of a pause and/or lengthening of the syllable before the boundary, pitch changes, such as an abrupt pitch movement (boundary tone) or a pitch break, and an overall drop in amplitude (Streeter, 1978; Wightman & Ostendorf, 1991; Wightman et al., 1992). In general, the higher the phrase is in the hierarchy, the more phonetic cues a speaker provides, enabling these higher level boundaries to be more easily perceived by a listener. Furthermore, there is some evidence that duration cues at prosodic breaks reflect the hierarchical structure of an utterance. Both pause lengths and preboundary lengthening have been
found to increase in degrees for successively higher levels of structure. Preboundary lengthening may be particularly important for marking lower levels of structure, where fewer pitch cues tend to occur.

While speakers increase both the number of phonetic cues and the size of duration cues with increasingly higher levels of structure, listeners can use these prosodic boundary cues to parse utterances into component phrases. In some experimental situations, listeners have been asked to choose between two meanings of a syntactically ambiguous sentence, based on the meaning they thought the speaker intended. More recently, listeners have been asked to label boundaries within utterances with values that reflect each boundary's perceptual strength. Both trained and inexperienced listeners have been found to label prosodic breaks consistently and with good intersubject agreement (Price et al., 1991; Terken & Collier, 1992; Pijper & Sanderman, 1993). Boundary strength appears to be a useful perceptual cue to the hierarchical structure of spoken language.

Speakers desire to communicate successfully. Ideally, we would like always to speak in a quiet environment, but unfortunately, conditions are often much less than ideal. Noise or reverberation will reduce the intelligibility of a speaker's message. A speaker may find it difficult to make his/her message clear when talking to someone who is hearing impaired. In order to be a successful communicator, a speaker must be able to
adapt the content and structure of his/her message to meet communicative and situational demands (Cutler, 1987; Lindblom, 1990).

There are a number of ways in which speakers consistently change the phonetic structure of speech when speaking in noise, and they appear to make these changes in order to maintain intelligibility. Speaking rate is slowed and pauses are inserted. Vowels and, in some cases consonants, are lengthened (although some studies have reported shortening of consonants in noise (e.g. Junqua & Anglade, 1990)). Both pitch and loudness are consistently elevated in noise as a speaker increases his or her vocal effort. Very similar results have been found for speech produced with the specific intention to be clear. In general, speech produced in noise and clear speech have been found to be significantly more intelligible to listeners than speech produced in a quiet environment with no effort to speak clearly.

While many studies have investigated the effect of noise on overall trends in the production of prosodic features, few have examined the way in which a speaker marks the phrasal structure of his/her message in the presence of noise. Since both pitch and loudness are elevated with increased vocal effort in the presence of noise, boundary tones and changes in amplitude may become less useful cues to prosodic boundaries in noise. It may be that a speaker has the greatest flexibility in using durational cues to enhance
prosodic structure in noise. Furthermore, it has been shown that duration is an effective perceptual cue to structure when pitch cues are ambiguous (Streeter, 1978; Beach, 1991).

The primary purpose of this study is to examine the way in which speakers mark prosodic structure in speech, and the effects of noise on how they mark that structure, looking particularly at the prosodic boundary phenomena of (1) preboundary lengthening and (2) pausing. A second goal of this study is to determine the effects of noise on speaking rate and overall trends in segmental lengthening (both vowels and consonants).
2. Literature Review

2.1 Introduction

Prosody is a term that is frequently used to describe a set of phonetic parameters that marks patterns of phrasing in speech. It is often said to consist of the features pitch, length and loudness (Cruttenden, 1986), although terminology may vary across researchers (cf. Lehiste, 1970; Ladefoged, 1982). Prosodic features are generally considered to be "suprasegmental", because they tend to affect the overall structure of utterances rather than just single segments. The realisation of individual segments is affected significantly by prosody, however (Beckman & Edwards, 1992). Vowels, for example, are longer when they occur in stressed syllables or in the final syllable before a pause, than they are elsewhere.

For many years the phonetic representation of an utterance was understood to reflect syntactic structure. While in many instances the two do coincide, they often do not. For instance, the sentence below is a frequently quoted example in which the syntactic form (example 1) and the phonetic form (example 2) differ significantly.
1) [This is [the cat that caught [the rat that stole [the cheese]]]].

Phonetically, this sentence is typically divided into three phrases as follows:

2) [This is the cat] [that caught the rat] [that stole the cheese].

Within the Standard Generative Theory of grammar (Chomsky & Halle, 1968) differences between syntactic and phonetic structures were accounted for by first altering the surface syntax through a series of readjustment rules, and then by adding a number of boundary elements to the linear string of segments in the phonological representation. These boundary elements defined the domains of application of phonological rules which could not be accounted for syntactically. The phonological component of the grammar was viewed as a type of interpretive device, projecting syntactic representations onto phonetic forms.

More recent non-linear phonological theories view prosody not simply in terms of common phonetic parameters which make syntactic structure explicit, but rather as a hierarchy of constituents which are defined rhythmically (or metrically). The phonetic
representation of an utterance is determined directly from this prosodic structure, and there is no need to insert boundaries into the surface syntactic representation. Prosodic structure is an important component of the phonological representation of language, and it is the prosodic constituents which form the domains of phonological rules.

Although prosodic and syntactic structure may differ, they are related, and prosodic cues will enable listeners to parse sentences into syntactic constituents. In some situations, a speaker's utterance may be structurally or lexically ambiguous. By analysing the prosodic cues that a speaker provides, listeners may be able to choose between two or more competing hypotheses to determine which is the intended interpretation. Prosody also gives a listener predictive power, in that it enables him/her to determine whether a speaker has more to add, or whether he/she is finished speaking.

Speakers use prosody to highlight new or novel information in what they are saying, and to focus the listener's attention on the most important aspects of what is being said. A number of pragmatic factors, such as attitude, and choice of speech act, will also influence a speaker's use of prosodic cues.
Without prosodic information, speech sounds mechanical and monotonous (Larkey, 1980). In recent years, research into prosody has expanded greatly, with a view to improving the quality of synthesised speech and automatic speech recognition systems.

2.2 The Prosodic Hierarchy

In phrasal phonology, the phonological representation of a language has a hierarchically arranged organisation. Words are grouped into small phrases, which in turn are grouped into larger phrases, in a hierarchy consisting of several levels (Selkirk 1978, 1980a,b).

The foundation on which phrasal phonology is built is the metrical theory of stress assignment (Liberman & Prince, 1977). In metrical theory, the relative prominence observed between syntactic constituents, and between syllables within words, is represented in binary branching trees. These trees consist of sister nodes alternately labelled strong and weak in relation to each other, but which have no independent meaning. In Liberman and Prince's (1977) original proposal, the phrases represented in metrical trees were seen as being identical to syntactic constituents at and above the word level.
Selkirk (1978, 1980a,b) was one of the first to elaborate on the ideas proposed by Liberman and Prince (1977). Although she agreed with Liberman and Prince in many respects, she felt that their theory could be improved upon. Specifically, she argued that metrical trees and syntactic trees were not equivalent. She also argued that each level of the hierarchical tree of an utterance could be labelled as a particular prosodic constituent. Although these prosodic constituents could be isomorphic to syntactic constituents, they would have their own structure motivated by the assignment of relative stress to sister nodes in the tree.

![Figure 1](image)

Figure 1. The metrical organisation of an utterance, with labelled prosodic constituents.

A look at Figure 1 should illustrate several key points about prosodic structure. Clearly, the prosodic hierarchy consists of different types of prosodic constituents. At each level, an utterance is exhaustively parsed into constituents. The prosodic constituents are strictly organised into layers, so that each layer immediately dominates a sequence of constituents in the level below it. This is known as the Strict Layer Hypothesis, and it is fundamental to most theories of prosodic constituency (Selkirk 1978, 1980a, 1986; Nespor & Vogel, 1986). Structures in this hierarchy are not
embedded as in the syntactic hierarchy, resulting in a generally flatter hierarchy than that of the syntax.

In addition to playing a role in defining the rhythmic structure of an utterance, prosodic constituents form the domains of application of many phonological rules whose applications cannot be formulated in terms of syntactic constituents. The fact that they do form the domains for rules is the primary motivation for the existence of the hierarchy.

2.2.1 A Description of the Hierarchy

Selkirk (1978) and Nespor and Vogel (1983) outlined four levels within the prosodic hierarchy: the prosodic (or phonological) word, the phonological phrase, the intonational phrase and the utterance. Others include an additional level, the clitic group (Nespor & Vogel, 1986; Hayes, 1989).

The lowest level in the prosodic hierarchy is the prosodic word. This level of structure was first proposed by Liberman and Prince (1977) to prevent overgeneralization of phrasal stress assignment to word level. They termed this level the mot, and essentially, it consisted of a single lexical category word. In Selkirk's (1978) description of prosodic theory, she equated the prosodic word with the mot, and Hayes (1989) agrees that the prosodic word is always as large as the terminal element of a syntactic tree.
Others have argued that the prosodic word may form a domain smaller than a lexical category word (Nespor & Vogel, 1986). This level of the hierarchy forms the domain of numerous phonological rules, including rules of stress assignment and vowel harmony (Hayes, 1989).

The clitic group, consisting of a content word and an adjacent function word, is perhaps the most controversial level of the prosodic hierarchy. The function word behaves as if it were a single weak syllable, attaching itself to a sister constituent next to it and being phonologically dependent on it (Hayes, 1989). The clitic group has been postulated as a prosodic category in order to explain certain phonological rules which do not apply in any other context (Nespor & Vogel, 1986; Hayes, 1989). The necessity of defining a category "clitic group" is questioned by others (Selkirk, 1986; Booij, 1983), who contend that clitics actually form prosodic words.

The phonological phrase is thought to play a role in the temporal organisation of an utterance, with an influence on its rhythmic properties (Selkirk, 1978, 1984), its division into pauses (Gee & Grosjean, 1983), and on the lengthening of boundary syllables within the utterance (Nespor & Vogel, 1986). It is the domain of application of the rhythm rule in English, a rule that retracts stress, such as that of thirteen in the phrase thirteen men (Selkirk, 1978). Some rules of external sandhi of Italian also operate within the domain of the phonological phrase (Nespor & Vogel, 1983, 1986).
The most complex constituent of the prosodic hierarchy is the intonational phrase. Traditionally, it has been recognised that there is a level of phrasing smaller than the utterance, with phonetically distinct boundaries and a single most prominent point (Ladd, 1986). The boundaries of the intonational phrase are marked by preboundary lengthening of the final syllable, pausing, and boundary tones (Liberman, 1975). Like the lower levels of phrasing, the intonational phrase forms the domain of phonological rules. However, it is also generally considered to serve as the domain of the intonational contour (Selkirk, 1978; Nespor & Vogel, 1986).

Although intonational phrases often do not correspond to syntactic constituents, they are frequently associated with specific types of syntactic constructions, including parentheticals, nonrestrictive relative clauses, preposed adverbials, tag questions, and vocatives. Example 3 illustrates the phrasing of a preposed adverbial (taken from Selkirk, 1978):

3) In Pakistan, Tuesday is a holiday:

   [[[In Pakistan]PPh][[[Tuesday]PPh][is a holiday]PPh]]U
Boundaries between clauses and between subject and verb phrases are also commonly thought to be marked by intonation breaks (Hayes, 1989).

More than any other lower constituent, the intonational phrase is subject to semantic influences (Selkirk, 1984; Nespor and Vogel, 1986, Vogel and Kenesei, 1990). The single prominence within an intonational phrase can mark semantic focus, and enable a speaker to treat new information differently than old (or given) information.

Phonological and performance factors can affect the distribution of intonational phrases. For example, long syntactic phrases may be broken down into shorter, rhythmical or symmetrically spaced intonational phrases, resulting in the formation of phrases that are not syntactically viable (Gee & Grosjean, 1983; Nespor & Vogel, 1983, 1986; Hayes, 1989). Speech rate may also affect phrasing; fast, colloquial speech tends to have longer intonational phrases (or fewer pauses), while slow, pedantic speech would be more likely to consist of shorter intonational phrases (Nespor & Vogel, 1983, 1986).

The utterance is the largest prosodic constituent; consisting of one or more intonational phrases, and forming the bounding domain of some phonological rules. Although the utterance usually corresponds to a full syntactic sentence, there are
situations in which two relatively short syntactic sentences uttered by the same person, in rapid succession, form one prosodic utterance (Nespor & Vogel, 1986).

2.3 Alternative Approaches to Phrasal Phonology

2.3.1 Recursive Structures

Ladd (1986) described an alternative means of defining prosodic constituents. He pointed out that the intonational phrase has traditionally been defined in two distinct ways: 1) it is marked by phonetic breaks (pausing and prepausal lengthening), and (2) it is assumed to be the domain of the intonation contour, which is assumed to have a nucleus (or single most prominent point). The problem, Ladd (1986) points out, is that these two definitions often conflict. For example, in example (4), a single intonation contour corresponds to a single intonational phrase, as expected, while in (5), a single intonation contour is spread over two intonational phrases, as defined by pause structure (examples taken from Gussenhoven & Rietveld, 1992):

4)  [But we're not telling John]

5)  [But we're not going][John]
Ladd (1986) proposed that two separate prosodic constituents could be defined on the basis of the two traditional criteria. One constituent would be defined by phonetic breaks, including pauses, prepausal lengthening, and boundary tones. The other would be defined on the basis of the intonation contour. This system was later expanded to include four levels of structure, to allow for the construction of compound phrases (Ladd & Campbell, 1991).

Unlike the prosodic hierarchy described above, Ladd's system allows for the construction of recursive, embedded prosodic constituents. A constituent of one type could dominate a constituent of that same type. This system allows pauses within intonation contours, as well as sequences of contours without intervening pauses (Ladd, 1986).

2.3.2. Tonal structures

Beckman & Pierrehumbert (1986) argue for (at least) two constituents based on intonational criteria: the intonational phrase and the intermediate phrase. They base their argument on earlier work by Pierrehumbert (1980), who analysed intonation contours into linear strings of one or more pitch accents, a phrase accent and a boundary tone. A pitch accent is always produced at a rhythmically strong syllable. A boundary tone always stays at the phrase boundary, and the phrase accent fills the space between the last accent and the phrase boundary. Thus sentence intonation is viewed as being "a linear sequence
of discrete phonological events" (Ladd, 1992:322). Beckman & Pierrehumbert (1986) argue that the phrase accent trailing the last pitch accent is a boundary tone for a level of phrasing smaller than the intonational phrase, the intermediate phrase. The intonational phrase consists of one or more intermediate phrases and a full boundary tone. Under this analysis, example (5), a vocative tag, can be interpreted as consisting of two intermediate phrases.

2.4 An Amalgamation of Hierarchies

As the discussion above illustrates, prosodic constituents have been defined in terms of both the metrical organisation and the tonal properties of an utterance. Initially, it was not clear how these two types of structural entities fit together. In the recent literature, an effort has been made to conflate these two approaches into a single theory of phonological structure, under the assumption that constituents defined on tonal grounds can be fitted into the prosodic hierarchy. When the two approaches are merged, the intermediate phrase (Beckman and Pierrehumbert, 1986) is seen as being equivalent to the phonological phrase (Wightman et al., 1992).

Within this theory, the metrical tree, the foundation on which prosodic theory was built, is associated with a number of tiers (see Figure 2). Prosodic constituents are defined by "heads" and "edges". An edge marking occurs at the end of a phrase which
consists of one or more smaller phrases, and marks the boundary of this larger phrase. Edge markings include boundary tones, preboundary lengthening, and pauses. Constituents defined by head markings are less clearly defined than those marked by edges. They stand out as constituents within larger phrases because they have a single prominent. The result is a rhythm of alternately more and less prominent subunits, with one prominent small phrase occurring within each large phrase (Beckman & Edwards, 1992; cf. Ladd, 1986, described above).

Figure 2. Representation of Prosodic Constituency (modified from Beckman & Edwards, 1992).

The diacritic “*” is used to mark a tone that is associated with an accented syllable, while the diacritic “%” is used to mark a boundary tone.
2.5 The Mapping Between Syntax and Phonology

The discussion thus far illustrates that the phonology of a language plays a significant role in mediating between the syntax of a language and its phonetic representation. What is not as clear, however, is the nature of the relationship between the syntactic and prosodic representations. Edges and heads are also seen as having important roles in defining this relationship. One approach (Nespor and Vogel, 1986) favours a "relation-based" (Chen, 1990) mapping between the syntax and the phonology. This would consist of a set of mapping rules, usually making reference to the lexical heads of particular syntactic constituents, that group together grammatically-related items to form a unit of speech production. Another approach argues that prosodic constituents are constructed by reference to the edges of syntactic constituents (Selkirk, 1986). Chen (1990) argues that both types of mapping are evident in the construction of prosodic domains.

2.6 Timing Phenomena and Linguistic Structure

2.6.1 Preboundary Lengthening and Pausing

The duration of a speech segment will be influenced by a number of factors. Its phonological identity, for example, will largely determine its length, as will the phonetic environment in which it occurs. A speaker's rate of speech will affect the length of
individual phones within an utterance. Prosodic effects will also influence duration.
Stressed or prominent vowels, which form the heads of prosodic constituents, are longer
in duration than the same vowels when unstressed (Klatt, 1975; Umeda, 1975; Beckman
& Edwards, 1992). Segments, particularly vowels, which occur at the edges of phrases
(in the preboundary syllable) are lengthened in relation to other positions. This
preboundary lengthening has been reported before utterance, phrase and word boundaries

Prior to the development of prosodic theory, researchers assumed that
preboundary lengthening was syntactically determined, and some suggested that the
degree of lengthening would vary systematically with the height of a constituent in the
syntactic hierarchy (Cooper et al., 1978; Larkey, 1980). One way of testing this
hypothesis is by using syntactically ambiguous sentences, sentences consisting of the
same words but differing structurally. In an experiment by Cooper and his colleagues
(1978), preboundary syllable durations were measured in pairs of ambiguous sentences
with alternative surface structure bracketings. Preboundary syllables in the version with
the structurally higher boundary were longer and generally accompanied by longer pauses
than those at the lower level boundary. The authors concluded that "lengthening is
applied at a number of successively narrow domains of grammatical coding." (p. 174).
As Larkey (1980) points out, however, the evidence did not offer direct support for this
claim. With the technique used, relative lengthening within a single sentence could not be determined.

It has been documented that speakers tend to insert pauses at major syntactic constituent boundaries (Grosjean and Deschamps, 1975; Cooper et al., 1978; Lea, 1980; Macdonald, 1976). The majority of these are significantly longer than pauses found within constituents (Grosjean & Deschamps, 1975). Clearly, however, speakers also pause at nonsyntactic constituent boundaries, as illustrated above in examples (1) and (2).

One study of pausing in speech suggests that it is the phonological phrase structure rather than the syntactic structure that influences the distribution of pauses in speech. Gee & Grosjean (1983) found that the prosodic hierarchy provided a much better explanation than did the syntactic structure for pausing in an oral sentence reading task. Speakers tended to divide sentences into a number of relatively small, pause-defined structures that could be represented hierarchically.

A combination of pausing and preboundary lengthening seem to be perceptually significant. Martin (1971) found that listeners located pauses in connected speech most accurately when the pauses were accompanied by a lengthening of the preceding syllable,
and that errors (hearing a pause when none is present) occurred at sites of pre-boundary lengthening.

2.6.2 The Relative Importance of Prosodic Cues to Structure

As well as trying to determine how speakers mark syntactic constituents phonetically, researchers have also investigated whether or not listeners can use these cues, and if so, which cues are the most important perceptually. Lehiste (1973) performed an experiment in which speakers read sentences that included both deep and surface structure ambiguities. Sentences that differed at the level of surface structure were much more successfully disambiguated by listeners than those that differed at the deep structure level. Speakers were able to signal the differences in surface structure so that listeners could use this information successfully to locate phrase boundaries.

An acoustic analysis of the speakers' productions indicated that duration cues played an important role in the listeners' abilities to disambiguate between the two alternative meanings. A sequence of words containing a syntactic constituent boundary was usually longer than that same sequence without a constituent boundary. Insertion of a pause or period of laryngealization between constituents also occurred in sentences which were successfully disambiguated.
In some cases, speakers varied in the way they used timing cues (some shortened where others lengthened). Lehiste (1973) concluded that duration cues are the most important perceptually, and that listeners use other cues only when duration cues are not very informative. It is important to note, however, that listeners did have all of the prosodic cues that the speakers used available to them, and may have been using them in combination to locate constituent boundaries.

Later work investigating the relative importance of prosodic cues in disambiguation tasks used artificially manipulated sentences to control for differences in fundamental frequency, duration and amplitude. The results of these investigations showed that both pitch and duration are important cues perceptually, whereas amplitude is a much weaker cue (Streeter, 1978; Beach 1991). In general, pitch had its strongest influence on listener judgements when duration cues were relatively uninformative, while duration had its strongest influence when the pitch contour was relatively ambiguous. In general, however, duration was the stronger cue. Clear duration cues in the presence of ambiguous pitch cues enabled listeners to correctly disambiguate between the two alternatives more often than ambiguous duration values in the presence of clear pitch cues. When duration and pitch signalled contrastive structures, the majority of subjects chose the sentence realised by the duration information available to them (Streeter, 1978; Beach, 1991).
2.6.3 Isochrony

The duration of segments provides strong cues to the phrasal structure of English. There has been some disagreement in the literature, however, as to how listeners use these cues. English has been referred to as a "stress-timed" or "isochronous" language, because the onsets of stressed syllables in utterances tend to occur at nearly equal intervals in time (termed "feet") (Pike, 1945). Supporters of isochrony have argued that preboundary lengthening is perceptually important because it affects the timing of the onsets of these stressed syllables. In support of this claim, it has been shown that when a foot is lengthened artificially while fundamental frequency is held constant, listeners can disambiguate sentences successfully (Lehiste et al., 1976). Scott (1982), however, found that listeners were able to use either lengthening of an entire foot or lengthening of only the final segment in a foot to locate phrase boundaries.

A problem with this type of perceptual experiment is that artificially manipulated sentences are used to determine what it is that listeners actually do. Although listeners are able to make use of the timing information available, it has been shown that speakers actually lengthen the final vowel and all following consonants in preboundary syllables, rather than an entire foot (Wightman et al., 1992). Furthermore, syllable-final lengthening is not confined to stress-timed languages, suggesting that its role is not confined to providing listeners with information about foot length (Beckman, 1992).
2.6.4 Prosodic Labeling

The experimental work described above was designed to investigate how speakers represent syntactic structure phonetically, and whether or not listeners are sensitive to the acoustic-phonetic cues provided. Speakers were successful in using phonetic cues to signal structure, and listeners could use this information to locate phrase boundaries.

With the development of prosodic theory, it became apparent to some researchers that listeners might be able to identify several levels of prosodic structure by using acoustic-phonetic cues. It was hypothesised that speakers would provide more phonetic cues at higher levels of structure than at lower levels. This led to the development of the "break index" or "perceptual boundary strength" used to rank the degree of decoupling between words in an utterance (Price et al., 1991; Wightman et al., 1992; Pijper and Sanderman, 1993). Presumably, the more phonetic cues available to a listener, the more perceptually distant the two words would be.

Price and her colleagues (1991) devised a seven-point break index labelling scheme, representing the maximum number of phrasal categories proposed in the literature: 0 (boundary within a clitic group), 1 (normal word boundary), 2 (boundary marking a minor grouping of words, 3 (intermediate phrase boundary), 4 (intonational phrase boundary), 5 (boundary marking a grouping of intonational phrases) and 6
(sentence boundary). It was found that trained speech scientists could label sentences consistently and with good agreement between themselves (Price et al., 1991).

Since the development of this comprehensive labelling system, an effort has been made by a number of researchers to construct a common transcription and labelling scheme. The "Tones and Break Indices" (Silverman et al., 1992) labelling system was developed to encourage researchers to adopt a common notation, so that data could be more easily compared and reproduced. The final version of the TOBI system has not yet been published.

Less formal systems of labelling perceptual boundary strength have also been developed. Terken & Collier (1992) asked experienced speech scientists to locate prosodic boundaries within sentences, making distinctions between boundaries of different strengths (major or minor boundaries) if they so wished. Naive listeners have also been shown to label sentences consistently and with good intersubject agreement using a ten-point scale of boundary strength (Pijper and Sanderman, 1993).

2.6.5 Phonetic Cues to Prosodic Structure

To investigate the mapping from phonological to syntactic structure, Price and her colleagues asked four professional speakers to read phonetically-ambiguous, structurally-
different pairs of sentences. The sentences were then labelled with the seven point break index system described under section 2.6.4. Listeners were generally successful in disambiguating the sentences, choosing the correct context 84% of the time. When a pair of ambiguous sentences differed only by the value of break index, those with major breaks (break index greater than 4) were more reliably identified than those with smaller breaks. Not surprisingly, then, breaks between intonational phrases or larger structures were particularly salient to listeners. For smaller constituents, the value of the break index or the position of the largest break index was often the sole prosodic information available to listeners for disambiguation. In a few sentences, prominence provided listeners with the only cue to the structural contrast between each pair of sentences.

Prosodic breaks were often marked by preboundary lengthening and pausing. Preboundary lengthening showed a tendency to increase with increasing break index. That is, duration was affected by constituents at many levels in the hierarchy (cf. Cooper et al., 1978). When these results were re-evaluated to statistically account for the effect of speaking rate, this gradational lengthening of preboundary syllables reached significance for four levels of prosodic phrasing (break indices 0-1, 2, 3, 4-6) (Wightman et al., 1992). Although listeners could perceive differences in boundary size for the major break indices (4-6), duration differences at these levels were found not to be significant. It was the presence of other cues at these boundaries, including pausing and boundary tones, which increased the boundary strength. Pauses were frequently found at the
boundaries of intonational phrases and groupings of intonational phrases (Wightman et al., 1992).

Intonational cues also marked boundaries. Boundary tones occurred at intonational phrase boundaries and higher level constituent boundaries. Other intonational cues included pitch accents and changes in pitch range. The role of prominence appeared to be related more to contextual focus rather than syntactic structure, as it was not associated with specific syntactic structures in any systematic pattern (Price et al., 1991).

The major phrase boundaries in the study described above were clearly marked by a number of phonetic cues, whereas the minor boundaries were predominately marked by preboundary lengthening. Terken and Collier (1992) reached a similar conclusion; 94% of all major prosodic boundaries in sentences read by a single speaker were marked by a pause, preboundary lengthening and an intonation boundary marker. Unlike earlier experiments (cf. Cooper et al., 1978; Wightman et al., 1992), they found that pause length and preboundary lengthening were inversely related. Their preboundary lengthening measurements were sparse, however, and were confined to a single speech sound (schwa) in a limited number of sentences.
In an experiment performed by Pijper and Sanderman (1993), three distinct levels of boundary strength were identified (no boundary, weak boundary, and strong boundary), based on average values of naive listeners' judgements of boundary strength. The relationship between three phonetic cues (pitch contour, pausing, and declination reset) and the perceptual labels indicated that the more cues provided by the speaker, the greater the perceptual boundary strength was perceived to be. There was a tendency for pause length to increase with boundary strength, although this finding did not reach significance. Pauses were frequently accompanied by a pitch break. Of the three features studied, a break in the pitch pattern was found to be the only cue used systematically in isolation and only for lower values of boundary strength (Pijper & Sanderman, 1993).

In their experiment, Pijper and Sanderman (1993) did not investigate the role of preboundary lengthening. Listeners were actually able to make perceptually finer distinctions than the three average values indicate. Interestingly, at the "no boundary" level, listeners had actually assigned a small amount of boundary strength. Perhaps at lower levels of phrasing where pitch and pause cues were absent, or where a pitch discontinuity was the only cue, preboundary lengthening was influencing listener judgements.

Four distinct levels of preboundary lengthening were found by Ladd and Campbell (1991) in an analysis of a pre-recorded narrative. In this study, sentences were
labelled according to prominence and pausing cues, using the recursive hierarchy proposed by Ladd (1986), and later expanded by Ladd and Campbell (1991).

The results of these boundary strength studies mark a return to the issue of the importance of phonetic cues at phrase boundaries. Two conclusions seem readily apparent: 1) the more acoustic cues at a boundary, the greater the perceived boundary strength will be; and 2) the importance of different acoustic cues seems to depend to some extent on the type of prosodic boundary being marked (Wightman et al., 1992). Preboundary lengthening appears to play an important role at smaller prosodic phrase boundaries, whereas intonational cues and pausing are more significant in marking the major prosodic boundaries. Furthermore, there is some indication that preboundary lengthening, and to a lesser extent, pause lengths, reflect the hierarchical structure of the prosodic constituency.

2.7 Successful Communication

When speakers talk to each other, whether it is to share information or to establish a social interaction, their primary intention is that the interchange be successful. Speakers must make a number of decisions to ensure that communication is successful. For example, a speaker must make judgements about how much knowledge he/she and the listener share (the presuppositional pool), and vary his/her content accordingly. As Cutler
notes (1987:24) "speakers draw on their knowledge of what their listeners already know in choosing what to say and how to say it". The speaker makes lexical, semantic, syntactic and pragmatic decisions based on what he/she determines to be the needs of the listener.

In some situations communication becomes more difficult due to characteristics of the speaker (e.g. voice problems), the listener (e.g. hearing loss or inadequate grasp of the language), or the environment (e.g. noise). Whenever a speaking situation is less than ideal, the speaker must alter his/her message or style of speaking in some way in order to ensure that communication remains successful. When viewed in this way it becomes clear that speech production is a very dynamic, decision-filled process. A speaker who is unable to revise decisions or adapt to changing environments is bound to be much less capable of communicating effectively than a speaker who can.

Lindblom (1990) likens this dynamic process of speech production to that of natural selection. Speakers wish to ensure that their messages are as discriminable and intelligible as possible, so they adapt their speech production to meet the needs of the listener and to suit the demands of the environment. A speaker must make judgements about a listener's prior knowledge and need for specific signal information, and make any modifications necessary to ensure that his/her message is understood.
According to the "H & H" model of Lindblom (1990), speakers vary their output along a continuum of hyper- and hypospeech. When environmental demands (output constraints) are minimal, speakers default to an economical form of speech called hypospeech. As output constraints increase, as when speaking conditions are not ideal, then demands on the speaker are high. In this situation, speech must be modified to ensure that it is discriminable and intelligible.

2.7.1 Investigations of Speech in Noise and Related Studies

One situation in which the demands on the speaker are high is in the presence of background noise. Noise affects speech understanding because it masks speech produced by the speaker. Masking reduces the discriminability of speech for the listener, and forces the speaker to increase his/her vocal effort in order to communicate successfully. Potentially, vocal effort could increase to the point that speech sounds would be distorted (Rostolland, 1985). The phenomenon of a speaker raising his voice in the presence of noise, and then returning to its original level when the noise is removed is known as the Lombard effect, after Etienne Lombard, who first documented it in 1911.

Since Lombard's original discoveries, increases in vocal intensity in the presence of noise have been documented time and again (Hanley & Steer, 1949; Dreher & O'Neill, 1957; Pickett, 1958; Webster & Klumpp, 1962; Lane et al., 1970; Lane & Tranel, 1971;
Van Summers et al., 1988; and Young et al, 1993). In a comprehensive review of the literature on the Lombard effect, Lane and Tranel (1971) note that in experimental situations, the degree of increase appears to depend on whether or not there is a listener involved in the experimental procedure. When a listener is present and has an active role to play (for example, the listener must repeat the speaker's words or write them down with a high degree of accuracy), the increase in intensity has been shown to be as high as 5 dB for each 10 dB increase in noise (Webster & Klumpp, 1962; Lane et al., 1970). In situations in which there is no listener, and the speaker merely has to read words or sentences into a tape recorder, the increase in intensity has been found to be much lower (Dreher & O'Neil, 1957).

Unfortunately, the great majority of studies examining speech in noise do not include listeners, and consequently, the results may not truly reflect how speakers attempt to remain intelligible in the face of perceptual degradation by noise.

By increasing speech intensity in noise, a speaker increases the signal-to-noise ratio of his/her message. Since noise can degrade the message considerably by masking, an increase in signal-to-noise ratio should increase the intelligibility of the utterance. It has been shown that utterances produced in noise are, indeed, more intelligible in noise than amplitude-matched utterances produced in quiet (Dreher & O'Neil, 1957; Pickett,
An increase in the fundamental frequency of utterances spoken in noise is well documented. This effect has been produced in both continuous and short bursts of noise (Ladefoged, 1967; Rastatter & Rivers, 1983; Rivers & Rastatter, 1985; Van Summers et al., 1988; Young et al., 1993). The effect of noise on the variability of fundamental frequency is less clear. In some cases a decrease in variability has been documented (Ladefoged, 1967; Rastatter & Rivers, 1985), while other workers have found an increase (Rivers & Rastatter, 1983). These discrepancies may stem from differences in the level of noise used. Decreases in variability were found in studies which employed a very high level of noise, at 100 dB SPL or higher. These results are consistent with those of Pickett (1958), who found that talkers were shouting as loud as possible in 95 dB SPL white noise. At very high levels of vocal effort, in very loud noise, a speaker would have to increase laryngeal tension to such an extent that variability would be very restricted (Rostolland, 1985).

Mixed results are also found when examining the effects of noise on speech rate and syllable or word duration. An overall decrease in speaking rate and an increase in word duration has been documented by several researchers (Hanley & Steer, 1949; Dreher & O'Neill, 1957; Webster & Klumpp, 1962; Hansen, 1988; Van Summers et al., 1958; Van Summers et al., 1988), except at very high levels of vocal effort (Pickett, 1958).
Rubenovitch and Pastier (1938), however, found that speakers increased their rate of speaking somewhat under noise conditions, while Charlip & Burk (1969) and Young et al. (1993) found little or no change in speaking rate in the presence of noise. Many of these studies looked only at the overall duration of words or sentences. Potentially, changes in segmental lengthening could occur, but compensated for in such a way that the net change in speaking rate would be negligible. For example, Junqua & Anglade (1990) found that speakers tended to increase the duration of vowels but slightly decrease consonant duration, for an overall slight increase in word duration.

Speech produced at very high levels of vocal effort, or shouted speech, is less intelligible than speech produced at lower levels of effort. This result has been found for both normal-hearing (Rostolland, 1985; Tschopp & Beckenbauer, 1991) and hearing-impaired listeners (Tschopp et al., 1992), when signal-to-noise ratio is matched for each condition. In addition to a significant increase in the overall amplitude of the speech signal, a number of other acoustic-phonetic changes occur in shouted speech. Fundamental frequency is elevated but variability is reduced, producing a relatively flat intonation contour. Vowel duration increases significantly; consonants, however, are actually shortened by as much as 20% (Rostolland, 1985). Also, in shouted speech, there is a significantly greater difference between the intensity levels of vowels and most consonants (Tschopp et al., 1992), resulting in reduced discriminability of consonants.

1 Generally, when a number of different noise levels are used, a significant decrease in speaker rate is observed only between the quiet and noise conditions, and not between different noise levels.
When shouting, then, speakers are attempting to maintain intelligibility by producing speech at a much higher amplitude, but these gains are limited by the distortion of the speech signal.

A related series of experiments focuses on the characteristics of "clear speech" as compared to conversational speech. Clear speech is often elicited by specifically asking speakers to produce speech naturally, as in conversation, and clearly, as if they are trying to communicate in a noisy environment or to a hearing-impaired listener (Lindblom, 1990; Picheny et al., 1985; 1986; Payton et al., 1994). The results of these studies are similar to those found in noise studies; speakers are able to adjust their speech to become more intelligible. When speakers speak clearly, they are consistently more intelligible to hearing-impaired listeners than when they speak conversationally, regardless of the level of presentation or the frequency-gain characteristics of the signal (Picheny et al., 1983). In fact, for both normal-hearing and hearing-impaired listeners, clear speech is significantly more intelligible than conversational speech in quiet situations and in conditions of noise and/or reverberation. As listening conditions deteriorate, this intelligibility advantage becomes greater and greater (Bond & Moore, 1994; Payton et al., 1994).

When speaking clearly, speakers decrease their rate of speaking by inserting or lengthening pauses, and by lengthening individual speech segments. Picheny and his
colleagues (1986) point out that this decrease in speaking rate is not accomplished by uniform lengthening or stretching of each speech sound. Instead, the amount of lengthening is dependent on the characteristics of each individual phoneme and the acoustic environment in which it is found. Compared to conversational speech, both consonants and vowels are lengthened; plosive duration is substantially longer, while fricatives, nasals, and semivowels are also lengthened. Tense vowels are lengthened more than lax vowels (Picheny et al., 1986).

Cutler and Butterfield (1990) investigated durational changes in clear speech specifically at word boundaries. They found that speakers lengthened syllables immediately before a word boundary in clear speech, and that the amount of lengthening was significantly greater before a weak syllable (a word with a weak syllable onset or an unstressed word) than before a strong syllable. Furthermore, in clear speech, talkers paused in places where they had not paused in conversational speech, and pause lengths were highly correlated with the lengths of the preceding syllable. As a result of their findings, the authors conclude that, when speaking clearly, speakers can use durational changes such as increases in preboundary syllable duration and the insertion of pauses to emphasize word boundaries. This effect is greater for word boundaries which would be particularly difficult for listeners to perceive.
Other prosodic changes also occur in clear speech. Values of fundamental frequency show a tendency to increase somewhat in clear speech relative to conversational speech, and these values tend to be more variable (Picheny et al., 1986; Cutler & Butterfield, 1992). The amplitude of the speech signal also tends to increase consistently by a small amount (Picheny et al., 1986)

Speakers tend to naturally vary in their speaking habits, and, according to Bond & Moore (1994), some speakers "inadvertently" speak clearly while others do not. In a comparison of utterances produced by 5 speakers, the authors found that some speakers were more consistently intelligible than others in both noise and quiet conditions. The most intelligible speaker consistently used prepausal and sentence-final syllable lengthening, and was the only speaker who inserted pauses within sentences. He also used relatively long and clearly distinguished vowels.

Speakers tend to use at least some of the same phonetic modifications when they produce speech that they intend to be easily understood, either in clear speech or in noise. The increases in amplitude and fundamental frequency of the signal are much greater in noise than in clear speech, as talkers attempt to maintain a constant signal-to-noise ratio

---

2 In addition to prosodic changes, talkers tend not to apply phonological processes common to conversational speech to clear speech. For example, in clear speech, stops tend to be fully released, particularly in word final position (Picheny et al, 1986), and vowels tend to be fully realized rather than reduced (Picheny et al., 1986; Cutler and Butterfield, 1990, 1991).
in order to remain intelligible. A decrease in speaking rate and an increase in the number and length of pauses are common to both situations. Perceptually, these changes give listeners quite an advantage when listening conditions are adverse. Both a slower rate of speech and the presence of pauses allow a listener more time to process the incoming message. Unlike shouted speech, clear speech is characterized by a lengthening of both vowels and consonants. This is particularly advantageous for consonants, because an increase in duration increases the threshold of audibility of these sounds which are already more difficult to perceive (due to higher frequency, shorter durations, and lower amplitudes than vowels) (Cutler & Butterfield, 1991).

The results of the Junqua and Anglade (1990) study seem to suggest that there is a difference between speech produced in noise and clear speech for the duration of consonants. As in the case of shouted speech, they found that consonants decreased in duration. However, unlike most other noise studies, they found that intelligibility of words remained unchanged or decreased in noise. On close inspection of the increase in intensity of vocal effort extended by subjects, it becomes clear that while some speakers increased intensity by very little (by as little as 4 dB), others increased intensity dramatically (by as much as 24 dB). Some of the subjects participating in this study were evidently producing speech at a shout while others were not. As all of the results were presented as a whole (results from all subjects were combined), the overall report of a
reduction in consonant length may be due to the fact that several subjects were in fact shouting.

No studies have specifically examined the effects of noise on prosodic phrase structure, and in fact, several have concentrated primarily on the production of single words. Clear speech studies, however, indicate that speakers can highlight the structural aspects of sentences with preboundary lengthening and pausing when necessary, with a resulting increase in intelligibility (Cutler & Butterfield, 1990, 1991; Bond & Moore, 1994).

Since fundamental frequency and amplitude variation would be relatively restricted in noise due to an increased level of vocal effort, durational cues would allow speakers greater flexibility in enhancing boundaries than would fundamental frequency or amplitude. In noise, boundary tones may be less pronounced than they normally would be, due to the restricted range of the fundamental. They also may be easily masked, as would changes in amplitude. In these conditions of relatively ambiguous pitch and amplitude cues, listeners would be able to use the exaggerated timing cues to parse sentences into their component constituents; that is, they would be able to use those cues that have been shown to be the strongest perceptually in early studies investigating the importance of prosodic cues in speech.
2.8 Hypotheses

Hypothesis 1:

Speakers will not change their rates of speaking across quiet and noisy speaking conditions.

The alternative experimental hypothesis:

Speakers will slow their rates of speaking in noisy conditions relative to quiet, lengthening both vowels and consonants.

Hypothesis 2:

In both quiet and noisy conditions, vowel durations (in the rhyme of preboundary syllables) and pause durations will not significantly increase with increasing boundary strength, as labeled by experienced listeners.

The alternative experimental hypothesis:

In both quiet and noisy conditions, vowel durations (in the rhyme of preboundary syllables) and pause durations will increase with increasing boundary strength, as labeled by experienced listeners.
Hypothesis 3:

The degree of preboundary lengthening at phrase boundaries will not be significantly different across quiet and noisy speaking conditions.

The alternative experimental hypothesis:

The degree of preboundary lengthening at phrase boundaries will be significantly greater in noisy speaking conditions relative to quiet.

Hypothesis 4:

The number and duration of pauses will not be significantly different across quiet and noisy speaking conditions.

The alternative experimental hypothesis:

The number and duration of pauses will be significantly greater in noisy speaking conditions relative to quiet.
CHAPTER 3

3. Methods

The methodology of this study included five steps:

1) construction of syntactically ambiguous sentence pairs,

2) the subjects' task - reading pairs of syntactically ambiguous sentences under two experimental conditions (noisy and quiet) to a listener,

3) perceptual labeling of these sentences by three trained listeners,

4) phonetic labeling and analysis of duration of speech segments produced by the subjects, and

5) normalization of duration measurements.

3.1 Subjects

Eight participants, four women and four men, were recruited for this study. Because of the large amount of data generated, however, the speech produced by half the subjects was eliminated prior to analysis. The four remaining subjects were chosen at random from counterbalanced groupings. Three were male, and one was female, between
education (5-11 years). All subjects were native speakers of English (with Western Canadian dialects) and none reported having a history of speech or hearing problems.

3.2 Materials

The materials used in this study were inspired by those used in the study by Price *et al.* (1991). Twelve pairs of syntactically ambiguous sentences were constructed, with each sentence preceded by one or two sentences of disambiguating context (see Appendix A). Four structural ambiguities were represented, as described below.

**Group 1:** the “a” version represents two sentences joined by a coordinating conjunction, while the “b” version represents a verb and its complement (where the complement is a noun phrase or embedded sentence), as shown in example 4:

4) a) He will not discuss her or deal in real estate any longer

   b) He will not discuss her ordeal in real estate any longer

**Group 2:** the “a” version represents far attachment of a final phrase, while the “b” version represents near attachment of a final phrase. In example 5, the bracketed phrase modifies “the judge” in (a), whereas in “b”, the bracketed phrase modifies “the woman”:
5) a) The judge sentenced the woman [in a foul mood]

b) The judge sentenced the woman [in a foul mood]

**Group 3:** the "a" version represents left attachment of a middle phrase, while the "b" version represents right attachment of a middle phrase, as shown in example 6:

6) a) Beginning Monday, afternoon aerobics classes will be extended to one hour

b) Beginning Monday afternoon, aerobics classes will be extended to one hour

**Group 4:** the "a" version represents a parenthetical sentence or phrase, while the "b" version represents a nonparenthetical embedded sentence, as shown in example 7:

7) a) John knows everything, you know, about breeding birds

b) John knows everything you know about breeding birds

The first member of each pair (version "a") in groups 1 and 4 contains a syntactic break which occurs higher in the syntactic hierarchy than the corresponding break in the "b"
version. Groups 2 and 3 contain words or phrases that differ in their place of attachment (see Appendix B).

Two of these groups (2 and 3), the far vs. near and left vs. right attachment ambiguities, represent structural distinctions identical to two of those found in the study by Price et al. (1991). The remaining two, conjoined sentences vs. verb complements and parentheticals vs. embedded structures, have been modified somewhat. For three of the four groups, both sentences in a pair were comprised of identical words, but in some cases differed in comma placement. In the remaining group (conjoined vs. verb complements) the paired sentences consisted of the same set of phones, but were lexically as well as syntactically different, as was illustrated above in example 4.

For each category of paired structural representations, 3 pairs of sentences were constructed, ranging in length from 6 to 12 words.

3.3 Test Apparatus and Procedures

Subjects were asked to attend two sessions. In the first session, they were familiarized with the purpose of the experiment (see Appendix C), and participated in one of the experimental conditions, either the quiet or noisy condition. Subjects were presented with the written stimulus sentences in their contextual paragraphs, with the
target sentence italicized. Examples of target sentences in their contextual paragraphs are provided in examples 8 and 9.

8) Ted was relieved that his prize-winning guppies did not succumb to a mysterious fish disease. *Although the guppies survived, for a while they looked sick.*

The guppies never looked healthy and it wasn't long before they died. *Although the guppies survived for awhile, they looked sick.*

9) After receiving his fourth speeding ticket on the way to court that day, the judge gave a woman 10 years for shop lifting. *The judge sentenced the woman in a foul mood.*

As the judge announced his decision in court, the woman was still fuming about being arrested. *The judge sentenced the woman in a foul mood.*

The two versions of each sentence pair had been randomized across two lists ("A" list and "B" list), each consisting of half the sentences. Both lists were read in each session that subjects attended. To minimize the effects of practise, the order of sentences (list A vs. list B) and experimental condition (i.e. quiet vs. noise) were counterbalanced across subjects, as illustrated in Table 1. Subjects eliminated from the study due to the large amount of data generated are shown in brackets.
<table>
<thead>
<tr>
<th>Subjects</th>
<th>Session 1</th>
<th>Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Condition</td>
<td>List Order</td>
</tr>
<tr>
<td>(1A1),1A2</td>
<td>Quiet</td>
<td>AB</td>
</tr>
<tr>
<td>1B1,(1B2)</td>
<td>Quiet</td>
<td>BA</td>
</tr>
<tr>
<td>2A1,(2A2)</td>
<td>Noise</td>
<td>AB</td>
</tr>
<tr>
<td>2B1,(2B2)</td>
<td>Noise</td>
<td>BA</td>
</tr>
</tbody>
</table>

*Table 1. Counterbalancing of materials and speaking conditions (quiet vs. noise) across subjects.*

Each subject spent 20-30 minutes reading over the materials and becoming familiar with them. Subjects were aware that there were two versions of each sentence represented. Each subject was instructed that he/she would be reading aloud only the italicized target sentence in the experiment, and that the preceding paragraphs were provided to set each sentence in context.

Subjects were seated in a sound attenuating booth with a microphone placed approximately 12 inches from the lips. In both the quiet and noisy conditions, subjects were fitted with TDH-39P headphones, to ensure that the occlusion effect was held constant under both conditions. In the noisy condition, multitalker babble was linefed from a Yamaha Natural Sound KX-500U stereo cassette recorder into an OB802 speech audiometer, and presented binaurally through the headphones at a level of 65 dB HL. Subjects' speech was recorded on a Marantz PMD 420 stereo audiocassette recorder.
Each recording session included a listener, seated outside of the sound attenuating booth, to whom the subject read the sentences. Subjects were instructed that the listener would be trying to disambiguate the sentences, using the contextual paragraph given with each sentence. Subjects were told that the listener expected the sentences to be completely randomized, so that the same version could occur twice in one session. In fact, the listeners were knowledgeable about the task, and their results were never intended to be scored. In the noise session, subjects were informed that the listeners were listening in the same noisy conditions in which the subjects would be producing the sentences. Listeners occasionally gave auditory feedback to the subjects, to indicate that they had heard a given sentence. Visual feedback was minimal.

Subjects were instructed to read the sentences as naturally as possible, and that should they feel that their productions were not acceptable in some way, they were to say "repeat" or "I'll try again". Occasionally, a subject produced a sentence incorrectly (substituted, added or omitted words). In these situations, the incorrect sentences were rerecorded at a later date.

3.4 Perceptual Labeling

Two of the goals of this study were to examine the variation in preboundary lengthening and pausing with the phrasal structure of utterances, and to determine
whether this phrasal structure is more exaggerated in noise. Therefore, it was necessary that the prosodic phrasal structure of the utterances be known. In this study, phrasal structure was determined by means of the perceptual labeling scheme developed by Price et al. (1991), as described above in Chapter 2. This particular labeling scheme was adopted because it has been shown to be used consistently and with good agreement within and across labelers, and has also been shown to correlate very well with automatically labeled break indices.

The prosodic labeling scheme used requires labelers to label three characteristics of an utterance: the degree of juncture between each word in an utterance, or break index; type and location of boundary tones; and prominence (Appendix D). Break indices were labeled on a seven point scale, used to represent the levels of prosodic phrasing proposed in the literature: 0 (boundary within a clitic group), 1 (normal word boundary), 2 (boundary marking a minor grouping of words, 3 (intermediate phrase boundary), 4 (intonational phrase boundary), 5 (boundary marking a grouping of intonational phrases) and 6 (sentence boundary). Boundary tones were labeled as final falls, continuation falls, continuation rises, and question rises. Every syllable of an utterance was labeled with some degree of prominence on a three-point scale, including major phrasal prominence, lesser prominence and no prominence.
The subjects' utterances were labeled by three speech scientists. Two of the judges were faculty members in the School of Audiology and Speech Sciences at the University of British Columbia; the third judge was a recent graduate of the Masters of Science program offered in that School. Each judge labeled the utterances individually in a quiet room, listening to each sentence as many times as they felt it was necessary.

Following completion of the labeling task, results were compiled to produce a cumulative break index score for each word boundary, ranging from 0 to 18. This is a similar method of scoring to that employed by Allen (1972) in his study of stress patterns. For the purposes of this study, however, this method of scoring proved to be disadvantageous, because there were very few occurrences of several of the cumulative break index scores, resulting in discontinuities between break indices. An alternate method of scoring was adopted, in which break indices were assigned on the basis of agreement between labels assigned by at least two of the three labelers. The following modified break index scale was then adopted for this study: a score of “1” was assigned when at least two of three labelers agreed that a boundary was within a clitic group, “2” was assigned to normal word boundaries, “3” was assigned to a boundary marking a minor grouping of words, “4” was assigned to intermediate phrase boundaries, “5” was assigned to intonational phrase boundaries, and “6” was assigned to sentence boundaries. There were no cases in which two of the three listeners assigned an original break index
of “5” (grouping of intonational phrases). For this reason, this category was excluded from the modified break index scores.

3.5 Phonetic Labeling and Segmentation

The recorded sentences produced by the subjects digitized into sound files using the NeXT SoundWorks software program. These sound files were then analyzed with the Sonogram speech analysis program. Individual speech sounds within the sentences were segmented and phonetically labeled using a combination of aural repetitions, broadband and narrowband spectrograms, and waveform displays. Parameters used in the spectrographic analysis are provided in Appendix E.

The segmentation criteria used consisted of the segmentation strategies described by Peterson and Lehiste (1960), and later summarized by Shoup and Pfeifer (1976). One significant modification was made for the purposes of the present study; for each stop consonant, a single measurement was made, which included both the hold (or closure) and the release portions. For non-plosive releases (e.g. lateral releases), the non-plosive release portion of the signal was included in the following segment (Crystal & House, 1982). These segmentation strategies are briefly summarized in Appendix F. Pauses were defined as any silent interval greater than 5 ms between words, excluding silent intervals preceding word-initial stop consonants (cf. Picheny et al., 1986).
3.6 Normalization of Raw Duration Measurements

As noted in Chapter 2, the duration of a speech sound will be influenced by several factors. Each speech sound has an intrinsic duration (Peterson & Lehiste, 1960), determined largely by articulatory constraints, but by other factors as well, such as whether the syllable in which a phone occurs is stressed or not, and by the position of the speech sound in a word or utterance. When the same amount of lengthening is measured between a phone which has a short intrinsic duration and a phone which has a large intrinsic duration, the lengthening will be much more significant for the shorter segment. Raw durations are unable to account for these differences and prevent meaningful comparisons between different phones (Campbell, 1992).

Recently, a method of normalization for duration data has become popular in the literature (Price et al., 1989; Ostendorf et al., 1990; Wightman & Ostendorf, 1991; Campbell, 1992; Wightman et al., 1992). When raw scores are normalized, segmental durations are expressed in terms of "z-scores", or distances from the mean of all measurements of a single phone. This reduces the variance across duration data, and allows for comparisons across phones of different types. The equation used is as follows:

1) \[ z\text{-score} = \frac{d_{\text{token}} - \mu_{\text{type}}}{\sigma_{\text{type}}} \]
where \( d_{\text{token}} \) is the raw duration of a particular phone, \( \mu_{\text{type}} \) is the mean duration determined from all measurements of that particular phone (calculated for each subject under each listening condition), and \( \sigma_{\text{type}} \) is the standard deviation of those measurements (for each subject under each listening condition). When durations are normally distributed, 68% of tokens will occur within one standard deviation of the mean, and 99% will occur within 3 standard deviations of the mean. Segments that are longer than average, as would be the case for phrase-final lengthening, have normalized durations that are positive in value. Segments that are shorter than the mean have negative values (Wightman et al., 1992).

Normalized durations were used in this study in order to compare: (1) the degree of preboundary lengthening between different types of phones across increasingly higher levels of phrasing, and (2) the degree of preboundary lengthening between different types of phones across the two speaking conditions, quiet and noise. Raw durations were used in this study to examine changes in segmental durations in all word positions across the two speaking conditions.
CHAPTER 4

4. Results

4.1. The Effect of Noise on Speaking Rate

The average speaking rates in words per minute for each speaker under both conditions are shown in Table 2.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Quiet</th>
<th>Noise</th>
<th>Percent change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2B1</td>
<td>223</td>
<td>201</td>
<td>-9.6%</td>
</tr>
<tr>
<td>2A1</td>
<td>176</td>
<td>182</td>
<td>+3.2%</td>
</tr>
<tr>
<td>1A2</td>
<td>212</td>
<td>192</td>
<td>-9.5%</td>
</tr>
<tr>
<td>1B1</td>
<td>208</td>
<td>203</td>
<td>-2.3%</td>
</tr>
</tbody>
</table>

Table 2. Average speaking rates in words per minute.

A wider range of speaking rates was observed in quiet (176 to 223 wpm) than in noise (182 to 203 wpm) across the four speakers. Two speakers, subjects 1A2 and 2B1, decreased their rate of speaking by almost 10% in noise. These two subjects were also the fastest speakers in the quiet condition. In contrast, the speaker who was the slowest speaker in the quiet condition, subject 2A1, increased his rate of speaking in noise. All of the observed speaking rates were well above the average rate of natural speech (140 wpm) (Picheny et al., 1986). As noted by Picheny and colleagues (1986) in their study of clear speech, sentence level materials generally do not include hesitation and breath pauses which would account for the overall faster rate of speech.
Speakers are able to alter their rate of speaking in two ways: by lengthening phones, and by adding or lengthening pauses. To determine which strategies were used by the subjects in this study, articulation rates (in syllables per second), which exclude pauses, were calculated (Table 3). Articulation rates computed for both the quiet and noisy conditions were found to fall within the range of articulation rates for natural speech (4.5 to 5.0 syllables/sec) (Picheny et al., 1986).

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Quiet</th>
<th>Noise</th>
<th>Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2B1</td>
<td>5.4</td>
<td>5.0</td>
<td>-7.4%</td>
</tr>
<tr>
<td>2A1</td>
<td>4.6</td>
<td>4.8</td>
<td>+3.0%</td>
</tr>
<tr>
<td>1A2</td>
<td>5.2</td>
<td>4.6</td>
<td>-10.8%</td>
</tr>
<tr>
<td>1B1</td>
<td>4.9</td>
<td>4.8</td>
<td>-2.0%</td>
</tr>
</tbody>
</table>

Table 3. Average articulation rates in syllables per second.

The similarities between changes in articulation rate and speaking rate indicate that a large proportion of the change is accommodated by variation in the duration of phones, rather than by pausing. The effect of pausing is seen most clearly in the data for Subject 2B1, who showed the greatest difference (2.4%) between changes in speaking rate and articulation rate. Very small differences (0.2 to 0.3%) between articulation rate and speaking rate changes are seen in the data for subjects 1A2 and 1B1.
4.2 The Effect of Noise on Segmental Duration

The total distribution of speech sounds present in the materials used in this experiment, including both speaking conditions, is provided in Table 4.

<table>
<thead>
<tr>
<th>Type of Speech Sound</th>
<th>Number of Each Type of Speech Sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowels</td>
<td>304</td>
</tr>
<tr>
<td>Glides</td>
<td>36</td>
</tr>
<tr>
<td>Liquids</td>
<td>60</td>
</tr>
<tr>
<td>Nasals</td>
<td>94</td>
</tr>
<tr>
<td>Fricatives</td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>74</td>
</tr>
<tr>
<td>Voiced</td>
<td>50</td>
</tr>
<tr>
<td>Stops</td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>66</td>
</tr>
<tr>
<td>Voiced</td>
<td>70</td>
</tr>
<tr>
<td>Affricates</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 4. The distribution of speech sounds in the stimulus materials.

Table 5 presents the mean durations of speech sounds, expressed in seconds, for the four speakers in quiet and noise, as well as the amount of change observed between the two conditions, calculated as a percentage of the duration observed in quiet.
<table>
<thead>
<tr>
<th>Subject 2B1</th>
<th>Mean Duration (secs)</th>
<th>Standard Deviation (secs)</th>
<th>Percent Change in Mean Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quiet</td>
<td>Noise</td>
<td>Quiet</td>
</tr>
<tr>
<td>Vowels</td>
<td>0.088</td>
<td>0.100</td>
<td>0.047</td>
</tr>
<tr>
<td>Glides</td>
<td>0.053</td>
<td>0.057</td>
<td>0.025</td>
</tr>
<tr>
<td>Liquids</td>
<td>0.061</td>
<td>0.065</td>
<td>0.032</td>
</tr>
<tr>
<td>Nasals</td>
<td>0.060</td>
<td>0.062</td>
<td>0.022</td>
</tr>
<tr>
<td>V/l Fricatives</td>
<td>0.094</td>
<td>0.094</td>
<td>0.033</td>
</tr>
<tr>
<td>V/d Fricatives</td>
<td>0.054</td>
<td>0.060</td>
<td>0.027</td>
</tr>
<tr>
<td>V/l Stops</td>
<td>0.071</td>
<td>0.071</td>
<td>0.030</td>
</tr>
<tr>
<td>V/d Stops</td>
<td>0.065</td>
<td>0.066</td>
<td>0.024</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject 2A1</th>
<th>Mean Duration (secs)</th>
<th>Standard Deviation (secs)</th>
<th>Percent Change in Mean Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quiet</td>
<td>Noise</td>
<td>Quiet</td>
</tr>
<tr>
<td>Vowels</td>
<td>0.112</td>
<td>0.110</td>
<td>0.064</td>
</tr>
<tr>
<td>Glides</td>
<td>0.065</td>
<td>0.054</td>
<td>0.030</td>
</tr>
<tr>
<td>Liquids</td>
<td>0.074</td>
<td>0.066</td>
<td>0.042</td>
</tr>
<tr>
<td>Nasals</td>
<td>0.069</td>
<td>0.066</td>
<td>0.030</td>
</tr>
<tr>
<td>V/l Fricatives</td>
<td>0.106</td>
<td>0.104</td>
<td>0.046</td>
</tr>
<tr>
<td>V/d Fricatives</td>
<td>0.073</td>
<td>0.072</td>
<td>0.064</td>
</tr>
<tr>
<td>V/l Stops</td>
<td>0.082</td>
<td>0.085</td>
<td>0.059</td>
</tr>
<tr>
<td>V/d Stops</td>
<td>0.075</td>
<td>0.069</td>
<td>0.033</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject 1A2</th>
<th>Mean Duration (secs)</th>
<th>Standard Deviation (secs)</th>
<th>Percent Change in Mean Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quiet</td>
<td>Noise</td>
<td>Quiet</td>
</tr>
<tr>
<td>Vowels</td>
<td>0.096</td>
<td>0.112</td>
<td>0.046</td>
</tr>
<tr>
<td>Glides</td>
<td>0.056</td>
<td>0.065</td>
<td>0.028</td>
</tr>
<tr>
<td>Liquids</td>
<td>0.064</td>
<td>0.070</td>
<td>0.035</td>
</tr>
<tr>
<td>Nasals</td>
<td>0.059</td>
<td>0.064</td>
<td>0.027</td>
</tr>
<tr>
<td>V/l Fricatives</td>
<td>0.092</td>
<td>0.097</td>
<td>0.032</td>
</tr>
<tr>
<td>V/d Fricatives</td>
<td>0.053</td>
<td>0.056</td>
<td>0.039</td>
</tr>
<tr>
<td>V/l Stops</td>
<td>0.074</td>
<td>0.074</td>
<td>0.054</td>
</tr>
<tr>
<td>V/d Stops</td>
<td>0.067</td>
<td>0.068</td>
<td>0.042</td>
</tr>
</tbody>
</table>
Table 5. Mean durations and standard deviations for the speech sound categories in quiet and noise. The percentage of change in the mean, with reference to the duration observed in quiet, is also provided.

In both the quiet and noisy conditions, vowels were longer than nonvocalic sonorants. Liquids were always the longest of the sonorants. Glides were always the shortest, except for one subject (2A1) in noise. Of the obstruent sounds, affricates were not well-represented in the corpus (a total of four affricates was produced by each subject). For this reason, the data generated is not considered to be as representative as that for the remaining obstruent sounds, and will not be considered here. Voiceless fricatives were the longest of the obstruent sounds, in both the quiet and noisy conditions, followed by voiceless stops. Voiced stops always exceeded voiced fricatives in length, except for those produced by one subject (2A1) in the noisy condition.
The pattern of change in the durations of the different types of speech sounds across the two experimental conditions tended to be quite complex and individual. However, some general trends were observed, as illustrated in Figure 3. For subjects 2B1 and 1A2, the two subjects who decreased their rates of speaking, all of the speech sounds showed an increase in duration in the noisy condition, with the exception of stop consonants for subject 1A2 and voiceless obstruents for subject 2B1, which showed no change. Of the vowels and nonvocalic sonorants, vowels showed the greatest amount of change, followed by glides, liquids and nasals.

While the mean segmental durations increased in noise for the productions of subjects 2B1 and 1A2, subject 2A1 tended to decrease the length of segments in noise, with the exception of voiceless stops (increased) and voiced fricatives (no change). Of the vowels and nonvocalic sonorants, the greatest percentage of change occurred for the glides, followed by liquids, nasals and vowels in the data for subject 2A1. For both subjects 2A1 and 1B1, very little change was observed in the duration of fricatives. Subject 1B1 showed the least variation in segmental duration across the two conditions.
Figure 3. Percentage of change in segmental durations for each subject in noise, with reference to durations observed in the quiet condition.
Table 6 illustrates the percentage of change in duration of vowels and all consonants for the four speakers. For each speaker, the direction of change was the same for both vowels and consonants. Subject 2A1 was the only speaker who demonstrated a decrease in mean segmental length for both vowels and consonants, as he was the only speaker to increase his rate of speaking in the noisy condition. For three of the four speakers, the mean change in consonant length was small compared to the change in vowel length. The exception to this was seen in the data for subject 2A1, in which the percentage of change for consonants exceeded that for vowels.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Percent Change in Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowels</td>
</tr>
<tr>
<td>2B1</td>
<td>+13.6</td>
</tr>
<tr>
<td>2A1</td>
<td>-1.8</td>
</tr>
<tr>
<td>1A2</td>
<td>+16.7</td>
</tr>
<tr>
<td>1B1</td>
<td>+4.8</td>
</tr>
</tbody>
</table>

Table 6. Percent change in mean duration of vowels and consonants, with reference to duration observed in quiet.

4.3 The Relationship Between Perceived Boundary Strength and Syntactic Attachment

Prior to analyzing the effect of noise on prosodic boundaries, it is of interest to determine whether the level of syntactic attachment has any effect on perceived boundary strength. The results of the perceptual labeling task in this study indicate that, in general,
syntactic attachment at a higher level in the syntactic hierarchy is more perceptually distinct than lower level syntactic attachment. The findings described here are very similar to those found by Price et al. (1991).

For the sentence pairs that differed by place of attachment (near vs. far attachment of final phrase and right vs. left attachment of middle phrase), clear differences were observed between the “a” version, which had the higher syntactic break, and the “b” version, which had the lower syntactic break. Larger break indices occurred at the critical boundary in the “a” version rather than the “b” version in 20/24 of the sentence pairs representing far vs. near attachment of final phrase (across all subjects and both speaking conditions), as illustrated in example 10. Usually, the break index at the attachment site was the largest in the sentence for the “a” version.

10) Far attachment: He 2 saw 2 the 2 man 4 with 2 binoculars 6

Near attachment: He 2 saw 2 the 2 man 3 with 2 binoculars 6

In the sentences which differed by place of attachment of middle phrase, the break index at the location of the attachment (the lower syntactic break) was smaller than at the opposite end of the word or phrase to be attached (the larger syntactic break) for 22/24
sentence pairs (across all subjects and both speaking conditions), as shown in example 11.

11) Left attachment: Beginning 2 Monday 5 afternoon 2 aerobics 2 classes 2 will 2 be 2 extended 3 to 2 one 2 hour 6

Right attachment: Beginning 2 Monday 2 afternoon 5 aerobics 2 classes 2 will 2 be 2 extended 3 to 2 one 2 hour 6

The structural ambiguities (sentences containing parentheticals vs. embedded sentences and conjoined sentences vs. complements) also showed marked differences between the “a” and “b” members of each pair. At least one of the boundaries of an inserted parenthetical phrase or sentence was labeled with a break index of “4” or higher. In 17/24 of the sentences (across subjects and speaking conditions), the boundaries at both ends of the parenthetical were labeled with break indices of “4” or above. In contrast, at least one of the critical boundaries in the nonparenthetical sentences was labeled with a break index of “3” or less. This difference between labeled break indices in the parenthetical and nonparenthetical sentences is illustrated in example 12.
12) Parenthetical: They will continue barring further interruptions in the afternoon.

Embedded Phrase: They will continue barring further interruptions in the afternoon.

In the conjoined vs. complement pairings, 16/24 of the sentence pairs were marked by a larger break index in the "a" version than the "b" version (across subjects and speaking conditions), as shown in example 13. No differences were observed at the critical boundary between the remaining eight pairs of sentences.

13) Conjoined sentence: I guessed that Nate Bureau would still be there.

Verb complement: I guessed that Nate Bureau would still be there.

4.4 The Effect of Noise on Perceived Boundary Strength

The distribution of break indices for each speaking condition is summarized in Table 7. The correlation between break indices across the two conditions is provided in parentheses.
Table 7. Distribution of break indices across speaking conditions.

A Chi-square test indicated that the frequency of occurrence of each break index was independent of speaking condition ($X^2 (df=4)=1.77, p>0.05$). Because prosodic breaks are affected by the rate at which a person speaks, it was thought that separating speakers into two groups, based on the direction of change of speaking rate between the quiet and noisy conditions, might reveal a dependence of break index on speaking condition. A Chi-square test of independence did not confirm this hypothesis ($X^2 (df=1)=0.86, p>0.05$ for subjects who spoke more slowly in noise; $X^2 (df=1)=0.32, p>0.05$ for the subject who spoke more quickly in noise).

Although the frequency of occurrence of each break index was found to be independent of speaking condition, the effect of noise on a speaker’s production of sentences appeared to have some effect on perceived boundary strength. The greatest differences are seen in the data for subject 2B1, while the least change is seen for subject 1B1. For subject 2B1, the greatest change between speaking conditions occurred for word
boundaries (break index=2) and boundaries at the edges of small groupings of words (b.i.=3). Eleven percent of boundaries labeled with a “2” in quiet increased to a “3” or “4” in noise; twenty-seven percent of boundaries labeled as small groupings (b.i.=3) in quiet, however, were decreased to a break index of 2 when produced in noise. Twenty-seven percent of boundaries labeled with a “3” also increased to “4” in noise. Major boundaries in sentences produced by subject 2A1 showed a tendency to decrease in strength in noise. Thirty-eight percent of boundaries labeled with a “5” in quiet were decreased to “4”, and 50% of boundaries labeled “4” were decreased to “3”. For subject 1A2, 83% of clitic group boundaries (b.i.=1) increased to word boundaries (b.i.=2) and 27% of intermediate phrase boundaries increased in strength to intonational phrase boundaries. No clear pattern of increasing or decreasing break index emerged in the data for subject 1B1.

Because the stimulus materials consisted of paired sentences, changes in perceived boundary strength across speaking condition were examined in relation to both members of each pair of sentences. A Chi-square test, however, indicated that the direction of change in break index was not dependent on sentence version ($X^2_{(df=1)}=1.48, p>.05$).
4.5 The Effect of Noise on the Duration of Vowels and Nonvocalic Sonorants in Preboundary Syllables

Figures 4a through 7a illustrate the relationship between break index and mean normalized duration of the vowel nucleus (in the final syllable before a boundary), in both the quiet and noise conditions. For all four subjects, duration increased rapidly between the break indices 2, 3 and 4, and then usually leveled off or decreased for the higher break indices. The largest increase in vowel duration across these three break indices occurred in the data for subject 2A1, while the smallest amount of variation is seen for subject 1B1. Differences between break indices 1 and 2 are more variable across subjects; the number of “1” judgements, however, was less than three for all subjects except 2A1. Little effect of noise on vowel duration can be seen in figures 4a-7a.

Analyses of variance (unequal sample sizes) indicated that there was a significant effect of break index on vowel duration for all four speakers ($F(4,400)=10.41, p<.001$ for speaker 1B1, $F(5,400)=12.70, p<.001$ for 2A1, $F(4,400)=10.81, p<.001$ for 1A2, and $F(4,400)=12.26, p<.001$ for speaker 2B1). There was no significant effect of noise condition ($F(1,400)=0.00, p>.05$ for 1B1, $F(1,400)=0.10, p>.05$ for 2A1, $F(1,400)=0.00, p>.05$ for 1A2, and $F(1,400)=0.07, p>.05$ for 2B1), nor was there a significant interaction between noise condition and break index ($F(4,400)=0.15, p>.05$ for 1B1, $F(5,400)=0.04, p>.05$ for 2A1, $F(4,400)=0.04, p>.05$ for 1A2, and $F(4,400)=0.82, p>.05$ for 2B1).

---

3 Differences in the degrees of freedom reflect the fact that in some cases it was necessary to eliminate boundaries labeled with a break index of 1, due to the limited amount of data available.
Prior to these analyses, the data were tested for homogeneity, as ANOVA is sensitive to departures from this assumption. Using Bartlett’s test for homogeneity of variance for unequal sample sizes, the data for each subject were found to be homogeneous \( (X^2 (df=5)=3.30, p>0.05 \) for subject 1B1, \( X^2 (df=5)=-2.04, p>0.05 \) for 2A1, \( X^2 (df=4)=0.87, p>0.05 \) for 1A2, and \( X^2 (df=4)=4.13, p>0.05 \) for 2B1 across conditions.

The data for each subject were analyzed further using Duncan’s new multiple range test (adjusted for unequal sample sizes) with a confidence level of 95%. Several commonalities were seen across subjects. The differences in duration between break indices 4, 5 and 6 (intermediate phrase, intonation phrase and sentence) did not reach significance for any of the subjects. For three subjects (excluding 2B1), differences in vowel duration between the lower break indices (2 and 3) and the higher break indices (4,5,6) were statistically significant. In the noise condition, however, the difference between break indices 3 and 4 failed to reach significance for subject 1B1. For subject 2B1, the vowel durations at higher break indices (4,5,6) were significantly longer than those at break index 2 (word boundaries), but not those at break index 3 (small grouping of words).

As described above in section 4.2, large changes occurred in the duration of nonvocalic sonorants between the quiet and noisy conditions. Because of this, the distribution of mean normalized durations of nonvocalic sonorants in coda position (in
the syllable before a phrase boundary) was determined, as illustrated in figures 4b-7b. Nonvocalic sonorants show a similar pattern of increasing duration across break indices to that of vowels, in that the most rapid increase in duration occurs between break indices 3 and 4. This rise is much more rapid for nonvocalic sonorants than for vowels. Between break indices 4 and 6, duration decreases for three of the four speakers. For subject 2B1, duration increases rapidly across all of the break indices to 5, and then drops off rapidly. Analyses of variance were not performed on the data for sonorants, due to small numbers of occurrences of sonorants at several break indices.
Figure 4. Break Index versus mean normalized duration for (a) vowel nucleus in the final syllable before a boundary, (b) nonvocalic sonorants in coda position; subject 2B1.
Figure 5. Break Index versus mean normalized duration for (a) vowel nucleus in the final syllable before a boundary, (b) nonvocalic sonorants in coda position; subject 2A1.
Figure 6. Break Index versus mean normalized duration for (a) vowel nucleus in the final syllable before a boundary, (b) nonvocalic sonorants in coda position; subject 1B1.
Figure 7. Break Index versus mean normalized duration for (a) vowel nucleus in the final syllable before a boundary, (b) nonvocalic sonorants in coda position; subject 1A2.
4.6 Pausing

4.6.1 The Effect of Noise on the Distribution of Pauses

Pause distributions are illustrated in Figures 8 and 9. Although subjects were highly variable in the number of pauses they employed, there was almost no difference in the number of pauses produced by each subject between the quiet and noisy conditions.

Mean pause length increased for subjects 1A2 and 2A1 in the noisy condition, but decreased for subjects 1B1 and 2B1. The standard deviations of pause lengths was greater in the noisy condition for three of the four subjects. The data for Subject 2A1 show the only decrease in range for the noisy condition.
Figure 8. Frequency plots of pause durations for the quiet and noisy conditions for subjects 2A1 and 1B1.
Figure 9. Frequency plots of pause durations for the quiet and noisy conditions for subjects 2B1 and 1A2.
4.6.2 The Distribution of Pauses by Break Index and Noise Condition

The distribution of pauses for each subject are shown in Figures 10 and 11. Pauses occurred more often at boundaries with higher break indices than those with lower break indices. Furthermore, the proportion of boundaries in which pausing occurs tends to increase with increasing break index. As indicated in Table 8, pausing occurs at 80% or more of boundaries labeled with a break index of “5”, but at less than 1% of boundaries labeled with a “2”. No clear pattern in the number of pauses occurring at each break index emerges across speaking condition. A Chi-square test confirmed that the number of pauses differed significantly across break indices ($X^2$ (df=3)=42.36, $p<0.005$). There was no significant effect of noise condition ($X^2$ (df=1)=0.01, $p>0.05$), and no significant interaction of break index and noise condition on the number of pauses ($X^2$ (df=3)=0.09, $p>0.05$) inserted by each subject.

<table>
<thead>
<tr>
<th>Break Index</th>
<th>Percentage of Boundaries with Pauses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quiet</td>
</tr>
<tr>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>0.9</td>
</tr>
<tr>
<td>3</td>
<td>8.6</td>
</tr>
<tr>
<td>4</td>
<td>56.3</td>
</tr>
<tr>
<td>5</td>
<td>80.0</td>
</tr>
</tbody>
</table>

Table 8. Percentage of each type of boundary that is marked with a pause.
Figure 10. Distribution of pauses by break index and speaking condition. (a) 1A2 (n=16 in quiet, n=17 in noise), (b) 1B1 (n=6 in both quiet and noise).
Figure 11. Distribution of pauses by break index and speaking condition. (a) 2B1 (n=13), (b) 2A1 (n=26).
4.6.3 Pause Durations at Each Break Index and the Effects of Noise

Figures 12 and 13 illustrate the relationship between pause lengths and break indices in both the quiet and noisy conditions. As shown in the figures, pause lengths tended to increase with increasing break index. In some instances, mean pause duration decreased from break index 4 to 5.

The effect of noise on pause durations varied from speaker to speaker. Mean pause length increased in noise for subject 2B1 across all break indices except those labeled with a 2. The data for subject 2A1 reveal an opposite effect; mean pause length decreased in noise at all break indices except those labeled with a 5.

Due to the limited amount of pause data available, the results from all of the subjects were pooled in order to test for statistical significance. As subjects tended to differ in their pausing styles, these results should be interpreted cautiously. An analysis of variance (unequal sample sizes) indicated that there was a significant effect of break index on pause duration \( F(3,114)=6.19, p<.001 \), a significant effect of noise on pause duration \( F(1,114)=5.17, p<.05 \), but no significant interaction between break index and noise condition duration \( F(1,114)=5.17, p>0.05 \). Prior to analysis, the data were subjected to a log transformation, in order to satisfy the conditions of normality and homogeneity \( \chi^2(df=3)=1.67, p>0.05 \), Bartlett’s test of homogeneity for unequal sample
sizes). Multiple comparison analysis, using Duncan's new multiple range test (95% confidence interval, unequal sample sizes), indicated that differences in pause durations reached significance between break index 3 and the higher break indices of 4 and 5 in quiet. In noise, the mean pause duration at break index 2 was found to be significantly different from the higher break indices of 4 and 5.
Figure 12. Distribution of pause lengths across break indices; (a) 1A2, (b) 1B1.
Figure 13. Distribution of pause lengths across break indices; (a) 2B1, (b) 2A1.
5. Discussion

The following hypotheses were tested in this study:

Hypothesis 1:

Speakers will not change their rates of speaking across quiet and noisy speaking conditions.

The alternative experimental hypothesis:

Speakers will slow their rates of speaking in noisy conditions relative to quiet, lengthening both vowels and consonants.

Hypothesis 2:

In both quiet and noisy conditions, vowel durations (in the rhyme of preboundary syllables) and pause durations will not significantly increase with increasing boundary strength, as labeled by experienced listeners.

The alternative experimental hypothesis:
In both quiet and noisy conditions, vowel durations (in the rhyme of preboundary syllables) and pause durations will increase with increasing boundary strength, as labeled by experienced listeners.

Hypothesis 3:

The degree of preboundary lengthening at phrase boundaries will not be significantly different across quiet and noisy speaking conditions.

The alternative experimental hypothesis:

The degree of preboundary lengthening at phrase boundaries will be significantly greater in noisy speaking conditions relative to quiet.

Hypothesis 4:

The number and duration of pauses will not be significantly different across quiet and noisy speaking conditions.

The alternative experimental hypothesis:

The number and duration of pauses will be significantly greater in noisy speaking conditions relative to quiet.
5.1 The Effect of Noise on Speaking Rate and Segmental Duration

In this study, three of the four speakers talked more slowly in noise than they did in quiet. The degree of increase was not uniform across subjects. While two speakers decreased their rate of speaking in noise by almost 10%, the third decreased his rate by only 2.3%. One speaker talked more quickly in the noise condition than in the quiet condition. Changes in speaking rate did not occur solely on the basis of speakers inserting or lengthening pauses. In fact, calculation of articulation rates indicated that a large portion of the durational change across speaking conditions was accommodated by changes in the duration of speech sounds.

Durational variation did not occur in a linear fashion across speaking conditions, but instead depended to a large extent on the type of speech sound being modified. For the speakers who decreased their rates of speaking, a large percentage of change was seen in the length of vowels. The subject who increased his rate of speaking in noise modified vowel lengths by very little. Glides, liquids, and nasal consonants also showed a large amount of change between the two conditions. Speakers were more idiosyncratic in the way they altered the durations of obstruent consonants between the two speaking conditions.
For all of the speakers, the direction of durational change was constant for both vowels and consonants. Thus for three of the subjects, both vowels and consonants increased in length, and for the remaining subject, both vowels and consonants decreased in length.

These results tend to support the experimental alternative to the first hypothesis above, that speakers will slow their rates of speech in noise, and lengthen both vowels and consonants. The results from the fourth subject indicate, however, that speakers do vary in how they adapt to having to speak in a noisy situation. Variation is also evident in the degree of durational change that speakers apply.

The experimental results reported here for three of the four subjects replicate the findings of earlier speech in noise studies, which indicated that speakers slow their rate of speaking in noise. Unlike the results of a study performed by Junqua & Anglade (1990) and experiments involving shouted speech, however, in which vowels were found to increase in length while consonants were decreased slightly, both consonants and vowels showed the same direction of change in this study. The amount of change in vowel duration was also found to be much smaller than that found in shouted speech. The largest increase in vowel duration reported here was almost 17%, compared to a 67% increase found by Rostolland (1985) in his examination of shouted speech.
The results found in the current investigation more closely resemble those of clear speech, in which increases in both consonants and vowels were observed in speakers who increased their rates of speaking (Picheny et al., 1986). The proportion of change in speaking rates observed here, however, was much smaller than when subjects were asked to deliberately speak clearly.

5.2 Timing and the Prosodic Hierarchy

5.2.1. Preboundary Lengthening

The data in this study indicate that preboundary lengthening, in the vowel nucleus of the rhyme of a preboundary syllable, consistently occurs in increasingly larger increments between three levels of structure: word boundaries, small groupings of words, and intermediate phrases. This increase in lengthening reached statistical significance between the hierarchically lower phrase boundaries (up to small groupings of words) and the higher level prosodic constituents (intermediate phrases and above). In general, this preboundary lengthening levels off between intermediate phrases and intonational phrases, and in two cases decreased between intonational phrases and the end of the sentence. A similar pattern of results is found in the data for nonvocalic sonorants in coda position in preboundary syllables.
These results support the experimental alternative to the second hypothesis above, and indicate that the depth of structure proposed in the prosodic hierarchy is reflected in preboundary lengthening, at least for the lower levels of structure. As proposed by Wightman et al. (1992), preboundary lengthening seems to play its most important role in marking the smaller phrase boundaries, where few other acoustic phrase boundary cues occur. At the higher phrase boundaries, speakers provide a number of other cues to phrasal structure, such as pausing and boundary tones, and therefore do not need to mark boundaries with increasingly larger amounts of lengthening.

The results for preboundary lengthening described here are similar to those of Ladd & Campbell (1991) who found four distinct levels of preboundary lengthening, and in the related studies by Price et al. (1991) and Wightman et al. (1992). In these latter studies, preboundary lengthening was found to increase with increasing break index for the lower prosodic phrases, but leveled off between the major phrase boundaries (intonational phrases and above). Differences in mean normalized duration were found to reach significance between each of the lower levels of structure when measurements were adjusted for speaking rate (Wightman et al., 1991). In contrast to these two studies, an increase between intermediate phrases and intonational phrases was generally not observed in the present study. Speaker 2B1 was an exception to this, and in his case, a larger difference between mean normalized durations for intermediate and intonational phrases was observed in the data for nonvocalic sonorants than for vowels. The
differences between the results discussed here may in part reflect the fact that previous studies used professional radio speakers, who have been shown to use more acoustic cues than nonprofessional speakers (Price et al., 1991; Pijper & Sanderman, 1993).

5.2.2. Pausing

The number of pauses which occurred in this study was not large, and was characterized by a large amount of intersubject variability. While one subject paused only six times under each speaking condition, another subject paused four times as often. From the limited data acquired, however, there was a significant dependence between perceived boundary strength and the distribution of pauses. Perceptual boundary strength tended to increase as the number of pauses increased. Furthermore, a much higher percentage of intonational phrase boundaries were marked with pauses than any other type of phrase boundary.

A clear pattern of increasing pause durations with each increase in the level of prosodic structure did not emerge across all subjects. Pauses at intermediate phrase boundaries and intonational phrase boundaries were significantly longer than those at boundaries between words and between small groupings of words in quiet. In noise, the distinction between pauses found at the boundaries between small groupings of words and those found between intermediate phrases was less clear. The hypothesis proposed in
this study, that pause durations would not increase incrementally with increasing boundary strength, was confirmed. However, there was some tendency towards longer pauses seen in the data for the two subjects who paused most often, subject 2A1 in both quiet and noise, and subject 2B1 in noise. Hypothesis two would be better tested if more pause data were available.

Differences between pause durations across intermediate phrase boundaries and intonational phrase boundaries did not reach significance, and in some cases, mean pause lengths were found to be shorter at intonational phrase boundaries than at intermediate phrase boundaries. The experimental results found here, then, indicate that the role of boundary tones, which occur at the boundaries of intonational phrases, must be significant in a listener's perception of boundary strength, since differences between pause lengths and preboundary lengthening do not explain perceived increases in boundary strength.

5.3 The Effect of Noise on Prosodic Timing Cues

In this study, it was proposed that speakers would exaggerate preboundary lengthening under noisy speaking conditions in order to maintain intelligibility.
Underlying this hypothesis was the assumption that speakers would use this strategy because pitch cues would be less useful to listeners, due to perceptual masking by noise. The results show that, in fact, little or no change occurs in the degree of preboundary lengthening of the vowel nucleus in noise, confirming hypothesis 3. In an experiment in which intersubject variability tends to be large, this is a robust finding across all of the subjects. Although speakers varied in the degree to which they altered speaking rate and in which speech sounds they altered, they all appeared intent on maintaining the prosodic timing relationships between phrases across both the quiet and noise conditions.

It is quite possible that while noise would mask much of the segmental information available to a listener, the overall rhythmic timing of prosodic phrases would remain a relatively strong and stable cue. It has indeed been shown in the literature that when speech is distorted by filtering or masked by noise, prosodic cues can be detected by listeners (Huggins, 1971), and furthermore, they can use this information to detect word boundaries (Smith et al., 1989). Rather than altering prosodic patterns to maintain intelligibility, then, it seems that speakers try to preserve prosodic timing patterns to maintain intelligibility under conditions of noise.

The effect of noise on pausing was much more variable across speakers than its effect on preboundary lengthening. Hypothesis 4, which stated that the number and duration of pauses would not be significantly different across quiet and noisy conditions,
was confirmed. In general, none of the speakers made an effort to pause more in noise than in quiet. One of the speakers who decreased his rate of speaking showed a tendency to pause longer at each break index in noise than he did in quiet; in contrast, the speaker who increased his speaking rate in noise showed a tendency to pause for shorter periods of time at each break index in noise. No clear pattern of changes in pause duration emerged in the data for the remaining two subjects. The results of changes in pause durations across speaking conditions are similar to those found for changes in segmental durations, in that the results for each subject are highly idiosyncratic.

5.3.1. Noise, Speaking Rate and Perceived Boundary Strength

Although speakers showed remarkably little change in the degree to which they lengthened vowels at phrase boundaries, boundaries were not always perceived by the labelers to have the same strength in noise as in quiet. It appears that speaking rate had a small effect on the perceived strength of some boundaries, although this did not reach significance. In the sentences uttered by the two speakers who slowed their rates of speaking in noise by almost 10%, more boundaries were perceived to increase in strength than decrease. The opposite effect was observed in the sentences produced by the speaker who increased his rate of speaking in noise. These changes do not appear to be related to the addition or lengthening of pauses at these boundaries, but instead may actually reflect local phonological changes in the production of segments. Changes
observed in spectrographic analysis included vowel modifications (vowels were occasionally observed to be reduced to schwa), insertion of a glottal stop immediately before a pause, deletion of segments in rapid speech, and stop consonant modification (e.g. a full stop replaced by an unreleased stop or substitution of a flap for a full stop).

5.4 Conclusion

This experiment has demonstrated that preboundary lengthening is a useful cue to prosodic phrase boundaries. Its role appears to be more significant at lower levels of prosodic phrasing, where fewer pauses or pitch cues occur. More of the higher level phrases were marked by pauses than were the lower prosodic constituents.

Noise was found to affect segmental durations, with resulting changes in speaking rate; however, speakers were highly individualistic in the ways in which they altered the segmental components of speech. Vowels and nonvocalic sonorants absorbed much of the change across speaking conditions. Pausing was also idiosyncratic.

The effect of noise on the overall timing relationships obtaining between prosodic constituents was minimal. The degree of preboundary lengthening at each break index was virtually held constant between the quiet and noisy speaking conditions for all four subjects who participated in this investigation. It is proposed that speakers keep prosodic
timing relationships stable in order to maximize the intelligibility of their utterances when segmental information could be lost by masking.

5.4.1. Future Considerations

The stimulus materials used in the current investigation consisted of syntactically ambiguous sentence pairs representing four different types of structural contrasts. A logical extension of this study would be an examination of the effects of noise and syntactic attachment on listeners' abilities to disambiguate the stimulus sentences. Furthermore, the data generated in the current study could be used to examine 1) whether the intelligibility of utterances produced in noise is greater than those produced in quiet, when matched for signal-to-noise ratio, and 2) the effect of noise on pitch cues in prosodic constituents.

As noted in Chapter 2, much of the research investigating the effect of noise on speech has concentrated on the production of single words. The research described in this investigation goes a step further, by examining the effects of noise on prosodic constituency in sentences. Nevertheless, the utterances produced in this investigation were highly constrained sentences that subjects were asked to read aloud. It would be of
interest to determine the effects of noise on acoustic cues to prosodic constituent structure in natural speech.
6. References


Appendices

Appendix A

List of Stimuli

Group #1 Conjoined Sentences vs. Verb Complements: Each "a" version consists of two sentences joined by a conjunction. The "b" versions consist of an NP or embedded sentence forming the complement of the verb.

1a) Davy brought his bicycle to me for repairs. I worked on the bike all afternoon, as Davy watched patiently. I repaired the bicycle and Davy rode home.

b) My neighbour, Anne Davy, found a bicycle lying in a roadside ditch. It clanked loudly as she rode towards home. I fixed it for her, so that it looked and sounded almost as good as new. I repaired the bicycle Anne Davy rode home.

2a) When the woman was found dead on my property, the city's chief homicide detective, Nate Bureau, was sent to the scene. After awhile he left, and a constable informed me that there was little evidence to be found at the scene. I guessed that Nate Bureau would still be there.

b) My neighbours were interested in refinishing old furniture, so I suggested they help themselves to any of the old furniture that they liked in my garage. I guessed that ornate bureau would still be there.

3a) Bob made thousands of dollars buying and selling real estate. When Sheila ran off with all of his money, Bob was left feeling bitter and angry. He will not discuss her or deal in real estate any longer.

b) After Jenny lost thousands of dollars on the real estate market, Barry soon got tired of hearing about her problems. He will not discuss her ordeal in real estate any longer.

Group #2 Far vs. Near Attachment of Final Phrase: The "a" versions have a final phrase (either a VP or PP) attached at a distance from the word being modified. The "b" versions have a final phrase modifying an immediately preceding word.

1a) After receiving his fourth speeding ticket on the way to court that day, the judge gave a woman 10 years for shop lifting. The judge sentenced the woman in a foul mood.

b) As the judge announced his decision in court, the woman was still fuming about being arrested. The judge sentenced the woman in a foul mood.
2a) When Henry looked through his binoculars, he spotted the injured man below. *He saw the man with binoculars.*

b) Tracy insisted that a man had been following them for at least 15 minutes. "Which one is he?" asked Henry. "The one with the binoculars." she whispered. Then he spotted him. *He saw the man with binoculars.*

3a) As Anne looked hopefully for a seat on the bus, she was surprised to see her brother sitting at the back. *She saw her brother looking for a seat.*

b) As Anne sipped her coffee in the crowded cafe, she was surprised to see her brother looking around hopefully. *She saw her brother looking for a seat.*

**Group #3 Left vs. Right Attachment of Middle Word/Phrase:** The "a" versions have a word or phrase attached to the left, whereas in the "b" versions, the word or phrase is attached to the right.

1a) Although morning aerobics classes are an hour long, the regular afternoon aerobics classes have been 1/2 an hour in length. Many people requested that the 1/2-hour afternoon classes be extended. *Beginning Monday, afternoon aerobics classes will be extended to 1 hour.*

b) Although the community centre offers 1/2-hour aerobics classes throughout the day, many people have found the classes to be too short. *Beginning Monday afternoon, aerobics classes will be extended to 1 hour.*

2a) Ted was relieved that his prize-winning guppies did not succumb to a mysterious fish disease. *Although the guppies survived, for a while they looked sick.*

b) The guppies never looked healthy and it wasn't long before they died. *Although the guppies survived for awhile, they looked sick.*

3a) I wasn't feeling very energetic right after lunch but I thought I should get some exercise before dinner. *I went swimming, later in the afternoon.*

b) It was too cold to go swimming in the morning. *I went swimming later, in the afternoon.*

**Group #4 Parenthetical vs. Embedded Sentences:** Each "a" version contains a parenthetical sentence or phrase inserted between the verb and its complement. Each "b" version contains an embedded sentence, either linked to a previous word by a deleted subordinate conjunction or as direct object of the main sentence verb.

1a) If you need some advice about breeding your canaries, you should talk to John. *John knows everything you know about breeding birds.*

b) You and John should go into business together breeding canaries - you are both experts on the subject. *John knows everything you know about breeding birds.*
2a) During the parliamentary debate this morning, environmentalists broke down the door and demanded to be heard. Environmental concerns, however, will probably not affect the MP's debate. They will continue, barring further interruptions, in the afternoon.

b) When environmentalists attempted to interrupt a parliamentary debate this morning, security teams were able to diffuse the situation quickly. They will continue barring further interruptions in the afternoon.

3a) Mom was pleased that the goldfish survived while she spent a month in Europe. Mom knows, by the way, you were feeding her goldfish.

b) Mom knows that you really want a puppy and that you're trying to prove you can take care of a pet. Mom knows by the way you were feeding her goldfish.
Appendix B

Group 1

1(a) I repaired the bicycle and Davy rode home.

1(b) I repaired the bicycle Anne Davy rode home.
2(a) I guessed that or Nate Bureau would still be there.

(b) I guessed that ornate bureau would still be there.
3(a) He will not discuss her or deal in real estate any longer.

(b) He will not discuss her ordeal in real estate any longer.
Group 2

1(a) The judge sentenced the woman in a foul mood.

(b) The judge sentenced the woman in a foul mood.
2(a) He saw the man with binoculars.

(b) He saw the man with binoculars.
3(a) She saw her brother looking for a seat.

(b) She saw her brother looking for a seat.
Group 3

1(a) Beginning Monday, afternoon aerobics classes will be extended to one hour.

(b) Beginning Monday afternoon, aerobics classes will be extended to one hour.
2(a) Although the guppies survived, for a while they looked sick.

(b) Although the guppies survived for awhile, they looked sick.
3(a) I went swimming, later in the afternoon.

(b) I went swimming later, in the afternoon.
Group 4

1(a) John knows everything, you know, about breeding birds.

(b) John knows everything you know about breeding birds.
2 (a) They will continue, barring further interruptions, in the afternoon.

(b) They will continue barring further interruptions in the afternoon.
3(a) Mom knows, by the way, you were feeding her goldfish.

(b) Mom knows by the way you were feeding your goldfish.
If you have any questions, the investigator will be pleased to provide further information to ensure that you fully understand the nature of the project and what you are being asked to do.

A copy of this consent form has been given to me and I have read it. I have received a thorough explanation of the project. I consent to participate in this research study.

Date: __________________________
Signature of Subject: __________________________
Printed Name of Subject: __________________________
Subject Code: __________________________
Appendix D

Sentence Labeling

Each sentence will be labeled in three ways: with a "break index", to indicate the degree of separation between each word, with symbols indicating boundary tones, and with symbols marking prominent syllables.

Break Index:

A break index value marks the degree of juncture you perceive between each word in a sentence. A number of cues will affect your perception of juncture (e.g. pausing, phrase final lengthening, the presence of prominent syllables, and the continuity/discontinuity in the melodic line of a sentence). Seven levels of juncture are to be labeled, as indicated below. Mark each break index at the right edge of a word. The following values are to be assigned

Break Index

0 Assign a break index of 0 between two orthographic words where no prosodic break is perceived (i.e. marking a clitic group).

Examples: "did he" produced as [dldi], palatalization, as for the [dj] in "did you", flapping of an alveolar, as when "want an apple" is spoken with the /t/ deleted and the /n/ produced as a flap.

1 Assign to normal word boundaries.

2 Assign to a minor grouping of words within a larger unit.

3 Assign to an intermediate level phrase. A phrase accent (i.e. a single high or low tone) is associated with the boundary of an intermediate phrase (you do not need to mark the phrase accent).

4 Assign to a boundary marking an intonation phrase. Each intonation phrase ends with a boundary tone.

5 Assign to any boundary which marks a group of intonational phrases.

6 Assign to sentence boundaries.
Boundary Tones:

Label boundary tones as final falls (using the symbol "FF"), continuation falls ("CF"), continuation rises ("CR"), and question rises ("QR").

Prominent Syllables:

Each syllable in a sentence is to be labeled with a degree of prominence: "P1" for major phrasal prominences, "P0" for lesser prominences, and "s" for syllables with no prominence.

Example Sentence:

1) The 0 men 3 won 0 over 2 their 1 enemies. 6
   s P0 s P1s s P0s FF

OR

2) The 1 men 2 won 4 over 1 their 1 enemies. 6
   s P0 P1CR s s s P0s FF

Occasionally dysfluencies occur in the sentences. These can be marked with a "#" symbol.
Appendix E

Spectrogram Parameters used in Sonogram (version 0.9)

Analysis Options

- Analysis method: FFT
- Window shape: Hanning
- Peak picking: SideBand 1000 Hz
- Analysis resolution:
  - Window size = 128 pts.
  - Time increment = 64 pts.
- Sampling frequency: 22.05 kHz
- Low cut frequency: 0 kHz

Display Options

- Gray scale: many
- Pixel size: Y=1
- Emphasis: High shape
- Display frequency range: Upper limit set at 5 kHz
- Display dynamic range: Varied from subject to subject
Appendix F

Segmentation Criteria

Plosives

- following Crystal & House (1982; 1988), both the hold portion and the release portion was measured for stop consonants. For non-plosive releases, such as lateral releases, the non-plosive release portion of the signal was included in the following segment.

a) Voiceless stops:

onset: beginning of the stop was determined by the abrupt cessation of all formants

offset: beginning of following vowel determined by the onset of voicing after aspiration

b) Voiced stops:

onset: beginning of the stop was determined by the cessation of formants and an obvious decrease in intensity of the waveform

offset: beginning of the formant transitions of the following vowel marked the offset of voiced stops, and by a rapid increase in intensity on the waveform
Fricatives

a) Voiceless fricatives:

onset: the onset of a voiceless fricative was determined by the beginning of noise on the wide band spectrogram and waveform

offset: the onset of voicing and cessation of a random noise signal (on the waveform) were used to determine the beginning of the following vowel or semivowel

b) Voiced fricatives:

onset: the onset of a voiced fricative was marked by the end of the formants of the previous vowel or semivowel, and by a marked decrease in energy on the waveform

offset: the end of a voiced fricative was marked by the end of the noise of the fricative and the onset of vowel formants.

Nasals

onset: the onset of nasal consonants was marked by an abrupt change in the formant pattern and an abrupt decrease in intensity of the spectrogram.
offset: the end of nasal consonants were marked by a change from a steady state nasal formant pattern to the on-glide of the following vowel, with a marked increase in intensity of the formants.

**Liquids**

a) /l/

onset: the onset of /l/ was occasionally very difficult to determine. Usually, the beginning of an /l/ was marked by an overall decrease in the intensity level of the formants, and the end of a rapid transition to a high value of F3

offset: the end of an /l/ was often marked by an increase in intensity of the formant structure, and by a sudden change in the harmonic structure on the narrowband spectrogram, from a steady state to the onglide of a following vowel

b) /r/

onset: the onset of /r/ was marked by a sudden decrease in the intensity of F4, and by the end of a rapid downward transition of F3, putting it in close proximity to F2

offset: the end of an /r/ was marked by the beginning of a rapid rise of F3 away from F2, and by a sudden increase in the intensity of F4
Glides

a) /w/

onset: the beginning of /w/ was determined by a sudden decrease in the energy level of the formants of the previous vowel, particularly F3

offset: the end of /w/ was marked by a sharp increase in energy at the initiation of the vowel, particularly F3, and often by the second formant acquiring a positive slope.

b) /j/

onset: the onset of /j/ was often difficult to determine, particularly when it occurred between two vowels. A decrease in energy of the formants and harmonics, and a decrease in intensity on the waveform, was used to segment /j/ from a previous vowel.

offset: the end of /j/ was marked by a minimum in the frequency level of the third formant, which was followed by a rapid upward transition in F3 for the following vowel.