THE ROLE OF SEGMENTAL AND SUPRASEGMENTAL CUES IN LEXICAL ACCESS

By

SUZANNE LYNNE MACDONALD

B.A., The University of British Columbia, 1995

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

UNIVERSITY OF BRITISH COLUMBIA

April 2003

© Suzanne Lynne Macdonald, 2003
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 School of Audiology and Speech Sciences

The University of British Columbia
Vancouver, Canada

Date April 19/03
ABSTRACT

Research in the area of speech perception has shown that prosodic, or suprasegmental, information contributes to the recognition of spoken words. However, the significance of this contribution in English has been debated. Some researchers have argued that suprasegmental cues are not exploited much in English because suprasegmental cues co-vary with segmental cues and, therefore, information at the suprasegmental level is redundant. In this study, it is argued that segmental and suprasegmental cues are only redundant in ideal listening situations. It is hypothesized that suprasegmental cues will make a significant contribution to word recognition under degraded listening conditions.

Sixteen normal hearing young adults listened to 100 common nouns presented in a gating paradigm. The 100 words were presented in four conditions: intact segmental and suprasegmental cues, degraded segmental cues, degraded suprasegmental cues, and degraded segmental and suprasegmental cues. Participants were asked to identify the word being presented.

Participants' responses were scored according to the isolation point and the accuracy of word identification. Results for both measures showed that participants were able to identify the word sooner and more accurately when both segmental and suprasegmental cues were intact than when either cue was degraded; furthermore, when either cue was intact word recognition was faster and more accurate than when both cues were degraded.

These results supported the hypothesis that suprasegmental cues can contribute significantly to spoken word recognition in English.
# TABLE OF CONTENTS

ABSTRACT .................................................................................................................. ii

TABLE OF CONTENTS ............................................................................................... iii

LIST OF TABLES ......................................................................................................... vi

ACKNOWLEDGEMENTS ............................................................................................ vii

1. Literature Review ........................................................................................................ 1
   1.1 Introduction ........................................................................................................... 1
   1.2 Definition of prosody ............................................................................................. 2
   1.3 Models of spoken word recognition ...................................................................... 3
      1.3.1 Major models .................................................................................................. 4
      1.3.2 Common features of models of spoken word recognition ......................... 7
      1.3.3 Distinguishing features of models of spoken word recognition ................. 7
      1.3.4 Critique of models of spoken word recognition .......................................... 8
   1.4 Prosody and speech perception ............................................................................. 9
      1.4.1 Lexical stress and lexical processing ............................................................. 10
   1.5 The current study .................................................................................................. 36
      1.5.1 Suprasegmental cues .................................................................................... 36
      1.5.2 Segmental cues ............................................................................................. 37
      1.5.3 The gating paradigm ..................................................................................... 38
      1.5.4 Hypotheses .................................................................................................... 39

2. METHODS .................................................................................................................. 42
   2.1 Overview ............................................................................................................... 42
   2.2 Participants ......................................................................................................... 42
   2.3 Materials ............................................................................................................. 43
4.3 General discussion

4.3.1 Relevance to models of spoken word recognition

4.3.2 Limitations of the current study

4.3.3 Clinical implications

4.3.4 Future research

REFERENCES

APPENDIX I

APPENDIX II

APPENDIX III
LIST OF TABLES

Table 3.1 ........................................................................................................... 53
Table 3.2 ........................................................................................................... 55
ACKNOWLEDGEMENTS

I would like to thank my supervisors, Kathy Pichora-Fuller and Jeff Small, for their guidance and support throughout this project, with extra thanks to Jeff for taking on a larger role than had originally been anticipated. Thank you also to Barb Bernhardt for her valuable input. I would like to express an extra special thank you to Anne Marie Bowes for her support and her voice in my thesis. I would also like to thank Trudy Adam, Nicky Pontin, Edward Storzer, and Alvilda Douglas for their help with my thesis and their friendship. I would like to express my appreciation to the faculty and staff at the School of Audiology and Speech Sciences, especially to Sue Bryant for all of the encouragement along the way. Thank you to my family and my friends for their patience and their confidence in me and to Aaron Hignell for always being unconditionally supportive and understanding.

This thesis was supported by a grant from the Natural Science and Engineering Research Council of Canada.
1. LITERATURE REVIEW

1.1 Introduction

Prosody has been shown to play a significant role in the perception of spoken language. Research in speech perception has established that the prosodic information contained in the speech signal contributes to the processing of speech at the lexical, syntactic, and discourse levels (review in Cutler, Dahan, & van Donselaar, 1997).

Despite these findings, current models of spoken word recognition have included only the segmental information contained in the input in explaining the process by which humans understand spoken words. Researchers have been critical of these models for neglecting to address the role of the suprasegmental, or prosodic, information contained in the input, which has been shown by numerous studies to contribute to the processing of spoken words (Cutler & van Donselaar, 2001; Lindfield, Wingfield, & Goodglass, 1997; Lindfield, Wingfield, & Goodglass, 1999; Soto-Faraco, Sebastian-Galles, & Cutler, 2001; Wingfield, Lindfield, & Goodglass, 1997).

While researchers agree that suprasegmental information does contribute to spoken word recognition, different researchers have reached different conclusions regarding the significance of its contribution to the processing of words in English (Cutler, 1986; Cutler et al. 1997; Cutler & van Donselaar, 2001; Soto-Faraco et al., 2001). In English, segmental and suprasegmental cues co-vary at the lexical level, in that vowels in unstressed syllables are usually produced in a reduced form (shorter in duration and more neutral in vowel quality) while vowels in stressed syllables are always produced in their full form. This redundancy has led some researchers to argue that suprasegmental cues are “not exploited in English where vowel-quality information essentially provides all the cues to word identity which prosodic processing could provide” (Cutler et. al., 1997, page 159). Other researchers have argued that prosodic information plays a more significant role in recognizing spoken words in English.
(Lindfield et al., 1997; Lindfield et al., 1999; Wayland, Wingfield, & Goodglass, 1989; Wingfield et al., 1997; Wingfield, Lindfield, & Goodglass, 2000). These authors have measured suprasegmental cues independently of segmental cues and concluded that suprasegmental cues are exploited in English during spoken word recognition.

In the present study, it is argued that suprasegmental cues are only redundant in English in ‘ideal listening’ situations in which both segmental and suprasegmental cues are readily available to the listener. In listening situations that are less than ‘ideal’, for example, situations in which the segmental information that is available to the listener is limited, suprasegmental cues likely play a more important role in lexical access.

The current study was carried out to explore the role of suprasegmental cues in lexical processing for conditions that may represent more typical listening situations, that is, conditions in which not all of the information contained in the signal is available to the listener. Lexical stress was manipulated to alter the suprasegmental information that is audible to the listener. The segmental structure of the speech input was manipulated in order to measure the contribution of suprasegmental cues when the segmental information is or is not fully available to the listener. It was hypothesized that suprasegmental information would not contribute significantly to lexical processing when these cues are redundant with segmental cues (i.e., in ‘ideal’ listening circumstances), but would contribute significantly once they are no longer redundant with segmental cues (i.e., in ‘non-ideal’ listening circumstances).

1.2 Definition of prosody

Prosody has been defined by different researchers in both broad terms, “the structure that organizes sound,” and in more specific terms, “a synonym for suprasegmental features,” which encompass “pitch, tempo, loudness, pause” (Cutler et al., 1997). In the current study, the term prosody will be used to refer to the more specific definition of ‘suprasegmental features’ and these two terms will be used interchangeably.
At the word level, prosody can be realized, in different languages, as lexical stress (realized through pitch, duration, and loudness over a polysyllabic domain), lexical tone (realized through pitch), and lexical pitch accent (realized through pitch over a polysyllabic domain). The current study was conducted to explore the role of prosodic structure in spoken word recognition in English. The prosodic structure of interest is lexical stress because English is a lexical stress language. Unless otherwise specified, ‘prosody’ and ‘suprasegmental features’ will be used to refer to the stress pattern of words.

1.3 Models of spoken word recognition

Different models of spoken word recognition have been developed in the attempt to understand and describe the process by which humans recognize spoken words. Although these models have addressed many aspects of the process of spoken word recognition, the role of suprasegmental cues has not been adequately included in these models. In the current study, it was hypothesized that suprasegmental information would have a significant role in processing spoken words in English; therefore, a model of spoken word recognition needs to be developed that will account for these cues in the input. The development of such a model will be discussed further in chapter four. A review of the current major models of spoken word recognition will be presented in this chapter.

In their review of research on speech perception and spoken word recognition, Jusczyk and Luce (2002) discussed four current major models of spoken word recognition, including the Cohort Model, the Trace Model, the Shortlist Model, and the Neighborhood Activation Model (and its more recent version, PARSYN). Summaries of these models will be presented in the next section, followed in later sections by a summary of their common and distinguishing features and a discussion of the critique that they have all received from researchers in the field of speech perception.
1.3.1 Major models

1.3.1.1 The Cohort Model

According to the original Cohort Model (Marslen-Wilson & Tyler, 1980; Warren & Marslen-Wilson, 1987), as the initial phonemes of a spoken word are heard, a ‘word-initial cohort’ is activated. This word-initial cohort includes all of the items in the lexicon that share the same initial phonemes as the input word. As more of the onset of the word is heard, the number of lexical items sharing the same initial phonemes decreases until only one ‘best-fit’ candidate remains and word recognition occurs. Early versions of the Cohort Model focused on the activation of the cohort and the competition between candidates as a function of bottom-up (acoustic-phonetic input) information, processed in parallel. A more recent version of the model (Gaskell & Marslen-Wilson, 1997) incorporates both bottom-up (featural) information and top-down (semantic, syntactic, morphological, and phonological) information in a distributed connectionist framework. In this more recent version of the model, the process of speech perception is conceptualized as involving “the direct mapping of low-level featural information onto a distributed representation of lexical knowledge and form” (p. 615). In the Cohort Model, competition between candidate words is based solely on the fact that these words are included in the cohort as candidates; competitor words in the cohort are not assumed to have any effect on the activation levels of other candidates (i.e., no lateral inhibition).

1.3.1.2 The Trace Model

The Trace Model of speech perception (McClelland & Elman, 1986) is based on the principles of interactive activation; this involves information passing through excitatory and inhibitory connections between processing units, which results in changes in the activation levels of the processing units. The Trace Model consists of a large number of processing units, which are organized into three levels that correspond to ‘features’, ‘phonemes’, and ‘words’. According to McClelland & Elman, “each unit stands for a hypothesis about a particular
perceptual object occurring at a particular point in time defined relative to the beginning of the word” (p. 8). At the feature level, several banks of detectors serve to process input that is presented to the feature-level units sequentially in successive time slices. At the phoneme and word levels, there is one detector for each phoneme and word, respectively. In the Trace Model, speech processing is conceptualized as occurring through bidirectional excitatory and inhibitory connections between the processing units. Between levels, activation is passed by excitatory connections between feature, phoneme, and word units that are mutually consistent. Within levels, units have inhibitory connections, such that input that is consistent with a specific candidate inhibits activation of a competitor item. Through this process of activation and inhibition, a ‘best-fit’ candidate can be selected over its competitors.

1.3.1.3 The Shortlist Model

The Shortlist Model (Norris, 1994) is a two-stage connectionist model of spoken word recognition in which bottom-up information flows in one direction from phonemes to lexical units; there is no top-down information from the lexical units to the phoneme units. In the first stage of the Shortlist Model, a lexical search produces a ‘short-list’ of word candidates that match the bottom-up input. In the second stage of the model, competition occurs between the lexical candidates by lateral inhibitory links. Specifically, candidates in the short-list enter into a network of word units, in which overlapping words inhibit one another in proportion to the number of phonemes by which they overlap. Recognition occurs for the ‘best-fit’ candidate.

1.3.1.4 The Neighborhood Activation Model and PARSYN

In the Neighborhood Activation Model (NAM) (Luce, Pisoni, & Goldinger, 1990; Luce & Pisoni, 1998), the spoken word input is assumed to activate a set of similar acoustic-phonetic patterns in memory. The activation levels of the patterns are a direct function of their phonetic similarity to the input; higher activation occurs for patterns that are the most similar to the target. In this model, words emerge when a system of ‘word-decision units’ is activated (i.e.,
‘neighborhood’ activation) by the acoustic-phonetic patterns. In the NAM, the ‘word-decision unit’ is the mechanism by which the word is chosen. The task of these units is to decide which of the acoustic-phonetic patterns best matches the input by monitoring bottom-up (acoustic-phonetic) and top-down information (higher level lexical information, including word frequency). Word-decision units monitor the level of activation for each pattern, the frequency of the associated word, and the activation levels and frequencies of all other activated words in the neighborhood. Word recognition occurs once the word-decision unit for a given acoustic-phonetic pattern reaches criterion. In this model, words in 'dense' neighbourhoods (those that are phonetically similar to many other words in the lexicon) are assumed to be more difficult to retrieve than words in 'sparse' neighbourhoods. Experimental research supports this distinction, in that words from sparse neighbourhoods are recognized more accurately and rapidly than words from dense neighborhoods (Luce & Pisoni, 1998).

A more explicit processing version of this model, called PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000), incorporates both a sublexical level of representation, which is constrained by probabilistic phonotactics (defined as the “relative frequencies of segments and sequences of segments in spoken words” by Vitevitch, Luce, Pisoni, & Auer, 1999), and a lexical level of representation. PARSYN has three levels of units: an input level, a pattern level, and a word level. All three levels contain interconnected units. The input level has position-specific allophone units in banks of receptors, which correspond to the temporal sequence of the input. Connections among allophones at the pattern level are facilitative across temporal positions. The pattern level has banks that replicate those of the first level, and units at this level receive facilitative information from the input level. Units at this level also receive facilitative input from other units at this level, in preceding or following temporal positions. Other connections between units within this level are inhibitory. Links among these units at this level are weighted according to position specific transitional probabilities, which represent the
probabilistic phonotactic constraints in English. The resting levels of nodes in this layer correspond to position-specific probabilities of occurrence, so that allophones occurring in common positions have higher resting activation levels. The word level units receive facilitative input from the position-specific allophones at the pattern level. Inhibitory information is sent down from the word level to competitor units at the pattern level, once a given word is more highly activated than the other candidates.

1.3.2 Common features of models of spoken word recognition

The current major models of spoken word recognition have certain features in common. In all of the models, the process of spoken word recognition is conceptualized as involving the initial activation of a set of lexical candidates, followed by a competition process between the activated items and the eventual selection of the 'best-fit' candidate. All of the current models of spoken word recognition focus on the match between the segmental information in the spoken word and stored phonological information in the lexicon. This emphasis on the segmental information in the word recognition process has led to some criticism from researchers in the field; this criticism will be discussed below.

1.3.3 Distinguishing features of models of spoken word recognition

Despite sharing the features outlined above, each of the major models of spoken word recognition has a combination of features that distinguish it from the others. Specifically, the models differ according to the dynamics of processing by which spoken words are recognized. Each of the models has a different combination of the following features: the type(s) of information that contribute to the recognition process (i.e., bottom-up versus top-down), the direction(s) in which information travels (i.e., unidirectional versus bidirectional), the manner of processing (i.e., parallel versus serial), the levels of processing involved (i.e., some combination of acoustic-phonetic patterns, features, allophones, phonemes, and words), the type(s) of connection between processing units and levels (i.e., excitatory versus inhibitory),
and the competition process between candidates (i.e., no effect versus direct effect on activation
levels of competitors). For instance, the Cohort model incorporates bottom-up and top-down
information that travels bidirectionally in a connectionist framework; it includes feature and
word levels of processing; it assumes no direct effect (excitatory or inhibitory) on the activation
levels of other candidates during the competition process. The Trace model incorporates
bottom-up information in a parallel connectionist framework; information travels
bidirectionally between units and levels that correspond to features, phonemes, and words;
candidates directly affect one another in the competition process via excitatory and inhibitory
connections. In the Shortlist model, bottom-up information travels in one direction from a
phoneme level to a word level in a connectionist framework; competition occurs through
inhibitory connections between candidates at the word level. In NAM, bottom-up and top-down
information contribute to the recognition process; this model incorporates a acoustic-phonetic
patterns level and a word level; competition occurs indirectly through comparisons between the
activation levels and frequencies of the candidate and its competitors. The PARSYN model is a
more explicit processing version of NAM. This model incorporates bottom-up and top-down
information that travels bidirectionally in a connectionist framework; excitatory and inhibitory
connections exist between the three levels of processing (input, pattern, and word); direct
competition occurs between the candidates.

1.3.4 Critique of models of spoken word recognition

Speech perception research has shown that suprasegmental information does contribute to
the recognition of spoken words. However, current models of spoken word recognition, such as
those outlined previously, have been framed solely in terms of the match between the
segmental information contained in the input and the phonological representations in the
lexicon. Given findings that suprasegmental information does contribute to the process of word
recognition, numerous authors have concluded that models of spoken word recognition should
take suprasegmental information into account (Cutler & van Donselaar, 2001; Lindfield et al., 1997; Lindfield et al., 1999; Soto-Faraco et al., 2001; Wingfield et al., 1999). A literature review of these studies, as well as other studies that have explored the role of prosody in spoken word recognition, will be discussed in the next section. The need for current models of spoken word recognition to incorporate these findings in their design, by addressing the suprasegmental information contained in the speech input, will be discussed in a later chapter.

1.4 Prosody and speech perception

Prosody has been shown to contribute to the comprehension of spoken language at the lexical, syntactic, and discourse levels of speech processing (review in Cutler et. al., 1997). For example, studies that have explored the role of prosody in processing discourse structure have found that prosody serves to accent or ‘highlight’ salient information, to determine reference and topic structure, to interpret semantic meaning, and to structure conversation (Cohen, Douaire, & Elsabbagh, 2001; Couper-Kuhlen, 2001; see Cutler et al. for a review, 1997; Pynte, 1998; Swerts & Hirschberg, 1998; Wennerstrom, 2001).

Studies of the role of prosody in processing syntactic structure have shown that prosody contributes to the processing of higher-level constituent structure and to the resolution of ambiguities in sentences, both globally (referring to the interpretation of the sentence as a whole) and locally (referring to the relation between lexical units in the sentence) (Carlson, 2001; Carlson, Clifton, & Frazier, 2001; review in Cutler et. al., 1997; Fox-Tree, & Meijer, 2000; Kjelgaard & Speer, 1999; Kjelgaard, Titone, & Wingfield, 1999). Studies have also been conducted that have explored the role of sentential prosody in word recognition (e.g., Hernandez, Fennema-Notestine, Udell, & Bates, 2001; Herron & Bates, 1997).

Speech perception research on the role of prosody at the lexical level of processing has explored how prosody aids in identifying word boundaries and in processing words (for a review see Cutler et al., 1997). Studies have explored the role of lexical tone, lexical pitch, and
lexical stress in word recognition, across different languages. Studies on lexical tone have been conducted in numerous languages, including Mandarin and Cantonese (Cutler & Chen, 1997; Cutler et al., 1997; Lee, Vakoch, & Wurm, 1996). Lexical pitch accent studies have been conducted in Japanese (Cutler et al., 1997; Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999). For instance, Cutler & Otake (1999) used pairs of Japanese words that differed only in lexical pitch accent, such as 'ame' (produced with a High-Low pitch accent) and 'ame' (produced with a Low-High pitch accent), to explore the role of pitch accent in spoken word recognition in Japanese. Research on the role of lexical stress in spoken word recognition has been conducted in many different lexical-stress languages, including English, Dutch, and Spanish (Cutler, 1986; Cutler & Clifton, 1984; Cutler et. al., 1997; Cutler & van Donselaar, 2001; Lindfield et al., 1999; Soto-Faraco et al., 2001; Wingfield, et al., 2000). Research in these languages will be reviewed in greater detail below. Although the findings of these studies provide support for the role of prosody in lexical processing, Cutler et al. (1997) suggest that more research in this area is necessary for a more complete understanding of the contribution of prosody at this level of processing.

1.4.1 Lexical stress and lexical processing

Most studies on the role of prosody in lexical processing have been conducted in English, which is a lexical stress language, and, therefore, most studies in this area have focused on the role of lexical stress in the processing of spoken words (Cutler et al., 1997). Lexical stress refers to the number of syllables in the word as well as the relative duration, amplitude, and pitch of each syllable. English is considered a ‘free-stress’ language, in that it does not have a fixed word-internal position for stress. The lack of an obligatory internal lexical stress pattern allows for the possibility that the stress pattern information contained in the word may contribute to the processing of that word, because stress could actually be used to distinguish between words that are identical in segmental form.
However, despite the fact that stress pattern could be used to distinguish between words in English, it is very uncommon for semantically distinct English words to differ on stress pattern information alone, as in the words ‘insight’ and ‘incite’ (Cutler et al., 1997). Furthermore, suprasegmental cues typically co-vary with segmental cues at the lexical level in English. Differences in stress pattern usually involve both segmental and suprasegmental changes, since vowels in stressed syllables are always produced in their full form while vowels in unstressed syllables are usually produced in a reduced form. For these reasons, the significance of the contribution that is made by suprasegmental information, beyond that provided by segmental cues alone, in processing lexical items in English continues to be debated by researchers.

Cutler and her colleagues have tended to discount the contribution made by suprasegmental information in the processing of spoken words in English (Cutler, 1986; Cutler & Cliffton, 1984; Cutler et al., 1997; Cutler & van Donselaar, 2001; McQueen, Norris, & Cutler, 1994; Soto-Faraco et al., 2001). Based on their findings from English and other lexical stress languages, the authors take the position that lexical stress is not exploited in English to the extent that it is in other languages. They have argued that suprasegmental information can usually be extracted from segmental structure in English and segmental analysis is, therefore, almost always sufficient to process words. Conversely, Wingfield and his colleagues have tended to take the perspective that suprasegmental structure does contribute significantly to the processing of spoken words in English, particularly when segmental information is not available (Lindfield et al., 1999; Wayland et al., 1987; Wingfield et al., 1997; Wingfield et al., 2000). In the following sections, studies conducted by both groups of researchers will be reviewed and discussed.
1.4.1.1 Cutler and her colleagues

Research conducted by Cutler and her colleagues on English will be presented first, followed by research conducted on Spanish and Dutch.

1.4.1.1.1 Lexical stress and lexical processing in English

Cutler and Clifton (1984) conducted three experiments that explored the use of prosodic information in word recognition. Their first experiment measured the ability to use anticipatory information about stress pattern, (i.e., ‘stress priming’) in spoken word recognition. Monosyllabic and bisyllabic words and non-words were presented visually to participants in matched stress lists (stress priming) and mixed stress lists (no stress priming). The matched stress lists consisted of either all monosyllabic items (such as ‘blaze’) or all bisyllabic items with a strong-weak stress pattern (such as ‘tiger’) or all bisyllabic items with a weak-strong stress pattern (such as ‘canoe’). The items in the mixed stress lists consisted of a combination of all three types of stress patterns. Lexical decision response times were compared for the matched versus mixed stress lists. The same experiment was then conducted with auditory presentation of the lists. Results showed no facilitative effect of stress priming. The authors interpreted these results as providing no evidence that prior knowledge of prosodic structure increases the speed of lexical retrieval. The authors concluded that these results indicate that lexical access is not constrained in such a way that the number of word candidates is reduced when prior information about stress pattern is provided.

The second experiment conducted by Cutler and Clifton (1984) explored whether stress pattern information could indirectly facilitate word recognition by providing the listener with information about the grammatical category of the word. The authors argued that, since nouns are typically produced with a strong-weak (Sw) stress pattern and verbs are typically produced with a weak-strong (wS) stress pattern, words conforming to these patterns should be recognized more quickly than wS nouns and Sw verbs if stress pattern information facilitates
word recognition. Listeners were asked to make a decision as to the grammaticality of phrases for monosyllabic and bisyllabic nouns and verbs, which were matched for frequency of occurrence. The nouns and verbs were presented in the context of "the _____" or "to _____". The words were either presented in grammatical phrases (such as "to punish" or "the sleeve") or ungrammatical phrases (such as "to lagoon" or "the crave"). For the bisyllabic words, equal numbers in each grammatical category had wS and Sw stress patterns. Therefore, half the verbs and nouns conformed to the typical stress pattern for their respective grammatical categories (such as the strong-weak produced noun "apple" and the weak-strong produced verb "await"). While the other half had 'atypical' stress (such as the weak-strong produced "cigar" and the strong-weak produced "borrow"). Response times for the grammatical decisions were measured and the results showed no facilitative effect for words with the grammatically typical stress pattern, compared to the words with the grammatically atypical stress pattern. As an additional measure, nouns and verbs that differed from one another only in lexical stress (i.e., CONduct vs. conDUCT), and therefore formed minimal stress pairs, were included in the experiment. For these words 'trick trials', in which the stress pattern of the word did not match the grammatical category, were administered. Results of the 'trick trials' showed that listeners were more willing to accept words that shifted to the inappropriate Sw pattern than the inappropriate wS pattern, particularly when the shift did not involve a change in vowel quality. The authors concluded that no evidence was found to support the argument that grammatical category is cued by stress pattern. However, based on the results of the 'trick trials', they suggested that certain forms of mis-stressing words may be more acceptable to the listener than others.

A third experiment conducted by Cutler and Clifton (1984) explored the effect of mis-stressing words on the word identification of bisyllabic nouns. Nouns were chosen that were clearly identifiable as having physical referents (such as "nutmeg") or mental referents (such as
"deceit"). Listeners were expected to decide whether words had mental or physical referents. According to the authors, this was an "easy and natural" decision requiring "full word recognition" (p. 191). The words were mis-stressed by shifting stress from a Sw stress pattern to an inappropriate wS stress pattern and from a wS stress pattern to an inappropriate Sw stress pattern. Half of the words in each of the groups were normally produced with unreduced vowels in both syllables (such as "nutmeg" or "typhoon"). For these words, the vowels remained unreduced in both syllables when the stress was shifted to the other syllable and, therefore, no change in vowel quality occurred with a shift in stress. The remaining words in each group were normally produced with an unreduced vowel in the strong syllable and a reduced vowel in the weak syllable (such as "wisdom" or "deceit"). For these words, a shift in stress resulted in a shift in vowel quality, in that normally unreduced vowels became reduced while normally reduced vowels became unreduced. The fact that some words shifted only in stress pattern while other words shifted in both stress pattern and vowel quality allowed for a comparison of the effect on lexical retrieval when suprasegmental changes occur in isolation and when suprasegmental changes occur in conjunction with segmental changes. Listeners were asked to 'ignore' any mispronunciations while making the 'physical' versus 'mental' decisions. Response times for the semantic judgements were measured. A strong main effect of mispronunciation was found; listeners took significantly longer to make judgements for mis-stressed words than for normally stressed words. This effect was found for both strong-weak and weak-strong words. This finding shows that placing incorrect stress on words does have an effect on word identification. Further analyses showed that, for weak-strong words, the mispronunciation effect was significantly smaller for unreduced words (which only had shifts in stress) than for reduced words (which had shifts in stress and vowel quality). This finding indicates that changes in suprasegmental information accompanied by changes in segmental information caused a greater impairment of word recognition than changes in suprasegmental
information alone, but that, even in the absence of segmental changes, changes in suprasegmental information did cause impairment to word recognition. The authors concluded that the results demonstrated that correct stress pattern is important for efficient word recognition to occur, since mis-stressing words was found to impair word recognition abilities, even in the absence of vowel quality changes.

In a later study, Cutler (1986) further explored the role of prosody in word recognition by seeking to establish the exact point at which prosody contributes. Cutler distinguished between lexical access and lexical retrieval as two different operations in the process of word recognition. She argued that prosody could either constrain lexical access by being part of the access code by which lexical entries are located or it could contribute to the retrieval process by being part of the phonological code in the lexicon, which is consulted only after access to the lexicon has been achieved. In her study, she sought to determine whether or not prosody contributes to word recognition by constraining the activation of lexical items. Cutler used a 'cross-modal priming paradigm' (Swinney, 1979). This task involved minimal stress pairs, in which the two members were segmentally identical, but each had a semantically distinct meaning depending on the stress pattern (i.e. 'forbear' means 'ancestor' if produced Sw, or 'be patient' if produced wS). Listeners were presented auditorily with sentences that contained one version (Sw or wS) of a minimal stress pair and were asked to make a lexical decision regarding an item that would appear visually on a computer screen at the offset of the target word during the presentation of the sentence. Of the words that were visually presented, some were related semantically to the version of the minimal stress pair contained in the sentence, some were related semantically to the other version of the stress pair, and some were unrelated control words. Cutler argued that "if prosody constrains lexical access in much the same way that segmental information does, then pairs of words which differ only in prosody should generate quite distinct lexical access codes and be, in practice, not confusable" (p. 204). In
other words, according to Cutler, if prosody constrains lexical access, then only the stress version semantically related to the visually presented word should prime that word and result in a faster lexical decision response time; the other version should result in a similar response time to the control. On the other hand, she would argue that, if both versions of a minimal stress pair can facilitate the lexical decision for either of the semantically related words to result in a faster response time than the control word, then it can be assumed that prosody does not constrain lexical access, according to Cutler. The results of the study were consistent with Cutler’s latter argument. Both versions of the minimal stress pairs were effective in priming both of the words that were each semantically related to one of the stress versions. Based on these results, Cutler concluded that, in English, lexical access is not directly constrained by suprasegmental information.

In their review article, Cutler et al. (1997) included the results of the preceding studies, as well as the results of other studies, in discussing the role of lexical stress in word processing in English. Based on the results of these studies, the authors concluded that lexical stress has been shown to have some effect on lexical processing in English. However, they suggested that many of the observed effects of lexical stress in lexical processing may be attributable primarily to the changes in vowel quality that typically accompany the suprasegmental changes. Cutler and her colleagues emphasized the need for studies to control for segmental cues in order to truly explore the effect of suprasegmental cues on word recognition. Furthermore, the authors discussed the distinction between lexical access, in which lexical candidates are activated by the stimulus input, and lexical retrieval, in which a specific candidate is selected as the perceived word. Lexical access is assumed to occur earlier in the word recognition process than lexical retrieval, which is assumed to occur only after lexical access has been achieved. They used the results of Cutler’s (1986) study to suggest that the evidence does not support a role of prosody in lexical access in the word recognition process in
English; however, they argued that it may have a role in lexical activation in other languages. In making this suggestion, Cutler and her colleagues included evidence from research that was conducted in other lexical stress languages, a topic that will be discussed in the next section. Overall, the authors concluded that “the accumulated evidence on lexical stress does suggest that it is exploited by listeners in a number of languages, and not exploited in English where vowel-quality information essentially provides all the cues to word identity which prosodic processing could provide” (p. 159). This conclusion was made despite the fact that the third experiment of Cutler and Clifton (1984) had provided evidence that changes in suprasegmental cues resulted in impairments to word recognition, even in the absence of segmental changes.

1.4.1.1.2 Lexical stress and lexical processing in other languages

To explore the contribution of prosody in languages in which suprasegmental cues are not fully redundant with segmental cues, Cutler and her colleagues conducted research in Spanish (Soto-Faraco et al., 2001) and Dutch (Cutler & van Donselaar, 2001). Spanish is a lexical stress language and, like English, polysyllabic words in Spanish have primary stress placed on one syllable. This syllable can occur in any position in the word. However, Spanish differs from English in that segmental and suprasegmental cues do not co-vary in this language. In contrast to English, vowel reduction does not occur in unstressed syllables in words in Spanish. Since both stressed and unstressed syllables in Spanish contain full vowels, it is possible for semantically distinct words to have identical segmental structure and differ only in suprasegmental structure. Despite this possibility, Spanish is similar to other lexical stress languages in that there are not many pairs of semantically distinct words differing only in suprasegmental structure. Soto-Faraco et al. (2001) used pairs of such words in their study.

Soto-Faraco et al. (2001) were interested in the effects of match versus mismatch between input and lexical representations. In conducting this study, they presumed that both match and mismatch contribute to spoken word recognition in normal listening situations.
Current models of spoken word recognition, as previously discussed, credit a perceptual matching process between the stimulus input and the lexical candidates as being the process by which words are recognized. In some of these models, such as the Shortlist Model (Norris, 1994), increases in activation occur for lexical candidates that match the input, making recognition of those candidates easier, while lexical candidates containing information that mismatches the input are inhibited, making these candidates more difficult to recognize.

In conducting their study, Soto-Faraco et al. (2001) used a technique called 'fragment priming'. In this task, auditory presentation of a sentence ending in a word onset fragment (consisting of the first two syllables of a word) was followed by visual presentation of a word or a non-word, for which the listener was required to make a lexical decision. According to the authors, the use of the lexical decision task in the priming paradigm allows the researchers to tap into automatic word activation processes that underlie speech perception; the accuracy of this assumption is debatable and will be discussed in a later section. The visually presented words were presented in three conditions, 'matching,' 'mismatching,' and 'control.' In the 'matching' condition, the visually presented word matched the preceding word onset fragment for segmental and suprasegmental structure. For example, the auditory fragment ‘PRINci’ matched the following visually presented word ‘PRINcipe’ (the capitalized portion represents a stressed syllable). For this condition, the word onset fragment was presumed to prime the visually presented word. In the 'mismatching' condition, the visually presented word mismatched the preceding word onset fragment in either segmental or suprasegmental structure. For example, the auditory fragment ‘prinCI’ suprasegmentally mismatched the following visually presented word ‘PRINcipe’. In this condition the suprasegmental structure of the auditory word onset fragments matched a different word than the word that was visually presented, creating a situation in which the participant was primed to expect a different word than the one that was presented. For instance, in the previous example the fragment ‘prinCI’
actually matches the onset of the real word ‘prinClpio’, which was not presented. It was presumed that the expectation of a different word, based on the word fragment presented, would inhibit the recognition of the visually presented word. In the control condition, words that differed in both segmental and suprasegmental structure from the auditory word onset fragments were visually presented. For example, the auditory fragment ‘mos’ segmentally and suprasegmentally mismatched the following visually presented word ‘PRINcipe’. Response times and accuracy of response were measured.

The first experiment focused on the effects of a mismatch of suprasegmental information between the word onset fragment and the visually presented word. The authors argued that, if suprasegmental information is involved in the lexical activation and competition process, then a fragment matching the following word for suprasegmental information would facilitate activation of that word and result in a faster response time than the control word. On the other hand, a fragment mismatching the following word (but matching a different word) on suprasegmental information would inhibit activation of that word and result in a slower response time, compared to the control word response. Results showed that, as hypothesized, the matched primes resulted in faster response times than did the control primes while the mismatched primes resulted in slower responses compared to control primes. Accordingly, the results of the ‘matching’ condition supported the argument of facilitation for lexical items that matched the information in the input. The results of the ‘mismatching’ condition supported the argument of inhibition for lexical items that mismatched the information in the input. Based on these results, the authors concluded that lexical stress information is used by Spanish listeners in lexical access.

In the second and third experiments of the study (Soto-Faraco et al., 2001), the same technique was used as in the first experiment to measure the effects of segmental mismatch on spoken word recognition. For these experiments, the stress pattern of the words was held
constant across conditions while the segmental structure was manipulated in the ‘mismatching’ condition. For example, in the matching condition the auditory fragment ‘aban’ segmentally and suprasegmentally matched the following visually presented word ‘abanDOno’, in the mismatching condition the auditory fragment ‘abun’ (from the word ‘abunDANcia’) segmentally mismatched and suprasegmentally matched the following visually presented word ‘abanDOno’, and in the control condition the auditory fragment ‘e’ (from eLAStico) segmentally and suprasegmentally mismatched the following visually presented word ‘abanDOno’. The results of these experiments again showed that the matched prime sped up the lexical decision response time while the mismatched prime slowed it down. According to the authors, the results again suggested that candidates were activated in both the matched and mismatched conditions, but that the mismatched segments led to the inhibition of the visually presented words while the matched segments led to facilitation of the visually presented words. The results from the second and third experiment were very similar to those obtained in the first experiment and the authors concluded that both segmental and suprasegmental cues have significant roles in constraining lexical activation for spoken word recognition.

Overall, the results of the three experiments (Soto-Faraco et al., 2001) indicated that segmental or suprasegmental matching primes facilitated word recognition while segmental or suprasegmental mismatching primes inhibited word recognition. The investigators concluded that their results provide evidence that, in Spanish, segmental and suprasegmental cues are both involved in lexical access and both have similar roles in constraining lexical activation.

Cutler and van Donselaar (2001) conducted a study to explore the role of suprasegmental cues in spoken word recognition in Dutch. Dutch is a lexical stress language with many similarities to English. Like English, polysyllabic Dutch words have stressed placed on one syllable, which can occur in different positions in the word. Segmental and suprasegmental cues co-vary with one another in Dutch, in that stressed syllables are produced
with full vowels while unstressed syllables tend to have reduced vowels. However, there are many more words in Dutch that contain full vowels in unstressed syllables, as compared to English. Therefore, the interdependency between the segmental and suprasegmental structure of words can be considered less pervasive in Dutch than it is in English. The authors suggest that, with regard to the interdependency of segmental and suprasegmental structure at the lexical level, Dutch can be considered to be on a continuum ‘between’ Spanish and English. It is possible, therefore, to conduct an experiment in Dutch that is similar to the previously discussed study in Spanish (Soto-Faraco et al., 2001), in which the contribution of the stress pattern in lexical processing can be measured independently of segmental structure.

In conducting this study, Cutler and her colleagues (2001) wanted to ascertain whether suprasegmental information plays as strong a role as segmental information in constraining lexical activation when the two levels of information are somewhat interdependent, as in Dutch, compared to when the two levels are entirely independent, as in Spanish. If so, the results of the study in Dutch should be more similar to the results obtained in Spanish, where segmental and suprasegmental information appear to contribute equally to lexical access (Soto-Faraco et al., 2001), than to the results obtained in English, where segmental cues appear to make a more significant contribution to lexical access than suprasegmental cues (i.e., Cutler, 1986).

Like other lexical stress languages, it is uncommon for semantically distinct Dutch words to differ in suprasegmental information alone and few pairs of such ‘minimal stress’ words exist in Dutch. In their first experiment, Cutler and van Donselaar (2001) used twelve minimal stress pairs as stimuli. The minimal stress pairs consisted of four bisyllabic pairs, seven trisyllabic pairs, one four-syllable pair that were recorded by a female native speaker of Dutch and then edited for presentation to the listeners as monosyllabic or bisyllabic word fragments. The word fragments containing stressed syllables were presumably louder than their
unstressed counterparts; the authors did not indicate whether or not the loudness of the stimuli was controlled for across the stressed and unstressed fragments. Participants were provided with an answer sheet containing the two words of each minimal stress pair and asked to circle the word they thought had been partially presented. For example, a listener presented with the auditory monosyllabic stressed fragment 'LO' could choose between 'DOORopen' and 'doorLOpen' on the answer sheet. Results showed that the participants scored significantly above chance in selecting the correct member of the stress pair. The authors concluded that Dutch listeners are able to judge whether a syllable or bisyllable presented in isolation comes from a word in which it is stressed or a word in which it is not stressed.

The second experiment explored word recognition ability using a repetition priming paradigm. In this paradigm, listeners are required to make lexical decisions for words and non-words presented in a list. Some of the items are presented more than once in a list. For repeated items, the first presentation of the item is expected to facilitate word recognition for the second presentation of the same item. Cutler and van Donselaar (2001) used the minimal stress pairs from the previous experiment, presented in their entirety, to explore whether or not the earlier presentation of one member of a minimal stress pair would facilitate word recognition for the later presentation of the other member of the minimal stress pair. For example, listeners were first presented with 'VOORnam' and then presented either with 'VOORnam' in the same list, or with 'voorNAM' in the same list, or with an unrelated control word in the same list. The listeners' lexical decision response times between conditions were measured. The authors were interested in whether or not the lexical decision response times for the later presentation of 'voorNAM' would benefit from the earlier presentation of 'VOORnam', to the same degree that response times for the later presentation of 'VOORnam' would benefit by the earlier presentation of 'VOORnam'. According to the authors, this experiment is similar to the study conducted by Soto-Faraco et al. (2001), in that the results allow for conclusions to be drawn.
regarding the role of suprasegmental information in lexical activation. They argued that, if the results showed that both members of a minimal stress pair can effectively prime either member of the pair, then it can be assumed that suprasegmental cues do not constrain lexical access. Conversely, they argued that if each member of a minimal stress pair is only effective in priming itself, then it can be assumed that suprasegmental cues do constrain lexical access. Results showed that response times were faster for identical items appearing earlier in the list than the response times for either the other members of the minimal stress pairs or the control words. In other words, each member of a minimal stress pair was only effective as a prime for itself and did not prime the other stress version of the pair. The authors concluded that this finding confirms that Dutch listeners can use stress pattern mismatches to rule out the activation of a lexical item that matches the input segmentally. They further suggested that Dutch listeners should also be able to use stress mismatch to rule out the activation of a lexical item that is not the other member of a minimal stress pair.

In the third experiment, a word-spotting task was used to explore this possibility. In this task, listeners were asked to detect real words embedded in nonsense strings of phonemes. This paradigm had been used in a previous study by McQueen, Norris, and Cutler (1994), in which they had found that English words were harder to detect when they were embedded in nonsense words that formed the onset of real English words than when they were embedded in nonsense words that did not form the onset of real English words. For instance, they found that 'mess' was harder to detect in 'domes', which forms the onset for 'domestic,' than in 'nemes,' which does not form the onset of a real word. This finding was attributed to inhibition of the word 'mess' due to competition from 'domestic,' which had been activated by the presentation of the onset 'domes'. In their experiment, Cutler and van Donselaar (2001) presented listeners with words that were embedded in the second syllable of two-syllable strings. Half of these strings formed the segmental onset of real trisyllabic Dutch words with stress occurring on the second
syllable; the other half did not form the segmental onset of a real word. For each of these strings two stress versions were created, one in which stress was placed on the second syllable (suprasegmentally matching the onset of the real trisyllabic word) and one in which stress was placed on the first syllable (suprasegmentally mismatching the onset of the real trisyllabic word). Listeners were asked to respond when they heard a real word embedded at the end of a stimulus. For example, the Dutch word 'muSEum' ('museum') has stress placed on the second syllable, which is homophous with the word 'zee' ('sea'); for the word 'museum', two different segmental strings were used 'muzee' (which formed the onset of a real word) and 'luzee' (which did not form the onset a real word), each of which were presented in two different stress conditions (matching and mismatching the stress pattern of the real word); altogether, the word 'zee' was embedded in four different bisyllabic strings: 'muZEE' (segmentally and suprasegmentally matching the onset of the real word), 'MUzee' (segmentally matching and suprasegmentally mismatching the onset of the real word), 'luZEE' (segmentally mismatching and suprasegmentally matching the onset of the real word), and 'LUzee' (segmentally and suprasegmentally mismatching the onset of the real word). Results showed a significant main effect of word onset versus nonword onset. Words (such as 'ZEE') were found to be harder to detect when embedded in strings that formed the onset of a real word (such as 'muZEE') than when embedded in strings that did not form the onset of a real word (such as 'luZEE'). The authors attributed this finding to inhibition of the embedded word due to competition from the real word that was activated by the onset. There was no significant main effect of stress pattern; no significant difference was found in the speed of detection of 'ZEE' when it was embedded in 'luZEE' compared to when it was embedded in 'LUzee'. This indicates that 'luZEE', which matches the onset of 'muSEum' on suprasegmental cues was no more effective in activating 'muSEum' than was 'LUzee', which does not match the onset of 'muSEum' for either cue. However, a significant interaction effect was found between the two conditions. The embedded
word 'ZEE' was detected significantly faster in the string 'MUzee' than it was in the string 'muZEE'. According to the authors, this indicates that 'MUzee' was less effective than 'muZEE' in activating 'museum'. The authors argued that this finding indicates that both segmental and suprasegmental information constrains lexical activation in Dutch. However, the fact that a significant main effect of word versus non-word (which was essentially a main effect of segmental information) was found while no significant main effect of stress pattern was found, suggested to the authors that segmental cues make a greater contribution in constraining lexical activation in Dutch than do suprasegmental cues.

The fourth experiment of Cutler and van Donselaar (2001) was designed to explore this possibility. In this experiment, some of the stimuli from the previous experiment were used in a cross-modal priming task. This task required participants to make lexical decisions regarding visually presented items that followed auditory primes. The auditory primes consisted of bisyllabic strings in four different conditions. In the ‘matching’ condition, the prime (for example, 'muZEE') matched the onset of the visually presented item (for example, 'museum') for segmental and suprasegmental structure. In the ‘segmentally mismatching’ condition, the prime ('luZEE') matched the onset of the item ('museum') for suprasegmental structure only. In the ‘suprasegmentally mismatching’ condition, the prime ('MUzee') matched the onset of the item (museum) for segmental structure only. In the ‘mismatching’ or ‘control’ condition, the prime (VIba) did not match the onset of the item ('museum') for segmental or suprasegmental structure. These conditions allowed the experimenters to compare the effects of a segmental mismatch with that of a suprasegmental mismatch on word recognition ability. Response times were measured and the results showed a significant main effect of prime condition. Word recognition was fastest when the prime matched the first two syllables of the target word both segmentally and suprasegmentally. It was slower when the prime mismatched the first two syllables of the target either segmentally or suprasegmentally. Word recognition was the
slowest when the prime mismatched the target both segmentally and suprasegmentally. Post-hoc analysis showed that all conditions were significantly different from one another, with the exception of the segmental mismatch and control conditions. The authors concluded that the finding of slower word recognition for words that suprasegmentally mismatched the primes, as compared with words that segmentally and suprasegmentally matched the primes, provided evidence that suprasegmental information does contribute to lexical activation in Dutch. However, they further argued that the statistically equivalent results obtained in the segmentally mismatching condition and the control condition suggests that segmental cues play a greater role in word recognition in Dutch.

Overall, the authors concluded that the results of their experiments showed that suprasegmental cues do contribute to spoken word recognition in Dutch. Specifically, they argued that Dutch listeners make use of suprasegmental information to constrain lexical activation. However, segmental cues were argued to play a stronger role.

1.4.1.2 Wingfield and his colleagues

Wingfield and his colleagues have also conducted numerous studies exploring the role of suprasegmental cues in spoken word recognition in English (Lindfield et al., 1999; Wayland et al., 1987; Wingfield et al., 1997; Wingfield et al., 2000). Based on their findings, these authors take the position that listeners do exploit suprasegmental cues for word recognition in English.

Wayland, Wingfield, and Goodglass (1987) conducted a study using the gating paradigm to explore the relationship between cohort size and spoken word recognition in English. Their stimuli consisted of 45 nouns (17 one-syllable, 15 two-syllable, 13 three-syllable words); the nouns were chosen so that the range of word frequency was balanced across the three different word-length groups. For the gating task, listeners were instructed to attempt to identify the stimulus word for each successive presentation of the initial portion of the word,
which increased in 50 millisecond increments until the entire word was presented. The findings of this study showed that word recognition can occur even when several words remain in the cohort. According to the authors, this finding implies that the cohort that was activated by the gated portion of the word may have been limited to words with matching stress. If this were the case, the authors argued that the size of the cohort relevant in spoken word recognition could actually be overestimated by counts that include only segmental information.

In a follow-up study, Wingfield, Goodglass, and Lindfield (1997) measured the effects of cohort size and syllabic stress on spoken word recognition for words that were gated either from the beginning of the word or the end of the word. The authors were interested in the differences in listener word recognition performance for forward gated versus backward gated words to explore whether word recognition is based on an overall goodness of fit between the stimulus word and the lexical candidates or whether special priority is given to word initial information. Of particular relevance to the current study, the authors were also interested in whether the stress of the word and the cohort size would influence these differences. The stimulus items included two-syllable words, equal numbers with first-syllable stress (such as 'window') and second-syllable stress (such as 'police'), and three-syllable words, equal numbers with first-syllable stress (such as 'radio'), second-syllable stress (such as 'apartment'), and third-syllable stress (such as 'engineer'). The results of their study indicated that, regardless of the direction of gating, the word recognition threshold was a function of cohort reduction, particularly when syllabic stress was taken into consideration. This finding suggests that listeners do use suprasegmental information in spoken word recognition. In support of their findings, Wingfield and his colleagues (1997) reported the results of a study by Luce (1986), in which it was found that the uniqueness point of a word (the point at which a word becomes segmentally distinct from all other words in the lexicon) does not necessarily occur before the end of the word. The probability of a word segmentally diverging from all other words, at some
point in its duration, was found to be only .39. Again, this indicates that other cues, such as stress pattern, may contribute to the word recognition process.

Lindfield, Wingfield, and Goodglass (1999) conducted a study that was designed to measure the ability to use prosody to facilitate word recognition. In this experiment, a gating task was used to measure the amount of word onset information required for word identification when prosodic information was available compared with a condition without prosodic information. Listeners were presented with two-syllable words (equal numbers with first- and second-syllable primary stress) and three-syllable words (equal numbers with first-, second-, and third-syllable primary stress) in three different experimental conditions. In the first condition, listeners were presented with only the gated onset of the word (onset-only condition). In the second condition, listeners were presented with the gated onset of the word followed by white noise, which provided information regarding the total duration of the word (onset-plus-duration condition). In the third condition, listeners were presented with the gated onset of the word followed by a bandpass-filtered recording of the remainder of the word. This provided information regarding the prosody of the word in the absence of any segmental information (onset-plus-prosody condition). Responses were measured by the mean onset gate size required for word recognition. Results showed a significant advantage in word identification ability for the onset-plus-prosody condition as compared to the onset-plus-duration and the onset-only condition. Listeners were able to identify words with less onset information when the prosody of the remainder of the word was available. Providing the listener with information regarding the duration of the word in the onset-plus-duration condition did not result in an improvement in word recognition ability compared to the onset-only condition. An analysis of pre-recognition errors also showed that earlier error responses sharing the same prosodic pattern as the target word were obtained for the onset-plus-prosody condition, as compared to the other conditions. Based on their results, the authors concluded
that their study “confirms the suggested importance of word stress in spoken word recognition” (p. 315).

Wingfield, Lindfield, & Goodglass (2000) conducted a follow-up study to measure the ability of young and elderly adults (with normal hearing) to use prosody in word recognition. Once again, the gating technique was used for two- and three-syllable words in three experimental conditions: onset-only condition, onset-plus-duration condition, and onset-plus-prosody condition. Results showed a significant main effect of age and condition. Further analyses showed that the significant main effect of age reflected a significant advantage of the younger adults over the elderly adults, for all three conditions. The significant main effect of condition reflected a significant advantage in the onset-plus-prosody condition over the other two conditions, for both age groups; this finding replicated the findings of the previous study (Lindfield et al., 1999). The absence of a significant age x condition interaction reflected the similarity in performance for both age groups in the onset-plus-prosody condition; both groups required significantly less word onset information for word identification in the onset-plus-prosody condition than in either of the other two conditions. The findings of Wingfield et al. (2000) again supported the conclusion that prosodic information contributes to the listener's ability to recognize words. Wingfield et al. (2000) concluded that “listeners have access to stress information in their store of knowledge” and that they “can use this information in conducting a perceptual match between the target stimulus and its perceptual alternatives” (p. 922).

1.4.1.3 Concluding remarks

In the preceding sections, evidence has been presented that supports the view that suprasegmental information does contribute to lexical processing. However, different researchers have taken different perspectives as to the significance of the contribution made by suprasegmental cues in lexical processing in English.
Based on studies conducted in English and other lexical stress languages, Cutler and colleagues (Cutler, 1986; Cutler et al., 1997; Cutler & van Donselaar 2001; Soto-Faraco et al., 2001) have concluded that lexical stress plays a less significant role in lexical processing in English than it does in other lexical stress languages. Soto-Faraco et al. (2001) concluded that, in Spanish, the contribution made by suprasegmental information in processing spoken words is similar to the contribution made by segmental information; both types of information are effective cues in constraining the number of candidates that are initially activated during processing. Cutler and van Donselaar (2001) reached similar conclusions regarding the role of lexical stress in spoken word recognition in Dutch. They concluded that Dutch listeners use suprasegmental cues to constrain lexical activation, though to a lesser degree than segmental cues. Based on their findings and those of Cutler (1986), the authors of both of these studies suggested that English listeners do not exploit suprasegmental cues in spoken word recognition to the degree that listeners in other languages do. Specifically, in contrast to the results of research by Soto-Faraco et al. (2001) and Cutler and van Donselaar (2001), Cutler (1986) concluded that the results of her study indicated that English listeners do not exploit suprasegmental information much in constraining lexical activation. On the other hand, Soto-Faraco et al. (2001) and Cutler and van Donselaar (2001) suggested that the difference across languages in the way that suprasegmental cues contribute to lexical processing can be attributed to the degree to which segmental and suprasegmental cues are independent in these languages. In English, suprasegmental cues tend to be redundant with segmental cues and, therefore, suprasegmental cues do not tend to provide additional useful information in the processing of words. In contrast, suprasegmental cues are less redundant with segmental cues in Dutch, and in Spanish the two types of cues are completely independent of one another. In these languages, therefore, suprasegmental information can potentially contribute additional useful information beyond that provided by segmental cues to the listener. Based on their review of studies on
lexical stress, Cutler et al. (1997) had previously concluded that the evidence suggested that listeners exploit the information that is useful to them in processing their language. Similarly, Soto-Faraco et al. (2001) concluded that “listeners of a language will use all available cues for lexical access that usefully serve to distinguish between words of the language in question” (p. 427). According to this perspective, prosodic information is not as useful to English listeners for lexical access, because it does not provide distinguishing information that is not already provided by the segmental information.

The research conducted by Cutler and her colleagues contributes to a greater understanding of the contribution made by suprasegmental cues in the processing of spoken language (Cutler, 1986; Cutler & Clifton, 1984; Cutler et al., 1997; Cutler & van Donselaar, 2001; McQueen et al., 1994; Soto-Faraco et al., 2001). Specifically, their research shows that suprasegmental cues do contribute to this process, across languages, and they emphasize the need for models of spoken word recognition to address this contribution. However, they discount the contribution that suprasegmental cues make to the processing of spoken words in English. This argument can be disputed based on some of the assumptions and methods adopted in their research.

The focus on the distinction between lexical access and lexical retrieval in the earlier research conducted by Cutler (1986) is not well motivated. Cutler provides no explanation of the precise form that these two presumed distinct operations take in the process of word recognition. Lexical access is presumed to occur earlier in the process of spoken word recognition, followed later by lexical retrieval. However, without a precise description of the timing by which these two different operations are presumed to occur, it is difficult to establish which operation is being measured. In Cutler’s (1986) study, and in later studies, it is assumed that the cross-modal priming task reflects an on-line measurement of lexical access (Cutler & van Donselaar, 2001; Soto-Faraco et al., 2001). However, it is possible that the lexical decision
task involves post-access processing. Specifically, it is likely that the candidates are activated and the target word is retrieved before the lexical decision is made. Therefore, the methods used do not permit the conclusion that lexical access, as opposed to lexical retrieval, is being measured. Furthermore, this perspective is very linear and does not allow for the possibility of interaction between the different stages of processing. Clearly, there must be a specific temporal point at which word retrieval occurs; however, the process leading to retrieval is not necessarily linear. Instead, parallel processing could allow for activation to continue until the very point at which a word is chosen. In addition to these limitations, Cutler's assumption that pairs of words differing only in prosody should generate distinct lexical codes is a very modular perspective; it assumes that there is no interaction between suprasegmental and segmental information in lexical processing.

Recent articles by Cutler and colleagues (Cutler et al., 1997; Cutler & van Donselaar, 2001; Soto-Faraco et al., 2001) focused more on the redundancy in segmental and suprasegmental information in English. Specifically, they concluded that suprasegmental information does not contribute as much in languages in which the redundancy between segmental and suprasegmental information is pervasive, such as English, compared to other languages in which this redundancy does not exist or is less pervasive. However, the argument that suprasegmental cues provide limited information in English is not necessarily reflective of everyday listening situations. While suprasegmental cues may be largely redundant with segmental cues in ideal listening situations, in which all of the signal cues are readily available to the listener, listening situations in everyday experiences are not always ideal. Whether due to environmental, speaker, or listener factors, listeners in everyday situations are often deprived of important information that would be available in the speech signal in ideal situations. In such non-ideal situations, listeners need to exploit information that may be considered redundant in more ‘ideal’ listening circumstances. For instance, background noise, hearing loss, or age-
related declines in temporal processing may limit the listener's ability to exploit segmental cues. In these circumstances the listener would then need to use suprasegmental cues to a greater degree in processing the speech signal; therefore, suprasegmental cues would no longer be a redundant source of information, but instead essential to word recognition. Furthermore, the 'minimal stress pairs' used in previous studies are not reflective of typical words that listeners encounter. In fact, segmentally identical, semantically distinct words that can be distinguished based on suprasegmental information alone are very rare in all three of the languages studied, regardless of the degree of interdependence or independence between the segmental cues and suprasegmental cues in the language. Instead, minimal stress pairs in these languages are more commonly semantically related with differences in stress serving to distinguish grammatical meaning. In fact, even segmentally identical, grammatically related minimal stress pairs would likely not differ only on suprasegmental structure in normal use in these languages, since syntactic and contextual cues would likely also be available and, therefore, provide other types of redundant information.

Some of the methods used in the studies conducted by Cutler and colleagues are also not necessarily reflective of normal listening circumstances. Many of the studies used some form of cross-modal priming. In cross-modal priming, only part of the information is delivered in the auditory modality. The information to which the listener actually responds is delivered in the visual modality, which is not typical of everyday listening. Furthermore, many of the tasks used in these studies required decisions on the part of the listener regarding, for instance, the existence of words or the grammaticality of phrases presented. Again, this is not a typical response required in processing speech in normal situations. The methods used in these studies, therefore, do not reflect the manner in which listeners typically use and process spoken language.
The research conducted by Wingfield and his colleagues (Lindfield et al., 1999; Wayland et al., 1987; Wingfield et al., 2000) is more representative of typical listening conditions, with the exception of the experiment using backward gated words (Wingfield et al., 1997). In their research, the stimuli were commonly used words that can be considered to be more representative of words typically encountered in normal use of spoken language. These words were presented in a gated format. While speech is not typically delivered in such incrementally gated portions, the gating paradigm does allow for a measurement of the listener’s response to information presented entirely in the auditory modality, which is more reflective of the manner in which speech is typically delivered. Whether or not the gating task is actually an on-line measure or whether it reflects post-access processes is still debated in the literature (Grosjean, 1996); this issue will be discussed further in a later section.

In addition to using methods that can be argued to be closer to normal language use, the study by Wingfield et al. (2000) focused on a greater understanding of spoken word recognition for a wider range of typical listeners. Specifically, their study was designed to explore the usefulness of suprasegmental information for two different populations of listeners, young and elderly listeners. Their findings can, therefore, be argued to have implications that are more generalizable to typical listeners.

Furthermore, the research conducted by Wingfield, Goodglass, and Lindfield (2000) explored the contribution of suprasegmental information when part of the segmental information is not available to the listener. This research is more representative of situations in which the listening conditions are not ideal and not all of the information in the signal is available to the listener. In other words, their research explored the contribution of suprasegmental cues when they are not entirely redundant with segmental cues, as might well be the case in everyday situations. Their finding that suprasegmental cues did enable the listener to recognize the word significantly sooner than when these cues were absent indicates
that listeners are able to exploit suprasegmental information, in the absence of segmental
information, in the process of spoken word recognition. This finding is not consistent with the
argument made by previous authors (i.e., Cutler et al., 1997) that the effects of lexical stress in
word recognition may exclusively reflect changes in vowel quality. However, while Wingfield
and colleagues did measure the effects of suprasegmental cues on spoken word recognition,
their research design did not allow for a comparison between the effects of suprasegmental
information alone, the effects of segmental information alone, and the combined effects of
suprasegmental and segmental information. The research design of the current study
incorporated these comparisons.

Another limitation to many of the reviewed studies that were conducted in English
(Cutler, 1986; Lindfield et al., 1997; Wingfield et al., 1997; Wingfield et al., 2000) relates to
the fact that English has three levels of syllable stress: primary, secondary, and unstressed.
Both primary and secondary stressed syllables contain unreduced vowels. Syllables that are
unstressed have reduced vowels. Therefore, a ‘weak’ syllable could actually contain either an
unreduced or a reduced vowel. In the third experiment by Cutler and Clifton (1984), the authors
did control for unreduced versus reduced vowels in the ‘weak’ syllable; half of the words in
this experiment had an unreduced vowel in the ‘weak’ syllable and half of the words had a
reduced vowel in the ‘weak’ syllable. Their results indicated that listeners performed
differently depending on whether or not the ‘weak’ syllable had secondary stress or was
unstressed; specifically, there was a smaller mispronunciation effect for the mistressed words
when the words contained unreduced vowels in both syllables. In other words, they found that
listeners were better able to tolerate shifts in stress when the shifts did not result in a change in
vowel quality. Despite this finding, the other studies reviewed did not control for secondary
versus unstressed syllables. For instance, Cutler (1986) included both words with secondary
stress and unstressed syllables in her stimuli. Similarly, Wingfield et al. (2000) did not control
for secondary stress versus unstressed syllables in their word lists; for example, they included both 'cathedral' (unstressed-primary stress-unstressed) and 'foundation' (secondary stress-primary stress-unstressed) in their lists of words containing stress on the second syllable.

1.5 The current study

The purpose of the current study is to explore the role of suprasegmental information in the recognition of spoken words in conditions that can be considered more representative of 'typical' listening situations; that is, situations in which not all the information contained in the speech signal is necessarily available to the listener. Four conditions will be constructed and used in a gating task: one in which both types of cues are available, one in which the suprasegmental cues are degraded, one in which the segmental cues are degraded, and one in which both types of cues are degraded. These conditions will allow for measurements of the relative contributions of both types of information in the processing of spoken language, in isolation or in combination with one another. This research will increase our understanding of the significance of the contribution made by suprasegmental information to spoken word recognition in English.

The following sections provide a rationale for the methods chosen to degrade the stimuli. These sections are followed by a discussion of the gating paradigm and finally by a discussion of the hypotheses of the present study.

1.5.1 Suprasegmental cues

The lexical stress of the stimuli will be manipulated to alter the suprasegmental information available to the listener. To manipulate the lexical stress information available to the listener, two versions of each stimulus word will be recorded. One will be recorded with the speaker using 'normal' prosody. The other will be recorded with the speaker producing the words in a 'robot-like' or 'monotone' manner. This method has been chosen in order to minimize the effect that altering the stress pattern will have on vowel quality, so that the
contribution of suprasegmental cues can be measured as independently as possible of the
collection of segmental cues.

This method of degrading the suprasegmental information by recording the stimuli in a
‘monotone’ manner has been used in other speech perception studies (Cohen, Douaire, &
Elsabbagh, 2001; Hellier, Edworthy, Weedon, Walters, & Adams, 2002). In the study by
Cutler and Clifton (1984) a similar method was used to mis-stress the speech stimuli. An
alternative method that has been used to degrade the prosodic structure is re-synthesis of the
speech stimuli to achieve a ‘monotone’ quality (Darwin, & Hukin, 2000; Laures, & Weismer,
1999).

1.5.2 Segmental cues

In speech perception research, there are a number of ways to degrade segmental
information. A commonly used method of degrading segmental cues is to present the speech
stimuli in noise, including white noise (Persson, Harder, Arlinger, & Magnusson, 2001;
Pittman, & Wiley, 2001; Studebaker, Sherbecoe, McDaniel, & Gwaltney, 1999) and
background speaker noise (Pittman, & Wiley, 2001; Tun, O’Kane, & Wingfield, 2002; Snell,
Mapes, Hickman, & Frisina, 2002; Snell, & Frisina, 2000). Lindfield et al. (1999) and
Wingfield et al. (2000) used an alternative method to using noise to degrade the segmental
information; they used band-pass filtering to eliminate the segmental cues in their speech
stimuli.

In the current study, a method called ‘jittering’ will be used to degrade the segmental
information available to the listener. Jittering interferes with information at the segmental level
of the stimulus by altering the fine structure of the recorded signal. Subjectively, this has the
effect of distorting the quality of the segmental information in the speech signal, while leaving
the suprasegmental information intact (Brown, 2000; Pass, 1998; Pichora-Fuller, Schneider,
Brown, & Pass, unpublished manuscript; see also Schneider & Pichora-Fuller, 2001). The jitter method will be discussed in greater detail in Chapter 2.

1.5.3 The gating paradigm

In the gating paradigm (Grosjean, 1980), the participant hears an auditory presentation of the initial portion of a word (50 milliseconds [msec] in duration for the current study) and then attempts to identify the word. The initial presentation of the word onset is followed by subsequent presentations that increase in (50 msec) increments until the entire duration of the word has been heard. For each presentation, the participant attempts to identify the word. This technique allows for a direct measurement of how much of the word’s duration needs to be heard before identification can occur. Research has shown that word identification, for words spoken in isolation, can occur within 330 msec of the onset (Grosjean, 1980). Identification of a word prior to its entire duration being presented is presumed to be possible due to a rapid decrease in the number of possible word candidates as increasing portions of the target word are heard. Word identification can then occur when all but one candidate has been eliminated (Marslen-Wilson, 1984; Wayland et al., 1989). This argument has been used by proponents of the Cohort Model as being due to a reduction of word-initial cohorts as more of the word onset becomes available (Marslen-Wilson, 1984). However, other researchers using the gating paradigm have shown word recognition to be a function of cohort reduction, regardless of whether gating is used from the offset in a ‘backward’ direction or from the onset in a ‘forward’ direction (Wingfield et al., 1997).

Presumably, altering the speech signal in a manner that degrades the quality of the signal, and therefore the amount of information it conveys to the listener, necessitates longer portions of the word to be presented before identification can occur. However, this would only be the case for information contained in the signal that makes a significant contribution to the word identification process. Using the gating technique in the current study will allow for a
comparison between how much of the word needs to be presented for identification when the segmental and/or suprasegmental information is degraded versus when the signal is intact. Such a comparison will allow for conclusions to be drawn about the contribution made by different types of information to the word recognition process. It is expected that greater portions of the word will need to be presented when the segmental information in the signal is degraded compared to when the signal is intact, because segmental information is expected to contribute significantly to the word recognition process. Of greater interest in the current study is the question of the role of suprasegmental information in word recognition. If greater portions of the word need to be presented for word identification to occur when the suprasegmental information is degraded, then it could be inferred that the suprasegmental information significantly contributes to the word identification process. However, Grosjean (1996) cautions that it is still debated as to whether or not the gating paradigm can be considered a real on-line paradigm, as gating “may reflect post-access operations” (p. 601). The gating paradigm, therefore, allows for conclusions to be drawn regarding the final outcome of word recognition, but does not necessarily reveal information regarding the intermediate levels of processing.

1.5.4 Hypotheses

It is hypothesized that suprasegmental information will not be found to contribute significantly to lexical access when these cues are redundant with segmental cues. However, it is hypothesized that the role of suprasegmental cues will become significant when these cues are not redundant with segmental cues. Specific hypotheses relating to each of the experimental conditions are presented below.

1.5.4.1 Condition A: Intact segmental and suprasegmental cues

The condition containing words with intact segmental and suprasegmental cues was designed to resemble 'ideal' listening circumstances. Therefore, listener performance for the three conditions containing degraded segmental and/or suprasegmental cues will be analyzed in
relation to the scores obtained in this 'ideal' condition. It is hypothesized that, for the condition with intact segmental and suprasegmental cues, participants will demonstrate the 'best' scores for isolation point (the point at which a word is correctly identified with no change in response thereafter) and for accuracy of word identification.

1.5.4.2 Condition B: Intact segmental and degraded suprasegmental cues

For the condition containing words with intact segmental cues and degraded suprasegmental cues, the degradation of the suprasegmental information is not expected to adversely affect the listener's ability to access lexical items. This expectation is based on the assumption that the contribution of suprasegmental information is not necessary for lexical access in situations in which all of the segmental information contained in the signal is available to the listener; as has been suggested by previous authors (Cutler et al., 1997; Cutler & van Donselaar, 2001; Soto-Faraco et al., 2001). It is, therefore, hypothesized that listener performance for isolation point and accuracy of word identification in this condition will be equivalent to the performance achieved for words with intact segmental and suprasegmental cues.

1.5.4.3 Condition C: Degraded segmental and intact suprasegmental cues

For the condition containing words with degraded segmental and intact suprasegmental cues, it is expected that the suprasegmental cues will be found to contribute significantly to the listener's ability to access lexical items. This expectation is based on research by Wingfield and colleagues (Wingfield et al., 1997; Wingfield et al., 2000), in which they found that providing listeners with information regarding the suprasegmental structure of a portion of the word resulted in faster word recognition scores than when neither segmental nor suprasegmental cues were available to the listener. It is, therefore, predicted that the intact suprasegmental information in this condition will allow the listener to compensate for the degraded segmental cues. This will be demonstrated through the attainment of higher scores when only
suprasegmental cues are available to the listener, compared with the condition in which both cues are degraded. It is also expected that the intact suprasegmental cues will not be sufficient to compensate entirely for the degradation of the segmental cues. It is, therefore, predicted that the degraded segmental information in this condition will result in significantly lower scores than those obtained in the condition with both cues intact.

1.5.4.4 Condition D: Degraded segmental and degraded suprasegmental cues

When both segmental and suprasegmental cues are degraded, it is hypothesized that listener performance for isolation point and accuracy of word identification will be significantly lower than the scores obtained in the conditions in which either cue or both cues are intact.

In summary, it is predicted that the ability to recognize spoken words will not significantly differ between the conditions in which both cues are intact and segmental cues alone are intact. Listener performance for the condition in which suprasegmental cues alone are intact is expected to be lower than the performance achieved in either of the previous conditions. Listener performance is expected to be the most adversely affected when both cues are degraded.
2. METHODS

2.1 Overview

The methods used in conducting the gating experiment will be discussed in this chapter. A description of the participants will be presented in the first section, followed by sections describing the experimental stimuli and procedures.

2.2 Participants

Participants were recruited by advertisements posted at the University of British Columbia and were required to attend one or two sessions, totalling two to three hours. Prior to testing participants completed two copies of an informed consent form, one for their records and the other for the records of the experimenter (see Appendix I). They were screened using a language and hearing questionnaire, the Mill Hill vocabulary test (Raven, 1938), and audiometry to ensure that there were no factors which might have a negative influence on their ability to recognize words. Participants were paid an honorarium of $10/hour for their participation.

The participants consisted of 16 adults, 7 males and 9 females, ranging in age from 18 to 35 years (M=25.4, SD=5.6). The group mean for years of formal education was 15.9 (SD=2.13). The group had a mean score of 15/21 (Range=11 to 18; SD=1.86) on the Mill Hill vocabulary test, a measure intended to test vocabulary in normal adults (Raven, 1938). This mean score is comparable to the mean scores found for participants in this age range in other studies (Brown, 2000; Pass, 1998). All participants were native speakers of English and all had normal hearing. Hearing ability was determined by a pure-tone air conduction test that was conducted prior to experimental testing. Normal hearing was defined as being 20 dB HL or better at the frequencies 250 through 4000 Hz. Therefore, the listeners were considered to be normal adults with above average performance on measures that might influence ability to recognize words.
2.3 Materials

2.3.1 Stimuli description

The stimuli used in the gating experiment consisted of 100 words (see Appendix II) recorded in two different conditions (normal prosody and degraded prosody), resulting in a total of 200 recorded stimuli. The 200 recorded stimuli were then manipulated to create two different segmental conditions (normal segments and degraded segments), resulting in a total of 400 stimuli. Each of the 400 stimulus words was then gated into 50 msec gates for presentation in a gating task.

All of the 100 target words were nouns and had frequencies of 75 or less in the Kucera and Francis corpus (1967). The 100 words consisted of two- and three-syllable words, some of which were taken from Wingfield, Lindfield, & Goodglass (2000) and the remainder of which were chosen from the MRC Psycholinguistic Database website (2002). Originally, there were 40 two-syllable words chosen (mean duration of 651 msec), with equal numbers having first and second syllable primary stress; the remaining 60 words chosen were three-syllable words (mean duration of 748 msec), with equal numbers having first, second, and third syllable primary stress. Due to limitations related to counterbalancing word frequency and word density across the different stress pattern categories, the ‘weak’ syllables in the words were not controlled for secondary stress versus no stress. The five different ‘primary’ stress patterns were chosen to include a variety of different stress patterns in the experiment, with equal numbers of words in each stress pattern category. However, there was some discrepancy as to the actual stress pattern of the recorded words in the experiment. Specifically, two of the words originally chosen (‘brochure’ and ‘registrar’) have alternate stress patterns than the category in which they were originally placed and it was decided that, in both cases, it was the alternate pattern that was used in the current experiment. Therefore, these words were moved to the appropriate list and, as a result, the number of words in each of the five stress pattern categories
varied from nineteen to twenty-one. Words across the five different types of stress pattern were matched for word frequency (Kucera & Francis, 1967). The 100 words were divided into four lists, each of which contained five words of each type of stress pattern, with the exception of the two ‘exception’ words explained above (see Appendix III). In total, there were 25 words in each list. The four lists were created to counterbalance presentation of the stimuli across the four experimental conditions. Each list was balanced for word frequency (Kucera & Francis, 1967). The 100 words were also matched for neighborhood density; all of the words had zero or one neighbours (Sommers, 2003).

2.3.2.2 Stimuli preparation

2.3.2.2.1 Recording the stimuli

The stimulus words were recorded by an adult female native speaker of English (not the experimenter) in the carrier phrase “the word is... today.” Each word was recorded in both the normal prosody and the degraded prosody conditions. For the normal prosody condition, the speaker used natural prosody. To achieve the degraded prosody condition, the speaker spoke in a ‘robot-like’ or ‘monotone’ manner, targeting equal pitch, duration and volume across syllables. A metronome was used to aid the speaker in achieving this goal.

The stimuli were recorded in the speech recording program of the Computer Speech Research Environment 4.5 (CSRE 4.5, 1995) in .adf format at a sampling rate of 20,000 Hz. The recordings were made in a double-walled, sound attenuating IAC booth, using a Sennheiser model K3U microphone positioned approximately 6 inches from the speaker’s mouth. The stimuli were converted from analog to digital format via the DD1 component of a Tucker Davis Technology system and saved on the computer hard drive using CSRE 4.5 (1995).
2.3.2.2.2 Editing the stimuli

Once recorded, the stimuli were converted to .bi files and saved to a compact disc, to be edited on a different computer. The recordings were edited in the Cool Edit 2000 (2000) computer software program to delete the carrier phrase in order to isolate the target word. Deletions were made at zero crossings at points judged by the experimenter to contain the entire target word with minimal effects of co-articulation by the preceding and following words of the carrier phrase. Decisions made regarding the deletion points were based on both the auditory stimulus and the visual representation of the waveform.

During the editing stage, several criteria needed to be met in order for words to be included as stimuli. If these criteria were not met, the words were re-edited or re-recorded. First, each target word was matched for duration across the two stress conditions. The duration of the two recordings (normal prosody and degraded prosody) of each word was targeted to be within 30 msec, of one another. The mean durations of the final stimuli were 706 msec for words in the normal prosody condition and 712 msec for words in the degraded prosody condition. A second criterion for inclusion as stimuli required that each word be matched across the two stress conditions for the vowel in the first syllable. The first vowel of each word needed to begin in the same 50 msec portion of the word in both recordings. This ensured that the first vowel in both tokens of each word would begin in the same gate, once the stimuli were gated. As a final criterion at this stage, words in the degraded prosody condition needed to be judged qualitatively by the experimenter to sound to have sufficiently equal stress (degraded prosody) across syllables. If any of the recorded stimuli did not meet the above criteria, they were re-edited or re-recorded.

2.3.2.2.3 Subjective judgements of the stimuli

Once 200 recordings were made that met the above criteria, the next stage in determining whether or not the recordings would be included as experimental stimuli involved
judgements by four naive listeners regarding the quality of these recordings. The listeners heard
the isolated words presented through headphones in a quiet room. Each listener heard all 200
recordings of the 100 words, in pseudo-random order. For each word, two listeners heard the
degraded prosody version first and two listeners heard the normal prosody version first. The
listeners were asked to first identify the word being presented and to then give a rating of the
prosody of the word from 1 (monotone) to 5 (normal). Listeners were also encouraged to
provide subjective comments about the recordings. In order to be included as stimuli, the
recordings needed to be judged to meet several criteria, administered in a series of steps. The
first step was to include all normal prosody stimuli rated by at least three of the four people as
5, with no ratings of 3, and all degraded prosody stimuli rated by at least three people as 1, with
no ratings of 3. The second step involved re-recording, re-editing, or replacing any stimulus
word for which one of the following occurred: a rating of 3 was given by at least one person,
the experimenter agreed about a negative subjective comment, more than one person made the
same negative subjective comment, correct identification had not been made by one or more
listeners. Next, the original four listeners again gave subjective ratings of the quality of the
newly recorded or re-edited words and two new naive listeners were asked to identify the new
words that had been recorded as replacement words. As the next step, the first step was again
administered for any additional words that now fit the criteria. The fifth step involved re-
assigning the words in the original four lists that had been created for presentation in the four
experimental conditions. The words in the four lists were re-assigned to balance the words not
meeting the above criteria across the four lists, while maintaining the balance of word
frequency across the four lists. Next, additional recordings previously not included were now
included as stimuli and equally distributed across the four lists. At this stage, stimuli were
included if a rating of 4 or 5 was given in the normal prosody condition by at least two of the
four people, if a rating of 1 or 2 was given in the degraded prosody condition by at least two
people, if the experimenter did not agree with a negative subjective comment that had been made, or if the experimenter had re-edited a stimulus word according to a subjective comment. At this point, few recordings were left that had not met the above criteria. These words were distributed evenly across the four lists; some words in the four lists were re-assigned so that the balance for word frequency was maintained across the four lists.

2.3.2.2.4 Jittering of the stimuli

After the recordings to be used in the experiment had been chosen, the stimuli were saved on compact disc for transfer back to the computer containing the CSRE 4.5 (1995) software. The stimuli were then saved on the hard drive of this computer and a second set of the 200 recordings was created, resulting in a total of 400 stimuli. The 200 words in the second set were created in order to alter the segmental information contained in the speech signal, using a jitter program (for a description see Brown, 2000).

The jitter program manipulates the temporal fine structure of speech. In speech, the information contained in the speech signal is delivered sequentially over time. When the speech signal is digitized, the sampling rate at which it is digitized dictates the number of amplitude values per second used to define the sound. In the current study, a sampling rate of 20,000 Hz was used in recording the speech stimuli. This translates to a signal with 20,000 amplitude values per second, or a duration of 0.05 milliseconds between amplitude values. When the digitized speech signal is jittered, a temporal asynchrony is introduced to the signal. This is accomplished by shifting amplitude values in the speech signal from their original temporal positions, which has the effect of altering the fine structure of the speech signal or degrading the segmental cues that are available to the listener. Jittering does not affect the overall duration, average amplitude, or envelope structure of the whole speech signal and, therefore, the suprasegmental cues are left intact. Subjectively, the jittered signal is perceived by the listener as a 'distortion' of the segmental cues. Note that the human voice is not perfectly
periodic so that, in effect, the jittering algorithm can be thought of as slightly increasing the amount of aperiodicity of the recorded voice.

In the current study, to control the amount of jitter applied to the stimuli, the following values were entered into the program: standard deviation (SD) of 10 (equivalent to 10 x .05 = 50 msec), frequency range of 10,000 Hz, and bandwidth of 100 Hz. These values were determined based on research conducted by A. Heinrich (personal correspondence) that found that young adults performed at 91% word identification accuracy with these values. These values also ensured that all segments with spectral information in the 0 to 10,000 Hz range (which included all of the segments) would be affected. Each of the stimulus words was individually jittered. After the words had been jittered, the 100 stimulus words for all four experimental conditions were complete.

2.3.2.2.5 Gating the stimuli

The next stage of stimulus preparation was gating. Each of the 400 stimulus items was individually edited into gates that increased incrementally in duration by 50 milliseconds. The gates were created in Cool Edit (2000) and saved on the computer hard drive as individual sound files. Most of the words had durations that were not evenly divisible by 50 msec and, therefore, had final gates that were less than 50 msec long; each of the final gates that was shorter than 50 msec in duration was also saved as an individual sound file.

2.3.2.2.6 Creation of the experimental protocol

The stimuli were then converted from .bi format back to .adf format. Sixteen different experimental blocks were created in the CSRE 4.5 (1995) Ecosgen program. Each of the blocks contained the gated stimuli of one of the four lists of 25 words in one of the four conditions. Each list was assigned a quasi-random order of words. The quasi-random order was created by randomly assigning one word of each syllable pattern type to each successive group of five
words. The same order was used for each list across all conditions. The experimental blocks were designed to present each 50 msec sound file of each word in the list in a sequential order.

2.4 Procedure

2.4.1 Participant instructions

Prior to beginning the experimental testing, participants were informed that they would be listening to words presented over headphones in a sound booth. They were told that these words would consist of nouns that were familiar to them. The words would be presented to them in portions, the duration of which would increase in increments until the entire word was presented. Their task was to guess the word that was being partially presented by making a verbal statement into a microphone after each presentation. Participants were asked to give some response with each presentation, even if the response was to inform the experimenter that they had no guess or if they were certain that they had already correctly identified the word. Participants were informed that they would be listening to four sets of words, some of which might be more difficult to understand than others. They were told that they would be given an opportunity to take a break between each list and that they could also request a break at any time during testing.

2.4.2 Experimental testing

Testing was conducted in a double-walled sound-attenuating IAC booth. The experiment was played through the Ecoscon program in CSRE 4.5 (1995). The stimuli were converted from digital back to analog format through a Tucker Davis Technology System II and presented to participants binaurally through headphones. The experimenter controlled the presentation of the stimuli to the participant. Each stimulus was presented to the listener and was followed by a pause, during which time the participant would verbally give a response. The experimenter listened to the participant's response through headphones and recorded it on an answer form. The experimenter then presented the next stimulus.
In total, participants were presented with four blocks of 25 gated words each. The experiment was designed so that each participant heard all four lists with each presented in one of the four different conditions. All participants heard the intact segmental and suprasegmental condition first, followed by the other three conditions in different orders counterbalanced across the participants. The four lists were counterbalanced so that across the 16 subjects each list was presented in each condition four times.

At the end of testing, participants were debriefed regarding the purpose of the experiment and were paid an honorarium of $10/hour.
3. RESULTS

3.1 Overview

A description of the scoring procedures used to calculate the results are provided in the next section, followed by results from the gating experiment.

3.2 Scoring Procedures

Participant responses to the gated stimuli in the four experimental conditions were scored for the isolation point. The isolation point has previously been defined as "the size of the segment (measured in msec or % stimulus) needed to identify the stimulus (without any change in response thereafter)" (Grosjean, 1996). In the current experiment, the isolation point was measured as the number of 50 msec gates needed to identify the word (without any change in response thereafter). For words that were not identified on the last gate, the missing isolation point was replaced by the full duration of the word (measured in total number of gates), as suggested by Grosjean (1996).

In scoring for isolation points, the results were found to include a substantial number of words with missing isolation point values (see accuracy results below). This finding was not unexpected. It seemed likely, prior to testing, that some words would remain unidentified even after being presented at full duration, particularly those words for which both segmental and suprasegmental cues had been degraded. Given the number of words with missing isolation points, response accuracy was also scored, as an additional measure of the role of segmental and suprasegmental cues in lexical retrieval. Whereas measuring the isolation point provided information about the different points at which words with varying segmental and suprasegmental cues could be identified, measuring the accuracy of responses provided information about the overall ability to identify words in these conditions. Response accuracy was measured as the percentage of words correctly identified on the final gate.
3.3 Isolation Point Results

3.3.1 Overview

The main effects of the experimental variables (prosody and jitter) on isolation point are described first, followed by a description of the interaction between these two variables. The analyses were conducted using a 2 (stress) x 2 (jitter) within-subjects analysis of variance (ANOVA).

3.3.2 Main Effects

3.3.2.1 Prosody

The mean number of gates needed to identify words containing normal prosody was 7.9 (SD = .16). In the degraded prosody condition, an average number of 10.3 gates (SD = .22) was required for word identification. The results indicate that listeners required significantly more gates to identify a word when the prosody was disrupted, as compared to when the prosody was intact $F(1, 15) = 108.42, p < .0005$.

3.3.2.2 Jitter

Results showed a mean isolation point of 8.4 gates (SD = .18) in the no-jitter condition. The mean number of gates needed to identify words in the jitter condition was 9.8 (SD = .16). Overall, listeners required significantly more gates to accurately identify words with degraded segmental information, as compared to words with intact segmental information $F(1, 15) = 117.17, p < .0005$.

3.3.3 Interaction Effects

3.3.3.1 Prosody x Jitter

Results of the repeated measures ANOVA showed a significant two-way interaction between prosody and jitter $F(1, 15) = 10.97, p < .01$. 

52
3.3.4 Follow-up planned comparisons

The results obtained through the ANOVA indicated that significant differences existed between the four experimental conditions. To explore these differences further, the mean isolation points in the four experimental conditions were compared (see Table 3.1).

Table 3.1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean Gates</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Normal prosody/normal segments</td>
<td>7.4</td>
<td>0.83</td>
</tr>
<tr>
<td>B. Degraded prosody/normal segments</td>
<td>8.4</td>
<td>0.55</td>
</tr>
<tr>
<td>C. Normal prosody/jittered segments</td>
<td>9.3</td>
<td>1.06</td>
</tr>
<tr>
<td>D. Degraded prosody/jittered segments</td>
<td>11.2</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Using paired samples T-tests, the results showed that the mean isolation point in each of the four experimental conditions differed significantly from the mean isolation points for all of the other conditions (in all conditions $p < .01$). In addition to the main effect of prosody and the main effect of jitter, the results of the paired samples T-test indicated that there was a significant difference between the two conditions in which only the segmental or the suprasegmental cues alone were degraded ($t(15) = 3.42, p = .004$). Specifically, the results of the T-tests indicated that listeners required significantly more gates to identify words when only the segmental information was degraded, compared to when only the suprasegmental information was degraded.

Overall, the results calculated for isolation point indicated that the number of gates required for word identification with intact segmental/suprasegmental cues was significantly fewer than the number of gates required when either or both of these cues were degraded. When only one cue was degraded, degraded segmental cues were significantly more detrimental to word identification than were degraded suprasegmental cues.
3.4 Accuracy Results

3.4.1 Overview

Results based on the percentage of words correctly identified are presented in this section. The main effects of two of the experimental variables (prosody and jitter) on accuracy of identification are described first, followed by a description of the interaction effect between these two variables. A 2 x 2 within-subjects ANOVA was conducted to assess the significance of the effects of prosody and jitter on the accuracy of identification.

3.4.2 Main Effects

3.4.2.1 Prosody

On average, words in the normal prosody condition were identified with an average accuracy of 97% (SD = .01). In the degraded prosody condition, words were identified with an average accuracy of 86.5% (SD = .02). These results indicate that listeners were able to accurately identify significantly more words with normal prosody, as compared to words with degraded prosody $F(1,15) = 83.73, p < .0005$.

3.4.2.2 Jitter

Words with normal segments were identified with 96.2% accuracy (SD = .01), on average. Words with jittered segments were identified with an average accuracy of 87.3% (SD = .02). According to these results, listeners accurately identified significantly more words with normal segments than words with jittered segments $F(1,15) = 50.63, p < .0005$.

3.4.3 Interaction Effects

3.4.3.1 Prosody x Jitter

Results of the ANOVA showed a significant two-way interaction between prosody and jitter $F(1,15) = 20, p < .0005$. These results indicate that listeners accurately identified significantly fewer words when both segmental and suprasegmental cues were degraded, as compared with dégradation of either cue alone (see below).
3.4.4 Follow-up planned comparisons

The results obtained through the ANOVA indicated that significant differences existed between some of the four experimental conditions. To further explore these differences, the average accuracy of word identification across the four experimental conditions was compared using paired samples T-tests. The results are presented in Table 3.2.

Table 3.2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean Accuracy</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Normal prosody/normal segments</td>
<td>98.5%</td>
<td>.01</td>
</tr>
<tr>
<td>B. Degraded prosody/normal segments</td>
<td>94.0%</td>
<td>.01</td>
</tr>
<tr>
<td>C. Normal prosody/jittered segments</td>
<td>95.5%</td>
<td>.01</td>
</tr>
<tr>
<td>D. Degraded prosody/jittered segments</td>
<td>79.0%</td>
<td>.03</td>
</tr>
</tbody>
</table>

Results of the T-tests showed that significant differences in accuracy scores ($p < .01$) existed between each of the conditions, with one exception. There was no significant difference in the accuracy of identification between conditions B (degraded prosody/normal segments) and C (normal prosody/jittered segments) ($t(15) = -1.57$ ($p = .138$). This finding differs from the previously discussed finding for mean isolation point, in which a significant difference was found between the two conditions in which only one type of cue was degraded. This difference in findings for the dependent measures suggests that the point at which a word can first be accurately identified is more adversely affected when segmental cues are degraded than when suprasegmental cues are degraded; however, the overall ability to identify words accurately does not differ according to which of the cues is degraded.

Overall, the results for accuracy of identification indicated that the ability to identify words accurately was significantly better when both segmental and suprasegmental cues were intact than when either or both cues were degraded. The ability to identify words accurately was found to be significantly worse with degradation of both segmental and suprasegmental
cues compared with words with intact cues. There was no significant difference found between
the ability to identify words accurately when only segmental cues were degraded compared
with degradation of only suprasegmental cues. Two intact cues resulted in significantly higher
accuracy scores than one intact cue (regardless of which cue) and one intact cue resulted in
significantly higher accuracy scores than two degraded cues.
4. DISCUSSION

4.1 Introduction

The current study was conducted to explore the role of suprasegmental information in spoken word recognition. To measure the contribution made by suprasegmental information, both segmental and suprasegmental cues were manipulated to create four different experimental conditions. The use of a gating paradigm allowed for measurement of listener performance for isolation point and accuracy of word identification when both cues were intact compared with when either cue or both cues were degraded. Specific hypotheses were stated in the first chapter regarding the expected results for listener performance in each condition. The following section discusses the results as they relate specifically to each of the stated hypotheses. Following this, a general discussion of the current study is presented.

4.2 Discussion of the results

4.2.1 Condition A: Intact segmental and intact suprasegmental cues

The condition containing intact segmental and suprasegmental cues was designed to resemble 'ideal' listening circumstances. Therefore, it was hypothesized that listeners would obtain the 'best' scores for isolation point and accuracy of word identification in this condition and that no other condition would result in scores higher than those obtained in this condition. The results for mean isolation point and accuracy of word identification confirmed this prediction.

The score for the mean isolation point for word identification in this condition was 7.4 gates, significantly lower than the mean isolation point for any other condition. This finding will be discussed further in the following sections as it relates to the stated hypotheses for the other experimental conditions. The score of a mean isolation point of 7.4 gates in this condition equates to a mean duration of 370 msec, which is similar to Grosjean’s finding that words presented in isolation can be identified within 330 msec of the word onset (Grosjean, 1980).
With respect to the score for accuracy of word identification (98.5%), the hypothesis stated for this condition was again supported. In all other conditions, the scores for accuracy of word identification were significantly lower. Again, this finding will be discussed further as it relates to other stated hypotheses.

4.2.2 Condition B: Intact segmental and degraded suprasegmental cues

For the condition containing words with intact segmental cues and degraded suprasegmental cues, the degradation of the suprasegmental information was not predicted to affect the listener's ability to recognize spoken words. This prediction was based on arguments made by Cutler and colleagues (Cutler et al., 1997; Cutler & van Donselaar, 2001; Soto-Faraco et al., 2001); they argued that because suprasegmental information is redundant with segmental information in English, suprasegmental information is not exploited by English listeners. Because segmental information was kept intact in this condition, it was assumed that degradation of the suprasegmental cues would not significantly affect the results. The hypothesis regarding this condition, therefore, stated that the scores obtained in this condition would be statistically equivalent to the scores obtained for the condition with both cues intact. The results of the current study did not support this hypothesis.

The score for mean isolation point of words in this condition was 8.4 gates. This number was significantly higher than the mean score of 7.4 gates obtained with both cues intact. This finding indicates that a significantly longer portion of the word needed to be presented for word identification to occur when the suprasegmental information was degraded compared with when both cues were intact. This finding contradicts the conclusions made by Cutler et al. (1997) discussed above. The fact that intact segmental cues in the current study were not sufficient to enable listeners to reach the scores achieved when both cues were intact indicates that suprasegmental cues are exploited by listeners in English and that these cues do
significantly contribute to the process of word recognition, beyond the contribution made by segmental cues.

The mean score for accuracy of word identification in this condition was 94%. Again, this score is significantly lower than the score of 98.5% that was obtained with both cues intact. This finding provides additional support to the argument that suprasegmental cues do, in fact, significantly influence the process of word recognition in English.

4.2.3 Condition C: Degraded segmental and intact suprasegmental cues

For the condition containing words with degraded segmental and intact suprasegmental cues, it was expected that the intact suprasegmental cues would be found to contribute significantly to word recognition. This expectation is contrary to the conclusions made by Cutler et al. (1997), in which they argued that English listeners do not exploit suprasegmental cues because they are redundant with vowel quality cues; however, their argument was based on listening conditions in which the segmental information contained in the speech signal is entirely available to the listener. In normal listening circumstances the information available in the speech signal is often limited. In situations in which there is limited segmental information available to the listener, suprasegmental cues are not redundant with segmental cues. In designing the current study, it was assumed that listeners do exploit suprasegmental cues for word recognition in these situations. Support for this assumption can be found in the research conducted by Wingfield, Goodglass, and Lindfield (1997) and Wingfield, Lindfield and Goodglass (2000), in which they found that providing listeners with information regarding the suprasegmental structure of the final portion of a word resulted in faster word recognition scores than when neither segmental nor suprasegmental cues were provided. A more extreme example of the usefulness of suprasegmental cues when all segmental cues are obliterated is demonstrated by experiments in which word recognition is maintained when the band-passed speech envelope is filled with white noise and no fine structure is retained (Shannon, Zeng,
Kamath, Wygonski, & Ekelid, 1995). In designing the current study, it was expected that degrading the segmental cues would result in lower scores than those obtained with both cues intact; however, it was also expected that the significant contribution of suprasegmental cues would be demonstrated by higher scores obtained when the suprasegmental cues were intact than when they were degraded. The hypothesis for this condition, therefore, stated that scores for isolation point and accuracy of identification would be significantly lower than the scores obtained in the condition in which both cues were intact and significantly higher than the scores obtained in the condition in which both cues were degraded. The results for this condition supported this hypothesis.

The score for mean isolation point for word identification in this condition was 9.3 gates. As predicted, this number is significantly higher than the mean number of gates required for word identification when both cues were intact. This indicates that intact suprasegmental cues were not sufficient for listeners to recognize words as early as when both cues were available. Also as predicted, words in this condition were recognized significantly faster than words in the condition in which both cues were degraded. This finding indicates that suprasegmental cues do significantly contribute to word recognition in situations where these cues are not redundant with segmental cues.

In addition to this finding, the results showed that the mean number of gates required for word identification with the degradation of one cue was significantly higher when segmental cues were degraded compared with when suprasegmental cues were degraded. This finding was not predicted, but it suggests that segmental cues may play a greater role in word identification than do suprasegmental cues; however, this interpretation cannot be conclusively stated due to limitations relating to the manner in which the stimuli were manipulated to achieve degraded segmental and suprasegmental cues, respectively. This issue will be discussed further below. This finding does potentially provide support to the arguments made.
by Cutler and colleagues that segmental cues play a greater role in lexical retrieval in English than do suprasegmental cues (Cutler et al., 1997; Cutler & van Donselaar, 2001; Soto-Faraco et al., 2001). However, as discussed in the previous and following sections, contrary to their assumptions, the current findings indicate that suprasegmental cues do significantly contribute to the process of word recognition in English.

The mean score for accuracy of word identification in this condition was 95.5%. This score was significantly lower than the 98.5% accuracy obtained in the condition with both cues intact. Again, this score indicates that intact suprasegmental information was not sufficient to allow listeners to perform with the same level of accuracy as when both cues were intact. However, this score was significantly higher than the score of 79% accuracy achieved with both cues degraded. This score indicates that suprasegmental information does contribute to the process of word recognition when segmental cues are degraded.

In addition to the predicted results, it was found that the mean score for accuracy of word identification in this condition was not significantly different from the 94% accuracy achieved when suprasegmental cues alone were degraded. This finding suggests that degrading either cue adversely affected word recognition abilities. Again, this cannot be stated conclusively due to limitations with the stimuli, as will be discussed below. This finding potentially contradicts the arguments made by Cutler and colleagues that segmental cues play a greater role than suprasegmental cues in word recognition in English (Cutler et al., 1997; Cutler & van Donselaar, 2001; Soto-Faraco et al., 2001).

4.2.4 Condition D: Degraded segmental and degraded suprasegmental cues

For the condition containing words with both segmental and suprasegmental cues degraded, it was expected that the combined effect of degrading both cues would significantly adversely affect the listener's ability to recognize spoken words. It was hypothesized that the scores obtained for isolation point and accuracy of word identification in this condition would
be significantly lower than the scores obtained in conditions in which either cue or both cues were intact. The results supported this hypothesis.

The score for mean isolation point in this condition was 11.2 gates, significantly higher than the mean number of gates for any of the other conditions. This indicates that the combined effect of degrading both segmental and suprasegmental cues results in significantly longer portions of the word being required for word identification than when both cues are intact or either cue is intact.

The score for accuracy in this condition was 79%, which was significantly lower than the percentage correct for any of the other conditions.

4.3 General discussion

The general pattern of results indicated that the highest listener performance, that is, the minimal number of gates required for word recognition and accuracy of word identification, was achieved when both cues were intact; the lowest listener performance was found when both cues were degraded. When either of the two cues was degraded, listeners scored between these two extremes on both measures. Despite the fact that this overall pattern of results does not entirely support the hypotheses stated in the introduction, as was discussed in the preceding sections, it seems plausible that the availability of both cues in normal listening conditions would allow for better word recognition than when one of the two cues was degraded. Similarly, it seems plausible that the availability of one cue would allow for better word recognition than when both cues were degraded.

The two different measures of word recognition, the number of gates required for word recognition and the accuracy of word identification, resulted in different findings between the two conditions in which only one cue was degraded. For the mean isolation point measure, results indicated that degrading the segmental cues was significantly more detrimental to listener performance than was degrading the suprasegmental cues. For the accuracy of
identification measure, there was no significant difference in scores achieved between the two conditions. These two different measures allow for different conclusions to be reached. The first finding allows for the conclusion that segmental cues do contribute more to the word recognition process than do suprasegmental cues. The second finding, on the other hand, suggests that both cues contribute equally to the word recognition process. Since the two measures reflect different tasks and aspects of processing, both findings may provide insight into the relative contributions of segmental and suprasegmental cues in normal listening circumstances. The results of the isolation point measure suggest that degrading the segmental information is more detrimental to the isolation point than is degrading the suprasegmental information. Since the isolation point measure reflects the earliest point at which the word can be consistently recognized, this finding suggests that segmental cues contribute more significantly than suprasegmental cues to early identification of the word. On the other hand, the accuracy of word identification measure reflects the overall ability to identify the word by its end. The fact that there was no significant difference between the effect of degrading segmental cues and that of degrading suprasegmental cues on the accuracy of word identification suggests that both cues may contribute equally to the overall ability to identify the word. However, these implications cannot be conclusively stated due to potential limitations with the manner in which the stimuli were degraded; these limitations will be discussed below. Despite the limitations involved, both measures clearly indicate that degrading either cue significantly adversely affects the ability to recognize words, compared with when both cues are intact; this provides strong evidence that both cues significantly contribute to lexical retrieval in English.

In previous articles by Cutler and colleagues (Cutler, 1986; Cutler et al., 1997; Cutler & van Donselaar, 2001; Soto-Faraco et al., 2001) the authors argued that listeners do not exploit suprasegmental cues in the process of word recognition in English because segmental cues are
sufficient for word recognition to occur. This argument was based on the fact that suprasegmental cues are redundant with segmental cues in English spoken words, in that vowels in unstressed syllables are usually produced in a reduced form. The findings of the current study contradict the general argument made by Cutler and colleagues and prompt some qualifications to it. While suprasegmental cues may not provide additional information in the word recognition process when these cues are redundant with segmental cues, it was predicted in the current study that suprasegmental cues would be found to contribute significantly in conditions in which segmental cues were not fully available to the listener. This is frequently the case in everyday life. Results of the current study confirmed this prediction. Furthermore, the results indicated that the suprasegmental cues also contributed significantly when these cues were presented with segmental cues. This contribution was demonstrated by the significant decrease in listener performance for the condition with intact segmental cues and degraded suprasegmental cues, compared to the condition in which both cues were intact. However, as was discussed previously, the scores for mean isolation point in the current study suggest that segmental cues may play a greater role than suprasegmental cues in spoken word recognition in English. These findings parallel those found in Dutch (Cutler & van Donselaar, 2001) that supported the conclusion that suprasegmental cues significantly contribute to spoken word recognition in Dutch, though segmental cues were argued to play a more significant role. The similarity in the results of the current study and those of Cutler and van Donselaar (2001) is interesting, particularly considering the fact that these two languages share many features, including redundancies at the segmental and suprasegmental levels.

Previous research by Wingfield and colleagues had indicated that suprasegmental cues contribute significantly to word recognition (Lindfield, Wingfield, Goodglass, 1997; Lindfield, Wingfield, & Goodglass, 1999; Wayland, Wingfield, & Goodglass, 1989; Wingfield, Goodglass, & Lindfield, 1997; Wingfield, Lindfield, & Goodglass, 2000). However, as was
discussed previously, the design of their research only provided direct evidence regarding the
contribution of suprasegmental information to the process of spoken word recognition. The
design of the current study allowed for a comparison between each of these cues presented in
isolation, as well as both cues presented together. The results, therefore, reflect the contribution
of suprasegmental cues when they are redundant with segmental cues and when they are only
partially redundant with these cues. The results of the current study support the findings of
Wingfield and colleagues. They are also consistent with those of Shannon and his colleagues
who found that speech recognition was largely preserved when spectral information (segmental
cues) was greatly reduced (Shannon et al., 1995). In the current study, the condition in which
the suprasegmental cues were intact resulted in significantly higher scores than the condition in
which both cues were degraded; these conditions are similar to those in the studies of
Wingfield and colleagues. However, as was discussed previously, the results of the current
study also indicated that suprasegmental cues significantly contribute to the processing of
spoken words when these cues are redundant with segmental cues, which the research of
Wingfield and colleagues was not designed to measure.

4.3.1 Relevance to models of spoken word recognition

The results of the current study provide further support for the argument made by
numerous authors that current models of spoken word recognition need to account for
suprasegmental information in their design (Cutler & van Donselaar, 2001; Lindfield et al.,
1997; Lindfield et al., 1999; Soto-Faraco et al., 2001; Wingfield et al., 1999). There is currently
no model of spoken word recognition that considers the prosodic structure of the input; instead
current models have focused entirely on the featural or segmental information in the input.
Furthermore, empirical findings from studies of special populations of listeners, such as older
adults or people who are hard-of-hearing, have provided ample evidence that alternative
models are required.
Although any of the current models of spoken word recognition could be modified to incorporate the role of suprasegmental information into their design, suggestions for specific modifications that could be made for one specific model will be presented in this section. For the sake of illustration, the PARSYN model will be discussed in regards to modifications that would allow for it to account for the suprasegmental information contained in the input. This model was chosen due to certain features that make it particularly adaptable to a design that would include prosodic structure; these features include a sublexical level of representation that considers the likelihood of segments occurring together and receptors that are sensitive to temporal position.

PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000) is a model that has moved beyond considering individual features or segments in the input by incorporating a sublexical level of representation in its design. In this model, receptors at the allophone level are sensitive to the temporal position in which allophones occur in the input. At the pattern level, units receive facilitative information from these temporally sensitive units at the allophone level, as well as from other units at the pattern level in the preceding or following temporal positions. The word level units then receive facilitative input from the position-specific allophones at the pattern level. While PARSYN is still entirely focused on segmental information in the input, it does begin to account for the interaction between information at different temporal locations in the input. Prosody, and more specifically lexical stress, is realized through the relative stress of portions of the word across the temporal signal. A model that is able to account for the interaction of information across the temporal signal can, therefore, be considered a step towards a model that could incorporate prosodic structure in its design. Furthermore, the PARSYN model's focus on the likelihood of segments occurring together can be considered a step towards potentially incorporating a level of syllable representation in a model of spoken word recognition. Since lexical stress is actually realized through the relative stress of syllables
in a word, PARSYN can also in this respect be considered a positive step towards a model that would incorporate suprasegmental information in the input.

The acoustic dimensions associated with stress are frequency, duration, and amplitude. Syllables that receive stress are higher in frequency and amplitude, and longer in duration, than unstressed syllables. These acoustic parameters are primarily carried by the vowel in the syllable. One possibility for accounting for suprasegmental information in the input would be a model of spoken word recognition that would be sensitive to the acoustic dimensions of frequency, amplitude, and duration that are placed on vowels. For instance, a model could be designed to have an allophone level that would have different receptor units for vowels, depending on whether they are stressed or unstressed in the input, based on the duration, amplitude, and frequency of each vowel, relative to the other vowels in the input. In the same manner in which PARSYN passes facilitative information between units at the pattern level that occur in preceding and following temporal positions, facilitative and/or inhibitory information could then be passed between vowels that occur in different temporal locations in the input. Facilitative information could then be passed to the word level, which could have word level units that carry the acoustic dimensions of relative frequency, amplitude, and duration carried on the vowels in the different syllables. Alternatively, an intermediate syllable level of representation could be incorporated into the model, in which a syllable is ‘built’ around each vowel and the consonants between vowels are placed in each syllable according to the phonotactic constraints and rules requiring onsets and/or codas in a specific language. Lexical stress could then be considered according to the relative frequency, duration, and amplitude of syllables in different temporal locations across the word.

4.3.2 Limitations of the current study

There are two limitations of the stimuli in the current experiment that relate specifically to the stimulus words that were chosen. First, as was discussed in the Methods chapter, two of
the words chosen had alternate stress patterns that were used in the recording of these words. It was therefore necessary to include these words in a different stress pattern category than the one for which they were each originally chosen; this resulted in an unequal number of words in each stress pattern category. However, it was decided that this was not a significant limitation, as the five different stress patterns were incorporated in order to include a variety of different stress patterns in the current experiment; this was achieved, despite the uneven numbers.

Another limitation relates to the three levels of stress that occur in English; these include primary, secondary, and no stress. In the current study, the words were chosen and placed in a stress category based on the syllable containing the primary stress; the stress of the other syllables was not considered in choosing the words. Therefore, some words contained unstressed syllables in the remainder of the word while others contained a syllable with secondary stress; the number of each type of word was not balanced across stress pattern categories. Because stress was not included as an experimental variable in the current study, the fact that different types of 'weak' syllables were in different words should not have affected the results significantly. However, ideally the words included in each of the different stress pattern categories should have been balanced for the number of words containing a syllable with secondary stress, across the different categories. This is a limitation that could be addressed in future research.

The design of the current study focused on measuring the contributions of segmental and suprasegmental cues in spoken word recognition in English. Theoretically, this required that each type of cue be presented and measured independently of the other. However, it was felt that measuring the contribution of one cue in the complete absence of the other cue would be less reflective of typical listening conditions than would degrading the stimuli to the degree that some information of each cue was still available to the listener. Thus, the segmental and suprasegmental cues in the stimulus words used in the current study were only partially
degraded in the conditions in which one or both cues were degraded, rather than being eliminated entirely. As a result, conclusive statements regarding the absolute contribution of each cue to the process of word recognition cannot be made.

The stimulus items used in the current study were initially created using natural speech. While natural speech was used in order to create as realistic a listening situation as possible, it also created certain limitations in the preparation of the stimuli. To create the two different suprasegmental conditions, two different tokens of each word were recorded. While the two tokens needed to meet certain criteria in order to be included as stimuli, it was not possible to ensure that the prosody of each word was degraded in exactly the same way, to exactly the same degree. Furthermore, while the speaker was instructed to degrade the suprasegmental information in her production of the words for the degraded suprasegmental tokens, there was no means of guaranteeing that vowel quality would remain completely unaffected. In fact, considering the interconnectedness of segmental and suprasegmental cues in natural speech, it is likely that the vowel quality of some of the tokens was somewhat affected. On the other hand, the jitter conditions were created using a computer program. Therefore, it was possible to degrade the segmental information across words in exactly the same way and to the same degree. However, since there is no quantifiable method of comparatively measuring segmental and suprasegmental information, there was no method of degrading the two different types of cues to exactly the same degree. The effect of degrading the two different types of cues appeared to be similar based on the accuracy results, which were comparable across the two conditions. Nevertheless, any strong conclusions drawn regarding the contribution of each of these cues, relative to the contribution of the other cue, would be premature. Despite these limitations, the findings that both cues intact result in significantly higher scores than when either cue is degraded, and that both cues degraded result in significantly lower scores than
when either cue is intact, indicates that both segmental and suprasegmental cues do contribute to the process of spoken word recognition in English.

4.3.3 Clinical Implications

The finding that suprasegmental information contributes to spoken word recognition in English has implications for clinical populations. For instance, the contribution of suprasegmental information is potentially significant for clients who have difficulties retrieving lexical items.

Goodglass, Wingfield, Hyde, Gleason, Bowles, and Gallagher (1997) conducted an analysis of the naming errors produced by aphasic patients for the Boston Naming Test. They found that erroneous responses that matched the target word for syllable number and stress pattern contained significantly more of the segmental information in the target word than errors that did not match the word for suprasegmental information. Lindfield et al. (1999) argued that these findings are consistent with “suggestions in the literature that the prosodic pattern of a word may be activated at an early stage of lexical retrieval, and that the word’s prosodic pattern may act as a framework to facilitate the retrieval of segmental phonology (Goodglass, 1993; Levelt & Wheeldon, 1994)” (p. 313). Based on the findings of Goodglass et al. (1997), as well as those of Lindfield et al. (1997), Wingfield et al. (2000) and the current study, it would be appropriate to attempt to use suprasegmental information as a cue for clients with word-finding difficulties in clinical settings.

The contribution of suprasegmental information in spoken word recognition is also significant for people with difficulties producing syllabic stress. Errors in syllabic stress production can occur for young children learning English as their native language as well as for people learning English as an additional language. An inability to produce natural prosody can also sometimes result from an acquired right hemisphere brain injury. The findings of the current study suggest that inaccurate or absent syllabic stress would decrease the ability of
others to activate the intended target word of the speaker and, therefore, would decrease the speaker’s ability to communicate. Given this suggestion, the findings indicate that it would be appropriate to directly target syllabic stress pattern production during clinical intervention with these populations.

The findings of the current study also have implications for people who are unable to fully exploit segmental information for speech perception, whether due to a high frequency hearing loss or due to declines in temporal processing abilities suggested to be associated with normal aging. These populations also experience a decrease in their ability to communicate with others. The findings of the current study indicate that clients in these populations are able to utilize suprasegmental cues to partially compensate for the limited segmental information available to them. Therefore, these findings suggest that it is important to maintain clear and unexaggerated prosody while communicating with these clients.

4.3.4 Future research

In the current study, only one modality of speech perception was considered. In normal listening environments, listeners typically rely on other cues in addition to auditory cues during speech perception. Massaro (1994) argued that there are "multiple sources of information supporting speech perception" (p. 252). For instance, listeners typically exploit cues that are processed in the visual modality, such as lip movements and hand gestures, during speech perception. Redundancies in the speech signal may, therefore, also occur between visual and auditory cues, and degradation in one modality could be potentially compensated for by another modality. This would be an interesting area for future research.

Future research should also focus on developing a model of spoken word recognition that takes suprasegmental information into account. In order to move towards a model of spoken word recognition that incorporates both segmental and suprasegmental cues in its
conceptualization, future studies in this area should focus on controlling for the frequency of words as it relates to both segmental and stress pattern information.

Further research could also be conducted in measuring the effect of different syllable patterns on the listener's ability to identify words. In the current study, syllable stress pattern was included as a means of including a variety of different stress patterns. In future research, syllable stress pattern could be incorporated as an experimental variable to explore whether or not the effect of degrading suprasegmental information differs according to the stress pattern of the word. In exploring this question, the researcher would need to control for the primary stress pattern of the word, as well as the pattern of secondary versus unstressed syllables in the remainder of the word.
References


APPENDIX I

Informed Consent Form

The Relative Contribution of Prosody in Word Recognition
for Ideal vs. Non-ideal Listening Conditions
(Graduate Thesis Research)

Principal Investigator:
Kathy Pichora-Fuller, Ph.D., Medicine/Audiology and Speech Sciences, UBC
Phone: 604-822-4716

Co-Investigator:
Suzanne Macdonald, M.Sc. candidate, School of Audiology and Speech Sciences, UBC
Phone: 604-822-9474

Purpose:

The purpose of this study is to investigate the relative contribution of different elements of the speech signal to our ability to understand spoken language.

The study is a research project at the UBC School of Audiology and Speech Sciences and is the topic of a graduate student thesis.

Study Procedure:

As a participant, you will first have a hearing screening test to determine your eligibility to participate in this study. If selected for the study, you will be asked to listen to a series of partial words through headphones in a sound booth. You will be asked to guess the full word that is being partially presented through the headphones. You will give a stated response for each partial word that you hear.

All sessions will be held at the UBC campus. Each session will last approximately 1 hour. You will be required to participate in a maximum of 4 sessions. These sessions may be scheduled separately or together at times that are convenient for you. Frequent breaks will be offered during each session. There are no known side effects from any of the experimental procedures.
Privacy and Anonymity:

Any information collected during this study will be kept strictly confidential. All documents on paper or computer disks will be kept locked in the lab. You will not be identified by name in any reports of the completed study.

Compensation:

You will receive an honorarium of $10.00 per session following completion of the experiment.

Contact:

If you have any questions concerning the procedures of the study or would like to request further information regarding the study, you may contact Suzanne Macdonald at 604-822-9474.

If you have any concerns about your treatment during this study or your rights as a research subject, you may contact the Director of Research Services at the University of British Columbia at 604-822-8598.

Consent:

I understand that my participation in this study is entirely voluntary and that I may refuse to participate or withdraw from the study at any time.

I have received a copy of this consent form for my records.

I consent to participate in this research study.

Subject Signature

Date

Witness Signature

Date
## APPENDIX II

### Two-syllable words

<table>
<thead>
<tr>
<th>Initial syllable stress</th>
<th>Final syllable stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot</td>
<td>Veneer</td>
</tr>
<tr>
<td>Mansion</td>
<td>Cartoon</td>
</tr>
<tr>
<td>Servant</td>
<td>Bouquet</td>
</tr>
<tr>
<td>Penny.</td>
<td>Cement</td>
</tr>
<tr>
<td>Campus</td>
<td>Advice</td>
</tr>
<tr>
<td>Sofa</td>
<td>Giraffe</td>
</tr>
<tr>
<td>Dentist</td>
<td>Canteen</td>
</tr>
<tr>
<td>Cabin</td>
<td>Brigade</td>
</tr>
<tr>
<td>Sugar</td>
<td>Disguise</td>
</tr>
<tr>
<td>Dolphin</td>
<td>Surprise</td>
</tr>
<tr>
<td>Scissors</td>
<td>Platoon</td>
</tr>
<tr>
<td>Cricket</td>
<td>Quartet</td>
</tr>
<tr>
<td>Menu</td>
<td>Balloon</td>
</tr>
<tr>
<td>Forest</td>
<td>Champagne</td>
</tr>
<tr>
<td>Penguin</td>
<td>Guitar</td>
</tr>
<tr>
<td>Cabbage</td>
<td>Caboose</td>
</tr>
<tr>
<td>Luggage</td>
<td>Trombone</td>
</tr>
<tr>
<td>Blanket</td>
<td>Shampoo</td>
</tr>
<tr>
<td>Soldier</td>
<td>Cigar</td>
</tr>
<tr>
<td></td>
<td>Disease</td>
</tr>
<tr>
<td></td>
<td>Brochure</td>
</tr>
</tbody>
</table>

### Three-syllable words

<table>
<thead>
<tr>
<th>Initial syllable stress</th>
<th>Medial syllable stress</th>
<th>Final syllable stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broccoli</td>
<td>Spaghetti</td>
<td>Kangaroo</td>
</tr>
<tr>
<td>Pyramid</td>
<td>Banana</td>
<td>Tambourine</td>
</tr>
<tr>
<td>Microphone</td>
<td>Umbrella</td>
<td>Violin</td>
</tr>
<tr>
<td>Pineapple</td>
<td>Conductor</td>
<td>Gasoline</td>
</tr>
<tr>
<td>Animal</td>
<td>Museum</td>
<td>Engineer</td>
</tr>
<tr>
<td>Spatula</td>
<td>Gorilla</td>
<td>Tangerine</td>
</tr>
<tr>
<td>Telescope</td>
<td>Safari</td>
<td>Lemonade</td>
</tr>
<tr>
<td>Vitamin</td>
<td>Potato</td>
<td>Refugee</td>
</tr>
<tr>
<td>Continent</td>
<td>Tobacco</td>
<td>Personnel</td>
</tr>
<tr>
<td>Furniture</td>
<td>Foundation</td>
<td>Halloween</td>
</tr>
<tr>
<td>Tricycle</td>
<td>Suspender</td>
<td>Millionaire</td>
</tr>
<tr>
<td>Sesame</td>
<td>Pajamas</td>
<td>Guarantee</td>
</tr>
<tr>
<td>Barbecue</td>
<td>Salami</td>
<td>Cigarette</td>
</tr>
<tr>
<td>Photograph</td>
<td>Computer</td>
<td>Divorcee</td>
</tr>
<tr>
<td>Medicine</td>
<td>Piano</td>
<td>Referee</td>
</tr>
<tr>
<td>Alphabet</td>
<td>Mascara</td>
<td>Chandelier</td>
</tr>
<tr>
<td>Cinema</td>
<td>Baloney</td>
<td>Pioneer</td>
</tr>
<tr>
<td>Carpenter</td>
<td>Cathedral</td>
<td>Magazine</td>
</tr>
<tr>
<td>Stadium</td>
<td>Physician</td>
<td></td>
</tr>
<tr>
<td>Festival</td>
<td>Professor</td>
<td></td>
</tr>
<tr>
<td>Registrar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX III

List 1
Kangaroo
Broccoli
Veneer
Spaghetti
Robot
Pyramid
Banana
Menu
Tambourine
Cartoon
Violin
Bouquet
Umbrella
Servant
Microphone
Penny
Sesame
Conductor
Gasoline
Cement
Animal
Museum
Engineer
Campus
Advice

List 2
Brochure
Tangerine
Suspenders
Platoon
Spatula
Pajamas
Sofa
Registrar
Quartet
Telescope
Dentist
Lemonade
Carpenter
Salami
Balloon
Continent
Refugee
Cabinet
Champagne
Computer
Furniture
Piano
Sugar
Personnel
Guitar

List 3
Penguin
Gorilla
Tricycle
Halloween
Giraffe
Millionaire
Vitamin
Safari
Dolphin
Canteen
Potato
Brigade
Pineapple
Guarantee
Cricket
Photograph
Tobacco
Mansion
Disguise
Cigarette
Forest
Submarine
Medicine
Surprise
Foundation

List 4
Alphabet
Mascara
Scissors
Referee
Caboose
Baloney
Divorcee
Cinema
Cabbage
Trombone
Chandelier
Cathedral
Luggage
Barbecue
Shampoo
Blanket
Physician
Cigar
Pioneer
Stadium
Professor
Soldier
Magazine
Disease
Festival