# Acoustic Correlates
# of
# Encoded Prosody
# in
# Written Conversation

Mark Stephen Schmidt

*This thesis is dedicated to my parents*

*and to my grandfather Charles Hinks*

# ABSTRACT

This thesis presents an analysis of certain punctuation devices such as parenthesis, italics and emphatic spellings with respect to their acoustic correlates in read speech. The class of punctuation devices under investigation are referred to as *prosodic markers.* The thesis therefore presents an analysis of features of the spoken language which are represented symbolically in text. Hence it is a characterization of aspects of the spoken language which have been transcribed or symbolized in the written medium and then translated back into a spoken form by a reader. The thesis focuses in particular on the analysis of parenthesis, the examination of encoded prominence and emphasis, and also addresses the use of paralinguistic markers which signal attitude or emotion.

In an effort to avoid the use of self constructed or artificial material containing arbitrary symbolic or prosodic encodings, all material used for empirical analysis was taken from examples of electronic written exchanges on the Internet, such as from electronic mail messages and from articles posted on electronic newsgroups and news bulletins. This medium of language, which is referred to here as *written conversation*, provides a rich source of material containing encoded prosodic markers. These occur in the form of 'smiley faces' expressing attitudes or feelings, words highlighted by a number of means such as capitalization, italics, underscore characters, or asterisks, and in the form of dashes or parentheses, which provide suggestions on how the information in a text or sentence may be structured with regard to its informational content.

Chapter 2 investigates in detail the genre of written conversation with respect to its place in an emerging continuum between written and spoken

language, concentrating on transcriptional devices and their function as indicators of prosody. The implications these symbolic representations bear on the task of reading, by humans as well as machines, are then examined.

Chapters 3 and 4 turn to the acoustic analysis of parentheticals and emphasis markers respectively. The experimental work in this thesis is based on readings of a corpus of selected materials from written conversation with the acoustic analysis concentrating on the differences between readings of texts with prosodic markers and readings of the same texts from which prosodic markers have been removed. Finally, the effect of prosodic markers is tested in perception experiments involving both human and resynthesized utterances.

# DECLARATION

I hereby declare that the work presented in this thesis is my own, unless explicitly indicated otherwise. I further declare that this thesis has not been submitted in this or in any other form to this or any other university.

Mark Stephen Schmidt

# ACKNOWLEDGEMENTS

The completion of a thesis requires the continuing encouragement and support from a great number of people, including family, friends, supervisors, workmates and fellow Ph.D. students. To all of you I express may deepest gratitude for the strong support you have given me over the years.

In particular, I would like to thank my supervisor Stephen Isard for his encouragement, criticism and willingness to discuss, often at length, the most obscure aspects of language and linguistics. These ranged from amplitude to zero-crossings, including punctuation, 'smiley-faces', statistical text-analysis, the relationship between semantics, syntax and phonology, down to acoustic phenomena and their statistical analysis - I greatly appreciated his thorough knowledge in all these areas. Furthermore, I would like to thank Bob Ladd for teaching me about phonology (at least I can now describe the way I speak), and Alan Kemp and Gerry Docherty for making the subject of phonetics such a fascinating one when I first started learning about Linguistics.

My further thanks go to all my friends and colleagues who I have had the pleasure to work with over the last six years, in particular Ian Nairn, Andrew Sutherland, Steve Love, Steve Hiller, Keith Edwards, Liam Parker, Nestor Yoma, Fabrizio Carraro, John Foster, Fergus McInnes and Bob Anstruther. My special thanks go to those of you who took part in countless experiments and to those who spent hours in the recording studio providing me with the necessary speech to carry out this work - in particular Bec Vaughan and Sally Bates.

However, without a stable working environment, without the space to develop my ideas and carry them out, and without the opportunities to develop the way I wanted to go, I would have not been able to complete this work, let alone stay in Edinburgh. I am therefore especially grateful to my mentor and friend Professor Mervyn Jack, who took an active interest in my work, career and development - and made everything possible. Thanks Merv!

Last but certainly not least I would like to thank Verena for giving up home to come to Edinburgh and live with a restless and bad-tempered Ph.D. student and for somehow managing to get us through the last year.

# TABLE OF CONTENTS

# 3 Parenthetical Constructions

# 4 Emphasis Markers

# Appendices

# List of Tables

# List of Figures

# List of Graphs

# Chapter 1

# Introduction

## 1.1 Aims, motivations and methodology

This thesis presents a characterization of the acoustic correlates of a set of
punctuation devices in read speech, which it is argued, function as *prosodic markers*.
These markers enrich the orthographic presentation of information in certain texts
by providing symbolically encoded prosody in aid of text interpretation and
reading aloud. The aim of this thesis is to show that these devices, which include
parentheses and type-face settings such as italics, capitalized letters, or bold print,
have prosodic correlates which can be measured, analysed and tested for their effect
in speech production and perception.

These markers, which seemingly inherit their meaning from the spoken
language, are represented symbolically in many written genres, for example in
informal personal letters, cartoons, signs, instructions and notices, and their use is
seeing a marked increase through the use of computers for interpersonal
communication. This, it is argued in chapter 2, dilutes the boundaries between
traditional notions of written and spoken language, resulting in a language form
which is described here as *Written Conversation*.

This genre, its exponents including Internet communications such as
electronic mail messages, electronic newsgroups and pages on the World Wide

Web, combines elements from the written and the spoken language and provides particularly typical and plentiful examples of the use of prosodic markers as text interpretation aids. The genre is influenced by a set of specific attributes, in particular the speed of exchange, the expectancy of an immediate reply, writing under pressure, communicating with strangers and one-to-many communications - all of which have some influence on how individuals communicate and especially on their style of writing (Kiesler et al. 1984, Yates 1992a; 1992b).

The most important function of these prosodic markers we argue, is that they indicate specific deviations from what might be expected, thereby facilitating the interpretation of a sentence or text. These deviations may include marked instances of linguistic concepts such as stress, emphasis or the status or importance of pieces of information and may even include the provision of paralinguistic information through clues to attitude or emotion relating to the contents of a text. Both are exemplified in the following example taken from a personal electronic mail message.

1       What an HONEST man :-)

Capitalization of the word "honest" signals a marked placement of the sentence stress and the provision of additional emphasis on the particular accent, the 'smiley-face' or 'emoticon'[1] points to a sarcastic undertone which may have been obvious had the sentence been spoken aloud, but which would otherwise be lost if it had not been specifically marked in the written form.

A related function is that these orthographic interpretation aids sometimes simply enhance a particular interpretation even in cases where there is no deviation

---

[1]'Smiley-faces' are read by turning the head to the left or moving the page to landscape orientation.

from a possible 'default reading'. In all cases they represent safeguards against misunderstandings which may occur through a discrepancy between intention and interpretation.

The frequent use of linguistic and paralinguistic prosodic markers to supplement and enrich information in written texts together with the importance of these markers for interpretation, render these powerful linguistic tools. In particular, their influence comes from their ability to visually indicate particular *choices* which have been made regarding the presentation of information. These choices contribute considerably to the variability, expressive power and unpredictability of language - examples are discussed in section 1.2. A detailed examination and exploitation of these choices, some of which are indicated by the use of prosodic markers, is justified for the following two reasons:

Firstly, as with any traditional scientific investigation, analyses tend first to concentrate on the examination of regularities before attempting to incorporate exceptions or irregularities into a developed theory. Here, the study of prosody and intonation is no exception. There are a number of linguistic phenomena which many approaches to the study of prosody have investigated only partially, due to their focus on the specification of phonological systems and rules describing whole languages. In particular, much detail concerning special cases and exceptions such as emphatic-, parenthetical-, quoted- and reported speech is either missing or treated only in passing.

Some accounts suffer from a lack of data and incomplete analysis and are sometimes based on self-constructed, anecdotal and hypothetical examples. This was and still is necessary, due to the fact that purely statistical accounts may overlook these special cases, which leads Bolinger to argue for the careful construction of examples (Bolinger, 1989: 394-7). However, statistical approaches

are on the increase, brought about by the widespread availability of computational analysis tools and sizeable corpora, in addition to the effort to agree on standards and notations for automatic prosody assignment (Silverman et al. 1992b; Pitrelli et al. 1994).

Nevertheless, although the experimental and statistical investigation of suprasegmental phenomena is a relatively recent undertaking, advances in the understanding of many aspects of prosody have been immense in recent years[2]. The analysis of the acoustic correlates of specifically indicated choices in readings containing textual encodings of some spoken form may deepen our understanding of the use and function of some of these specialized cases, but can only hope to provide a small contribution towards the further understanding of prosody in speech.

The experimental analysis in this thesis is neither based on constructed examples nor is it a large scale statistical corpus based effort. However, a large corpus of electronic news articles, ranging from archery to zoology, as well as a variety of personal electronic mail messages was used to extract examples showing the particular punctuation devices that were of interest. This corpus is listed in Appendices A, B, E and I. Experimentation is therefore based on a set of examples from real language usage, containing the phenomena of interest in their written form and used for analysis in their spoken, read out representation. Results from these analyses are then tested in speech perception experiments whose main purpose is the evaluation of re-created stimuli with respect to perceptual adequacy.

---

[2]For thorough discussions see, for example, O'Connor & Arnold (1961), Cutler & Ladd (1983), Couper-Kuhlen (1986), Cruttenden (1986), Monaghan (1991).

The second motivation for this study comes from the field of Speech Technology. In particular, Text-to-Speech conversion has now reached a level where it can be considered fully intelligible, however continuing to lack the fundamental requirement of expressive naturalness which is a prerequisite for its future acceptance (Monaghan 1991, Terken & Lemeer 1988). Required are increases in naturalness obtained through more precise, more variable, more expressive, and in some instances more conversational intonation and prosody. Furthermore, a computational implementation of a theory often provides evidence for its applicability and validity, and many scholars have successfully implemented aspects of their theories in working systems (Collier 1990; Monaghan 1991; Campbell 1992; Collier & Terken 1987; O'Shaughnessy 1976).

Although it may seem that the main application of the results presented in this thesis lies in the design of a Text-to-Speech conversion system for electronic mail and news applications, this is not the case. It is hoped that the work presented here benefits text processing for Text-to-Speech conversion in general, as there are countless unresolved issues, discussed in section 2.4 in chapter 2. Furthermore, should complete textual analysis become available in the future, showing pragmatic relations, complete syntactic analysis and semantic descriptions, the requirement for knowing how to represent their effect in the acoustic speech signal will be even greater. At the more theoretical level, the implications of this work for prosody research in general have already been highlighted.

A further motivation from the field of Speech Technology is provided by the increased importance of prosody for current research into speech recognition, made possible mainly through high quality real-time pitch analysis (Waibel 1988, Silverman et al. 1992a). One particular area of speech recognition, that of key-word spotting, is particularly likely to benefit from the research presented here. The aim

of key-word spotting is the extraction of relevant information for a particular domain or interaction while ignoring the superfluous. Especially the knowledge concerning the acoustic correlates of parenthetical constructions, which signal background information, and emphasis markers, which signal important information, could benefit this process. These, in the absence of full speech understanding, would provide acoustic clues as to what is important and what can be left aside. Ultimately, being able to detect and then correctly interpret prosodic phenomena has enormous implications for intelligent speech understanding and information processing.

The ultimate goal, therefore, is the characterization of the acoustic and perceptual correlates of all linguistic and paralinguistic textual markers which may influence the interpretation and reading aloud of texts. This thesis concentrates on the two most frequently used textual markers in written conversation, parenthesis and emphasis markers, which are discussed in chapters 3 and 4 respectively. Before moving on to a detailed discussion of written conversation in chapter 2, the remainder of this introduction provides further background information by investigating in more detail the influence of 'choice' in the presentation of information and by outlining some traditional approaches to the study of the correspondences between punctuation and prosody.

## 1.2    'Choice' in the presentation of information

> *"The distribution of sentence accents is not determined by syntactic structure but by semantic and emotional highlighting"*
>
> (Bolinger 1972a:644)

This section exemplifies the kind of choices that are available in the presentation of information, setting the scene for the discussion in chapter 2 which shows that some of these choices are textually encoded in written conversation.

Bolinger's comments are specifically related to the placing of sentence accents, but his views are part of a philosophy of viewing language not only from a point of view of *what* is said or written but *how* it is said or written. It represents an alternative to purely syntactic accounts of language and attributes many aspects of surface form to the higher levels of pragmatics and semantics. Bolinger's approach reflects a view of language found in many psycholinguistic accounts which are concerned with the investigation of the underlying mental information processing necessary to produce language (Chafe 1979; Levelt 1989). In these accounts great importance is placed on the speaker, who not only decides or chooses what to say, but also selects how to present the information to best convey his or her intentions. It is even argued that the output is constantly monitored in order to make sure that the intentions are getting across, leaving the possibility of making immediate modifications if necessary (section 2.2.3.1).

This approach to the study of language is supported in this thesis, which deals not only with the acoustic analysis of textual prosodic markers, but also examines the background issues leading to the introduction of these markers. Part of this thesis is, therefore, the application of a Bolingerian approach to a particular genre of the written language by examining phenomena where not the speaker but the writer indicates some of the decisions or choices made to convey his intentions. This is explored in depth below in section 2.2.3.

The following investigates instances of intentions, decisions or choices represented in text in a fairly general manner as background to chapter 2, which argues that, as far as conversational writing is concerned, the same or very similar

mechanisms to those found in the generation of spoken language are in operation in the generation of written conversation.

## 1.2.1 Linguistic, extralinguistic and paralinguistic choices

Communication, whether in written or in spoken form, is a matter of the presentation and exchange of information. In each such exchange, for example in a spoken conversation, a multitude of different types of information are transmitted between the participants as part of a particular communicative act. Consider the simple monosyllabic utterance at the start of many spoken interactions:

*"Hi"*

In the production of this utterance we can distinguish broadly between the verbal communicative act, physically realized as a speech waveform with associated duration, intensity and pitch, and the non-verbal communicative acts which may involve gesture such as an excited hand movement, a smile, or a raising of the speaker's eyebrows. This may suggest a strict division into what has been referred to as linguistic, extralinguistic and paralinguistic information (Laver 1992), in other words, the linguistic act and any associated gestural, attitudinal or emotional signs. In the world of transmitted physical speech waveforms this distinction is a blurred one, resulting from the interaction between the above mentioned types of information and their influence on the speech signal.

Consider all the possible ways the utterance "Hi" could be expressed. Variations are possible in the realisation of pitch - a large fall, maybe from an extended pitch range with associated increases in duration and intensity may indicate increased excitement, inviting possible interpretations such as "I didn't expect to see you here" or "Nice to see you after such a long time". A flat, monotone

8

and short "Hi" may indicate that people are not particularly happy to see each other, although it also may also indicate something completely different, such as "What an awful day" or "Hello, I'm still tired - don't speak to me". All levels of information influence the way in which something is said and consequently the way in which it is interpreted and all types of information bear in some way upon the actual speech signal that is transmitted between participants in communication, often doing so simultaneously.

Perhaps the central underlying factor influencing the way in which something is said, is that a considerable number of parameters are available for conscious manipulation, which means that speakers have the ability to use particular mechanisms for conveying information essentially as a matter of *choice*.

The speaker may have consciously chosen *how* to say "Hi" in order to achieve a particular effect. A speaker may make choices of which syllables to accent, and how much emphasis to put on them or which information to present as important or central to the conversation as part of choosing what is important at a particular point in the discourse.

These choices may be based on who the speaker is talking to, how he or she chooses to present the information, influenced by the particular situation or setting the interlocutors are in, dependent on an assessment of their shared knowledge or information previously exchanged as part of the discourse, and based on which particular effect the speaker is trying to achieve with the discourse. For example, in much the same way as deciding which words to give special emphasis, a speaker could decide to fake a cold by blocking off the nasal cavity through raising the velum, maybe in order to avoid having to go to school or simply for attracting sympathy.

Choice has an influence on how information is presented with respect to matters of prominence and relations between various aspects of the informational content of an utterance and with respect to subjective judgements (attitudes) and general frame of mind (emotions) of the participants involved. The above examples were related mainly to the influence of choice from extralinguistic and paralinguistic information, but consider the following simple example which illustrates how different prominence relations can be expressed.

2      Jim stroked the dog.

3      JIM stroked the dog.

4      Jim STROKED the dog.

In sentence 2, which is the unmarked case and hence contains no indication of prominence, we would expect the nucleus of the utterance to fall on the last accented syllable of the utterance. (Halliday 1967, Chomsky & Halle 1968). In sentences 3 and 4 which depict contrastive scenarios, the nucleus would be expected to shift from the last accented syllable of the utterance to the marked word and within the marked word be realized on its lexically stressed syllable. Consequently, sentence 3 could be an answer to the questions

a) Who stroked the dog? or b) Did Mary stroke the dog?

and 4 could be the answer to

c) What did Jim do to the dog? or d) Did Jim kick the dog?

In this case, the choice of *what* to accent is somewhat restricted by the provided context but the choice of *how* to accent the particular syllable is still potentially subject to large variation. This is brought about by varying degrees of emphasis

which the speaker is free to exploit. It is not unreasonable to suggest for example in sentence 3, that although the accent falls onto the same syllable 'Jim', the realization of that accent may be acoustically different (i.e. higher F0 maximum, increased duration) as a result of answering question b) rather than question a). Precise differences and acoustic correlates are discussed in chapter 4 below.

Another example relates to whether the provided information can be regarded as adding any new information to the discourse which has any importance for its interpretation, or whether it simply constitutes a point of digression, containing information that may or may not influence its interpretation. Consider for example the difference between restrictive and non-restrictive relative clauses such as:

> 5      The book which I bought yesterday is quite good.

> 6      The book, which I bought yesterday, is quite good.

In sentence 5 the relative clause restricts the interpretation to a particular book purchased the previous day, suggesting the existence of more than one book. Sentence 6 carries the connotation that the conversational participants know which book is being talked about, with the information given by the relative clause being of supplementary nature.

As the examples above have shown, choice is an important factor in the way information is presented in the spoken medium, and it is equally, if not more important with regard to special punctuation devices in the written medium, because it is reasonable to suggest that writing is a more 'conscious' activity than speaking. Consider the following example taken from an electronic mail message:

"Among other advantages, choosing a Creole for an interlanguage would mean that linguists wouldn't poopooh it as "not being a *real* language". :-)  "

This example shows that the writer is conscious of using an invented but stereotypical quote, and goes as far as to even provide his preferred accentuation. He also expresses his attitude towards the statement by inserting a disclaiming 'smiley-face' (emoticon) which may be an indicator of sarcasm or may simply say, "Don't take this too seriously."

Written conversation provides countless instances where presentational choices are indicated by manipulation of the textual form.

## 1.3   Traditional notions of punctuation and prosody

The previous section has shown how certain linguistic choices, made for example by a speaker, can influence the way information is presented and interpreted. As it is suggested below, that some of these choices are indicated by graphical means influencing interpretation and prosody in reading, it is useful to summarize a debate concerning the relationship between punctuation and prosody which has been carried out as part of a more fundamental argument relating to the linguistic status of written- in relation to spoken language.[3] The debate relates to traditional punctuation marks such as commas, full-stops, quotation marks, parenthesis, exclamation marks and colons and it may have to be reinterpreted if an extended set of graphical- or punctuation devices are considered, such as those found in

---

[3]For extensive discussions on this subject the reader is referred to Nunberg (1990), Levinson (1986), Markwardt (1942), Smith (1987), Sampson (1985), Bolinger (1975), Waller (1988).

written conversation, or if it becomes evident that traditional punctuation marks are emerging in certain genres with new meanings.[4]

The argument is between those who examine the written language primarily with reference to the spoken language and basically deny it its own linguistic status, and those who regard the two language forms as distinct linguistic systems. Nunberg refers to the former approach as "transcriptional" (Nunberg 1990:12), reflecting the attitude that writing is considered to be merely a particular language dialect which is visually represented. The introductory quote to the following chapter is possibly one of the strongest expositions of the transcriptional view.

Nunberg on the other hand argues in favour of an approach which sees the study of written language and in particular that of punctuation in its own right, by showing that punctuation is subject to grammatical constraints and can be described by grammars similar to the kinds found for the description of natural language itself. By that account, punctuation is considered to be a linguistic subsystem which nevertheless has some "functional overlap" with the spoken language. Nunberg's main concern is to show whether there is "anything *more* to written language than the features it shares with speech" (Nunberg 1990:4).

Possibly the most contentious aspects concerning the linguistic status of the written language is the topic of punctuation, and especially its relationship to prosody. Nunberg argues that the study of punctuation in particular has suffered from the transcriptional approach which offers a "theoretically uninteresting account of what is in any event not a very good correlation" (Nunberg 1990:15).

---

[4]This is happening with the traditional quotation mark which is now often used to mark real or imaginary quotations with strong sarcastic undertones.

Punctuation marks such as parenthesis, commas and exclamation marks are considered to be transcriptions of spoken language intonation by transcriptionalists, and are considered to function as intonation instructions in the reading of texts. Nunberg grants a certain degree of overlap between the language forms but would prefer to say that certain intonational changes are made as pronunciations of punctuation marks in certain genres, rather than saying that those punctuation marks *de facto* signal particular intonations.

Arguments in favour of - and against the transcriptionalist view are plentiful and refer to historical, psychological and practical considerations. Supporting arguments include the fact that historically, at the time when the reading of a text was commonly accompanied by its oral recital, punctuation reflected intonation more closely, a fact supported by punctuation norms at the time (Levinson 1986). A similar argument may function as a counter suggestion in that non-transcriptionalists claim that it is futile to provide markers of prosody or intonation in texts which are never intended to be read aloud. Further evidence in support of the transcriptionalist theory is found in references to 'acoustic images', 'inner voices' or 'good ears' which many writers claim to make use of for deciding on difficult punctuations, but as Nunberg points out, congenitally deaf people can learn to punctuate perfectly adequately. Nunberg provides further arguments against the transcriptional view by suggesting that some punctuational differences signal a difference in meaning without having an intonational equivalent, and that some punctuations have no intonational realization despite their presence.

The whole argument is reminiscent of the syntax-prosody debate, which will be touched on below, and may be resolved by adopting a parallel model of language generation by which communicative ideas and their semantic representations are *simultaneously* reflected at various levels without necessarily

determining each other - but nevertheless showing a certain degree of overlap due to their common semantic roots.

The purpose of this brief account concerning the relationship between punctuation and prosody was to show that traditional punctuation devices can be argued to be of transcriptional nature although the correspondences, as pointed out by Nunberg, are not always convincing. In particular, it serves to separate punctuation in traditional terms from the set of graphical devices such as italics, bold-script, capitalization, asterisks and 'emoticons' with which we will be concerned in the next chapter, some of which are argued to be precise markers of prosody. It should be born in mind, however, that strict separation is not always obvious as some traditional devices have taken on new, perhaps genre specific functions.

The following chapter is concerned with the development of a new genre, that of *Written Conversation*, which has its own particular set of graphical and punctuation devices considered to be transcriptional in nature.

## 1.4 Thesis overview

Chapter 2 investigates in detail the genre of written conversation by examining its main characteristics and in particular the nature and function of certain transcriptional devices which are argued to function as markers of prosody. Their relationship to the spoken language is examined together with the motivations for their use.

Chapters 3 and 4 present the acoustic correlates in readings of texts containing the two most frequently used textual markers in written conversation,

parentheticals and emphasis markers respectively. Both chapters contain literature surveys of their particular field of concern and in both cases the presentation of acoustic correlates is supported by series of perception experiments which either serve to underline the presented results or test their application for use in automated systems.

Finally, chapter 5 summarizes the findings and contributions of this thesis, points to its shortcomings and identifies areas of further research.

# Chapter 2

# 'Written Conversation'

*"Writing is not language, but merely a way of recording language by means of visible marks."*

(Bloomfield 1933: 21)

The aim of this chapter is to examine the background to *Written Conversation*, a comparatively new and fast spreading genre used in Internet communications such as electronic newsgroups, electronic mail messages and texts on the World Wide Web. Its status with respect to the traditional division between spoken and written language is examined (sections 2.1.1, 2.1.2) and the reasons behind the introduction of certain transcriptional devices, which seem to function as *prosodic markers*, are investigated. This chapter therefore, attempts to provide the justification for the acoustical analysis of the two most used textual markers in written conversation, parentheticals and emphasis markers, presented in chapters 3 and 4 respectively. Possible reasons for the insertion of prosodic markers are given (section 2.2.2) together with hypotheses on when in the process of generating written utterances they may be introduced (section 2.2.3).

The term Written Conversation was coined in order to suggest that this genre combines elements from the two major language forms, written and spoken language. 'Having a conversation' does not normally have the connotation of communicating in the written medium and the term *conversation* is more suggestive

of immediate, interactive spoken dialogue. Having a written conversation, if taken extremely literally, would be the equivalent of transcribing a spoken dialogue as best as possible in the written form, indicating facial movements, hesitations, re-starts, intonation, emotions and attitudes. Another crucial factor for having a written conversation is interactivity, which is given in e-mail and news conversations and is one of the central distinguishing factors between written conversation and more traditional means of communication such as books or newspaper articles.

The evidence below shows that written conversation is indeed a hybrid between spoken and written language and also demonstrates that there is some internal variance, in that some written conversations are more 'spoken' in character than others. It is suggested here, that the more 'spoken' exchanges are also more transcriptional, which was the term used in the introduction to this thesis to refer to the translation or transcription of certain - mainly prosodic aspects of speech - into visible textual annotations. Hence, the more a written exchange tends towards the spoken end of the continuum in character, the more characteristics from the spoken language it will retain. In those instances we would expect more prosodic markers to appear than in messages which tend more towards the written, providing some justification for using the introductory quote to this chapter to describe certain types of writing. In search of evidence for the status of written conversation a variety of accounts from a number of disciplines are examined.

Written Conversation, also referred to more generally as "Computer Mediated Communication" or "Computer Conferencing", has attracted substantial interest from a variety of fields such as psychology, psycholinguistics, general linguistics and speech technology. The interest for psychologists lies mainly in the effect of computer conferencing on interpersonal communication, behaviour and

relations (Adrianson & Hjelmquist 1985; Kiesler 1978; Kiesler et al. 1984, 1985; Kiesler & Sproull 1987; Rice et al. 1987; Sproull & Kiesler 1986, 1991; Sorensen 1991). For linguists the interest is in the relationship of this sublanguage to the traditional division between spoken and written language as outlined above (Halliday 1985; Yates 1992a, 1992b), and for this thesis, written transcriptions of prosody in the form of prosodic markers are the central topic.

It was suggested above that writings with a more spoken character are also likely to be more transcriptionalist in that they contain graphical pointers to speech events. Chapter 1 examined a traditional point of view, suggesting that punctuation does not only function to distinguish different types of utterances, for instance questions, exclamations or quotations, but that punctuation as such is a transcription of certain phenomena from the spoken language, pausing being the most obvious example.

In order to advance the argument presented here, that prosody related factors are indicated in certain genres and used in reading, it is important to find evidence supporting the suggestion that the markers found in written conversations could conceivably be prosodic markers, and if so, to investigate the reason for their existence. On a theoretical level this could be supported by showing that written conversations do have characteristics resembling the spoken language and that similar processes operate in the generation of both spoken and written language. This theoretical justification is presented here in chapter 2.

On a practical level, it can be shown that markers do have prosodic manifestations in reading by presenting readers with pairs of texts that differ only with respect to the presence of these markers. Any significant difference in prosodic realization would then be a direct result of their introduction or omission. In

chapters 3 and 4, detailed acoustic analysis of recorded readings quantifies some of these differences.

Written Conversation is of interest to speech technologists with respect to textual analysis and annotation (Carvalho et al. 1994; Guaitella & Santi 1990; House & Youd 1990, 1991, 1992; O'Malley et al. 1991), especially the textual encoding of linguistic principles and their representation in speech. The technological angle is explored in section 2.4 below.

## 2.1 The main characteristics of 'Written Conversation'

### 2.1.1 General observations

Example 7 at the end of chapter 1 has already provided a glimpse of some of the idiosyncrasies we might find in Written Conversation. This section examines some general observations about this medium collected mostly from users of electronic mail, electronic news bulletins, chat columns, as well as corporate and other electronic networks. This is intended to provide at least a preliminary characterization of this emerging new genre.

Some of these views are collected in an article that appeared in the Washington Post, titled "JUST PUT ON A HAPPY FACE" (Garreau 1993) where the main caption reads:

"Head-Turning 'Smileys' Give New Tilt to Computer Chatting"

It is not intended here to treat this article as a scientific base for discussion, but it is worth examining some of the points raised, which deal mainly with one particular

aspect of electronic writings, the inclusion of paralinguistic markers as signals of attitude and emotion such as humour, sarcasm, irony, happiness and sadness.[1] In the article, many of the crucial characteristics of electronic messages are listed and the reasons for their existence are examined. For example, reference is made to the instantaneous nature of the exchanges, the fact that they have "the speed of the telephone without any of the clues" and that "computer mail is more like talk than mail".

This is a characteristic commented on by both Spitzer (1986) and Sorensen (1991) who attribute a great deal of the character of written conversations to the influence of speed. Speed they argue, is one of the main influencing factors in determining at which point on the continuum between written and spoken language certain electronic writings can be placed.

> "... the higher the "speed" in the exchange of messages, the more the character and features of the linguistic interaction resemble the prototypical spoken interaction and language."
>
> (Sorensen 1991:52)

Yates (1992a) considers the possibility that off-line compositions of text may be more 'written' than on-line productions where the time pressure, and hence the speed of composition, is higher. A further interesting point of analysis would be the

---

[1]The Washington Post article prompted an international debate on the network, and I am grateful to Cathy Ball, Professor of Computational Linguistics at the University of Georgetown, for making available to me the article and her collection of replies received after posting a request for information to the LINGUIST newsgroup enquiring about 'emoticons' in the context of earlier writing systems such as hieroglyphs.

difference between an initial message or request - the starting point of a written conversation - and the follow up messages or replies which are composed under differing circumstances and are likely to be reflected in the text compositions. Particularly in an on-line discussion amongst multiple users, where any type of turn-taking found in ordinary human dialogues is clearly impossible, the pressure to be the first to reply, the first to get a point across, may have an influence on the speed of composition and hence the degree of spokenness of a message.

The Washington Post article attributes the inventiveness shown by the users of electronic mail to the "inhuman nature of silicon messaging" which requires clarification of intentions and the degree of "seriousness" attributed to individual remarks, and emoticons are seen as "short hand ways to add tone of voice or body-language to keyboard messages". The necessity to insert these markers arises from the absence of facial expressions and intonational clues present in natural speech, which convey attitudinal and emotional signs. In the same article, Cathy Ball implies that sophisticated literary skills are required in order to correctly convey the intended meaning or to avoid offending somebody, especially under the circumstances of writing under pressure.

In a small internal survey conducted in our institute with about forty researchers, all of whom are frequent users of electronic mail and news networks, colleagues were asked to give their views on the use of capitalized words and asterisks as markers of emphasis in electronic mail messages. Those surveyed did not believe that their usage was entirely due to individual preference but that clear functional distinctions could be made. Most responses suggested that capitals gave the impression of a word being spoken loudly, possibly in annoyance or anger, as marking points of particular importance and urgency or indicating a general raising of the voice. Asterisks on the other hand, were more used to signal contrastive

emphasis, gave more "local" attention to a word, did not involve "a change in volume" and were generally more subtle than words written in capitals.

The survey also asked the participants to compare electronic mail to other forms of written exchange and to consider italics as a further device for drawing the readers attention to a particular word or phrase. A particularly interesting response was the following:

> "...*caps* help the *eyes* to read where *italics* help
> the *mind* ... in that they tell the reader where to put
> the stress as they read the sentence in their mind."
> (original emphasis)

The concept of the *silent reader* is one which will be examined in more detail in section 2.2.3 below where we will examine whether an internal representation of speech could be responsible for the existence of prosodic markers in written conversation.

The responses also showed that there was an awareness of the differences between forms of writing in that other texts, particularly academic publications, are "more formal and less like speech", or to say it differently, electronic mail was described as "less formal and more conversational". A few subjects remarked that in "well written material"[2] emphasis was achieved by the manipulation of sentence structure or by explicit statements such as "It is important to note that ..."

With respect to differences in the use of textual devices, our small survey also showed that many people believe that techniques are being used differently in

---

[2]This seems to suggest that electronic writings are not perceived as being well written.

written conversation than in other written materials. Italics in other texts were thought to be more similar to asterisks in electronic mail than to capitals, and capitals were thought of mainly as functioning in titles or on signs and notices, and some remarked that in a similar way to electronic mail, capitals were used to transcribe shouting in dialogues in novels. Italics were seen as functioning in the introduction of new terms and as markers of contrastive emphasis of a subtle kind, or as transcriptions of direct speech.

In summary, it can be said that there seems to be a convergence of opinion about the function of certain textual devices which may signal aspects of spoken language, but talking of strict and well formed conventions would be an overstatement. It should be pointed out further that a survey within a single institute should be viewed with caution, due to the fact that particular conventions may develop within a group and may not apply more generally.

This section has given some indication that a new form of language may be developing, that regular contributors seem to be aware of its peculiarities and that general awareness of this development is increasing. Recently, the whole front cover of the Sunday Times supplement was covered with a gigantic smiley face with the associated article remarking that "...the Internet, ... has spontaneously generated a symbolic language of its own.[3]

## 2.1.2 "Lexical density" as a measure of textual complexity

This section explores whether the more general remarks and impressions presented above can be supported by a more detailed analysis of text corpora belonging to the

---

[3]The Sunday Times - 4th December 1994

genre of written conversation. Although the literature is unsurprisingly sparse, we find a statistical treatment concerning the place of computer conferencing on the continuum between written and spoken language (Yates 1992a; Adrianson et al. 1985). Yates compared written conferencing interactions from a corpus of some 1.5 million words and undertook a comparative analysis of the data with the written Lancaster-Oslo-Bergen (LOB) corpus (Francis 1980; Garside et al. 1987), and the spoken London-Lund corpus (Svartvik et al. 1980). The comparison was done with respect to three features:

1.    *Textuality*, for which an adapted version of Halliday's "lexical density" measure was used (Halliday 1985).

2.    *Subjectivity*, measured through the use of pronouns, and

3.    *Modality*, measured through the use of modal auxiliaries.

Lexical density is a measure which reflects the complexity or information density of an utterance. It does so by showing the relationship between the grammatical and the lexical elements of an utterance and is therefore an effective measure for quantifying the difference between spoken and written language. According to Halliday this is possible, because spoken utterances are characterized by large numbers of grammatical elements, or *function* words, whereas written utterances contain more densely packed lexical items, the *content* - or information providing words. The higher complexity of words and structure in the written language is therefore partially a result of the more tightly packed information - the higher lexical density - found in the written language (Halliday 1985).

For measurements on lexical density, three different types of interactions were examined by Yates. The first were from a follow-up conference on a day seminar on the use of electronic media, conducted by members of academic staff

(Academic). The second were student - tutor interactions (Course 1-3), part of a distance education course, and the third were chat messages (Chat 1-2). Results are presented in form of weighted-lexicals-per-clause,[4] where weighting has the function of distinguishing between high frequency and low frequency content words. This is necessary because frequent, predictable words do not contribute to lexical density in the same way as infrequent and therefore unexpected words. Table 1, adapted from Yates (1992a: p.8), shows clause based lexical density scores for a number of corpora.

Table 1:     Weighted lexicals per clause for conferencing and non-conferencing corpora (adapted from Yates 1992a)

| Conferencing corpora | | | | | |
|---|---|---|---|---|---|
| Academic | Course 1 | Course 2 | Course 3 | Chat 1 | Chat 2 |
| 4.2 wlc* | 3.8 wlc | 3.5 wlc | 3.3 wlc | 2.7 wlc | 2.1 wlc |
| Non-conferencing corpora | | | | | |
| Course Materials | | Teenage Magazine | Hansard | | Interview Transcript |
| 6.5 wlc | | 4.3 wlc | 5.9 wlc | | 2.6 wlc |

* weighted lexicals per clause

---

[4]Yates presents two measures for lexical density. Lexical density ratio, which gives a measure of lexical items to the total number of items in an utterance, and lexicals per clause which measures the lexical density in clauses. The latter is seen as the preferred measure due to the fact that spoken utterances are easier to segment into clauses rather than sentences, and secondly, because clauses have semantic importance and are therefore somewhat independent of the language medium.

The data from the conferencing examples show clearly that there is a continuum with respect to the degree of lexical complexity of the messages. This complexity could be described as the 'degrees of writtenness' of the interactions, with the chat messages evidently tending towards the spoken language. Comparing these results to non-conferencing text, it becomes clear that with respect to textuality, conferencing messages lie somewhere in-between speech and writing.

The non-conferencing examples, however, also show that there are ambiguities with respect to certain varieties of speech. The figure for the Hansard transcripts clearly shows that many parliamentary speeches are pre-written and that the difference between speech and writing may not be as obvious as one would expect - even in the more established genres. For the two other measures reported by Yates, subjectivity is high (see 2.2.2) in conferencing texts and is characterized by peculiar pronoun use, in the sense that there is little third person reference. Modality is very high, which leads Yates to conclude that conferencing contains more ambiguity relating to the knowledge available to the interlocutors than either speech or more traditional writing.[5] This may be the result of many written interactions being addressed to a large number of often unknown addressees.

For the moment it suffices to say that most examples of this genre fall somewhere between spoken and written language with respect to their linguistic status, and that some exponents of *written conversation* definitely tend towards the spoken language in their characteristics, which justifies the search for the existence of prosodic markers. Examples of prosodic markers are examined in the next section, which is followed by a discussion on the mechanisms by which prosodic markers might be inserted from a language generation point of view in section 2.2.2.

---

[5]Yates [personal communication] 1993

## 2.1.3 Transcriptional devices as indicators of prosody

A further characteristic of written conversation is the use of orthographic devices which go beyond standard punctuation and which function as indicators of prosody. Chapters 3 and 4 show, through the comparison of read utterances with and without markers, that these markers change the prosody of utterances which contain them.

The exact processes responsible for their insertion however, are unclear. The question is whether their insertion is triggered by prosodic processes themselves, such as the writer silently speaking the utterance before writing it down[6], or whether the actual meaning of the utterance directly determines its form. This may depend on the representation the writer has in mind at the time of inserting a punctuational device. If they represented a direct transcription of prosody, a phonological or acoustic representation would be responsible for their insertion. If, in contrast, the concept or concepts which a phonological representation actually depicts such as emphasis or contrast was the underlying cause, their insertion may be directly invoked from semantic or pragmatic representations (see 2.2.3 below). The following gives some examples of this extended set of punctuation providing linguistic and paralinguistic information.

### 2.1.3.1 Paralinguistic markers

Paralinguistic markers relate to the expression of attitude and emotion in written texts. These *smiley faces* or *emoticons* consist of a sequence of characters available on standard QWERTY keyboards capable of producing the full range of ASCII characters (Sanderson 1993). They are read by turning the head to the left or

---

[6]which corresponds to the "acoustic images" mentioned in chapter 1

alternatively, by turning the page so that the lines appear vertically. Whole dictionaries of emoticons have been compiled containing literally hundreds of different smiley faces. The following is a small extract from one of these collections (see Appendix N for an extended set):

:-)     standard smiley

;-)     winking smiley

:-(     un-smiley

:-x     'my lips are sealed' smiley

:-|     'have an ordinary day' smiley

*>--<-  'Bill the cat run over by a truck' smiley

It is quite clear that there is a great difference between collections of possible smileys, which exploit every possible combination of characters, and smileys which are actually used to serve a purpose in communication, as the following example illustrates. Out of context and in the absence of additional clues, the following sentence is almost impossible to interpret beyond the actual propositional content itself.

1       That was a nice dress you were wearing yesterday !

At the beginning of a conversation it may be part of what Laver et al. (1972) describe as phatic communion, which is merely an utterance for the purpose of starting a conversation.[7] Although unlikely in a written exchange, this still leaves

---

[7]Phatic communion in the context of electronic writing is an interesting research topic but it is of no further concern here.

two completely opposing interpretations, one honest, the other sarcastic -
undertones which can only be detected by means of intonation and gesture. Adding
an emoticon in a written conversation will alert the reader to a marked
interpretation.

2       That was a nice dress you were wearing yesterday ! :-)

This is likely although not necessarily meant in a sarcastic way, whereas the
following version with a winking smiley face is more likely to be meant honestly.

3       That was a nice coat you were wearing yesterday ! ;-)

Although the interpretation is facilitated a definite meaning is nevertheless difficult
to extract. This is due mainly to the ambiguity present in the use of emoticons and
the lack of context.

By far the most used emoticon is the simple smiley face, which is used in a
variety of contexts indicating happiness, a joke or a sarcastic remark. As a result,
emoticons often 'only' function as general markers for alerting the readers attention
to an interpretation which may not be obvious. Furthermore, the range of attitudes
that can be expressed is potentially large, and with as yet no established
conventions their use is likely to depend on individual writers' preferences. In the
following example there is little doubt about the intentions of the writer, who is
clearly not impressed by his new software package and a sarcastic undertone is
certainly also likely here, this time provided by the un-smiley face.

4       That's not a bug, it's a feature ! :-(

A further crucial observation is that emoticons are language - and to a certain
degree culturally independent, because they can be understood by people that do
not speak the same language or do not speak it well enough to catch less obvious

undertones. This distinguishes them in an important way from other icons such as certain abbreviations which have their origin in the spoken language and which are found frequently, for example, BTW (by the way), or IMHO (in my humble opinion). It is even arguable whether some of the phrases that the abbreviations depict were used with the same frequency in the spoken language before electronic mail became more widely used. "In my humble opinion" is a typical e-mail phrase saying "I really want to make this point - but don't shoot me for it", or even "I know a lot more about this than you think". It could be argued that it is indeed an icon rather than an abbreviation.

At this point it should also be remembered that other phrases which have been directly adopted from the spoken language are used in many exponents of informal writing and have been so widely accepted that their spoken origin may eventually be forgotten. This is particularly obvious in contractions such as 'don't' 'doesn't', I'm, she's, they're, to give only a few examples.

Although smileys clearly contribute to the character of written conversation, there are many more aspects which are worth examining and which may not be so easy to identify. This may be the case because they are not as obvious or because they make use of techniques which have always existed in the writing system. In these cases it is not the invention of a new typographical device which is characteristic, but the increased use of an already existing device. This is particularly true for the use of parenthesis and quotes which are used very frequently in electronic conversations.

## 2.1.3.2 Linguistic markers

This section has been named "linguistic markers" not because the respective markers are considered *only* to have a linguistic function, but because their linguistic function is our main concern.[8]

Linguistic markers relate to the way information is presented and to the relative importance attributed to these pieces of information for a particular communicative purpose. The most common of these markers indicate marked accentuation for the purpose of emphasis and contrast, concepts which will be discussed in more detail in chapter 4 below. Although all punctuational devices reflect *choice* decisions taken by the writer, as discussed in the previous chapter, the following are particularly clear examples of how choice manifests itself, endorsing Bolinger's view that "...the speaker adjusts the accents to suit his meaning" (Bolinger 1972a:635).

| | |
|---|---|
| 5 | "If you *really* want to live, move north to the *real* California !" |
| 6 | "It doesn't say that I WILL go there, just that I would be the happiest there." |
| 7 | "When I went back home (rural Northern New York) for vacation EVERY bar had those damn electronic dart boards." |
| 8 | "I spent a fair amount on calling the States and the people there (sorry, forget who, it's been a year or so) weren't really sure ... " |

---

[8]Chapter 4 discusses the differences between linguistic and paralinguistic functions of emphasis markers.

**9**  "Boy, doesn't *THAT* make you feel like this is an *HONEST* man!" :-)

Parenthetical constructions, as discussed in chapter 3 below, function to supply supplementary information which is not central to the proposition of an utterance. Their extremely common use in written conversation can be explained by the frequent lack of contextual information, and in particular the lack or uncertainty about shared knowledge between the interlocutors[9]. The use of parentheticals is also likely to be highest in one-to-many postings where some or all of the recipients are not known to the writer. Parentheticals, therefore, often function to provide this contextual information or supplementary knowledge which is necessary to place the utterance in the right context, but which is not part of the propositional content of the utterance itself. This is exemplified in sentence 7 above. Had the writer been talking to a close friend either by electronic mail or face-to-face, information about where his home was would have certainly been omitted, but it may add information for recipients who are not aware of the surrounding context.

Another linguistic marker to be mentioned is the quotation mark, which has come to take on rather different meanings in certain contexts. In its original sense it often occurs in replies to postings preceded by the phrase: On <date> <writer> wrote to newsgroup <x>: In this context the quote is often replaced by the symbol '<' or in some cases by the symbol '>'.

In other contexts it is used as an imaginary quotation, often with a sarcastic undertone, where a writer may envisage criticism from recipients or show his or her own criticism towards a proposition made by another writer. This is exemplified in

---

[9]Their function is not dissimilar to the use of footnotes in academic writings

the following examples which are clearly sarcastic, further supported by the use of the smiley face at the end of the utterance in 10.

10    "Among other advantages, using a Creole for an interlanguage would mean that linguists wouldn't poopooh it as "not being a *real* language". :-)  "

11    "I avoided it on POPULARITY grounds for months, but finally was in a position where it was the only "beer" available."

Quotes are also used to a great extent to indicate concepts, terms or definitions, which is similar to the function of *italics* in other written material, such as in:

12    "It is true that people, including professional linguists, occasionally use "dialect" and "language" in a way in which the former denotes a closer relationship than the latter..."

13    "But surely the relation, "are dialects of the same language" must be transitive."

14    "To use your own "bin" files, the best procedure we have seen is to have a "bin" directory structure ..."

This section has given examples of some of the most common additional punctuational devices found in the exponents of written conversation examined for the purpose of this thesis. These included paralinguistic markers in the form of emoticons and linguistic markers in the form of emphasis markers, parenthesis and quotation marks. Although this list is unlikely to be exhaustive, good acoustic correlates for these markers and their successful application is likely to benefit any automatic system which reads text containing these markers, or which has to provide acoustic output for the linguistic concepts these markers stand for.

Chapters 3 and 4 present the acoustic and perceptual correlates for possibly the two most used markers, parenthesis and emphasis markers.

## 2.1.3.3 Inventing new punctuation

Inventing new punctuation is, of course, not exclusive to users of electronic communication. Personal letters, for centuries, have been scattered with drawings of the sun, raindrops, teardrops, broken hearts and the like, or have been annotated by underlines, emotional markers or afterthoughts. The language used in comics must be one of the best examples of inventive punctuation and the use of a large variety of textual devices, and even scribes of ancient documents are reported to have annotated important sections by inserting a pointing finger into the margin next to the relevant section, or by using gold leaf lettering to highlight certain words or phrases.[10]

The extended set of punctuational devices used in written conversation is mainly a mixture of using sequences of characters to make new, character-based icons on standard ASCII computer keyboards and using characters which have always been part of the written language. Some others markers, such as bold script or italics, come from more sophisticated word processing tools, and as high quality colour copying and colour printing becomes more widely available the importance of colour will certainly increase.

With respect to this thesis, the interest in these markers lies in their function of translating linguistic devices, such as emphasis, contrast, stress, information

---

[10]A scribe's emotional reaction to the texts was often stated in the colophon or marginalia, which provides evidence for the existence of emoticons even in medieval texts. A comprehensive collection of medieval colophons by the Benedictines of Bouveret is in existence and Charles Plummer (1926) describes the use of colophons and marginalia by Irish scribes. Thanks to Charles Wright for pointing this out in a network discussion.

structuring and paralinguistic influences on the speech signal into a visible, textual representation. As discussed in chapter 1, written conversation provides a rich source of examples, although it is by no means the only form of writing which contains linguistically relevant textual annotations. A further advantage is that these texts can be directly used as input to automatic systems in order to test how well the acoustic output represents the linguistic principles encoded in the texts.

## 2.2   The function of transcriptional devices in 'Written Conversation'

This section is concerned mainly with three questions. The first question concerns the difference between traditional punctuation and the identified graphical devices in written conversation with respect to their function in the linguistic system (section 2.2.1). The second question asks why writers of written conversations introduce punctuational devices above the conventional standard set into their texts (section 2.2.2), and the third question deals with when during the production of a written utterance their insertion is triggered (section 2.2.3). Suggested answers are merely hypotheses, as complete answers to the above questions would require the precise and whole understanding of the semantic, syntactic and phonological building blocks of language and their interaction.

### 2.2.1   Graphical indicators of prosody - their relationship to the spoken language

Figure 1 below presents a model for distinguishing traditional punctuational devices from the kind of textual markers discussed above, which seem to operate a more direct link between the spoken and the written language. This separation is

made possible in the model by distinguishing punctuation which functions directly as a descriptor of semantic or pragmatic categories from punctuation which can be considered as mainly functioning as a descriptor of syntax. These are contained in the two large square boxes. Although this model fails to show the relation between the syntactic level and the semantic/pragmatic level and is therefore a simplification, the model helps to distinguish two classes of punctuational devices.

One class, described so far as traditional punctuation, fulfils a syntactic function and has a rather unclear relationship to phonological structure. In other words, it is unclear whether their insertion is motivated by phonological (and ultimately semantic) processes or whether their use is purely syntactic. For example, the convention used in German of starting each noun with a capital letter has no phonological motivation and is simply a morpho-syntactic convention. On the other hand, the use of the comma in English to delimit non-restrictive relative clauses is a case where there is a clear overlap between syntax and phonology (see examples 5 and 6 in chapter 1 above). It can be said for this class of punctuation that the reasons for their insertion are obscured by the fact that their origin is ambiguous. The transcriptionalists would see punctuation as having a purely phonological origin whereas writers like Nunberg (1990) would put their origin more within the development of writing (and its syntax), although they would grant a functional overlap between syntax and phonology.

The second class of punctuation is not obscured by an ambiguity of origin. The level of syntax plays no part and the relation between phonological origin (which itself is governed by semantic/pragmatic processes) and textual representation is seen as direct. These markers, such as capitalization on whole words or phrases and other emphasis markers are argued to have a transcriptional function in much the same way as transcriptionalists would attribute to standard

punctuation. In the Saussurean sense they can be considered *signs*, depicting a direct relationship between meaning and form (Saussure 1974, reprinted in Baskin et al.)

Figure 1:   Graphical indicators of prosody in 'written conversation'

**Descriptors of**

**Semantics/Pragmatics**

4    2

SL    3    WL

6

5    1

**Descriptors of**

**Syntax**
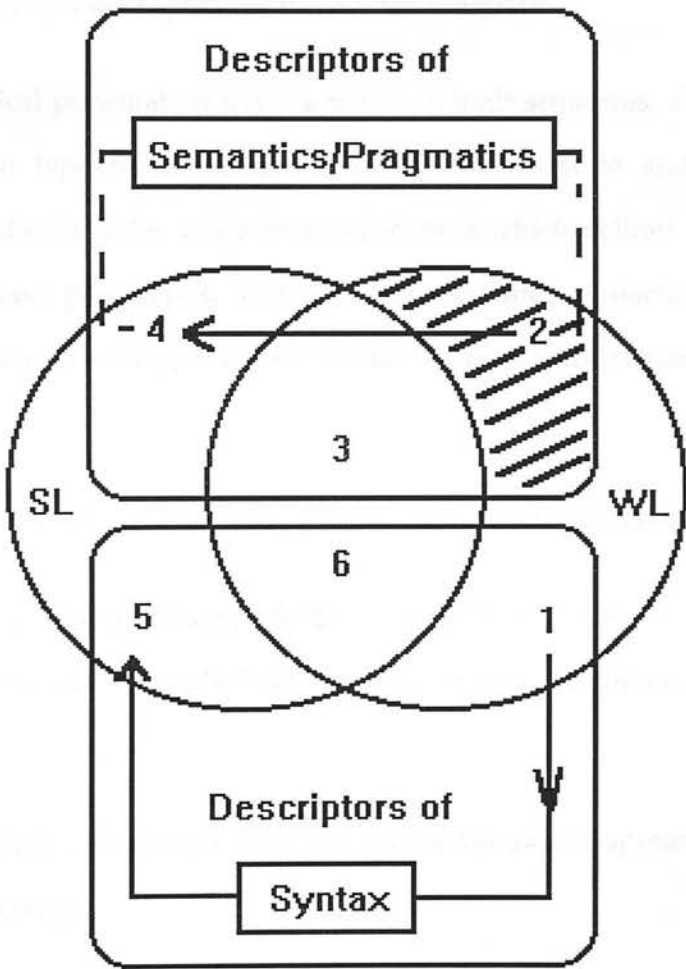
The two overlapping circles represent the spoken and written language respectively, which shows the view presented in this thesis of the two main language forms overlapping each other. Spoken and written language are not

38

considered to be separate linguistic systems but results of the same language generation system, which produces two outputs that have commonalities and differences. This overlap allows for a hybrid form of language such as written conversation and avoids taking sides with either the transcriptionalists or those who argue that written language is a separate linguistic system.

The individual fields in Figure I represent the following information:

**Field 1:**    Descriptors of syntax in the written language.

These are graphical punctuation devices which delimit sentences, clauses, phrases and indicate the type of sentence, for example declarative and interrogative sentences. This also includes white space characters which delimit chunks of text such as words and paragraphs. This field also includes syntactic constructions which are exclusive to writing or hardly appear in the spoken language, such as the subjunctive.

**Field 2:**    Descriptors of semantics/pragmatics in the written language.

These are graphical devices and type-face settings such as italics, capitalized words, quotes, underlines, small capitals, bold print, emoticons, parenthesis, and in some cases colour.

**Field 3:**    Descriptors of semantics/pragmatics common to spoken and written language.

This field contains commonalities between the two language forms at the semantic level such as the lexicon, the conceptualizer (section 2.2.3), as well as the message generation and monitoring facilities, which we argue here are the same for the two language forms (see Figure 2 below).

**Field 4:** Descriptors of semantics/pragmatics in the spoken language.

This is the spoken realization of the graphical devices in Field 2, in other words, the acoustic correlates of prosodic markers.

**Field 5:** Descriptors of syntax in the spoken language.

This field symbolizes the spoken equivalents of a syntactic form, for example pausing at utterance boundaries. This also includes syntactically triggered phonological phenomena such as clause and phrase final lengthening. This field also contains syntax which is exclusive to speaking, for example double negatives.

**Field 6:** Descriptors of syntax common to spoken and written language.

This field describes syntax which is common to speaking and writing, for example constituent ordering, and specific constructions, for example the passive construction or topicalization.

The model presented above allows some textual devices to be purely transcriptionalist, in the sense that aspects of a spoken representation of an utterance are translated into a textual annotation. Others, in contrast, also fulfil a syntactic function and would be part of the functional overlap which 'separatists' would see between graphical devices in the written form and prosodic devices in the spoken form.

The view taken here is that textual markers are not features of the written language which happen to overlap with certain features in the spoken language, but that they show quite deliberate attempts to transcribe prosodic features of the spoken language. It is the identification of which aspects of prosody are affected by these markers, and in what way, that is the main aim of this thesis.

## 2.2.2 Motivations for the use of transcriptional devices

*In cyberspace, no one can hear you scream.*
*Or laugh.*
*Or feel your kiss.*
*Or see you wink.*

(Joel Garreau, Washington Post, Aug. 4, 1993)

Much of what we communicate in spoken exchanges is not achieved by *what* we actually say, but by *how* we say it and by what accompanies an utterance in terms of gestures and body movements. Much of this information is lost when an utterance that may otherwise be spoken is written down instead. What remains is the propositional content of the utterance itself, supported by literary techniques which help to add certain undertones and clarify the author's intentions. The detection of these undertones or intentions, together with the detection of implicatures, is commonly referred to as 'reading between the lines', which may be the equivalent of guessing how something would have been said had it been uttered aloud.

In addition, there is little doubt that some writers are better at expressing their views than others, but even the interpretation of highly esteemed classical works, the authors of which would have possessed the highest literary skills, is no trivial task.

The motivations for the introduction of additional punctuational devices can be seen in the conversational nature of the exchanges, the speed with which these exchanges take place, and the lack of spoken- and gestural-, that is acoustic and extralinguistic cues. As Kiesler and Sproull report (1987; 1991) the lack of these cues can have rather serious consequences.

In the study by Kiesler and Sproull, which examined the efficiency of decision making in a networked environment, students were asked to solve a particular problem under time constraints. Groups either communicated verbally face-to-face or through networked terminals. Kiesler and Sproull reported that communication in the computerized groups frequently broke down, in one case resulting in physical threats which led to separate escorts from the premises for the involved parties. Similar incidents were not reported for the groups conversing naturally. Kiesler and Sproull also report a tendency towards harsher and more impolite exchanges on corporate electronic mail systems, resulting in over-emphatic prose with words indicating shouting and smiley faces indicating sarcasm. They attribute this to the absence of normal conversational cues and the anonymity of the conversational context.

The view presented in this thesis is that special textual markers such as emphasis markers and smiley faces are not a result of impolite and harsh exchanges but an attempt to avoid or pre-empt misunderstandings, which in addition to the anonymity of the context seem to be the cause for these impolite exchanges in the first place. The remaining problems in terms of conversational breakdowns and misunderstandings can nevertheless be explained either by the gross inadequacy or incompleteness of the transcriptional devices used or by the anonymity of the context.

To summarize, textual markers are seen as precise attempts to encode linguistic, extralinguistic and paralinguistic cues into a context from which they are naturally absent. In natural spoken language these cues are realized linguistically as intonation, stress, rhythm and voice quality, accompanied by gestures, nods and facial expressions.

## 2.2.3 Transcriptional markers - a reflection of internal speech ?

> *"Actually, the writer utters the speech-form before or during the act of writing and the hearer utters it in the act of reading; only after considerable practice do we succeed in making these speech-movements inaudible and inconspicuous"*

> (Bloomfield 1933:285)

This section attempts to hypothesize at which point in the generation of a written utterance an author may be led to the introduction of textual symbols which may represent prosodic concepts, and could therefore be argued to constitute prosodic markers.

The topic of speech generation[11], that is the underlying cognitive processes as well as the acoustic phonetic ones, is a vast field of study involving cognition, psychology, linguistics, psycholinguistics, computational linguistics and artificial intelligence, although there has been little collaboration between the disciplines. Before looking at the introduction of prosodic markers in text, it is important firstly to examine prosodic processing with reference to the spoken language, and especially at which point in the speech generation process phonological processing, and in particular prosody, is introduced. Levelt's model of the speech generation process (Levelt 1989) forms the basis for this discussion. This section does not

---

[11]For cognitive, psycholinguistic, computational and AI approaches the reader is referred to Levelt (1989); Bierwisch & Schreuder (1992); Garrett (1988); Günther (1992); Günther et al. 1993); Herweg & Maienborn (1992); McKeown & Swartout (1988); Paris, Swartout & Mann (1991).

For linguistic phonetic treatments to language production see, for example, Abercrombie (1967); Laver (1992); Ladefoged (1975); Couper-Kuhlen (1986); Docherty (1989).

describe the speech generation process in depth, it merely attempts to show that some parallels can be drawn between that process and the process of generating a written message. It is intended to show that the processes are not necessarily to be viewed as separate, but can be conceived as one - the process of generating language - resulting in speech on the one hand and written text on the other.

### 2.2.3.1 Levelt's blueprint for the speaker

Levelt presents a model of the speech generation process from the conception of an utterance to its articulated audible form (Levelt 1989). The five main components are a *conceptualizer*, a *formulator*, an *articulator*, a *speech comprehension system*, and a *monitor*.

The conceptualizer constitutes the pre-verbal stage of the speech production chain and is responsible for creating a *pre-verbal message*. The relevant information for the realization of an intention is selected by accessing various knowledge sources, such as knowledge about the world, knowledge about the discourse context and the discourse record, i.e. the history of what information has already been exchanged by the participants as part of that particular interaction.

*Macroplanning*, according to Levelt, is the planning stage responsible for selecting this information for the realization of a communicative goal and its subgoals. *Microplanning*, on the other hand, is concerned with the presentation of the informational perspective of an utterance, consisting mainly of its topic and focus structure and selected mainly with reference to the discourse record.

The pre-verbal message forms the input to the formulator, which outputs a linguistic structure by creating a *phonetic plan*. The first stage in this structuring process is the *grammatical encoding* procedure which accesses a semantically and syntactically subcategorized lexicon (lemma information), and together with

"syntactic building procedures" generates the *surface structure* of an utterance. This is translated into an articulatory plan by the *phonological encoding* component which accesses the lexical forms of the lemmata and creates an internal representation of the utterance to be articulated. Levelt calls this representation *internal speech*, a concept of major importance for the argument presented in this chapter. The implications of this concept are discussed further in section 2.2.3.3.

The articulator executes the phonetic plan and produces *overt speech*, which is processed by the speech comprehension system - the reverse process of the speech generation system allowing understanding and interpretation of the spoken utterance. The resultant *parsed speech* is available to a *monitoring* process which is responsible for comparing "...the meaning of what was said or internally prepared to what was intended" (Levelt 1989:13). This entails that the speech comprehension system also has access to the internal spoken representation facilitating the re-formulation of an intended utterance at the conceptual level before it is articulated, as well as being responsible for the early detection of articulation errors. Figure 2 below shows Levelt's proposed architecture. Annotations in blue are concepts of major importance for the argument presented here, annotations in red form part of Levelt's theory but are not shown in the original graph and have been added here for clarification.

Figure 2: A blueprint for the speaker (adapted from Levelt 1989:9)



CONCEPTUALIZER

MACROPLANNING

MICROPLANNING

message generation

monitoring

discourse model
situation knowledge
encyclopedia

preverbal message

FORMULATOR

grammatical encoding

surface structure

phonological encoding

LEXICON
lemmas
forms

parsed speech

SPEECH
COMPREHENSION
SYSTEM

phonetic plan (internal speech)

phonetic string

ARTICULATOR

overt speech

AUDITION

## 2.2.3.2 Extensions and modifications

Figure 1 above (section 2.2.1) suggests a direct mapping process from certain punctuational devices, which are suggested to be semantic or pragmatic descriptors, to their spoken equivalents, by by-passing any kind of syntactic processing. It is this direct mapping of parts of the semantic structure onto the phonological structure which Levelt's procedural model fails to incorporate or at least specify overtly. This is commented on in detail by Günther et al., who attempt to show that there is a direct relationship between semantic and phonological structure for certain linguistic phenomena, particularly ones concerned with information structuring (Günther et al. 1993). They argue that information structure influences the meaning of an utterance and that this is reflected by prosodic variations which are not necessarily determined by syntax. The examples used are concerned with the realization of linguistic focus in sentences with identical syntactic structure but which differ in their acoustic realization in terms of accent assignment due to contextual influences, such as:

15   JOHN has left

16   John has LEFT

We argue here that this is exactly what writers are indicating when inserting prosodic markers depicting emphasis or contrast. Contextual information, which is either not available to the reader at the time or requires re-reading previous passages, is openly indicated by textual means. The adapted model proposed in Günther et al. distinguishes between conceptual and contextual information at the pre-verbal stage which feed a semantic encoder and which correspond generally to Levelt's concepts of macroplanning and microplanning. For Levelt, however, the conceptualizer which creates a pre-verbal message provides all the semantic

information, aided by the lexicon at the formulation stage. A more structural representation of the linguistic components as suggested by Günther et al., which allows for a direct link between a semantic component and the phonological encoder may be of benefit in accounting for certain linguistic data. In addition, a more explicitly stated semantic component allows for the introduction of certain textual markers which influence phonological processing but have no syntactic motivation. A suggested modification to the model relating to a semantic component is introduced in Figure 3 below, indicated by the purple annotations.

### 2.2.3.3 A blueprint for the 'conversational writer'

Figure 2 above presented a model of the speech generation process from intent to articulation with a built in monitoring process. The following attempts to show the parallels between the *speech generation* and the *text generation* process, providing further support for the argument presented so far in this chapter, which portrays conversational writing as a partial transcription of speech.

One of the most interesting aspects of Levelt's model for the purpose of the argument presented here is the concept of *silent internal speech*, which was also commented on by scholars such as Bloomfield (1933). In Levelt's model, an internal, silently spoken representation of an utterance exists for the purpose of monitoring and consequent optional re-formulation of the utterance before the message is passed to the articulator. If we adopt this concept for the generation of written rather than articulated messages, the existence of certain punctuational devices which depict aspects of the spoken representation comes as no surprise. An investigation into the validity of this concept is far beyond the scope of this thesis, but it allows the hypothesis that certain textual markers seemingly have their origin in speech prosody, because prosody would be part of this internal spoken representation.

In an alternative model to Levelt's "Blueprint for the speaker", which, by analogy, may be named "A blueprint for the conversational writer", the articulator in Levelt's model can be replaced with a symbolic component responsible for creating written text rather than speech, a *symbolizer*. This is possible because there is no reason to believe that the facility for monitoring, for which internal speech is a prerequisite, should not be available in the writing process. It is even arguable that the monitoring process is more important in writing than in speech, because false starts or output that does not match the authors intention are more permanent and more irreversible than in speech - at the very least the editing of writing is a more cumbersome activity than the 'editing' of speech. In addition, immediate feedback from potential readers which would provide additional information on how well the message is getting across is not available to the writer. In such a model the conceptualizer generates a *pre-symbolic message* which undergoes the same formulation stages as the pre-verbal message, resulting in internal speech, but is then symbolized rather than articulated. The internal speech is monitored in the same way as before, by passing through the speech comprehension system, whereas the symbolized message passes through the *symbol comprehension system* which forwards a *parsed string* to the monitor.

The audition system responsible for dealing with overt speech is replaced by a *visual perception* component which passes *character strings* to the symbol comprehension system. The articulator in Levelt's model, the mechanism responsible for verbalizing of all the ideas, intentions and formulations which have been structured and monitored previously has now been replaced by a symbolizer which creates text rather than speech based on the same principles of language generation. This model is shown in Figure 3 below. The annotations in green show the changes and additions made to Levelt's model in order to accommodate the symbolic mechanisms operative in writing.

Figure 3: A blueprint for the 'conversational writer'
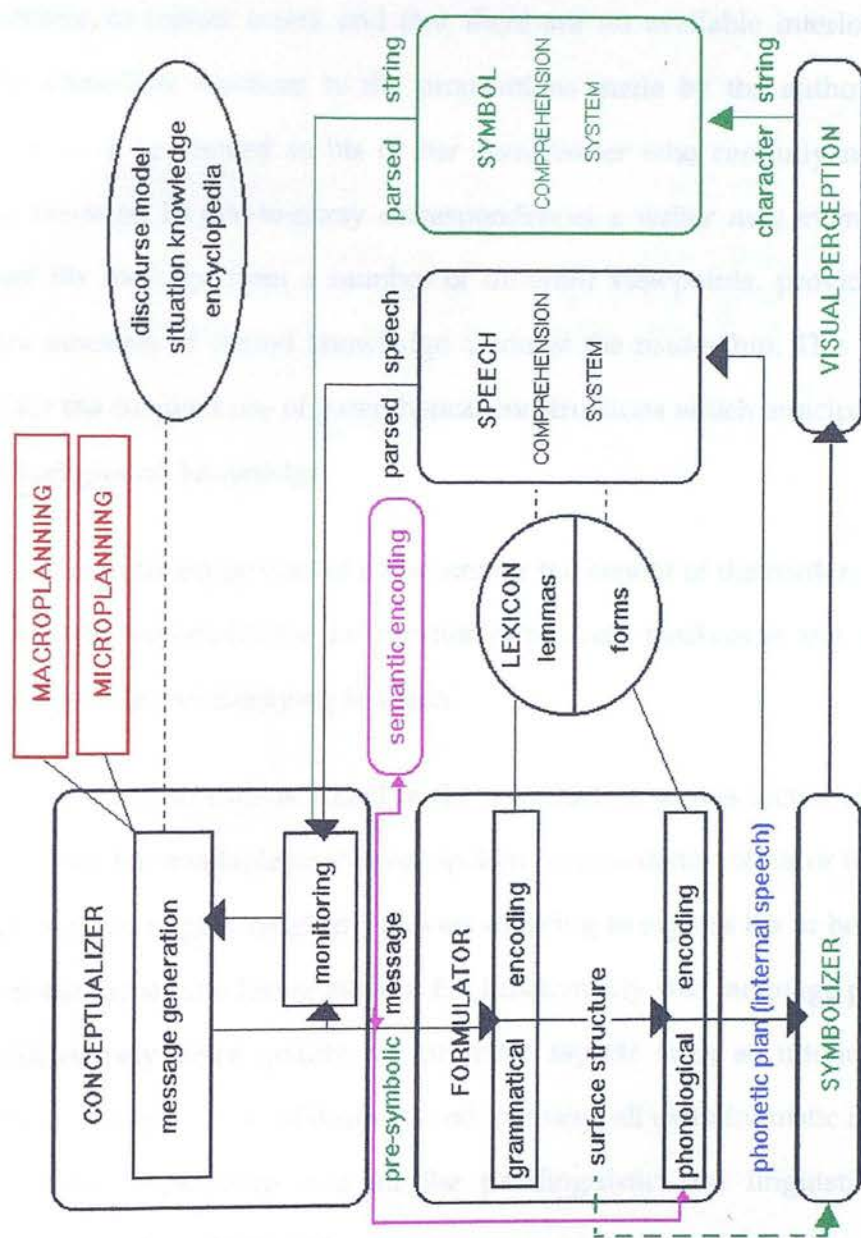
### 2.2.3.4 Interlocutors in written conversation

According to Levelt's theory concerning the monitoring process, "a speaker is his own listener" (Levelt 1989:13), who monitors both internal speech and overt speech. It was briefly mentioned above that the monitoring process may be even more important in writing than in speech, due to the fact that it is more difficult and cumbersome to correct errors and that there are no available interlocutors who provide immediate reactions to the propositions made by the author. A writer, therefore, may be viewed as his or her own reader who carefully monitors the created message. In one-to-many correspondences a writer may even attempt to interpret his message from a number of different viewpoints, providing for the different amounts of shared knowledge amongst the readership. This may be the reason for the frequent use of parenthetical constructions which anticipate a certain lack of background knowledge.

The monitoring process of the writer for the benefit of the reader, it is argued here, is chiefly responsible for the insertion of prosodic markers in text, which have a disambiguating and clarifying function.

As Bloomfield's quote stated in the introduction to this section, it is argued that the writer has available an internal spoken representation of his or her intended message and has to go through the process of trying to express his or her intentions whilst at the same time losing most of the functionality that language provides for doing so, namely voice quality, all prosodic aspects such as intonation, stress assignment, manipulations of duration and intensity, all extralinguistic mechanisms such as facial expressions and all the paralinguistic and linguistic tools for expressing attitude and emotion.

The theory presented here can be seen as a possible explanation for extensions to existing traditional punctuation devices used by authors in spontaneous written interactions. Devices which can be argued to be transcriptional have been observed in the past, most obviously punctuation itself, and relate to the crucial observation that there is a relationship between the ideas and concepts which the author of a text is trying to convey to their reader and the concrete form of the text itself. The following quote applies this observation to texts as a whole:

> "In a piece of expository prose, the author's indication of the development of the argument contributes to the reader's understanding of the text. Thus titles, chapter headings, sub-divisions and sub-headings all indicate to the reader how the author intends his argument to be chunked."
>
> (Brown & Yule 1983:7)

The task of the reader is the interpretation of the written symbolic message, which is identical to the task of the symbol comprehension system described above. The same mechanisms apply in the automatic conversion of text into speech which involves the interpretation of a set of symbols in the form of written text and punctuation, and their translation into a spoken form. These issues are examined more closely in section 2.4.1 below.

This section has provided a possible theoretical account of why prosodic markers are inserted in certain types of text and at which point in the language generation process they are likely to be introduced. This was achieved by examining Levelt's model of speech generation from the cognitive to the articulatory stages, and by modifying the model to the process of generating written

text. Both were presented not as separate, but as part of the larger process of language generation.

## 2.3. The linguistic status of written conversation - a summary

The aim of the previous sections of this chapter was a characterization of the nature of a relatively new written genre named *written conversation*. This form of writing was examined from a number of angles, including a corpus-based statistical treatment together with impressions and exemplifying material from people who regularly communicate in that medium. From these accounts if was possible to characterize the genre's linguistic status with respect to more traditional notions of spoken and written language.

Statistical methods allow the differentiation of exponents of this genre according to their "degrees of spokenness", and it was argued that the more 'spoken', and therefore more conversational messages contained textual devices which may function as indicators of prosody. It was hypothesized that these prosodic markers are inserted in written exchanges to counterbalance the loss of most of the linguistic devices available in the spoken language for conveying intentions, attitudes and feelings.

In addition, the question was examined whether there is any theoretical foundation for the hypothesis that representations of prosody exist - and are available during the cognitive and linguistic processes operative in writing. A model was presented which was based on the speech generation architecture proposed by Levelt (1989) which argues for the availability of an internal spoken representation of the utterance before it is symbolized in the written form, and that

this representation forms the basis for the introduction of paralinguistic and linguistic textual devices found in written conversation.

## 2.4 Implications for Text-to-Speech conversion

Following the previous section, which proposed a model of the processes involved in writing, this section briefly examines the task of reading aloud in order to show how an analysis of the acoustic correlates of textual markers may benefit the automatic conversion of text into speech.

### 2.4.1 Categories of knowledge used for converting text into speech

The process of reading aloud essentially involves five types of knowledge. The first type is determined by *overt* information in the text, such as the string of characters, delimiters like spaces, tabs, line breaks or paragraph markers, as well as punctuation- and other textual markers such as those found in written conversation. Overt information may also lead to the use of heuristic information such as the frequency with which a particular word or item occurs in a given text, which has been shown to influence the way it is pronounced (Fowler & Housum 1987).

The second type of knowledge comes from *covert* information present in the text and requires linguistic processing before it becomes available. This involves, for example, the analysis of sentence structure, establishing the relationship between pronouns and their antecedents and extracting information about pragmatic relations such as "given" and "new" or "foreground" and "background" information (Chafe 1979). At a higher level the processing of covert knowledge may involve the analysis of the argument structure of a particular text.

The third category of knowledge involves linguistic knowledge about the language as such and includes, for example, rules for the construction of sentences or rules specifying the correspondences between graphemes and phonemes. This category also involves other knowledge bases required for reading, for example lexical knowledge about the pronunciation of exceptional items such as abbreviations or proper names (Schmidt 1989; Schmidt et al. 1993, Schmidt et al. 1994).

Textual analysis for the extraction of overt and covert knowledge together with available linguistic knowledge bases are used in the more sophisticated Text-to-Speech systems to determine the assignment of prosodic structure. It is especially the knowledge extracted from linguistic analysis that expresses the relations that hold within a text, which, through the assignment of prosody demonstrate apparent understanding of what is expressed in the text (Monaghan 1991).

The fourth category of knowledge involves any additional knowledge that is used in text interpretation such as the particular topic with which the text is concerned, the context in which the text has been written, knowledge about the author of a text and most generally, knowledge about the world. The fifth category of knowledge concerns the mapping between an interpreted meaning and the acoustic representation of this meaning, following the application of the first four types of knowledge .

Categories one, two, three and five can be argued to constitute the linguistic competence necessary to provide an adequate reading of a text, whereas category four may contribute to a full and thorough *understanding* of the particular text. Some scholars, however, argue that a full understanding is not always necessary for providing an adequate reading (Monaghan 1991:4). After all, it may be possible for a linguistically competent reader to give a perfectly adequate reading of a text

concerned with quantum mechanics although he or she may know nothing about the subject. A reader may 'only' make use of the knowledge categories responsible for linguistic competence, but they bring to the task of reading aloud a vast array of linguistic knowledge which in the medium to long term will simply not be available to automatic reading machines (Schmidt 1990).

## 2.4.2 Making the best use of available information

Text-to-Speech systems rely to a large extent on overt knowledge and have rather limited access to the covert knowledge in the text, as the automatic linguistic analysis of texts is at best partially successful, even for aspects of linguistics which are theoretically fairly well understood. Syntactic parsing, for example, has seen some success after scholars accepted that complete parses may not be necessary for adequate assignment of prosodic structure (Fitzpatrick & Bachenko 1989; Willemse & Boves 1991), but little advance has been made in the automatic analysis of pragmatic and discourse structure suitable for use in TTS systems. The task for the textual analysis modules in TTS systems is therefore the analysis of as much overt and covert knowledge as possible - interpreting the physical layout of a text, applying as much linguistic processing as currently possible and developing heuristic rules which may bring some short-term benefit to the output of such systems (Monaghan 1988, 1991b). The physical layout of the text together with the limited linguistic analysis currently available, therefore, provides the only information available to the assignment of prosody, which Monaghan describes as "the listener's main index of a TTS system's apparent understanding" (1991:5).[12]

---

[12]Note that there is a difference between *extracting* linguistic information from previously written text and having available linguistic information for dialogue systems which involve speech synthesis in their output. In these cases, which often involve restricted domains, more

In the absence of more sophisticated linguistic analysis for running text, the enrichment of the overt knowledge of a text through textual prosody markers has an immediate benefit. For the text genres that contain these markers they provide more overt information which can be directly mapped to an acoustic representation, provided the acoustic correlates of these markers are known.

There is also a more long-term benefit of the analysis of textual prosody markers concerning the fourth category of knowledge necessary for reading aloud. This involves the mapping process from certain meanings, which may be expressed by phonological representations, to the actual acoustic signal which is generated to represent these meanings. The benefit lies in the availability of the knowledge of the mappings necessary to express linguistic phenomena such as emphasis, contrast and the importance of pieces of information. In short, if all the information which is ideally required for the adequate reading of a text becomes available, the system needs to know what to do with this information in terms of producing an acceptable acoustic output. In that sense, the analysis of parentheticals, which expresses background information, and emphasis, which signals important information may be of benefit in the future, even for texts which do not contain textual prosody markers.

In the meantime it is necessary to concentrate on the immediate benefits of the analysis of punctuation and other textual markers - a need which developers of TTS systems have recognized in recent years. Guaitella & Santi (1990) apply an analysis of how subjects punctuate read and spontaneous speech for the

---

sophisticated linguistic processing is used in order to produce the text which may later be read out by the automated system. Because these systems 'know' what they are going to say the provision of prosodic information is often done by *inserting* either phonological markers or direct instructions to the synthesizer into the text the system produces (House & Youd 1990; 1991; 1992, Vonwiller et al. 1991).

development of rules responsible for pausing and intonation contour specification. Carvalho et al. (1994) apply an analysis of punctuation to the conversion of e-mail to voice-mail, and O'Malley et al. (1991) carried out an investigation into the kinds of textual analysis most beneficial for the improvement of prosody in Text-to-Speech systems for unrestricted text.

O'Malley et al. were concerned with establishing the frequency with which particular textual surface phenomena occurred in e-mail messages, newsgroups and newspaper articles. Calculations were carried out for dashes, emphatic questions, exclamations, lists, quotes, parenthesis, sentence adverbs and afterthoughts. They concluded that the implementation of reliable prosody assignment rules for these markers would improve prosody about every one hundred words of text in e-mail. They further found that parenthesis was by far the most used textual device, occurring around every two hundred words in running text in electronic communications. They also remark, however, that the most common phenomena are also the least reliable with respect to the mapping between marker and prosodic effect - a finding which is investigated in detail in chapters 3 and 4 below.

Taking into account other textual markers such as emphasis markers or emoticons, it would be expected that the relevance of textual markers to the assignment of prosody would increase even further for certain genres. Cooper and Paccia-Cooper (1980) see the analysis of parenthetical constructions "which occur quite frequently in running discourse" as an important step and predict furthermore that "the role of emphatic stress will probably play a major role." They further remark that:

> *"Speakers utilize emphasis quite liberally in normal conversation; this feature lends much of the "color" to speech required to maintain a listener's attention. For long-term applications of speech synthesis,..., the proper use of emphatic stress could make a big*

*difference* ...." (Cooper and Paccia-Cooper 1980:224)

The introduction to this thesis has identified the study of the acoustic correlates of textual markers as potentially contributing to the achievement of more natural, more expressive and more conversational prosody for Text-to-Speech conversion systems. A detailed examination of the genre of *Written Conversation* has shown that there is a case for arguing that certain textual markers have some prosodic significance for the task of reading aloud texts. Exactly how this prosodic significance is manifested is examined in the following two chapters, which investigate the prosodic correlates of two of the most frequently used textual markers in written conversation, parenthesis and emphasis markers.

# Chapter 3

# Parenthetical Constructions

**parenthesis**

**1.** A phrase, often explanatory or qualifying, inserted into a passage with which it is not grammatically connected, and marked off by brackets, dashes, etc.

(Collins English Dictionary).

## 3.1  Introduction

Chapter 2 considered the prosodic significance of certain textual markers by arguing that authors translate a phonetic plan, or internal speech, into textual annotations which function as prosodic markers and that readers use these cues in order to attempt to provide the particular prosodic form reflecting the authors intentions.

This and the following chapter are concerned with examining readers' performance in translating prosodic markers back into a spoken representation - thereby measuring the *acoustic correlates* of these prosodic markers. This thesis concentrates on the two most frequently occurring prosodic markers in written conversation, parenthetical constructions, examined in this chapter, and emphasis markers, the topic of Chapter 4.

Although the most obvious application of the results lies in the specification of rules for Text-to-Speech conversion, it is expected that results support basic research in the respective areas, and more generally contribute to a better understanding of the relationship between pragmatics and phonetics.

This chapter examines the relationship between parenthetical constructions in the written language and their acoustic representation in the spoken language, that is the correspondences and differences of what may be considered *parenthetical* in the written and parenthetical in the spoken language. Parentheticality in the spoken language is often considered to involve noticeable juncture, lowering or contraction of pitch range and sometimes increases in speech rate. These characteristics are used by some scholars, as will be examined in detail below, to argue that parentheticals in the spoken language are *prosodically independent* units. The central goal of this chapter is to determine through acoustic and perceptual analysis if the claim of prosodic independence can be upheld and if so, how this notion of spoken parentheticality applies to the variety of 'parentheticals' found in written conversation.

Some of the acoustic correlates investigated below for their contribution to possibly prosodically independent structures include fundamental frequency contour effects such as F0 range and possible effects of parentheticals on the first F0 peak of an utterance (P1). Duration measures include pausing, pre-clausal lengthening and possible effects of the length of the parentheticals on the prosodic structure of the surrounding utterance. However, prior to an acoustic analysis of parentheticals the background literature is examined in section 3.2 followed by the results from a perceptual study into the prosodic independence of parentheticals carried out for this thesis.

Let us first consider the definition at the beginning of this chapter, which makes a number of statements which will be examined at various stages throughout this chapter. From a functional semantic point of view parentheticals are characterized as being of qualifying or explanatory nature and orthographically delimited by brackets or dashes. Grammatically speaking, Collins defines parentheticals as grammatically "unconnected" and "inserted" phrases. The functional characteristics are fairly uncontroversial, as discussed below in section 3.2, but the issue of grammatical unconnectedness bears vital implications for any claims about the prosodic behaviour of parentheticals in relation to their surrounding clauses.

A few examples taken from electronic news and mail messages will show that it is necessary to extend the above definition of parenthesis on a purely orthographic basis, thus providing for the possibility of differences between spoken and written parentheticals. Henceforth, a parenthetical construction is understood to be any piece of text whether a single word, a phrase, clause or even a whole sentence delimited from its surrounding text by brackets or other devices such as dashes or commas in certain instances. By far the most widely applied symbol signalling separation or delimitation of parentheticals are parentheses themselves, and due to the large amount of available data the selection of sentences for the experimental corpus has been restricted to parentheticals marked by brackets.

## 3.1.1 Parentheticals in written conversation

The following examples show the diversity with which parentheticals are used in written conversation and the problems this may present for their analysis.

1     ... In Australia it is very common for neighbouring speech forms ( a term I will adopt as neutral between "dialect" and "language" ) to be mutually comprehensible, but for intelligibility to drop off rapidly as distance increases. ...

2     ...There are corresponding non-sonorant codes files ( which in fact are identical for the two accents ) in the same directories. ...

3     ... For those whose memory is worse than mine - in 1989, after upsetting "Big Mac", Haarhuis beat someone ( I have a vague feeling that it was Sampras!! ) but lost to Krickstein ( who lost tamely to Becker then, right ?? ) . Basically, my forecast is based on my (whimsical ) conclusion that the maximum number of matches that Haarhuis can last on his top-class-player beating spree is 2. ...

4     This is *not* intended to start a flame war ( and I am not defensive ), but it's funny that my parents ( shrinks ) always warned me about the kids of school teachers, especially of the public school variety . :-)

5     ... I have several in my collection, and my favorite shows the loyal Indian volunteer ( complete with turban ) manning an MG-34 against those frightful Brits. ...

6     ... Sure, they would've found some crack-pot supporters ( Bose wasn't the only one who believed in violence as a means to freedom ), but the Gandhi/Nehru leadership of the Congress Party ( which led India's freedom struggle ) was fairly intelligent and sensible ( at that time ); ...

7     ... If you sent a message saying, "Look I just measured an effect which seems to point to a problem in your model of physics," and you could

back it up and it didn't apparently contradict innumerable other well-documented observations to the contrary ( "Hey I just dropped an apple and it fell up" ), I'd treat you ...

These examples show the diversity with which writers use parenthesis in this genre, indicating that it is possibly one of the most used techniques for signalling aspects of information and prosody in written conversation. This results in parentheticals consisting of single words, whole sentences, relative clauses with and without pronouns, conjunctions, adverbial phrases and more. Also note other markers, for example those for special emphasis, examined in detail in Chapter 4, and those giving some attitudinal or emotional clues.

From a purely functional standpoint parentheticals may be described as providing background or explanatory information in form of an 'immediate footnote', the relevance of which depends on the amount of contextual information available to the reader or listener at the time. The reason for the extended use of this technique lies in the nature of the medium where messages may be sent to hundreds of people at a time, especially in on-line discussion groups. Anticipating the lack of context available to certain readers, either through having missed earlier parts of the discussion or through a lack of knowledge on the subject, authors continuously provide contextual information in this way in case it is required. This aspect will be discussed below in Section 3.2.3 in conjunction with an extended functional account of parentheticals provided by Nunberg (1990).

## 3.2 Background

The study of parenthetical constructions has been approached mainly from two angles - firstly, as part of the examination of phenomena from the written language

which includes the analysis of written text and discourse, and in particular treatments concerning the linguistics of punctuation (Levinson 1986, Nunberg 1990; Meyer 1987). Parentheticals have also been approached with a view to their syntax, especially with respect to discontinuous constituent structure (Downing 1970; Emonds 1976, 1979; McCawley 1982; Postal 1964; Ross 1973; Wells 1947), as discussed below.

Secondly, and not as well represented in the literature, are accounts of parentheticals regarding their spoken form, such as their prosodic, intonational or acoustic representation. Although many different types and classifications are presented in the textual analysis of parentheticals, some writers on prosody or intonation tend to provide more generalized accounts. Bolinger (1964:25) refers to parentheticals as one of the most "conspicuous instances of separation" marked by a lowering of pitch, the insertion of pauses and a reduction in volume. Accents in parentheticals, according to Bolinger, occur in their expected places but are "flattened somewhat", suggesting the contraction of overall pitch range.

Crystal (1969:144) makes similar observations in that longer parentheticals "will be introduced with low onset" provided they cover more than one tone unit. In a section on contrasts of loudness Crystal suggests that "piano" is frequently used "with a potential for making important structural contrasts (for example parenthesis) ..." (p.160). This structural contrast, he believes, is further underlined by an increase in speech rate "in parenthetical speech and excited speech" which he suggests correlates with marked pitch range which is wide for excited but narrow for parenthetic speech (p.174). As far as accent realization and placement is concerned, Crystal points out that a high proportion of parentheticals carry the final element of a fall-rise tone as, for example, in the utterance "He said, you see, that ..."

and that the nucleus in parenthetic constructions, provided there is one, carries rising rather than falling tones.

Other writers on prosody generally mention a narrowing of fundamental frequency range in parentheticals (Monaghan, 1991) or the "lowering of intonation" (Silverman, 1987), effects which are mirrored also in other languages like German, Italian (Chapallaz, 1964) and French (Delattre, 1966). In this area, differences between classes of parentheticals are generally not drawn, with the exception of Cruttenden (1986) and Couper-Kuhlen (1984), who at least distinguish between sentence medial and sentence final parentheticals with respect to the insertion of pauses.

Studies investigating the acoustic nature of parentheticals are few and far between. The most notable contribution is found in Kutik, Cooper and Boyce (1983), who argue for independent declination of F0 in parentheticals together with an independent durational structure and thus independent mental programming of these constructions. Their work was principally motivated by earlier studies on declination by Cooper and Sorensen (1981). Other relevant results concerning duration, pausing and boundary tones are discussed in a study by Garro and Parker (1982) on suprasegmental characteristics of relative clauses, although they were more concerned with prosodic events at clause boundaries rather than within. These studies will be discussed in more detail below.

The at best marginal interest in parentheticals in studies of spoken language, however, is not surprising. Firstly, it is probably considered that the acoustic correlates suggested above sufficiently describe the acoustic nature of parentheticals. Secondly, the important function of parentheticals to serve as indicators of *information structure* may have been underestimated, although Bolinger (1964:26) refers to the marked separation of parentheticals, "where lowered pitch

suggests a lower ranking element in the discourse". Parentheticals and other textual devices can be regarded as small but safe pointers to some discourse structure, which is important in translating between orthographic and prosodic structure. Silverman (1987) recognizes this function and lists parentheticals among other pragmatic discourse factors such as "given" and "new" information, topic and comment etc. as examples where speakers encode information structure prosodically. It is exactly this important function which is reflected and probably enhanced somewhat by the extended use of parentheticals in written conversation.

The resulting situation is that there are functional classifications from textual analysis, grouping parentheticals into different types, but acoustic accounts tend to generalize from a particular type of parenthetical to the class of all parentheticals. It is only reasonable to suggest that functionally different parentheticals may also have different acoustic representations, and it is hoped that the analysis of parentheticals as they appear in written conversation will show that not all parentheticals are the same with respect to their acoustic realization.

This section gives equal prominence to accounts based on written and spoken language, because the acoustic and perceptual analysis of a large variety of parentheticals (section 3.3) may have to rely on groupings or classifications of parentheticals based on syntactic, orthographic or functional semantic accounts in order to explain identified acoustic differences. The following section summarizes the claim that parentheticals may by be prosodically independent, sections 3.2.2 and 3.2.3 search for analogous arguments by examining syntactic and functional accounts of parentheticals.

## 3.2.1 The independence hypothesis - Part I

Some preliminary hypotheses of the acoustics of read parentheticals, can be made by considering what the reader has to do on encountering an utterance containing a parenthetical string.[1] The reader has to achieve a separation between the background information contained in the parenthetical and the main proposition of the utterance, which may contain important new information. This means that the structure and meaning of the main proposition must only be interrupted but not destroyed by the presence of bracketed or otherwise delimited text. This is most difficult if the parenthetical occurs in sentence medial position, as this has the strongest effect on the intonation contour of the utterance which is interrupted by the insertion of the parenthetical. It is relatively uncontroversial, if hardly ever quantified in terms of acoustic measurements, that this is mainly achieved by contracting the fundamental frequency range, by inserting pauses before and after the parentheticals, and by increasing the speech rate, thereby shortening the duration of the segments or syllables within (Bolinger 1964, Crystal 1969, Ladd 1980).

If we accept in principle a certain relationship between syntactic form and prosodic form as advocated, for example, by Chomsky & Halle (1968), Halliday (1967) and Selkirk (1984), grammatical unconnectedness could, in principle, suggest the prosodic independence of parentheticals.[2] This does not require us to accept, a

---

[1]A competent reader is assumed who is capable of transmitting adequately and unambiguously their perceived meaning of the sentence or text. Because this is difficult, if not impossible to measure, an ideal situation would call for the use of a professional reader.

[2]For a comprehensive review on the syntax-prosody relationship, the reader is referred to Monaghan (1991) where as part of the discussion on "designer" approaches to intonation research, a distinction is drawn between syntax-based and focus-based accounts and historical reasons are offered for the almost blind acceptance of syntax based approaches amongst the "producers" in intonation research.

priori, that syntax determines prosody, because one could equally convincingly argue that a functional semantic notion such as "Parentheticals provide supplementary information" has influenced both syntactic and prosodic information in a similar way. This corresponds to the "Parallel Model of Prosody Generation" advocated by Monaghan (1991, p.34). By that account syntactic form may simply be mirrored in prosodic form without necessarily determining it.

Nevertheless, aspects of syntactic classification of parentheticals are worth exploring because it may turn out that there is indeed a strong correlation between syntactic form and prosodic form in parentheticals for whatever reason. This is done below in a discussion on discontinuous constituent structure and tone group assignment (section 3.2.2).

The two related observations, that readers attempt to preserve the structure of the parenthetical surrounding material by singling out the parenthetical and that parentheticals may be grammatically and prosodically disconnected from the remainder of the utterance, can be summarized in the *independence hypothesis*. This hypothesis has been investigated to some extent by Kutik, Cooper and Boyce (1983), who have shown for one particular type of parenthetical that a separate declination which is independent from the main clause can be attributed to the parentheticals. Furthermore, they suggest that the first post-parenthetical peak is neither a local effect nor a resetting of the declination of the main clause but rather a resumption to the previously interrupted declination of the utterance. Their work is described in detail below.

Prosodic separation or even independence as an indicator of grammatical unconnectedness is important with respect to processing these constructions in a Text-to-Speech system. For example, prosodic rules for parentheticals could be kept separate from the main set of text conversion rules. This would allow parallel

processing of text with the removal and later re-insertion of the parenthetical constructions resulting in simpler syntactic structures for the residual text, thus facilitating syntactic parsing of the textual input.

Therefore, prosodic independence and the acoustic characterization of parentheticals is the main theme of this chapter and the aim is to identify the classes of parentheticals - provided that a sensible classification is possible - for which these claims can be upheld.

## 3.2.2  The syntax of parentheticals

This section addresses the question of whether there is any evidence from syntactic theory which would provide a parallel argument to the one presented in the previous section, in terms of the structural independence of parentheticals. In other words, we are looking for a syntactic argument for the independence or grammatical unconnectedness of parentheticals - which may provide one example of the interdependence between syntactic and prosodic structure at least as far as parentheticals are concerned.

Some evidence which indirectly suggests the independence of parentheticals is present in writings concerned with *discontinuous constituent structure*, in particular McCawley (1982), and Wells (1947), but also Ross (1973) and Emonds (1976 & 79).

Consider the following structure taken from Wells (1947):[3]

---

[3]The circled node is a discontinuous constituent which dominates items without dominating intervening constituents.

His father according to John is the richest man in Scarsdale
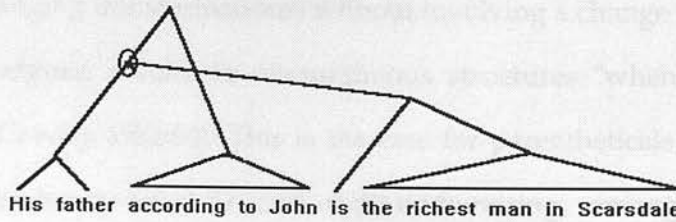
Here, it can be argued that a movement operation on the last constituent of the sentence "His father is the richest man in Scarsdale, according to John" separating the NP from the VP by an intervening constituent, results in a discontinuous constituent structure. As a result, "father" dominates "is the richest man in Scarsdale" but not the intervening parenthetical clause. As a result of such movements it is reasonable to suggest that the resulting discontinuous surface structure will also result in a *discontinuous prosodic structure* when the sentence is spoken aloud.

Although, for reasons which go beyond the scope of this chapter, some transformationalist grammarians have rejected the concept of discontinuous structures[4], even its proponents disagree with respect to one important technical detail - what exactly does the movement transformation do, that is, where is the structure attached and what is its constituency.

The main assumptions upon which McCawley bases his arguments are that the deepest relevant syntactic structures are ordered continuous trees and that movement transformations can be distinguished with regard to whether they change syntactic relations (relation-changing transformations), sometimes

---

[4]For a precise account on the rejection of discontinuous structures the reader is referred to Postal (1964, pp.69-70) or McCawley (1982, pp. 91-93)

accompanied by a change in order, or whether their sole purpose is a change in order (order-changing transformations) without involving a change in constituency. The latter, he argues, results in discontinuous structures "when nonsisters are permuted" (McCawley 1982:94). This is the case for parentheticals, non-restrictive relative clauses, heavy-NP-shift and right-node-raising, amongst others. The differences between the analyses offered relate to constituent structure and are displayed in the following trees (McCawley 1982).



1. Input tree
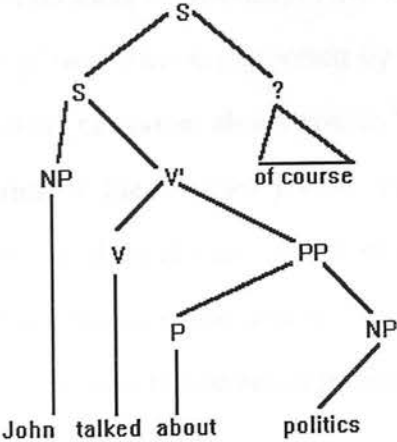
Emonds [76;79]
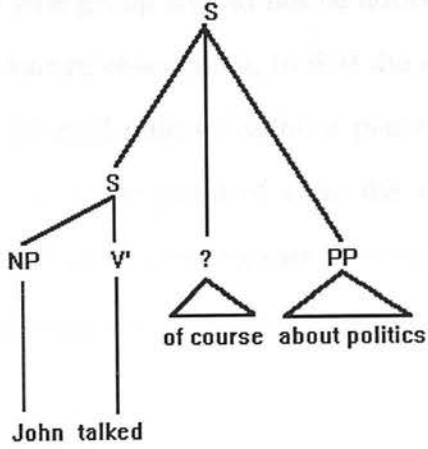
Ross [73]
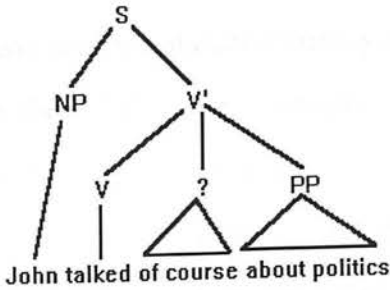
McCawley [82]
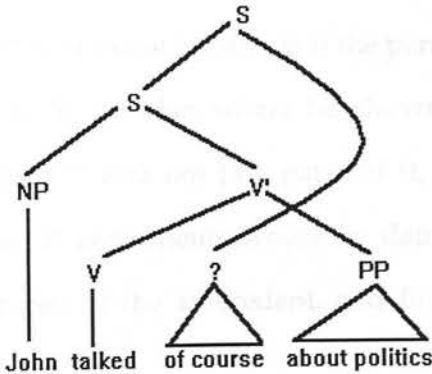
For Ross, the parenthetical is moved into the Verb Phrase and becomes one of its constituents, whereas for Emonds, the Prepositional Phrase is extracted and placed after the parenthetical with no subordinate relationship to the verb phrase. For McCawley, the structure of the verb phrase does not change at all, resulting in the mentioned discontinuity. For the prosodic realization of these structures we could therefore hypothesize that the structure proposed by McCawley would simply result in a temporary interruption of the intonation phrase "talked about politics" by an independent inserted tone group "of course".

In the analysis offered by Ross the verb phrase gains new material which would suggest its prosodic integration with its sister constituents, although with this analysis there is no reason why a new tone group should not be added inside the verb phrase. This is supported by an intuitive observation, in that the sentence "John talked of course about politics" can be read with or without pauses at the boundaries of the inserted phrase. From a prosodic point of view the structure proposed by Emonds seems the most counter-intuitive because it involves the complete syntactic separation of the verb phrase and hence the separation of the intonation phrase "talked about politics".

In addition to his analysis on constituency, McCawley offers further arguments which provide evidence from syntactic theory which support the independence hypothesis. He goes as far as saying that in fact, "all grammatical phenomena to which the constituency of V' is relevant behave as if the parenthetical were not there." He offers examples from V' deletion where he shows that the controller of deletion is the whole original V' and not just parts of it, from the relationship between pronouns and their V' antecedents where he demonstrates that the parenthetical does not count as part of the antecedent, and finally from

topicalization where he argues that a V' derived from Emond's analysis can not be topicalized .[5]

In summary, it can be said that McCawley's analysis of phrases appearing in surface structure as discontinuous constituents, causing no change in constituency in the clauses in which they appear, does provide evidence from syntactic theory which supports the independence hypothesis and which supports the hypothesis that there may be a correlation between syntactic discontinuity and prosodic discontinuity or independence. McCawley's analysis supports findings on the acoustic realization of parentheticals presented section 3.2.4 below, as well as perceptual results reported below which both involve syntactic structures where the parenthetical independently attaches to the root sentence.

## 3.2.3 The function of parentheticals

The purpose of this section is to complement the hypotheses laid out in the two previous sections on the independence of certain parentheticals by examining some of their functional characteristics. Similarly to the way in which a particular syntactic analysis provided a structural analogy to that of prosodic independence, an examination from a functional semantic point of view may provide further evidence.

Nunberg (1990) provides a comprehensive functional account of parentheticals as part of a study on the linguistics of punctuation in which he discusses, and refutes, the claim that written language - and specifically

---

[5]For examples the reader is referred to McCawley (1982)

punctuation - is merely a means for transcribing the intonation of spoken language. He does so by suggesting various functional categories, positional occurrence restrictions and some co-occurrence restrictions with other orthographic devices such as commas and dashes. The following concentrates on categories because his occurrence restrictions apply in many cases to literary text and are likely to be violated in the more colloquial genre of written conversation.

### 3.2.3.1 Textual versus lexical parentheticals

For Nunberg there is a fundamental, and basically syntactic, distinction between lexical and textual parentheticals which are subject to differing constraints and interpretations (Nunberg, 1990, p.21), although they can take on the same or similar functions. This distinction relates mainly to the syntactic mechanisms by which parentheticals are introduced. Lexical parentheticals are considered to be introduced by the "lexical grammar" and generated by lexical phrase structure rules and mostly function as "alternate text parentheticals", as the following example shows (Nunberg 1990:111)

8      He made his remarks to (then Secretary of State) Henry Kissinger.

Here it is possible to construct an alternate text which would consist of the same utterance as in 8 but without the parenthesis, addressing the needs of a reader who possibly does not know who Henry Kissinger was.

Textual parentheticals on the other hand are generated by a "text-grammar", where the parenthesis is one of a set of graphical devices which Nunberg calls "text-category indicators" (Nunberg 1990:17). A textual parenthetical is therefore a text category or structural entity in the written language just as for example a quotation, a paragraph, a section or even a chapter, delimited by its corresponding text

category indicator. The following example shows a textual parenthetical from written conversation.

9        ... Sure, they would've found some crack-pot supporters ( Bose wasn't the only one who believed in violence as a means to freedom ), but the Gandhi/Nehru leadership ...

Nunberg's distinction between lexical and textual parentheticals is a matter of grammar rather than function and is included in this discussion because it may help to explain the prosodic independence of certain parentheticals. Furthermore, this distinction may be easier to apply than some of the more fine grained functional distinctions which are made below.

The lexical grammar describes the syntactic dependencies between lexical items in what Lyons (1968) would traditionally refer to as the "sentence" which is "a grammatical unit between the constituent parts for which distributional limitations and dependencies can be established, but which can itself be put into no distribution class" (p.173). In contrast to this notion of the sentence, which Brown and Miller (1980) describe as having "a certain sort of unity", grammatical completeness and the ability to "stand on its own" with a "degree of semantic independence" (p.149), Nunberg posits a category which he calls the "text-sentence". This is a string of written characters delimited by a capital letter and full stop and can be made up of single or multiple lexical sentences but also of clauses which do themselves not constitute lexical sentences. The notion of a text sentence is illustrated in the following example (Nunberg, 1990, p.22) which also contains a textual parenthetical.

10       "The poor (we would now call them the disadvantaged) cried; Caesar wept - what else would you have had him do ?"

This text sentence is made up of the two (lexical) sentences, 'The poor cried' and 'Caesar wept' which are independent, and the parenthetical together with the question which are dependent on the surrounding sentences for interpretation. Nunberg's distinction is exactly what Brown and Miller (1980) describe as the difference between grammatical rules and text formation rules. Text formation rules can produce sentences which can only be understood in relation to other sentences because the rules allow for ellipsis as well as the insertion of proforms and binding expressions.

The essential difference between lexical and textual parentheticals concerns the level at which they are introduced, textual parentheticals are introduced by a text grammar at the discourse level, lexical parentheticals are introduced at the level of phrase structure grammar.

3.2.3.2 Functional distinctions for presumed and alternate readers

From a functional rather than a grammatical angle, Nunberg distinguishes parentheticals which provide "alternate text" (example 11), restrict the context of the proposition (example 13), or provide "in case you're interested" type information thus providing alternative readings, elaborations and points of digression.

11      Franz Beckenbauer (the Kaiser) was one of the best footballers ever.

Sentence 11 would be classified as a lexical parenthetical by Nunberg, but textual parentheticals are also used to provide alternate readings as in 12.

12      Among these will be the unit set (that is, the set with one member)
        (Nunberg 1990, p.112)

13      The new Ford is selling very well (in America)

Further functional distinctions are made by Nunberg based on the notions of *presumed* and *alternate reader*[6] which is a parallel to the remarks made earlier about the assessment made by authors on how much context or knowledge is available to certain readers. Most of these differences involve "in case you're interested" parentheticals.

In some cases parentheticals are inserted because the alternate reader is thought to know less about a particular subject than the presumed reader, such as 14, in other cases, the alternate reader may know more than the presumed reader and additional knowledge is offered (example 15), although this does not mean that the presumed reader necessarily knows less than the alternate reader.

> 14    He made his remarks to (then Secretary of State) Henry Kissinger
>        (Nunberg 1990:111)

> 15    Capital profit was that of a medium sized company (minus tax
>        deductions, standing charges and dividend payments) and was
>        expected to grow.

Other parentheticals of elaboration may be inserted in order to pre-empt possible criticism from literal minded readers such as the following

> 16    Clearly (any utterance of) the sentence "It is raining" will be true only
>        if it is raining when the (utterance of the) sentence is spoken.

---

[6]The distinction between presumed and alternate reader is an important one for written conversation and explains the heavy use of parentheticals in this genre. This is the case because an author often does not know who his audience is. A presumed reader has to be 'identified' or fixed in the authors mind and provisions (in the form of parentheticals) have to be made for possibly a large number of alternate readers.

The grammatical distinction between lexical and textual parentheticals is a useful one and may have implications for the prosodic independence of parentheticals. There is certainly a strong case for textual parentheticals, which do not actually interfere with the phrase structure of the sentence they relate to, to result in unambiguously discontinuous structures which are prosodically independent units, introduced at a high level of linguistic processing.

Lexical parentheticals, on the other hand, are introduced at a lower level, in the case of lexical alternate text parentheticals simply by copying a phrase structure and attaching different lexical items to it, as shown in sentence 11 above. This may result in syntactic discontinuity, dependent on the analysis, but the boundaries are likely to be weaker, affecting the prosodic independence of the parenthetical. This point is best illustrated by comparing example 11 with example 7 in 3.1.1 above.

### 3.2.4 Accounts on the spoken realization of parenthetical clauses

The most comprehensive study into parentheticals from read speech is found in Kutik et al. (1983). Their main concern in analysing the production of parentheticals and main clauses is the phenomenon of *declination*, which is the continual decline of fundamental frequency contours over the course of an utterance (Pierrehumbert 1980; Lieberman 1967; O'Shaughnessy 1976). The setting for the study in Kutik et al. is the continuing controversy of whether involuntary physiological or precise cognitive processes are responsible for this phenomenon.

The particular model of declination used in the study is known as the *topline rule*, which claims to predict F0 values for intermediate peaks in single declarative read sentences of English. Their main motivation comes from the finding that the presence of parentheticals does not seem to disturb the main clause topline (Cooper

and Sorensen 1981). More specifically their aim is to measure F0 within the parentheticals and to establish whether the length of the parenthetical has any influence on the declination of F0 in the main clause. Furthermore, they investigate the influence of parentheticals on the first peak of an utterance (P1). These accents have been shown to be higher for longer main clauses (O'Shaughnessy 1976; Cooper and Sorensen 1981), but not affected by the insertion of parentheticals and the resultant increase in sentence length.

For that purpose, seven varieties of one base sentence were recorded, where a parenthetical was inserted between the Noun Phrase and the Verb Phrase of the main clause and which in each sentence was successively lengthened by inserting at least one extra stressed syllable into the parenthetical clause. Thus the shortest sentence is the one in 17a below and their longest sentence the one in 17b.[7] Measurements were taken from recordings obtained from six speakers of the F0 peaks of key stressed segments (italicized), the durations of the main clause, the parenthetical clause, clause final syllables within each, and the pre- and post-parenthetical pauses.

17a    The *clock* in the *church*, it oc*cur*red to *Clark, chimed* just as he began to *talk*.

17b    The *clock* in the *church*, it *never* in a *mil*lion years would have oc*cur*red to the absent-minded *Clark, chimed* just as he began to *talk*.

---

[7]The italicized words contain the syllables from which F0 and duration measurements were taken by Kutik et al. (1983)

Topline predictions were also calculated for each sentence with the following topline prediction equation[8]:

$$F0 = Pn + \frac{2}{3} \frac{(P1 - Pn)}{(T1 - Tn)} (t - Tn)$$

The following table shows mean peak values from key locations in the seven sentences. Standard deviations ranging from 36.0 (Hz.) - 56.5(Hz.) have been omitted from the table.

Table 2: Mean peak values (in Hz.) from key locations in sentences (1)-(7)

| Sentence | P1 (1st stressed syllable of m.c.) in Hz. | parenthetical-final segment in Hz. | P'(return to m.c.) in Hz. | PN (m.c.-final segment) in Hz. |
|----------|-------------------------------------------|------------------------------------|---------------------------|--------------------------------|
| 1 | 204 | 155 | 175 | 153 |
| 2 | 195 | 151 | 171 | 149 |
| 3 | 196 | 149 | 172 | 150 |
| 4 | 199 | 145 | 170 | 143 |
| 5 | 206 | 147 | 171 | 149 |
| 6 | 200 | 146 | 167 | 141 |
| 7 | 197 | 149 | 175 | 140 |

It was noticed that declination proceeded normally from the main clause into the parenthetical until the point of return to the main clause after the parenthetical (P') which is characterized by a sharp jump well above the predicted declination topline. This can be seen clearly by comparing columns 3 and 4 of the above table. After calculating new topline values for the main clause by omitting the parenthetical clause from the calculation, they found that the predicted main clause

---

[8]P1 is the first peak in the utterance at the time T1, Pn is the last peak of the utterance at the time Tn, t is the time at which the intermediate peak value is predicted by the rule

topline and in particular P' was much more accurately predicted than a topline computed on the basis of the whole sentence length.

For example, for an initial peak (P1) of 202 Hz., an utterance final peak (Pn) of 140 Hz. and a peak of 176 Hz. where the parenthetical returns to the main clause (P'), the topline rule predicts the following. It the calculation is based on the total utterance length including the parenthetical the predicted value for P' is around 158 Hz., 18Hz. lower than the actual value of P'. If, however, the calculation is based on the total utterance length minus the length of the parenthetical, i.e. the parenthetical is treated as if it were not present, the predicted topline value of P' is around 179Hz. The latter is a considerably more accurate calculation since the predicted value is only 3 Hz. higher than the actual value.

This behaviour was taken as evidence that the jump to P' represents a return to the declination norm of the main clause, supporting a separate processing hypothesis which "entails the assumption that the presence of the parenthetical has no effect on the computation of declination for the main clause" (Kutik et al. 1983:1734). The topline function was also applied to parenthetical peaks only, in sentences where the parenthetical was of sufficient length for the application of their particular model and clearly showed separate declination and reduced pitch range for these clauses as predicted by the rule. P1 did not increase with the length of the parenthetical.

Further evidence for the independent suprasegmental programming of parentheticals and main clauses comes from a durational analysis of the data (Kutik et al. 1983), in which no effect is reported of the length of the parenthetical on the duration of either part of the main clause. Previously, the length of upcoming

material had been shown to influence pre-clausal lengthening[9] (Cooper and Paccia-Cooper 1980). Unfortunately, a token of the sentence which contained no parenthetical at all was omitted from the recordings, because it would have provided more reliable durational analysis of the main clause duration. In addition, parenthetical length did not influence the duration of parenthetical-final and main clause-final syllables. The fact that the clause final syllable of the first part of the main clause did not vary with increased parenthetical length was taken as evidence for parentheticals not constituting an environment for pre-clausal lengthening. The findings are based on measurements of the duration of the phonologically long central vowel of the word *church* in the seven sentences. Considering the intrinsic length of that vowel and the fact that it is the most likely location for nucleus placement in that context, it is likely to be close to its maximum length despite its position. As will be discussed below, Garro and Parker (1982) show clearly that a pre-clausal lengthening effect (before non-restrictive relative clauses) can be identified.

For pausing, the only consistent finding was that post-parenthetical pauses are consistently longer than pre-parenthetical pauses but no correlation between the length of the parenthetical and pause durations was found.

From an analysis of clause durations, Kutik et al. also infer that "speaking rate did not increase with greater parenthetical length" (Kutik et al. 1983:1737). Presumably they were referring to the total duration of both parts of the surrounding main clause which remain essentially constant because speech rate inside the parentheticals was not measured. This is an unfortunate omission

---

[9]Pre-clausal lengthening here is used to describe what is essentially clause-final lengthening. The former is a more correct term for cases where parentheticals interrupt main clauses as at the time of interruption the main clause is not finished.

because a marked increase in speech rate inside the parenthetical could have pointed to the existence of some temporal constraint on post-parenthetical F0 resetting, because it might have suggested that a reader would have to get the parenthetical over with quickly in order to return to the main clause. This would have also supported the intuitive observation made earlier that the structure of the main clause is preserved by the prosodic demarcation of the intervening parenthetical in terms of F0, duration, intensity, pitch range and speech rate. Moreover, if such an increase in rate were not identified it would support one strand of their argument further, in that the ability of a speaker to reset F0 following a parenthetical would be shown to be independent of the amount of intervening material and thus intervening time.

Further evidence relevant to the acoustics of parentheticals can be found in a study concerning the acoustics of relative clauses (Garro and Parker 1982). Non-restrictive relative clauses such as 18, show great similarity in function to parentheticals by providing "in case you're interested" or "by the way" type information.

18      My brother, who is a lawyer, lives in Germany

Observations concerning the spoken realization of non-restrictive relative clauses are also similar to those made about parentheticals, and argue especially for separate intonation and separation through the insertion of pauses (Halliday 1967; Lee 1956).

In order to collect further, non-anecdotal evidence, Garro and Parker recorded sixty tokens (thirty from each of two speakers) of a relative clause (see 19), which varied in type (restrictive/non-restrictive), word order (SVO/OSV), and type

of relative pronoun (who(m)/that). The restrictive (RRC) and non-restrictive (NRRC) readings are shown in 19.

19 Gigolos(,) who prefer blondes(,) are fickle

These were recorded in isolation, in direct contrast, and in short paragraph conditions. They measured pitch changes in the word preceding each clause boundary, the length of the last vowel preceding each clause boundary and the duration of the pauses if present. The following table shows results for pause length and vowel length:

Table 3: Average vowel and pause lengths (ms.) for RRC's and NRRC's within each condition and across conditions (Garro and Parker 1982 p.157)

| Condition | Clause type | V1 | V2 | P1 | P2 |
|---|---|---|---|---|---|
| Context | RRC | 180 | 218 | - | 24 |
| Context | NRRC | 264 | 303 | 118 | 202 |
| Isolation | RRC | 173 | 199 | - | 15 |
| Isolation | NRRC | 269 | 276 | 44 | 95 |
| Contrast | RRC | 167 | 192 | 4 | 19 |
| Contrast | NRRC | 258 | 277 | 53 | 97 |
| All | RRC | 173 | 203 | 1 | 19 |
| All | NRRC | 263 | 285 | 71 | 131 |

One of the main findings is that the last main clause vowel and the last relative clause vowel seem to be subject to pre-clausal and clause-final lengthening respectively, and that this effect is enhanced in NRRC clauses for which the vowel lengths are approximately 50% longer as those in RRC's. This seems to be in direct contrast to the claim in Kutik et al. who found no durational effect in clause-final syllables in their data. This either points to contradictory evidence in the two studies, or shows fundamental differences between the non-restrictive relative clauses investigated in Garro and Parker and the parentheticals examined in Kutik et al. with respect to duration. This possibility will be investigated in the experimental section below.

The pausing results support Kutik et al., with post-relative clause pauses being consistently longer than pre-relative clause pauses. F0 was only measured at boundaries where restrictive relative clauses exhibit falling pitch in pre-clause and clause-final position as opposed to rising pitch in NRRC's. F0 of the relative clauses as such was not part of the study reported by Garro and Parker.

The study by Kutik et al. (1983) was possibly the first real attempt to describe the prosody of parenthetical constructions in English based on acoustic measurements, rather than purely on hypothetical or impressionistic judgements. Although it certainly supports many of these judgements by supplying analyses from spoken data, the study suffers from generalizations which are based on the analysis of relatively small amounts of data, and very specific, in many cases specifically designed sets of sentences. That of course has the advantage that experiments are more easily constrained and specific hypothesis are more readily testable. The disadvantage is that in this case the results based on nine specifically designed sentences are generalized to parenthetical constructions in English, although only one type of parenthetical was investigated. The investigation in Garro and Parker (1982) was also based on one single carrier sentence.

One of the aims of the experiments discussed below, dictated on the one hand by the need to investigate different types of parentheticals and on the other hand by the type of application that is envisaged, is the examination of as many different parentheticals as possible taken from written conversation, that is from real language use. As previously discussed, the notion of a 'bracketed string' from written conversation could be very different from the notion 'parenthetical constructions in English'. The aim of this chapter can therefore be described as the determination of which of these bracketed strings behave like parentheticals in the sense that they are prosodically independent, and which ones do not.

Furthermore, it is necessary to establish whether parentheticality correlates with any surface, syntactic or functional semantic classification, and if it is possible that some constructions which may appear are more 'parenthetical' than others. The methodology is that of applying perceptual experiments in order to test for prosodic independence.

## 3.2.5 The independence hypothesis - Part II

We can now begin to create a more complete picture which shows how the above accounts from different levels in the linguistic hierarchy provide possible explanations for why certain types of parentheticals seem to be prosodically independent and how these explanations interact. One observation is that the parentheticals described both in Kutik et al. (1983) and McCawley (1982) would classify as textual parentheticals in Nunberg's (1990) terms.

Nunberg argues that the syntactic processes for inserting textual parentheticals are much more complex than those required for the insertion of lexical ones because they involve operations at the discourse level (Nunberg 1990:112). This may suggest that the insertion of the former produces discontinuous structures at the highest level in the constituent structure. Lexical parentheticals may also have discontinuous structures but they attach to lower nodes in the tree, for example as phrasal constituents such as in 20.

20 Mussels (in season) are a delicious treat.

The following figure summarizes the above accounts by showing at which levels of the linguistic structure we find arguments supporting the prosodic independence of certain types of parentheticals.

Figure 4:    Linguistic dependencies in prosodically independent
            parentheticals

```
                    ┌─────────────────────────────────────────────────┐
                    │  DISCOURSE & PRAGMATICS                          │
                    │                                                  │
                    │  Text formation rules     Brown & Miller 1980    │
                    │  Textual grammar          Nunberg 1990           │
                    │  Information status       Chafe 1974             │
                    └─────────────────────────────────────────────────┘

┌──────────────────────────┐              ┌──────────────────────────┐
│  SYNTAX                   │              │  PHONOLOGY               │
│                          │  ⟵────────⟶  │                          │
│  Discontinuous structure │              │  Prosodic discontinuity  │
└──────────────────────────┘              └──────────────────────────┘
                                                        │
                                                        ▼
                                          ┌──────────────────────────┐
                                          │  PHONETICS               │
                                          │  Acoustic demarcation    │
                                          │  F0; F0 range; rate; juncture; │
                                          │  durations               │
                                          └──────────────────────────┘
```
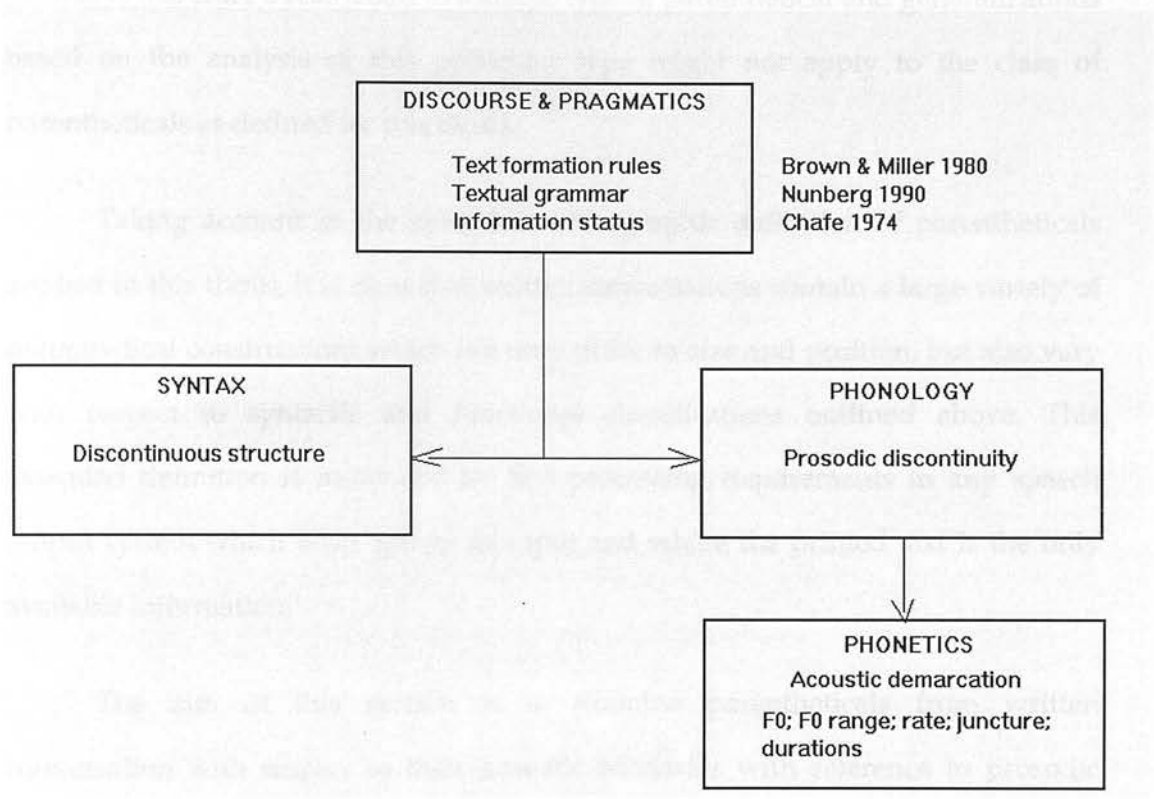
## 3.3   The perceptual analysis of parentheticals

The preceding sections have shown that whilst different types of parentheticals are
identified in functional and syntactic accounts, studies relating to the spoken
realization of parentheticals seem to concentrate on relative clauses. The reason is
that parentheticals are often defined by the way they are spoken rather than with
reference to the orthography, and non-restrictive relative clauses provide possibly
the clearest example of parentheticality. As discussed above, non-restrictive relative

clauses seem to be more easily identified by the way they are spoken than by the way they are written.[10] The analysis of spoken parentheticals in terms of relative clauses is therefore a restriction to a single type of parenthetical and generalizations based on the analysis of this particular type might not apply to the class of parentheticals as defined for this thesis.

Taking account of the extended, orthographic definition of parentheticals applied in this thesis, it is clear that written conversations contain a large variety of parenthetical constructions which not only differ in size and position, but also vary with respect to syntactic and functional classifications outlined above. This extended definition is motivated by text processing requirements in any speech output system which takes text as its input and where the printed text is the only available information.

The aim of this section is to examine parentheticals from written conversation with respect to their acoustic correlates with reference to prosodic independence and their possible influence on surrounding clauses.[11] With respect to the ideas and results discussed in the literature above, an adequate account of parenthetical constructions from written conversation requires answers or at least partial answers to the following three main questions:

A)      What are the acoustic correlates of parentheticals?

---

[10]Although non-restrictive relative clauses are marked by obligatory commas on either side.

[11]To put this into phonological terms we may say that this chapter is concerned with finding further evidence for a feature we may call [+parenthetical] which would have the described acoustic effects, thereby creating prosodically independent units from parentheticals. The main aim of the chapter could then be described as determining to which surface classes of parentheticals from written conversation this feature applies and to which ones it does not apply.

B)   Do these correlates point to parentheticals as independent acoustic
     units which seem to be acoustically unrelated to their surrounding
     text?

C)   Can any established acoustic or perceptual difference between
     different types of parenthetical be explained by their membership
     to a particular syntactic, functional or surface class, in particular
     their length and position in the utterance?

Two experiments are described below which have been designed to investigate the
above questions. Experiment I gathers perceptual evidence for the independent
prosodic processing hypothesis by investigating different readings of fourteen base
sentences containing parenthetical non-restrictive relative clauses. The aim of this
experiment is to provide perceptual evidence for the claims made in Kutik et al.
(1983) and to extend the analysis to more than one basic sentence. Experiment II
attempts the same for a larger variety of parentheticals.

## 3.3.1  Experiment I: Evidence for the prosodic independence of relative-clause parentheticals

The acoustic data provided by Kutik et al. for read sentences containing
parentheticals, together with various analyses from the literature examined here,
suggest that parentheticals may indeed be prosodically independent. However, the
difficulty with examining data from speech production only is that it can be
insufficient with respect to the amount of data available and also with respect to the
number or variety of measured acoustic parameters. Furthermore, measured
differences may not be perceptually salient, because they may be too small or
because they may not contribute to the category which they were considered to

contribute to in the first place. For this reason we ask whether a detailed perceptual investigation can provide further information on the prosodic independence of parentheticals. Before explaining how this can be carried out a short summary of the results presented above in Section 3.2.3 is necessary. Examined accounts of parentheticals argue for:

1.  the *resumption hypothesis,* which states that the fundamental frequency returns to a value which corresponds to the declination of the main utterance as it would be without the parenthetical which has its own independent F0 declination. However, their lowest values are not significantly different.[12]

2.  variations in length of the parenthetical seem to show no durational effects, such as pre-clausal lengthening on the main clause, and no such effects on parenthetical final and main clause final syllables which may have been expected as the interruption may be more severe for longer parentheticals.

Support for the prosodic independence of parentheticals at the perceptual level is investigated by the following hypothesis, which has the advantage of being independent of any particular prosodic model such as the topline rule.

If it is possible to remove a parenthetical construction from its surrounding sentence frame without causing obvious editing errors such as missing segments, unnatural pausing, unnatural breathing or clicks, *it should be impossible for listening subjects to determine whether a parenthetical has been removed from the utterance or whether the utterance was simply recorded naturally without the parenthetical.*

---

[12]See also Graph 4 below

In other words, if parentheticals are completely independent acoustic units, their removal should not adversely affect the prosodic naturalness of the surrounding speech. The purpose of this and the following experiment is to investigate this hypothesis. The most interesting results are expected from the comparison of the differences or similarities of parentheticals of different types and especially their behaviour in different positions in the utterance. However, it has to be acknowledged that the assumed correspondence between prosodic independence and failure to detect the removal of a parenthetical is not without difficulty. For example, an effect not mentioned so far which may influence the subjects' perceptual judgements is whether an increase in subglottal air pressure following a post-parenthetical breathing pause will result in increased prominence of the first few syllables following the parenthetical. If this is the case, subjects might perceive this as unnatural in the sentences where the parenthetical has been edited out and especially in cases where the pause has also been removed.

For this reason, the design of the experiment attempts to focus the subjects attention on the intonation of the utterance, as is explained in section 3.3.1.2 below.

3.3.1.1 Data collection and preparation

The testing of the above hypothesis in a perception experiment requires the comparison between utterances from which parentheticals have been removed, and recordings of the same utterances without parentheticals. The data design, collection and editing procedure is described in this section. For both Experiment I and II, all sentences containing parentheticals of various types are read examples

taken from written conversation. A total database of 120 utterances from 40 base sentences was recorded for one competent female speaker.[13]

For Experiment I, which is designed to test the behaviour of relative clause parentheticals, fourteen base sentences or short paragraphs were recorded, most of which contain one bracketed relative clause in either medial or final position. These base sentences were recorded in their original form as well as in two alternative conditions. In the second condition brackets were removed from the text and replaced by commas, yielding mostly standard non-restrictive relative clauses. In the case of an elided relative pronoun in the original, this was not inserted in this condition. In a further condition the sentence frame was presented to the subject and recorded entirely without the parenthetical. This was possible because in Nunberg's terms most of these lexical parentheticals provide alternate text, either by paraphrasing something previously mentioned, elaborating on such or providing additional pieces of information. Removing these bracketed strings therefore results in neither syntactically nor semantically ill-formed residuals. The following three sentences show the three conditions under which each base sentence was recorded. Sentence 21a is the original.

21a    On the other hand there is equally no sign of it in the one tape
       (rebroadcast a couple of years back ) that I possess.

---

[13]The variety of data exemplified in written conversation required the collection of substantial speech databases for the experimental investigation in this thesis. To provide an adequate analysis of this variety of data, it was necessary to capture the breadth of the data by collecting as many different types of examples as possible. This means that restrictions had to be made in order to keep the database to a reasonable size. It was therefore not possible in this chapter to investigate inter-speaker differences in the spoken production of parentheticals by collecting and analysing parentheticals from different speakers.

**21b** On the other hand there is equally no sign of it in the one tape, rebroadcast a couple of years back, that I possess.

**21c** On the other hand there is equally no sign of it in the one tape that I possess.

Twelve base sentences were recorded in these three conditions; the remaining two contained both medial and final parentheticals.[14] The total number of fifty utterance tokens recorded for this experiment were presented to the one female subject for recording in random order as determined by a random number generator. The database is listed in Appendix A.

The database was recorded in a professional anechoic recording booth, digitised at 16 bits and sampled at 10 kHz. Digital signal processing was performed using the OROS AU-22 board, which collected two channels. Channel 1 acquired the speech signal from a SHURE SM-10 microphone, Channel 2 acquired a laryngograph signal through a laryngograph strapped around the speaker's neck.

For a reliable analysis of the fundamental frequency patterns of parentheticals and their surrounding sentence frames it was decided to use a laryngograph signal which was post-processed after acquisition in order to produce readable F0 contours. First, a pitch detection algorithm determined the instances of glottal closure from the recorded sampled data, storing the results as absolute pitch mark times. Then the durations between the pitch mark times were calculated and stored as Hertz values. A non-linear smoothing algorithm was then applied to the data which, depending on its level of smoothing, resulted in a more or less smooth

---

[14]For one of the sentences a restrictive reading was also possible resulting in eight tokens for that sentence but these were not used further. Furthermore, two sentences contained full bracketed referring sentences in medial position rather than relative clause parentheticals.

F0 contour (Bagshaw et al. 1993). Determining the F0 contour from a laryngograph signal was preferred over pitch-tracking the actual speech signal due to its precision and reliability.

In order to prepare the utterances for comparison, the portion of speech containing the parenthetical construction was cut out of the speech waveform by means of a waveform editor, leaving only the non-parenthetical residual clauses. Figure 5 shows two speech waveforms with their associated F0 contours below and frequency values on the y-axis.[15] The top two contours show the original waveform with the parenthetical marked for cutting between two vertical bars. The bottom two contours show the same sentence recorded without the parenthetical.

A number of editing criteria were used. Any pauses not present in the reading without the parenthetical were cut out. However, for sentence medial parentheticals preceding a comma rather than parenthesis for instance, natural breathing was not edited out because it is not unnatural to occur in that environment and may have been triggered by the presence of the comma rather than by the parenthetical. In doubtful cases, pausing was modelled on the sentences recorded without the parenthetical which are the ones used for comparison in the perception test. This was legitimate in order to avoid perceptual judgements on intonation and duration being influenced by unnatural pausing, and because pausing as such was not part of the perceptual investigation. Any unnatural jumps or discontinuous intonation contours, together with any unnatural segment durations should have therefore been detected in their own right.

---

[15]The same technique was applied to Experiment I and Experiment II described below. This example is taken from an utterance used in Experiment II and therefore does not contain a relative clause parenthetical

Figure 5: Speech waveforms and F0 contours for the sentence:

"Along the way he tells of the obsession (including his) of the bodybuilding lifestyle" (top two contours).
"Along the way he tells of the obsession of the bodybuilding lifestyle" (bottom two contours).
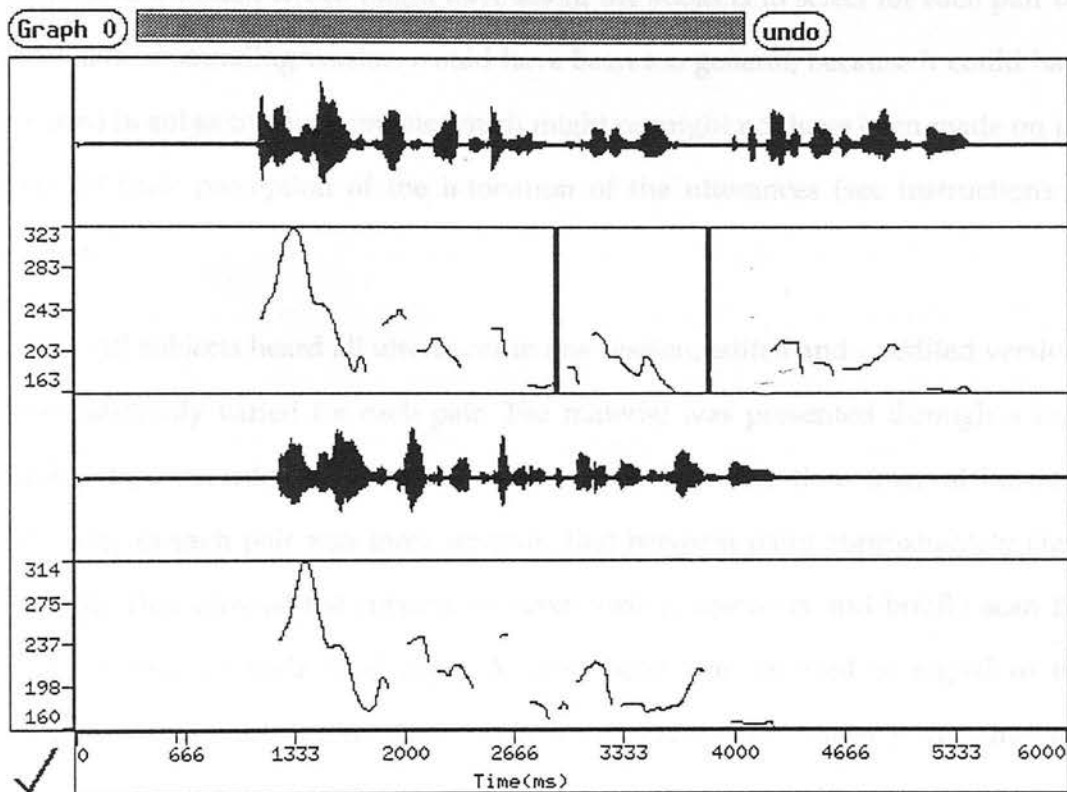


## 3.3.1.2 Methodology

Twelve subjects, all native speakers of varieties of British English, served as unpaid volunteers in a perception test. Two subjects were naive listeners, three were Linguistics graduates, four subjects were phonetician speech-technologists and three were experts in intonation with one intonational phonologist.

The subjects were presented with fourteen pairs of read sentences. In each pair, one of the sentences was edited as discussed above, the other was the

equivalent recording made without the parenthetical. After informing the subjects that for each pair one read token was edited and that this may affect the natural flow or continuation of the *intonation* of the utterance, they were asked to indicate for each pair which sentence version they had perceived as being the edited one. An alternative question which might have asked the subjects to select for each pair the most *natural* sounding version would have been too general, because it could have resulted in subjective judgements which might or might not have been made on the basis of their perception of the intonation of the utterances (see instructions in Appendix C).

All subjects heard all utterances in one session, edited and unedited versions were randomly varied for each pair. The material was presented through a high quality tape recorder attached to an active speaker. The time interval between sentences in each pair was three seconds, that between pairs approximately eight seconds. This allowed the subjects to make their judgements and briefly scan the next sentence on their hard copy. A short beep was inserted to signal to the participants the start of a new pair. This would prevent the subjects from missing the start of a new pair while still scanning the sentence on the hard copy. The provision of a hard copy was thought to reduce the cognitive load, because the meaning of the sentence did not have to be decoded from the speech but was known in advance, leaving the subjects to concentrate on the intonation of the utterance to which they had been specifically pointed in their instructions.

It was decided to focus the subjects' attention on the flow of the intonation specifically since this was thought to provide them with the best chance of finding unnatural contours resulting from the edits. A pairwise comparison with mostly non-naive subjects who were pointed to the exact nature of the task, therefore, provided the toughest possible test for the above hypothesis.

The twelve subjects made one judgement for each of the fourteen sentence pairs which yields a total of 168 judgements. A forced choice of which sentence was the edited one resulted in either correct or incorrect answers, which means that if subjects were merely guessing which sentence in each pair was edited they would be expected, according to chance, to detect the edits in 50% of all cases. A result in the region of 50% would support the above hypothesis that cutting out this type of parenthetical can not be detected in comparison to the same sentence recorded without the parenthetical.[16] This in turn would support the independence hypothesis as discussed in the introduction to this chapter. A significant positive detection rate however, would indicate that the intonation of the remaining clauses had been somehow disrupted, presumably because an integral part of a complete, whole contour has been cut out, thus undermining the independence hypothesis.

### 3.3.1.3 Results and discussion

Results were pooled across subjects and of the 168 judgements, 94 edits were correctly identified, which corresponds to 55.9% of the total number of judgements. This is not significant by binomial analysis.

A closer examination of the results with respect to individual subjects and position of the parenthetical provides some interesting further results. Of the fourteen sentences, eight had parentheticals only in medial position, four only in sentence final position and two had one in medial and final position. Out of a total of 48 judgements for the sentence final condition, edits were detected in 30 cases

---

[16]This seems at first glance to support the null hypothesis and it would have been more correct to show *first* that cutting out certain parentheticals does disrupt the continuous flow of the intonation. This is shown below in Experiment II, but for reasons of developing the argument in this chapter it was decided to present the evidence for prosodic independence of relative clause parentheticals first.

and not detected in 18 cases. Binomial analysis shows that this is a significant result (p < .05). This positional effect is examined further in Section 3.4.1.1 below which investigates F0 values at the boundary between the main and the parenthetical clause.

From a total of 96 judgements for pairs with parentheticals in medial position, 52 edits were correctly identified whereas 44 were not, a result which is not significant. The two utterances with multiple parentheticals were not further analysed because subjects were not asked to indicate where they had detected an edit in the sentence, which made it impossible to include them in the more detailed analysis. This was done for some sentences in Experiment II described below.

Significant differences between the subject groups were not identified and there was no bias towards selection of either the first or the second sentence in the pairs.

Examining the results for individual subjects shows that eleven of the twelve subjects could not detect the fact that a parenthetical had been cut out to a significant degree. One of the subjects however identified twelve out of the fourteen edits correctly, which is significant (p < .01). Not surprisingly, this was the one intonational phonologist on the panel. He, however, pointed out that a lot of his judgements were based on durational cues. Kutik et al. claim that the duration of the main clause remains essentially constant even if interrupted by parentheticals of differing lengths. They also suggest that parentheticals do not trigger phrase final lengthening in the first part of the main clause. Furthermore, they show that there is no relationship between the length of the parenthetical and segmental durations in either the main clause or the parenthetical.

The observation on durational clues made above, however, possibly suggests the presence of pre-parenthetical lengthening, an effect that was identified for non-restrictive relative clauses by Garro and Parker (1982). This is examined further in Section 3.4.

### 3.3.1.4 Conclusion

This experiment provides conclusive evidence for the *intonational* independence of relative clause parentheticals with the possible qualification of some *durational* processes which might operate at clause boundaries. The editing out of these parentheticals for a perceptual comparison with sentences recorded completely without the parentheticals failed to be detected to a significant degree by a panel of listeners familiar with speech research. The only exception were the results from one intonational phonologist who did detect the edits to a significant degree.

The overall result of this experiment provides further evidence for the intonational independence investigated also in Kutik et al. who posit separate declination lines for parentheticals and for their surrounding main clauses.

A further experiment was conducted in order to investigate a wider range of parentheticals found in written conversation, and section 3.5 provides the results of a pilot experiment which attempted to test whether the amount of informational content inside a parenthetical influences the way a reader treats these parentheticals prosodically. Section 3.4 attempts to find explanations for the perceptual significance of the duration parameter for one of the subjects.

### 3.3.2 Experiment II: Extending the analysis

Experiment I was mostly concerned with providing supporting evidence for the prosodic independence of relative clause parentheticals. In this experiment, the analysis is extended to a larger variety of parentheticals found in written conversation by adopting the same methodology as above.

#### 3.3.2.1 Data preparation and methodology

A further database consisting of twenty six base sentences was recorded by the same speaker, providing a version with parentheticals and a version without parentheticals. The conditions were the same as above. The database contains different types of parentheticals such as parentheticals containing conjunctions, participial phrases, isolated words and whole sentences, and is listed in Appendix B.

Twenty five sentence pairs were presented to a panel of six subjects. The instructions were the same as for Experiment I, except that for certain marked sentences subjects were not only asked to identify the edited version but also to indicate on their hard copies where in the utterance they had thought the edit occurred. The purpose of this experiment was to investigate a wider range of parentheticals in order to establish whether they can all be treated the same, as independent acoustic units, or whether finer distinctions can be made with regard to certain classes of parentheticals.

A further aim was to determine that subjects were not making judgements in a more global impressionistic way, but that if they detected the edits at all, they were actually detecting the edits at the points where they occurred. Therefore, subjects were asked to indicate at which point they thought the edit had been made (see Appendix D for instructions to subjects). In other words, this experiment

investigates parentheticals or classes of parentheticals for which the independence hypothesis may not hold and the reasons why this should be so.

3.3.2.2 Results and discussion

In this experiment the pooled result across subjects points to a significant result in identifying the edited versions of the presented sentences (.0005), in contrast to the results from Experiment I above. The edits in five of the twenty five sentences were identified by all subjects - one of these was an utterance containing a parenthetical question. This shows that the removal of parentheticals does have the potential of disrupting the intonation pattern of the carrier sentence. The only significant result for the detection of edits on a subject by subject basis was again scored by the same intonational phonologist as in Experiment I, although all other subjects also detected more than 50% of the edits correctly. As a control, another intonational phonologist took part in this experiment and scored exactly the same as the first, which may also suggest that the results from that particular subject in the first Experiment were not accidental.

Looking at individual utterances, for fourteen of the twenty five utterance pairs edits were identified on the correct utterance by at least four out of the six subjects, five utterance pairs resulted in a split decision, and for six utterance pairs the edit was thought to have occurred on the utterance where no edit had been made. Due to the use of parentheticals of various surface types, differing in length, differing in their position in the utterance and constituted of various phrase and clause types it is difficult to say whether the removal of any of those types was more readily detected than the another. There are some obvious cases such as the parenthetical question, or cases where there are clearly identifiable continuation rises preceding the parenthetical phrases, disrupting the main clause contour (utterances 004; 009; 011; 020; 021 - Appendix B). Further tendencies can be

102

observed in that some detected edits involved participial phrases (014; 021; 024) and conjunctions (001; 009; 011), whereas some of the utterances where edits were less likely to be detected included whole clauses or sentences (008; 013; 018) and also included relative clauses (012; 013; 025) supporting the results in Experiment I .

Of the specially marked sentences which contained more than one parenthetical, pooled results show 24 correct and only 6 incorrect judgements. Furthermore, of the 24 correct ones, the location of one of the edits was also identified in 22 cases. For only two correct judgements the location was not identified and this was for the same sentence. With one exception the same edits were pointed to by all subjects who identified the right version in the first place. This seems to suggest that in the sentences containing multiple parentheticals, the removal of one parenthetical was noticed more than the removal of the other. However, this has to be treated with caution because subjects were not aware of the possibility of sentences having more than one parenthetical and were not asked to mark more than one, although they might have detected more than one edit at the time. Nevertheless, the high degree of success in marking the exact points where edits were perceived shows that choices were not based on some global impressionistic measure, such as for example a naturalness judgement or individual preference for any other reason.

This seems to suggest that not all parentheticals as defined in the extended definition applied to written conversation are treated as independent intonational units, although even for Experiment II some clearly are. On the other hand, some parentheticals or classes of such may be acoustically not independent at all and hence not set aside from the surrounding material.[17] As a result, edits on these

---

[17]At least for this particular reader.

utterances are perceptually detected. The purpose of section 3.4 below is to provide evidence from acoustic measurements carried out on the utterances used in the two perceptual experiments in order to determine if correlations can be established between subjects behaviour in the perception experiments and the acoustic realization of the utterances used. This is carried out with respect to the following three questions which were identified above as being central to a thorough account of parentheticals from written conversation.

A)    What are the acoustic correlates of parentheticals?

B)    Are there any surface classes dependent on parenthetical length or position?

C)    What is their influence on their surrounding clauses?

## 3.4  Acoustic correlates of parentheticals

This section presents measurements of the acoustic correlates of parentheticals and their possible influence on the prosody of the surrounding main clause. Parentheticals are examined with respect to their position and length, juncture preceding and following the parenthetical, pitch range, and their effect on intonation and duration of the main clause.

### 3.4.1  Positional effects

This sub-section investigates whether parentheticals can be categorized with respect to their surface position in an utterance and examines the reasons for the differences in the detectability of edits between utterance medial and utterance final parentheticals.

3.4.1.1 Utterance-final parentheticals

For Experiment I, a positional effect was identified whereby the removal of sentence final parentheticals was detected to a significant degree by the panel of listeners. This may be due to the reader not pre-planning the utterance final F0 drop at the end of the main clause preceding the parenthetical. Graph 1 shows the differences in final F0 values for the sentences which contained final parentheticals and the same sentences read without the parentheticals.

**Graph 1:**
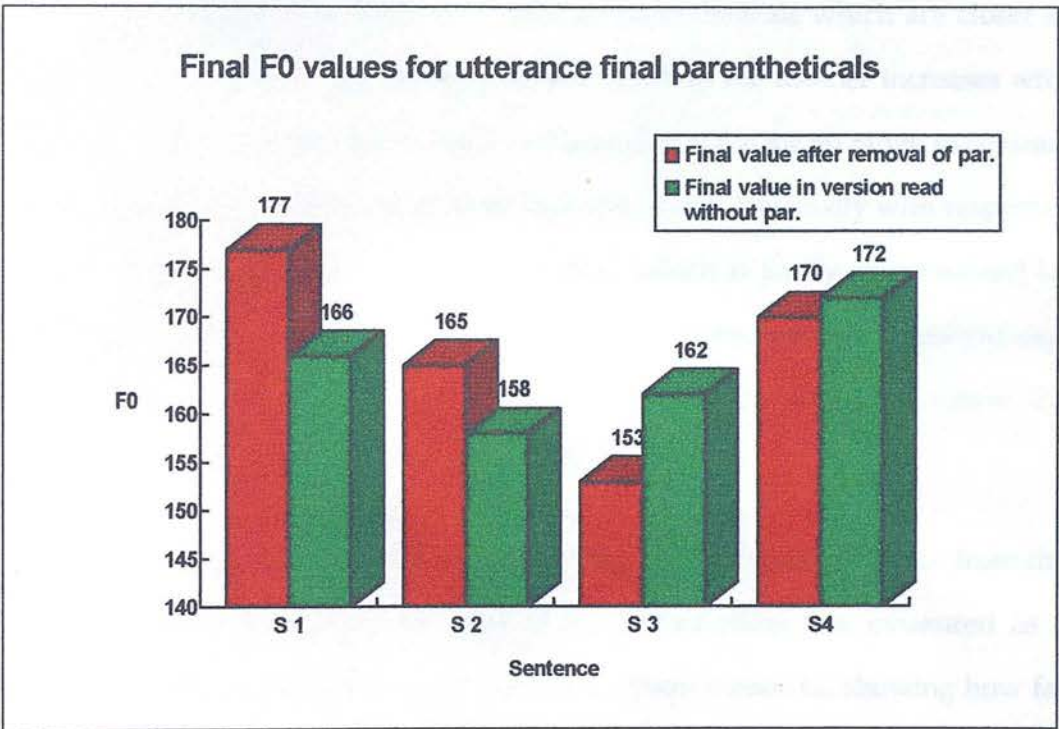


For utterances one and two (Experiment I), the edit detection rate out of a possible 12 edits was 10, whereas for sentences three and four only five from a possible 12 edits were detected. The above graph shows clearly that the final F0 values for utterances one and two are considerably higher for the readings containing the parentheticals than for the readings without the parentheticals. Sentence three has a very low F0 value for the utterance containing the parenthetical and the values for

sentence four are very similar. A rank correlation between the difference in final F0 value in the two conditions and the detection score is significant at .025 (Correlation Coefficient=0.90). A possible effect which could have been caused by the length of the utterance final parentheticals on their detectability was not identified.

3.4.1.2 Utterance-medial parentheticals

Although the overall detection rate for edits on sentence medial parentheticals was insignificant, some were detected more readily than others. This section examines the possibility that edits on parentheticals which are closer to the beginning of the utterance are noticed more easily than edits on parentheticals which are closer to the end. This is plausible, because the cognitive load on the listener increases with increasing utterance length, and because of the tendency for the F0 range to contract as a function of time (Pierrehumbert 1980; Vaissière 1983). Especially with respect to the latter point, the extraction of a parenthetical, which is partly characterized by reduced pitch range, from a portion of speech which is inherently reduced in range may be less conspicuous than the extraction of a parenthetical from nearer the beginning of the utterance where the F0 range is wider.

In order to find evidence for this hypothesis the temporal distance from the beginning of the utterance to the start of the parenthetical was measured as a percentage of the total duration of the utterance. These measures, showing how far through the utterance the parenthetical begins, were correlated with the detection rates for each of the sentence pairs from Experiment II (Correlation Coefficient = 0.12).

The calculated correlation coefficient shows that a positional effect in terms of the distance of the start of the parenthetical from the beginning of the utterance could not be identified for sentence medial parentheticals, suggesting that the only

distinction in terms of position for the examined data can be drawn between sentence medial and sentence final parentheticals.

## 3.4.2  Duration measures

This section examines the role of duration and its influence on the prosodic status of parentheticals. It is examined whether the duration of the parenthetical clauses themselves determines prosodic independence. Duration is measured at clause boundaries in order to establish if pre-parenthetical lengthening does occur and if its presence correlates, as suggested by one of the subjects, with the perception of edits. First, however, measurements of the duration of pre- and post parenthetical pauses are presented since this is one of the most likely contributors to the prosodic independence of parentheticals.

### 3.4.2.1 Pausing

Pre- and post-parenthetical pauses were measured for the relative-clause parentheticals in Experiment I and the extended range of parentheticals used in Experiment II. Graph 2 shows that pre-parenthetical pauses tend to be considerably shorter than post-parenthetical pauses and points to a difference between pre-parenthetical pauses for relative-clauses in comparison to other types of parenthetical. The mean pause length for the pre-parenthetical pause (pause1) at the boundary of relative clause parentheticals is 201ms. whereas it is merely 129ms. for the extended set. The extended set also contains a considerable number of pre-parenthetical pauses shorter than 100ms. which may not be perceived as a juncture by subjects. This could be a significant factor in explaining the differences in the outcome of Experiments I and II, because there were fewer instances of short pre-parenthetical pauses in Experiment I and a mean of around 200 ms. A strict

correspondence between the length of the first pause and the detection of edits could not be established although there seems to be an overall tendency for edits on parentheticals with shorter first pauses to be detected more readily. In addition, the editing process was also more complex in examples with no or very short pre-parenthetical pauses.

**Graph 2: (Arrows point to mean scores)**

## Pre- and post-parenthetical pauses in relative clause parentheticals and other types of parenthetical
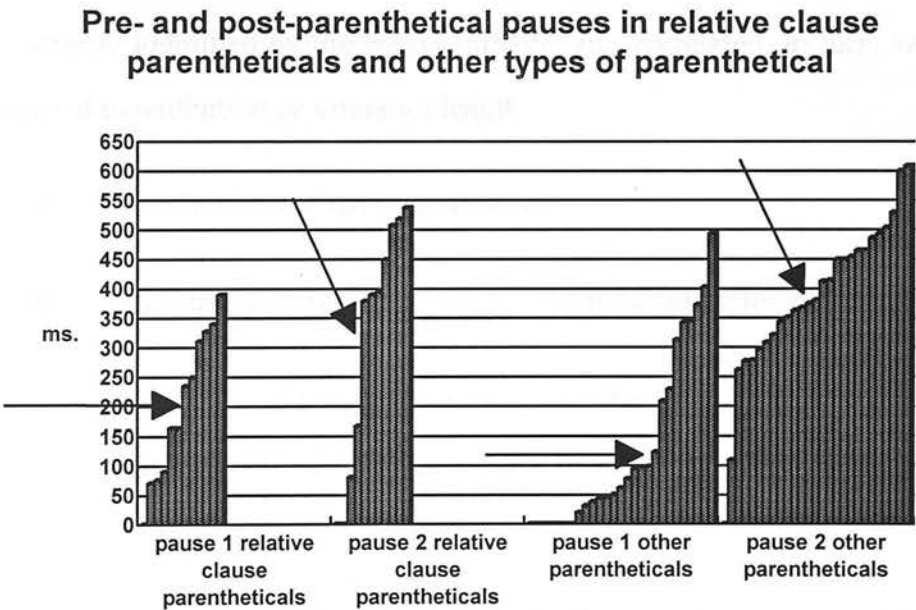


The mean duration of post-parenthetical pauses is 311ms. for pauses preceding relative-clause parentheticals and 385ms. for pauses in the extended set. This difference however is not likely to affect the perception of edits since the pauses in both conditions are long.

### 3.4.2.2 Parenthetical length

This section examines the influence of the length of the parenthetical on the detectability of an edit. After all, removing a larger chunk, a longer parenthetical, may cause more disruption than the removal of a shorter speech segment. The difficulty for a speaker of prosodically connecting the second half of an interrupted utterance to the first half may increase with the length of the parenthetical, increasing the length or severity of the interruption.

Sentences from Experiment I were ranked with respect to the number of edits correctly identified by the twelve subjects and correlated, in turn, with three measures of parenthetical or utterance length.

a)    The duration of the parenthetical

b)    The absolute duration of the sentence including the parenthetical

c)    The length of the parenthetical as a percentage of the length of the sentence. This percentage ranged from 12% to 48% of the total length of the sentence.

No correlation was identified between the edit detectability and any of the length conditions. This would suggest that even for sentences where nearly half of the whole sentence was edited out, the remaining - and after the edit adjacent - intonation contours of the two main clause halves are not perceived as unnatural.

### 3.4.2.3 Pre-parenthetical lengthening

One subject stated after taking part in Experiment I that some of his judgements had been based on durational rather than intonational clues in the speech signal. One

phenomenon which may affect the duration at the particular part of the speech signal from which the parentheticals had been removed is pre-clausal lengthening.

This possibility is examined here, although Kutik et al. (1983) conclude that "parenthetical clauses do not constitute an environment for preclausal lengthening" (p.1736). Their observation was based on measurements of the vowels in the words "church", "Clark" and "talk" (in sentences 17 above), together with data on total main clause length and total parenthetical length. They could therefore not support earlier findings by Cooper et al. (1980) who found a small effect of the length of upcoming material on pre-boundary lengthening, and the results are in direct contrast to those in Garro and Parker (1982), who report pre-clausal lengthening preceding non-restrictive relative clauses.

Considering that the vowel in the word "church" is the only possible point of analysis in Kutik et al. and that it is likely to take the nucleus of the first part of the main clause - in which case it would already be marked by increased duration - their claim seems too strong. Furthermore, their data did not include a condition with no parenthetical.

Data from Experiment II was taken in order to investigate whether pre-parenthetical lengthening could be identified for the data, and if so, whether a connection between the amount of pre-parenthetical lengthening and the detectability of an edit in the perception experiment could be established.

Pre-parenthetical lengthening was measured in terms of the difference mainly in the duration of words between the condition where they preceded the parenthetical clauses and the condition where the clauses were read without the parentheticals. Where this was not possible due to words with very short durations preceding the parenthetical, small phrases or phrase fragments were used. Due to

the differences in the utterances, a common unit for measuring pre-parenthetical lengthening could not be used. Pre-parenthetical lengthening is expressed as the percentage by which the pre-parenthetical units are lengthened with reference to the duration of the same units in the utterances without parentheticals.

Following the quantification of pre-parenthetical lengthening by comparing each utterance pair, the percentage of pre-parenthetical lengthening was correlated with the number of subjects who detected an edit in the utterance from which the parenthetical had been removed. Each of the six subjects made one judgement for every utterance with the total number of detected edits (a maximum of six) determining the detection rate for each utterance.

The correlation between the amount of pre-parenthetical lengthening and the detection rate is shown in Graph 3. The correlation coefficient (Spearman's Rank Correlation) is 0.524, which is significant at .005.

**Graph 3:**



Correlation between pre-parenthetical lengthening and perceptual edit detection

111
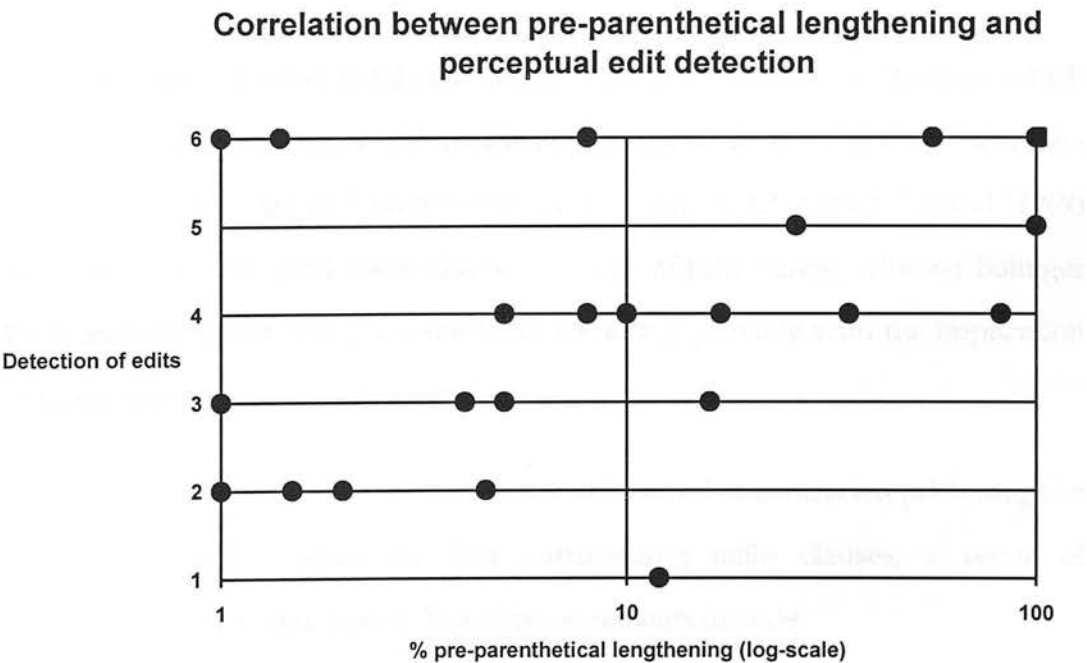
It is also possible to note a tendency consisting of the relationship between the amount of pre-parenthetical lengthening and the length of the parenthetical, in that pre-parenthetical lengthening is longer for shorter parentheticals (Rank Correlation Coefficient=-0.38), and furthermore, that the edit detection rate is higher for shorter parentheticals (Rank Correlation Coefficient=-0.25). This is an interesting result which may benefit from further investigation on larger samples of data. The significance of this tendency is that shorter parentheticals tend to be lexical in Nunberg's terms providing alternative readings or simple adjuncts. They therefore do not tend to consist of clauses or even phrases and their potential for prosodic independence may be reduced. This may result in short parentheticals becoming part of the prosodic unit of the main clause, making their deletion more noticeable to the listener.

### 3.4.3 Accentuation and contour effects

3.4.3.1 Pitch range in parentheticals

This section aims to provide data relating to statements made in the literature which include in their description of the acoustic realization of parenthetical constructions reference to contracted or lowered pitch range (section 3.2 above). Crystal (1969) and Monaghan (1991) talk about the "narrowing" of pitch range, whereas Bolinger (1964) and Silverman (1987) use the term "lowering" possibly with the implication of contraction.

In order to investigate the possibility of lowered or contracted pitch range in parentheticals with respect to their surrounding main clauses, a series of measurements were undertaken. These measurements include:

a. F0 at the first peak (P1) of the utterance ............................ (P1)

b. F0 at the maximum peak inside the parenthetical ............... (Max. Par)

c. F0 at the lowest trough inside the parenthetical ................ (Min. Par)

d. F0 at the maximum peak in the main-clause other than P1 ... (Max. MC)

e. F0 at the lowest trough in the main clause ........................ (Min. MC)

f. The lowest identifiable F0 trace ....................................... (Floor)


Table 4 shows the exact F0 values (Hz.) for the measured categories of pitch target for the test utterances:

Table 4: Frequency values (Hz.) for Max. PAR, Min. PAR, Max MC, Min MC, P1 and Floor for the test utterances

| Utterance | 001 | 002 | 003 | 004 | 005 | 006 | 007 (1st par) | 007 (2nd par) | 008 | 009 | 010 | 011 | 012 | 013 (1st par) | 013 (2nd par) | 014 | 015 | 016 | 017 | 018 | 019 | 020 | 021 | 022 | 024 | 025 | 026 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Max. Par | 242 | 199 | 223 | 221 | 204 | 226 | 199 | 225 | 235 | 257 | 236 | 225 | 216 | 256 | 213 | 208 | 221 | 186 | 230 | 228 | 212 | 240 | 216 | 205 | 214 | 226 | 215 |
| Min. Par | 165 | 155 | 166 | 163 | 160 | 154 | 141 | 151 | 167 | 162 | 163 | 170 | 164 | 165 | 159 | 173 | 151 | 159 | 159 | 162 | 170 | 157 | 175 | 166 | 154 | 160 | 165 |
| P1 | 318 | 314 | 323 | 323 | 298 | 313 | 313 | 313 | 314 | 317 | 327 | 289 | 328 | 323 | 323 | 286 | 330 | 271 | 316 | 324 | 326 | 324 | 271 | 271 | 271 | 327 | 324 |
| Max. MC | 261 | 266 | 241 | 242 | 270 | 256 | 263 | 263 | 275 | 301 | 263 | 236 | 265 | 265 | 265 | 269 | 269 | 167 | 269 | 290 | 253 | 257 | 252 | 236 | 261 | 263 | 277 |
| Min MC | 164 | 163 | 166 | 163 | 156 | 142 | 165 | 165 | 167 | 167 | 167 | 166 | 166 | 169 | 169 | 166 | 166 | 164 | 162 | 162 | 159 | 161 | 162 | 160 | 165 | 167 | 157 |
| Floor | 139 | 123 | 151 | 163 | 137 | 142 | 128 | 128 | 143 | 142 | 150 | 155 | 136 | 125 | 125 | 156 | 139 | 135 | 147 | 162 | 159 | 146 | 150 | 154 | 140 | 150 | 157 |

On the basis of these measurements, pitch range for main clauses and parentheticals was calculated as follows:

The lowest observed trough values were subtracted from the highest observed peak values separately for main clauses and parentheticals, excluding the value of P1 in the calculation of main clause pitch range. Pitch range was therefore defined as the absolute difference between the highest peaks (except P1) and the lowest troughs at the relevant points in the intonation contour.

By comparing the means of the six categories of pitch target in Graph 4, it can be seen that the maximum F0 values inside parentheticals are substantially lower than the maximum values in the surrounding main clause. Marginally lower

values can also be observed for the minimum values inside parentheticals. However, the differences between minimum values for parentheticals and the minimum values for main clauses is much smaller than the difference in their maximum values.

**Graph 4:**

**Differences in the Maximum and Minimum F0 Values for Parentheticals and Main Clauses**
**(Means and Standard Deviations (+;-1); n=27)**

F0 (Hz.)

340
320
300
280
260
240
220
200
180
160
140
120
100

P1    Floor    Max MC    Max PAR    Min MC    Min PAR
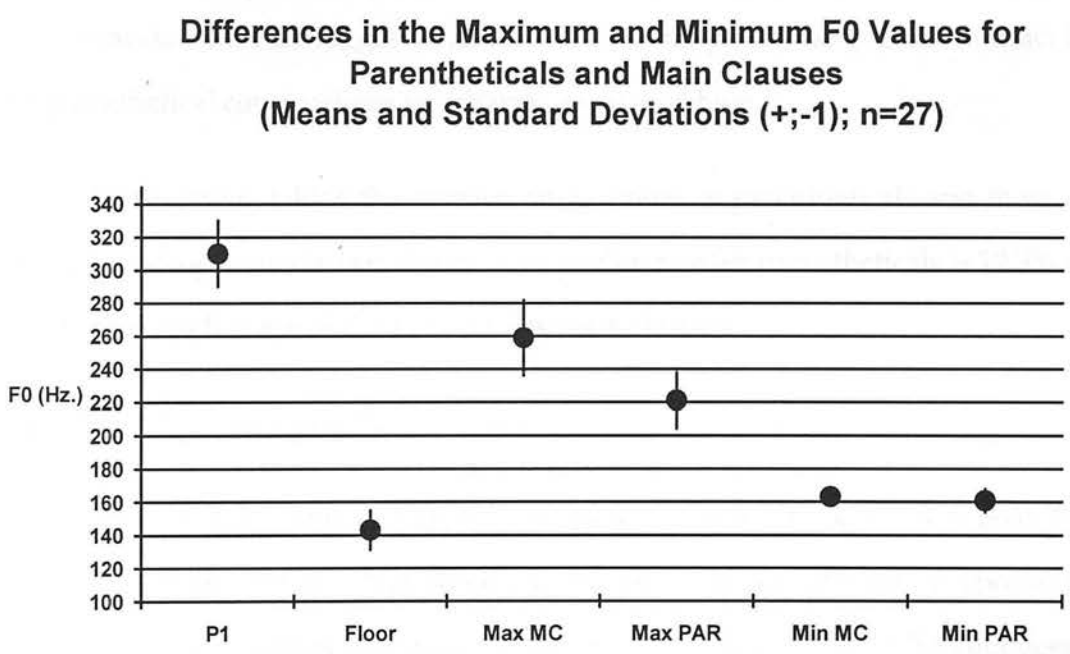
The following addresses the question of whether the observed data represents a *contraction* of pitch range rather than a simple *lowering* of both the maximum and the minimum values. In other words, is the available pitch range inside the parenthetical narrower in comparison to that of the main clause, or is it simply shifted downwards. In order to answer this question, each range value for each

114

parenthetical and main clause was divided by the mean of the range of that particular clause in order to normalize the data with respect to the absolute values.

$$\text{Range Value} = \frac{\text{Absolute Range}}{(\text{Max. CLAUSE} + \text{Min. CLAUSE}) / 2}$$

Range values were calculated for each of the main clauses and each of the parentheticals in the 27 test utterances. The significance of the differences was analyzed in a a t-test (t=7.67, N=27). This points to a highly significant result (.0005), in that the range values for parentheticals are significantly smaller than those for their main clauses. This suggests that fundamental frequency range does contract in the parenthetical constructions which were measured here.

Furthermore, taking the average range values of parentheticals and those of the surrounding main clauses, the average pitch range for parentheticals is 72.3% of the average pitch range of the surrounding main clauses.

3.4.3.2 Parentheticals and P1

Kutik et al. (1983) found no effect of increased sentence length on the first peak P1, although previous studies had shown an influence of the amount of upcoming material on the height of that peak. (Cooper & Sorensen 1981:36; O'Shaughnessy 1976). Kutik et al. took this as further evidence for the independent prosodic programming of parentheticals in that parentheticals do not add to the total length of the utterance in the speech production process of the speaker. In our case, it is important to measure this variable in order to investigate whether it constitutes a possible contributing factor in the detection of edits in certain utterances. This was done for Experiment II where a significant number of edits were detected.

Measurements were carried out on utterances from Experiment II which had no marked difference in accentuation which could have affected P1, and which were not subject to pitch tracking errors in the region of P1. These measurements confirm the results in Cooper et al. and Kutik et al. There was only an insignificant correlation between the difference of P1[18], in the sentences with and without the parentheticals and the length of the parentheticals as such. This suggests that increased sentence length resulting from the presence of a parenthetical has no influence on the height of P1.

A significant correlation between the absolute value of P1 and the total length of the sentences was also not found, where sentences containing parentheticals and the sentences not containing any parentheticals were examined independently and together as one set.

### 3.4.3.3  Pre- and post-parenthetical accents

The presence of a pitch accent on the word immediately preceding a parenthetical together with an accent on the word following the parenthetical could potentially violate rhythmic constraints (Monaghan 1991) in conditions where the parenthetical is removed. This violation could play an important role in the detection of edits, particularly in Experiment II.

The data shows however, that these "clashes" rarely occurred. Of 26 examined sentences, 18 had an accent on the word preceding the parenthetical, but only three had an accent on the word following the parenthetical. Only two utterances contained accents on both, and edits in these utterances were detected positively. A Chi-squared test on a three by two matrix with the variables (+;-) pre-

---

[18] as calculated by $(d = P1_{(par)} - P1_{(non-par)})$

parenthetical accent and a three way detectability rating (yes; ?; no) which was possible for the six subjects, was insignificant. This provides some indication that an accent on the word immediately preceding the parenthetical is unlikely to influence the detectability of an edit.

The sentences used in the experiments carried out by Kutik et al. however, do constitute an environment for a stress clash if read out as indicated. It is therefore unfortunate that an example of the sentence which had no intervening parenthetical was omitted from their investigation.

A further factor which contributed to the detection of edits in Experiment II for some sentences, was the presence of continuation rises on the last portion of the pre-parenthetical main clause, thus indicating the interruption by the parenthetical. This was particularly prominent in utterances 009, 011, 020 and 021 (see Appendix B) which contained parentheticals which were close to the beginning of the utterances,[19] as the following example shows:

> 014    And (putting on my flame-proof suit) there's no solid reason for
> treating Norwegian, Danish, and Swedish as three languages ...

On all these utterances, edits were detected to a highly significant degree, all subjects identified edits on utterances 009 and 011 and five of the six subjects detected the edits on utterances 020 and 021.

### 3.4.3.4 Pre-parenthetical breakoff and post-parenthetical resumption.

A further likely event in the speech signal affecting the intonation of an utterance after the removal of a parenthetical would be an unnatural discontinuity in the

---

[19]Although the distance from the beginning was not in itself identified as a significant factor

intonation contour at the point at which the main clause contours merge. Measurements were taken precisely at the point where pre- and post-parenthetical contours merge after the editing out of parentheticals. They concentrated on the distance of the last identifiable F0 trace before the parenthetical and the first identifiable trace after. An unnaturally large gap or jump in either direction would certainly disrupt the flow of the intonation and result in the identification of edits.

The calculation was based on the difference between the natural gap measured from sentences recorded without the parentheticals and the gap resulting from the edits.[20] The resultant gap values were correlated with the detectability of the edits for each utterance resulting in an insignificant correlation (RHO=0.26). Therefore, we can conclude that the difference between the F0 value at the end of the first part of the main clause and the F0 value at the beginning of the second part of the main clause did not contribute to the detection of edits.

## 3.5 A pilot analysis of parentheticals and information content

The above has shown that certain acoustic events such as pre-parenthetical lengthening contribute to the detection of edits. This section addresses the question whether there are degrees of independence or integratedness, pointing to *degrees of parentheticality,* and whether there is a threshold for the perception of edits which may correlate with a proposed classification based on the pragmatic notion of *information content.*

---

[20]as calculated by (gap= gapsize(par) - gapsize(non-par))

The following examines whether, in addition to the acoustic evidence presented above, a pragmatic classification can be used which may explain the results from the perceptual experiments. For example, to use Nunberg's terms, alternate text parentheticals may be more independent from surrounding clauses than parentheticals which elaborate or explain, because they may contribute less to the meaning or interpretation of a proposition. Therefore, one way of describing integratedness may be in terms of the contribution to the content of a message or proposition which the parenthetical construction makes. As discussed, such constructions contain secondary information, but there may be finer subdivisions. As Nunberg points out further, some information may be important or helpful to one reader and superfluous to another. It follows that, dependent on the degree of importance a reader attributes to a parenthetical, the acoustic presentation of that material may vary. A parenthetical with a higher degree of information content may not be set aside or separated prosodically as much as one with less information value. Practically this means that accents more in line with the speaker's normal register may be placed on certain content words inside the parenthetical preventing the contraction of F0 range as presented in section 3.4.3.1.

As a first step a small pilot study was conducted in order to establish that some sort of grading of the importance of parentheticals in relation to the proposition of the whole message was at all possible. Two subjects were asked to rate their subjective opinion of the importance of the information content of parentheticals contained in 26 sentences as:

0 =     unimportant, unnecessary

1 =     might be useful to some readers, should remain in brackets

2 =     important for interpretation of the whole sentence; could or should have been integrated into the text itself.

This test was designed to investigate whether subjects would be able to grade parentheticals at all. The overlap in grading between the two subjects is of secondary interest only, because they may have differing background knowledge and opinions which influenced their interpretation, and moreover, they were not the writers of the sentences. However, the grading of 13 of the 26 sentences were identical, eleven out of twenty six judgements differed by one point on the scale and only two differed by two points.

The subjects reported finding the task straightforward, one suggested they should have been presented with whole paragraphs or texts which would have provided more context. Furthermore, a finer grading scale was also suggested by both subjects, one subject suggesting a five point-, the other a four point scale. This suggests that a grading of parentheticals is possible.

Following the testing of a grading scale, the reader from which the parentheticals had been recorded was asked to grade the parentheticals on a four point scale for their informational content, of course, after the sentences had been recorded. A correlation between the detectability of the edits - which may have suggested degrees of parentheticality -, and the subjective importance classification by the reader could not be established (Rank Correlation Coefficient=0.04). We can conclude that although subjects were able to grade parentheticals with respect to their information content, there was no correlation between this grading and the detection of edits.

## 3.6 Conclusions

The aim of this chapter was to describe the prosodic correlates of a wide variety of parentheticals from written conversation. This was attempted by investigating the

claim that parentheticals are prosodically independent units, firstly by presenting analogous arguments from the literature concerning the independence of parentheticals from syntactic, pragmatic and acoustic analyses.

Prosodic independence was investigated not only because results could be compared to earlier accounts, but because it would be important with a view to being able to convert parentheticals from their written form to an adequate representation in the spoken form by an automatic system. The significance of prosodic independence would be that parentheticals could be processed independently and in parallel to the surrounding text by their removal and later reinsertion, which would have the considerable added benefit of simplifying the analysis of the remaining text.

Secondly, prosodic independence was investigated through perceptual experiments, followed by the examination of the acoustic correlates of parentheticals with a view to explaining the results from the perception trials.

Perceptual evidence was gathered and showed the prosodic independence of non-restrictive relative clause parentheticals, but this finding could not be repeated for an extended set of parentheticals from written conversation, including single words, whole sentences and other types of clauses. The acoustic analysis was carried out in order to determine the acoustic correlates of parentheticals and to find explanations for the differences between the two experiments.

The acoustic analysis pointed to a number of factors by which it can be argued that parentheticals are prosodically independent but also showed a number of ways in which they are not independent.

Factors by which parentheticals are independent were identified as the following and quantified in the relevant sections above:

1.      Juncture at parenthetical clause boundaries

2.      Contraction of pitch range inside the parenthetical constructions

3.      No effect of parenthetical length on the height of P1

4.      Relative continuity at the intonation contour breaking points

5.      No overall effect of parenthetical length

Ways in which parentheticals are not independent were identified as the following:

1.      Pre-parenthetical lengthening

2.      No F0 drop to floor preceding some sentence final parentheticals

3.      Continuation rises at some pre-parenthetical boundaries

4.      Possible rhythmic stress clashes in the main clause[21]

The differences between the outcome of the two perception experiments can be explained by arguing that prosodic independence of parentheticals is a result of the contribution of a significant number of the above listed factors. Relative clauses were particularly strongly separated by juncture, especially pre-parenthetical pauses were longer than for other types of parenthetical. In addition, they provide an example of syntactic discontinuity, and due to their prominent use in the language as a whole their function and acoustic representation may be well established.

The other types of bracketed strings contained more of the characteristics which seemingly counteract the independence of parentheticals, in particular pre-clausal lengthening and the presence of continuation rises on the final tones of the main clauses, signalling the upcoming interruption. In addition, the contributing

---

[21]Although these stress clashes hardly occurred in the data which was analyzed for this thesis they are nevertheless possible.

factors may have been less well represented, for example the juncture at the pre-parenthetical boundary was on average considerably smaller than for relative clause parentheticals.

Rules for the conversion of text containing parentheticals in automated systems would therefore include instructions implementing all the contributing factors, as quantified in the relevant sections, together with the rules making the necessary adjustments to the main clause chunks surrounding the parenthetical constructions.

# Chapter 4

# Emphasis Markers

## 4.1 Introduction

The aim of this chapter is to identify a set of acoustic parameters that speakers manipulate as a result of encountering punctuation devices which seem to function as markers of emphasis in written conversation. Following the identification of these acoustic parameters, we demonstrate that manipulation of these parameters alone is sufficient to create the percept of emphasis and that different degrees of emphasis created with these parameters are appropriate for conveying certain functions. Since most of the functional reasons for the insertion of these markers (Chapter 2) relate to the disambiguation or clarification of the author's argument, we aim to create perceptually adequate emphasis for functional categories such as specificity[1] and contrast.

Other functions involving emphasis may be more paralinguistic than linguistic since they carry connotations of attitude or emotion, such as shouting, anger or annoyance. These may alter parameters in the speech production process such as vocal cord tension or sub-glottal pressure and influence additional parameters in the speech signal such as amplitude or energy which are not part of

---

[1]'Specificity' is used in this chapter to refer to the function of 'picking from a set of possibilities'. This is different from 'contrast' which is seen here to involve binary relationships, although both notions come under the general description of 'contrastive stress' in the literature.

the investigation here.

The goals of the chapter therefore are the characterization of the relationship between typographic and acoustic emphasis at the linguistic level, and the exploitation of this relationship for the provision of contextually appropriate emphasis for speech synthesis. In addition, an understanding of acoustic emphasis triggered by typographic emphasis may support research into acoustic emphasis per se, which has recently attracted considerable interest in the area of intonational phonology and speech science (Gartenberg & Hertrich 1988; Godfrey & Brodsky 1986; Gussenhoven & Rietveld 1988; Ladd 1993; Hayes 1993; Kohler 1987, 1988; Terken 1991; Terken & v.d. Hombergh 1992).

The particular markers most associated with typographical emphasis in written conversation are the asterisk marker '*' and capital letter spellings, which are sometimes combined, although other markers such as the underline character, bold script, italic type script, or even colour, are also used by some authors when available. A preliminary study into the acoustic realization of italic script was carried out by Panasyuk & Panasyuk (1991).

The domain under investigation is the word, although markers are sometimes applied to larger domains such as the phrase or the clause and in exceptional circumstances the sentence. The following shows an example of each marker and an example of combined markers for the word domain.

1    We will reimburse each student *after* the workshop, so all
     registration fees and travel should be paid by the student now ...

2    It now looks OK to me - we don't need to provide TOO much detail ..

3    I've used their technique for statistical modelling of words in a speech recognizer, and it works *VERY* well...

Restricting the analysis to the domain of the word means that this chapter is concerned with instances where markers indicate local emphasis or *narrow focus* (Ladd 1980) for linguistic purposes. This contrasts with overall changes in the pitch range of an utterance, although the term *focus* is used in the literature to describe both local and global changes. (Chomsky 1969; Jackendoff 1972; Quirk et al. 1972; Gussenhoven 1983).

In order to avoid confusing terminology, the discussion here attempts to draw a clear line between acoustic and pragmatic or semantic notions. The term *emphasis* is used exclusively to denote acoustic prominence[2] which may occur in a variety of contexts and may therefore take on a variety of pragmatically or semantically determined functions. Hence, acoustic emphasis is seen as a vehicle for signalling functional notions such as contrast, specificity (picking from a set of possibilities), warning, the repair of a false assertions and so on. Which function acoustic emphasis denotes in the end depends heavily on the degree of emphasis employed and on the actual propositional content of the utterances which carries it.

The following examples give an indication of the variety of communicative functions portrayed by typographic emphasis markers. Dependent on the larger context, examples in 4 and 5 could be specific or contrastive, in the sense that it is "pairs" of bytes rather than single bytes which should be swapped, and that words

---

[2]The term emphasis here describes what Pierrehumbert (1980) refers to as *prominence*. This is a term used to indicate differences in the height of pitch accents which are not phonologically conditioned (i.e. through downstep) and not structurally conditioned (i.e. through metrical structure), but are attributed to local gradient variability of pitch range which is under paralinguistic influence.

in "all" corpora should be included rather than including words from only some, a few, or from a single corpus.

4    Maybe *pairs* of bytes need to be swapped ...

5    If you are doing cross-corpus evaluations, it should include all the
      words in *all* the corpora.

Examples in 6, 7 and 8 provide instances where emphasis markers reinforce something which already carries emphasis. The insertion of "do" in 6, sometimes referred to as "DO support" (Brown & Miller 1980), which is a syntactic device for signalling emphasis, is reinforced further by the asterisk markers.

6    We *do* have a way to reduce problems such as this.

Repetition of the noun-phrase "answering machine" in Sentence 7 requires contextual deaccentuation on its second mention and a shift of the accent onto the pre-modifier, which is again reinforced by textual emphasis. Typographic emphasis in 8 reinforces the temporal adverb "now" which would carry a certain amount of emphasis in any case.

7    The answering machine seems appropriate to the nature of the
      project -- preferably an *intelligent* answering machine.

8    Please tell me NOW if any of the information is incorrect.

Examples 9 and 10 clearly relate to a wider, previously established context with sentence 9 repairing a false assertion or assumption of the person holidaying after the conference.

9    I am *not* holidaying, so I am looking to return ASAP after the
      conference.

10      I wonder what will happen to *this* message ...

Emphasis in 11 both reinforces or strengthens the question while at the same time possibly implying ridicule, which would at the same time introduce an attitudinal element into the discourse. Example 12 depicts a reinforcement of doubt.

11      25,000 pounds for **that** ???

12      I *SHOULD* be back at ~6 p.m.

Some examples even seem to transcribe the way the author may speak the particular word and some durational cues may be given explicitly by repeating letters as in 13 and 14.

13      I'm in a funny mood this afternoon. Well, not so much *funny* as

        WWEEIIRRDD.

14      That was toooo good to be true!

As discussed above, the genre of written conversation is only just being noticed by linguists as a rich source of examples, with the signalling of accents and emphasis possibly being the most crucial and interesting area. This is discussed further in section 4.2, which looks at the literature mainly in order to show how an understanding of the acoustic correlates of typographic emphasis relates to aspects of intonation theory, and to show how an analysis of textually encoded emphasis may shed new light on existing theories. Section 4.2 also serves to delimit the phenomena with which we are concerned, summarized in section 4.2.2.

Section 4.3 presents the measured acoustic parameters of these textual markers. Section 4.4 then presents the results of experiments investigating the perceptual salience of the measured parameters. The outcome determines, on the

one hand, whether subjects associate created stimuli with textually annotated words, and if so, whether the percepts are naturally associated with linguistic functions such as contrast.

## 4.2  Background

This section gives a brief account of how the particular analysis of textual emphasis markers presented in this thesis relates to theoretical aspects of intonational phonology. In particular, we point to the areas of the gigantic puzzle - which we may name *accentuation* - to which the analysis is relevant, in the hope of providing one of a large number of missing pieces.

### 4.2.1  "Determinism" versus "Free Choice"

The topic of accentuation was the battleground for one of the most fundamental, long-lasting and fiercely contested debates carried out in the field of intonational phonology, the debate which Bolinger refers to as "the debate over determinism and free choice" (Bolinger 1987:51), manifested in classic articles such as Gussenhoven's "Focus, mode and the nucleus" (Gussenhoven 1983), Bolinger's "Accent is predictable (if you're a mind reader)" (Bolinger 1972a), as well as Bolinger's "Two Views of Accent" (Bolinger 1985) and the associated replies (Gussenhoven et al. 1987).

At the heart of this argument is the question of which phenomena to describe by rule, if possible, and which phenomena to attribute to speaker choice, leaving them beyond the realm of rule based classification. Gussenhoven attempts a rule based account, whereas Bolinger would argue that "speakers are perfectly free

to highlight word A rather than word B or word C" (Gussenhoven 1983:378) and that an attempt to determine which one they will choose is a rather futile exercise.

Although this argument was concerned specifically with the placement of nuclear accents, similar debates are reappearing in various forms, as phonologists are beginning to divert their attention away from determining primarily the fundamental building blocks of their theories to the extension and testing of those theories in areas which initially only received marginal attention.

One such area concerns emphatic accents and their acoustic realization. A significant factor in this territory, which was exemplified by the above examples, is that it is fuzzy with respect to the boundaries between strictly linguistic and paralinguistic phenomena. In this fuzzy area, a similar debate to the above is being carried out, this time not about "determinism" versus "free choice", but about the relationship between "categories" and "gradience"[3].

It can be argued that the proponents of the gradience theory of emphasis (Hayes 1993; Godfrey & Brodsky 1986; Pierrehumbert 1980) are in line with supporters of the free choice argument, in that the acoustic realization of emphasis is seen to be entirely under the control of the speaker, resulting in a potentially infinite number of possible emphasis values which are merely constrained by the mechanics of the speaker's speech apparatus. Proponents of the category theory on the other hand would argue, that there is a linguistically meaningful category <emphatic accent>[4] which speakers and listeners can identify as such, which is part

[3]This debate originated in Bolinger's critique (Bolinger 1951) of the 'phonemic' treatment of pitch levels in accounts by Pike, Wells and Trager-Smith (Pike 1945; Wells 1945; Trager-Smith 1951).

[4]Corresponding to Ladd's "Pitch Level 4" (Ladd 1993).

of the phonological representation of a language and as such part of the phonological inventory. Therefore, this argument is not one of *rules* versus *free choice* but one concerning the relationship between phonological *representations* and free choice.

For many phonologists, free choice is seen largely as a matter of the *phonetic realization* of a phonological category. Phonological categories are established by showing that boundaries can be defined which signal a point in a continuum of possible acoustic values at which a listener perceives something as belonging to category <a> rather than category <b>, or for example, judges a signal to be emphatic rather than simply nuclear. This is demonstrated by the suggestion that, possibly due to paralinguistic factors, the category <emphatic accent>, which may be part of the linguistic system, is not considered to prevent the incorporation of a gradient element. Ladd proposes (Ladd 1993) that once the category is in use, that is the category boundary has been overstepped, speakers are indeed free to choose how much emphasis to apply[5].

The argument concerning rules versus free choice and categories versus gradience, however, is an instance of a much more global argument concerning the relationship between *meaning* and *form*, and particularly relevant to the discussion here, the relationship between meaning and phonetic form. The real issue is how *direct* this relationship is, which aspects of this relationship can be expressed by rule, and if and how this relationship should be *represented* at intermediate levels

---

[5]Ladd argues against the free gradient variability of pitch range on *any* accent. He bases his argument on the prominence relationships between accents, exemplified by the Gussenhoven-Rietveld effect, and concludes that free variability only applies to global pitch range. Gradient pitch range can only apply to an individual accent if that accent is nuclear, and if it has 'Pitch Level 4' (in which case it is not subject to the G-R-effect). Our approach here is not at odds with Ladd's, since it is assumed here that textually marked words attract a nuclear accent.

such as those of syntax or phonology.

Bolinger adopts a Saussurean approach which sees language as direct combinations between meanings and forms (Saussurean *signs*), whereas Gussenhoven adopts a more indirect approach by postulating the intervening level of "focus domains" which the speaker is free to create and upon which the rule based placement of accents operates.[6] Thus, it can be argued that the two positions are in fact not as dissimilar as they may seem at first, the difference being that Gussenhoven's approach requires an intermediate level from which to capture rules, therefore being more indirect. Although this is often overlooked, the choice factor remains, as the focus domains are chosen by the speaker. With respect to representations rather than rules, Ladd (1993) does the same, by arguing that there is a linguistically meaningful intermediate representation called "Pitch Level 4".

A further approach involving the use of an intermediate level was an approach which attempted to account for accentuation in terms of syntactic structure. This approach led to the positing of the *Nuclear Stress Rule* (Chomsky & Halle 1968) and the following debate concerning the suggested modifications suggested by Bresnan (Bresnan 1971;1972, Lakoff 1972, Berman & Szamosi 1972). However, even here, the influence of semantics and pragmatics could not be ignored, exemplified by allowances made for "contrastive stress", and references made to accent-determining information status in terms of "given" and "new" information, which in turn was determined by the structure of the sentence (Chafe 1970).[7]

---

[6]For a similar argument, see Keijsper 1992.

[7]I will not replicate a detailed discussion of this approach to accent placement but instead refer to Ladd's discussion. (Ladd 1980, pp.78-ff.).

In summary, the central point of this argument is that in the grey area where linguistic and paralinguistic influences can operate on the same acoustic parameters, such as (emphatic) accentuation, there is less space for the stipulation of intervening levels between meaning and form, which results in debates such as that between Ladd and Hayes (Ladd 1993; Hayes 1993), on whether to put emphatic accents into the (intervening) level of phonology or whether to treat it as a direct relationship between meaning and form. As Hayes remarks, "in the study of phonetic rules, paralanguage often interacts closely with the linguistic system." (Hayes 1993:11). The analysis of emphatic accents, it seems, is an area where "reconciling the goals of *phonetic specification* and *linguistic generalization*" (Ladd 1983, p.725) is difficult.

The following sections provide some support for stipulating one or more phonological categories relating to that of <emphatic accent>, although it should be stressed that perceptual experiments reported below were designed to test the perceptual adequacy of rule generated emphasis rather than for the determination of phonological category boundaries in the acoustic space. Support can be found in the results of section 4.4, which point to the possibility of category boundaries as explained below, and further, by arguing that emphasis can take on clearly specified linguistic functions, such as specificity or contrast, with clearly identifiable acoustic manifestations for both.

As a result it could be argued that linguistic functions should be described by a linguistic category in phonological theory. The disadvantage of this approach is that generalisations may be difficult to make, due to strong differences in phonetic realization which require the stipulation of categories such as <specific accent> and <contrastive accent> at which point any generalization is lost and the relationship between meaning and form is again a direct one.

This chapter is not primarily concerned with investigating whether there is indeed a phonological category which we may call emphatic accent, or to use Ladd's terms "Pitch Level 4". Breaking the problem of accentuation into its central sub-components, which are the specification of accent inventories, phonetic accent realization, and accent placement, this chapter is primarily concerned with the phonetic realization of emphatic accents. In keeping with the general direct approach prevalent throughout this thesis[8], this is investigated by the provision of meaning in terms of functional contexts and the direct measurement of speaker performance, establishing a direct link between meaning or function and the associated form. This is achieved by providing contexts requiring specificity on the one hand and contrast on the other (section 4.4) and by finding acoustic settings which best represent each.

This approach does not deny the possibility of positing a phonological category such as <emphatic accent>, but the adopted methodology is a direct one - whether the relationship is one of meaning <---> form or whether it is one of meaning <---> phonology <---> form in the description of language is not of central concern here. Even if a phonological category or categories were specified, exact realization rules would nevertheless be necessary.

As far as accent placement is concerned the approach assumes that a typographical emphasis marker attracts an emphatic nuclear accent. This is treated in more detail in section 4.3.1, suggesting that written conversation potentially provides an infinite corpus for research into the placement of emphatic accents.

---

[8]Chapter 2 argued for a direct relationship between prosodic performance in the spoken language and prosodic markings in the written language. This is again a direct relationship; however, in the context of Chapter 2 it is a relationship between one form and another form.

## 4.2.2 Constraining the task

Emphasis markers are some of the most universally used textual markers in written conversation and as a result appear with high frequency on almost any possible word and word class in an utterance. They appear in different sentence positions on nouns, adjectives, verbs, adverbs, and even on otherwise deaccented words such as prepositions and, as the numerous examples above show, may fulfil a variety of communicative functions. This reflects the amount of choice speakers or writers have in presenting their information, as discussed in the introduction to this thesis.

This flexibility makes an exhaustive account of the acoustic effects of emphasis markers difficult, due to the large number of variables involved. The combination of sentence position, word-class of the marked word and function of the marker renders a considerable number of categories in which the acoustic realization of emphasis may or may not differ. This leaves two main avenues for attempting to build a model of the acoustics of textual emphasis. The first would follow the hypothesis that different positioning and function of emphasis indeed have different acoustic manifestations. This would require the careful selection and categorization of examples prior to analysis and could result in a rather complex model which specified precise acoustic values for particular cases.

The second option would follow the hypothesis that acoustic events for creating emphasis encompassing various functions and positions may be rather similar. In particular, differences may be a matter of *degree* rather than a difference in the type of acoustic parameter used, particularly since the scope of the task has already been restricted to the linguistic rather than the paralinguistic aspects of emphasis. For example, it is conceivable that exactly the same acoustic prominence operates on the relevant syllables in the following, functionally different examples, allowing for the difference in position and provided the first is not uttered with

particular anger or annoyance.[9]

    **15**    JOHN, don't do that!

    **16**    I didn't give the book to Bill, I gave it to JOHN.

Alternatively, differences in degree may be found in examples such as the following:

    **17**    JOHN went to the beach
            (As an answer to "Who went to the beach?")[10]

In this case John is picked out of a set of possible people known to the interlocutors who could have possibly gone to the beach.

    **18**    JOHN went to the beach
            (As an answer to "Did Bill go to the beach?")[11]

Here John is in contrast to Bill and the speaker may interpret the question as an implied false assertion by the questioner and provide the answer with particularly strong emphasis. Functional differences, therefore, may be accounted for by the combination of degree of emphasis and propositional content of the utterance.

The latter avenue is the one pursued in this thesis, because it allows the specification of a simpler model of emphasis. This model is presented in section 4.4.2 in this thesis and is identified as the *Accent-Boost-Model*. The model specifies the changes in fundamental frequency and duration which are required in order to

---

[9]For support of this argument, the reader is referred to Ladd 1980, pp. 101-102.

[10]This is treated below as a *specific* context where an entity is selected from a set of possibilities.

[11]This is treated as a *contrastive* context where one entity is in direct opposition to another.

turn 'ordinary accents' into emphatic accents in cases where these have been textually annotated. The model itself is independent of particular functions or word-classes because the analysis carried out in order to derive the model contained a wide variety of examples. Functional differences can be indicated by changing the parameter values in the model in order to achieve different degrees of emphasis required for different functions, and it can subsequently be evaluated with respect to its ability to create appropriate emphasis for various functions or contexts.

The acoustic analysis in section 4.3 below is concerned with the identification of the acoustic parameter values with which the model has to be provided. The acoustic prominence triggered by textual markers in terms of the *scale* and *range* of certain acoustic parameters is measured independent of particular linguistic functions or connotation.

To summarize, this Chapter is concerned with *measuring* acoustic parameters of emphasis, stipulating those parameters in a *model* of emphasis, and *testing* the model by attempting to resynthesize contextually appropriate emphasis.

## 4.3   Acoustic correlates of textually marked emphasis

This section identifies a set of acoustic parameters that speakers manipulate when they read text containing emphasis markers and presents measurements showing the scale and range within which these acoustic parameters are used.

This was achieved by adopting the general methodology applied in this thesis of measuring prosodic differences in readings of textually marked and unmarked utterances. In this chapter, readings are compared of utterances

containing typographic emphasis markers and readings of the same utterances from which these markers have been removed, as shown in examples 19 and 20.

19     I *SHOULD* be back at ~6 p.m.

20     I should be back at ~6 p.m.[12]

The approach adopted for the investigation of emphatic accents is generally in line with a target based rather than a contour based approach.[13] (Pierrehumbert 1980; Liberman & Pierrehumbert 1984; Bruce 1977), which sees the intonation contour as a string of tones which are aligned with the segmental string, and which are phonetically realized by phonetic realization rules. The difference is that we are concerned directly with F0 targets rather than with their phonological description. The target based approach was of advantage both for the investigation into the acoustics of local emphasis as well as at the practical level for the manipulation of individual pitch targets on the frequency and the time axis.

### 4.3.1 Scope of the Analysis

A fundamental assumption adopted in this chapter for the analysis of emphasis markers is that the syllables of textually marked words will always become the nucleus in their respective tone group and will be realized with an emphatic nuclear accent. This may result in the marked version having the same distribution of accents as the unmarked version in the case where the nucleus is the same in both versions, or it may result in the shift of the nuclear accent and consequent

---

[12]The symbol '~' translates into 'around' or 'approximately' in most e-mail messages.

[13]For a detailed review on both approaches, see Cutler & Ladd 1983 and Ladd 1983.

changes in the distribution of accents of the marked sentence.

The analysis concentrates on marked accents which are nuclear in their unmarked versions providing an account of the differences between unemphatic and emphatic nuclear accents, as the following example shows:

21    John kicked the dog

22    John kicked the DOG

Nuclear accents in the unmarked versions correspond to terms such as "tonic" (Halliday 1967), "sentence stress" (Schmerling 1976), "nucleus" (Gussenhoven 1983) and "primary stress" (Chomsky & Halle 1968) amongst others.

Making a distinction based on whether an accent is nuclear or not in the unmarked cases also reflects an ongoing debate which concerns the accent-inventories of pre-nuclear and nuclear accents. (Bing 1979; Silverman & Pierrehumbert 1991; Pierrehumbert 1980). Should these be different, they may differ in the way they are turned into emphatic accents and provisions for this should be made. This is discussed further in section 4.4.3.5.

The particular parameters under investigation are fundamental frequency, in terms of pitch target height and pitch target position - and duration, measured in terms of syllable duration, vowel duration and the duration of the pre-vocalic continuant consonants [m, n, f, v, s, ʃ, z, ʒ, l, w, r].[14] These parameters have been shown to contribute to acoustic prominence by phoneticians, phonologists and speech scientists. For example, duration changes for signalling emphasis are

---

[14]Measurements were not taken for each of the consonants as they did not all occur in the required environment in the data.

applied in the MITalk speech synthesis system, where emphasized vowels have a duration of 140% (Allen et al. 1987) in relation to their unemphasized counterparts. In a study carried out by Godfrey and Brodsky (1986), speakers were able to produce different degrees of emphasis by manipulating target height, syllable duration and amplitude, and Ladd, as discussed, argues in favour of a pitch level for emphatic accents.

Target position has been shown to contribute to acoustic emphasis in German, in that emphatic accents reach their F0 maximum later in the syllable than non-emphatic accents (Kohler 1987, 1988; Gartenberg & Hertrich 1988). The possible influence of the duration of continuant consonants was discovered during the recording of the corpus for this chapter. One of the two subjects seemed to employ particularly heavy lengthening on these segments - measurements of this parameter are given in section 4.3.4.

### 4.3.1.1 Data collection and preparation

A corpus of 92 utterances (Appendix E) was recorded for one male (JE) and one female (SB) speaker. Some of these utterances had been previously recorded for another female speaker (SF) and served as pilot data. In the same way as for the database recorded for chapter 3, speech acquisition was carried out on two parallel channels, collecting speech and laryngograph data. Again, the speech was digitized at 16 bits and sampled at 10 kHz. Processing of the laryngograph signal was carried out as for the data in chapter 3, described above in section 3.3.1.1 Smoothed laryngograph tracks were used to measure F0 maximum in Hz. and F0 maximum position for the emphatically stressed syllables in the textually marked version and the corresponding syllable in the unmarked version of the utterances. F0 maximum position was measured as a percentage of the total duration of the syllable. A waveform editing program for X-Windows (xmg) was used for all measurements.

Duration measurements were carried out on the relevant words and segments in the emphatic and unemphatic versions. For accurate duration measurements all data was hand-segmented[15] and used to calculate the duration of stressed syllables, vowels, and continuant consonants.

In summary, measurements concentrated on the differences in fundamental frequency and duration between 'ordinary' nuclear and emphatic nuclear accents, by comparing accents which remained nuclear in the versions from which emphasis markers had been removed, with their emphatic versions. The comparison between nuclear accents and emphatic nuclear accents was considered to be the best initial step in attempting to provide an acoustic model of 'emphaticness', that is a model of the differences between the most prominent accent of a tone group and the same accent realized with boosted prominence. The most likely success of the application of the accent-boost model furnished with values based on these differences would therefore be in the generation of perceptually adequate emphatic nuclei from unemphatic nuclei.

Restricting the analysis in such a way does mean that further analysis may be necessary if the accent-boost model fails to provide adequate emphasis for other accent positions such as textually marked accents which are not nuclear in their unmarked utterances. Creating emphatic nuclei from other accents could then either be done by applying the same changes as for nuclear accents, or by applying something like a 'double boost' which would involve turning the particular accent into an accent which would acoustically resemble a nuclear accent prior to applying the emphasis rules. This is discussed further in section 4.4.3.5. In all cases the result

---

[15]Although good automatic segmentation algorithms are available, it was decided to hand-segment the data to obtain the highest possible accuracy. For an evaluation of automatic speech segmentation see Schmidt et al. (1991).

would be an emphatic nuclear accent as the acoustic realization of a textual emphasis marker. Experiment II below investigates the performance of the accent-boost model on various accent positions in the test utterances.[16]

## 4.3.2 Fundamental frequency

### 4.3.2.1 F0 maxima in emphatic accents

Graph 5 shows the percentage increases in F0 maxima in textually marked words as compared to their unmarked versions for one male (JE) and one female (SB) speaker. Most of the data ranges from between 0-5% to between 20-25% peak heightening on emphatic accents. For speaker SB the average F0 peak value for nuclear non-emphatic accents is 258 Hz., that for nuclear emphatic accents is 276.2 Hz., an average difference of 18.2 Hz., corresponding to just over one semi-tone.

For speaker JE, on the other hand, the average non-emphatic value for nuclear accents is 107.4 Hz. with that for emphatic accents reaching 120.8 Hz. This is on average only a difference of 13.4 Hz., but the mean F0 heightening corresponds to just over two semi-tones.

The data spread evident in this and the following graphs gives some indication of the gradient nature of acoustic emphasis with the possibility of different degrees of emphasis being employed to signal different linguistic functions. As will be shown later, it is precisely this spread of values which can be characterized statistically and used to furnish the accent-boost-model with values necessary to create various degrees of emphasis for various purposes.

---

[16]As mentioned above, the argument on whether the accent inventory is different for nuclear and pre-nuclear accents is discussed in Pierrehumbert & Silverman (1991).

**Graph 5:**

## Percentage F0 max. increases in emphatic words (JE & SB)

Legend: ■ Speaker JE  ■ Speaker SB

Y-axis: Instances (6, 5, 4, 3, 2, 1, 0)

X-axis bins: -10 - -5 | -5 - 0 | 0 - +5 | +5 - +10 | +10 - +15 | +15 - +20 | +20 - +25 | +25 - +30 | +30 - +35 | +35 - +40

Percentages in 5-point bins
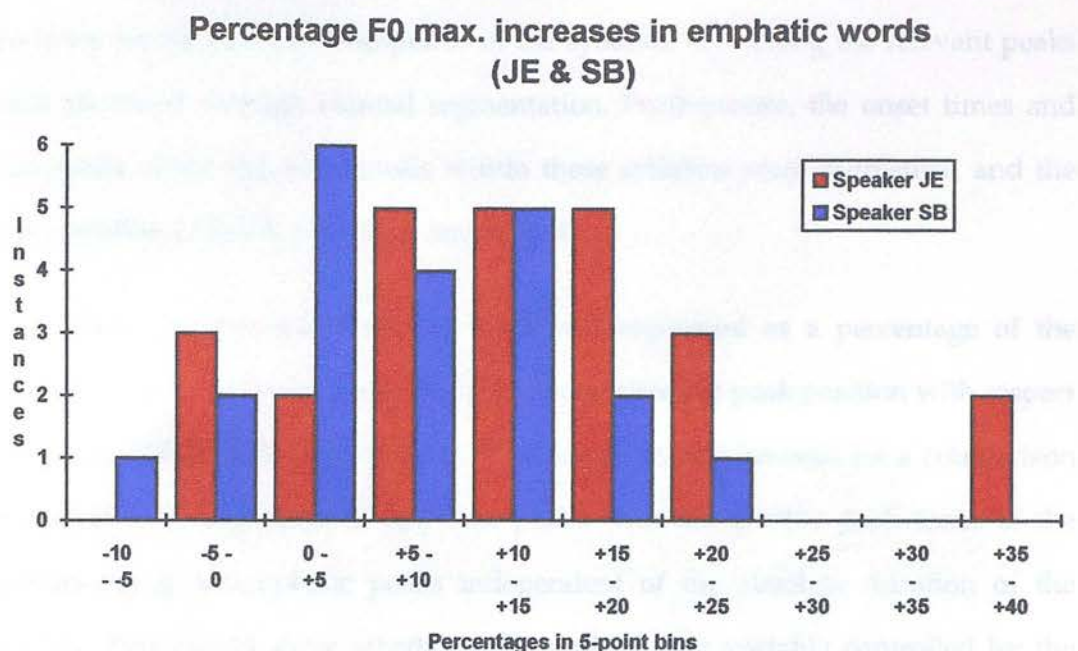
### 4.3.2.2 F0 maximum placement and contour shapes in emphatic accents

The investigation into F0 maximum placement was motivated by a series of studies carried out for German and English which suggest that more emphatic accents reach their F0 maximum later in the syllable than less emphatic accents and that some are even realized on the following syllable (Kohler 1987, 1988; Gartenberg & Hertrich 1988; Pierrehumbert & Steele 1989). Should these observations hold for the data presented in this chapter, F0 maximum placement may turn out to be an important variable in the generation of perceptually adequate emphasis undertaken here. It was also argued for German that the placement of the F0 target could indicate whether the entity was considered *given* or presumed, in which case the pitch target was placed early, as opposed to *new* or foreground information, for which the pitch target would be placed late.

The following investigates the possible influence of the degree of emphasis

on the placement of the F0 maximum by examining relative and absolute differences in peak position for emphatic and unemphatic syllables and vowels. For this purpose the start and end-points of the syllables containing the relevant peaks were identified through manual segmentation. Furthermore, the onset times and end-points of the stressed vowels within these syllables were segmented and the exact position of the F0 peak was determined.

First, the position of the F0 peak was expressed as a percentage of the elapsed syllable. This was done in order to normalize the peak position with respect to the absolute length of the syllable. This normalization allowed for a comparison of the relative peak times of emphatic peaks with the relative peak times of the corresponding un-emphatic peaks independent of the absolute duration of the syllable. This would show whether peak position is a variable controlled by the speaker independent of any durational changes made to emphatic segments or syllables. By subtracting the value of the relative peak position of the unemphatic peak from that of the emphatic peak, it was possible to express a quantity by which the F0 maximum is reached later in emphatic syllables as compared to their unemphatic versions.

Graph 6 shows the percentage differences between peak positions in the emphatic- and unemphatic syllables in the readings of Speaker JE. For example, the F0 maximum occurred between 0% and 10% later in the emphatic syllables for 8 examples and was more than 40% later in the emphatic- rather than the unemphatic version for one example.

Overall, the results indicate that the F0 maximum position in the syllable is on average 5.6% later for speaker JE and 4.04% later for speaker SB, which is significant at .05 for the former, but is merely a tendency for the latter (.10)

**Percentage differences in peak position between emphatic and unemphatic syllables (JE)**
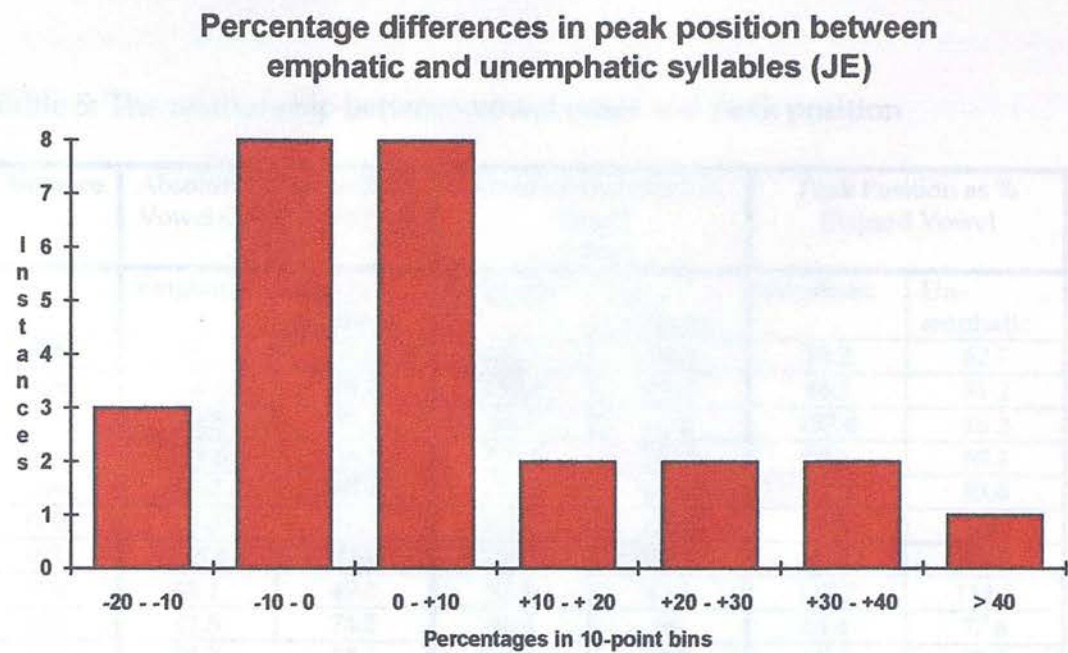


In a further analysis, the absolute durations from vowel onset to the F0 peak were calculated for stressed vowels of the emphatic and unemphatic test syllables in the readings of Speaker JE. These absolute durations are given in Columns 2 and 3 in Table 5 below. As Table 5 shows, these differences are significant. It is important to note that some of the durations from vowel onset to F0 peak are longer than the duration of the vowels themselves, which suggests that some very late peaks are, as observed in the data for German, realized on the following syllable. In order to investigate whether there is a correlation between the differences in the absolute values from vowel onset to F0 peak and the absolute differences of the vowel durations for the two conditions, a Spearman's Rank Correlation was calculated. The result turned out to be insignificant. (Rank Correlation Coefficient=0.226). This is an important result which suggests that there is no significant interplay between the amount by which a vowel is lengthened for emphasis and the amount by which the F0 peak is positioned later in the vowel. This lack of correlation may suggest

that the F0 peak position is a variable which speakers consciously control, and which they do so independently of duration.

Table 5: The relationship between vowel onset and peak position

| Utterance | Absolute Duration from Vowel Onset to F0 Peak (ms.) | | Absolute Duration of Vowel (ms.) | | Peak Position as % Elapsed Vowel | |
|---|---|---|---|---|---|---|
| | Emphatic | Un-emphatic | Emphatic | Un-emphatic | Emphatic | Un-emphatic |
| 011 | 83 | 52.6 | 140.1 | 83.8 | 59.2 | 62.7 |
| 018 | 139.4 | 79.7 | 157.8 | 155.7 | 88.3 | 51.2 |
| 019 | 111.9 | 30.4 | 56.7 | 67.2 | 197.4 | 45.2 |
| 065 | 191.6 | 74.5 | 237.6 | 117.5 | 80.6 | 63.4 |
| 066 | 56.7 | 55.9 | 259.1 | 166.3 | 21.9 | 33.6 |
| 070 | 88.3 | 97.5 | 115.9 | 128.2 | 76.2 | 76 |
| 072 | 135.4 | 114 | 204.1 | 136.1 | 66.3 | 83.8 |
| 074 | 61.1 | 49.5 | 51.9 | 43.3 | 117.7 | 114.3 |
| 075 | 61.5 | 74.5 | 94.9 | 96 | 64.4 | 77.6 |
| 076 | 91.2 | 35.3 | 104.1 | 99.7 | 87.6 | 35.4 |
| 077 | 111.7 | 57.3 | 245.1 | 176.9 | 45.6 | 32.4 |
| 078 | 112 | 56.7 | 107.1 | 88.9 | 104.3 | 63.8 |
| 081 | 68.8 | 77.6 | 186.7 | 129.7 | 14.6 | 59.8 |
| 082 | 133.1 | 65.7 | 129.4 | 96.3 | 102.9 | 68.2 |
| 083 | 39.1 | 34.3 | 37.2 | 61.9 | 105.1 | 55.4 |
| 084 | 57.5 | 47.7 | 74 | 89.1 | 77.7 | 53.5 |
| 085 | 41.3 | 17.5 | 341.2 | 234.6 | 12.1 | 7.5 |
| 088 | 178.9 | 32.4 | 55.4 | 50.8 | 50.1 | 63.7 |
| 089 | 89.3 | 87.4 | 115.1 | 102.1 | 77.5 | 85.6 |
| 090 | 72 | 95.9 | 118.1 | 92.8 | 61 | 103.3 |
| 091 | 103 | 32.7 | 91.2 | 69.5 | 113.5 | 47 |
| 092 | 105 | 42.3 | 382.3 | 140.5 | 27.5 | 30.1 |
| Means | 91.1 ms. | 63.9 ms. | 150.2 ms. | 110.3 ms. | 75.07% | 59.72% |
| T-values | t = 3.31; sig@.005 | | t = 8.3; sig@.005 | | t = 1.70; non significant tendency | |

A further relational measure is also introduced in Table 5 which, in the same way as the data for emphatic and unemphatic syllables presented in Graph 6, expresses the position of the F0 peak as a percentage of the elapsed duration of the stressed vowel. This normalization again allows the examination of the position of the F0 peak independent of the amount by which the vowel has been lengthened in the

emphatic versions - factors which have been shown to be independent in the results involving absolute measures as presented above. This relational measure shows that emphatic peaks are positioned on average 15% later in the stressed vowel than their unemphatic counterparts (not significant), which compares to around 5% observed in the syllable data presented in Graph 6. As this result is insignificant we can conclude that although a tendency for later peak positions in emphatic vowels can be observed, speakers do not manipulate F0 peak position with reference to a relational measure involving the total duration of either stressed vowel or syllable. The results point to the independent nature of F0 peak position and duration as acoustic markers of emphasis although it was impossible to clearly identify rules underlying the placement of F0 peaks in emphatic accents.

Descriptions of late peak positions have found their way into phonological theory in terms of the feature specification [delayed peak] (Ladd 1983). Ladd argues that in English as well as German this feature changes a fall-accent to a fall-accent which begins with a rise on the accented syllable and may not reach its peak until the following unaccented syllable. These contours are referred to a 'scooped' contours (Ladd 1980; Youd 1992). Some analysts, however, treat them as different tones altogether, by referring to scooped contours as 'rising-falling' tones (Halliday 1967; O'Connor & Arnold 1961).

Graph 6 and Table 5, however, also point to a few instances where the peak in the emphatic version is placed considerably earlier in the syllable as compared with the same peak in the unemphatic version creating a flattened and lengthened fall. This occurred, for example, in the reading of the utterance "We *do* have a way to reduce problems such as this". These earlier peaks up to and including moderate later peaks can all be generally considered "medial" peaks which show considerable variation in their alignment. This is reflected in some theories which distinguish

between truly early peaks - which Grice (1995) describes with the feature [anticipated peak] - and all other peaks (Kohler 1987) positing no strict boundaries between medial and late peaks. It is suggested here that emphatic nuclear tones are realized by a variety of pitch accent shapes, mainly H*, L+H* and L*+H, with the first constituting an environment for medial- and the latter two for delayed peaks with the latter two also resulting in the mentioned scooped contours.

Possible regularities in the occurrence of any of the accent shapes were not observed in the data, hence this parameter was not manipulated in the experiments reported below.

### 4.3.3 Duration

Graph 7 shows the differences in syllable duration between textually marked and unmarked words.

**Graph 7:**



Percentage syllable duration increases in emphatic words (JE & SB)
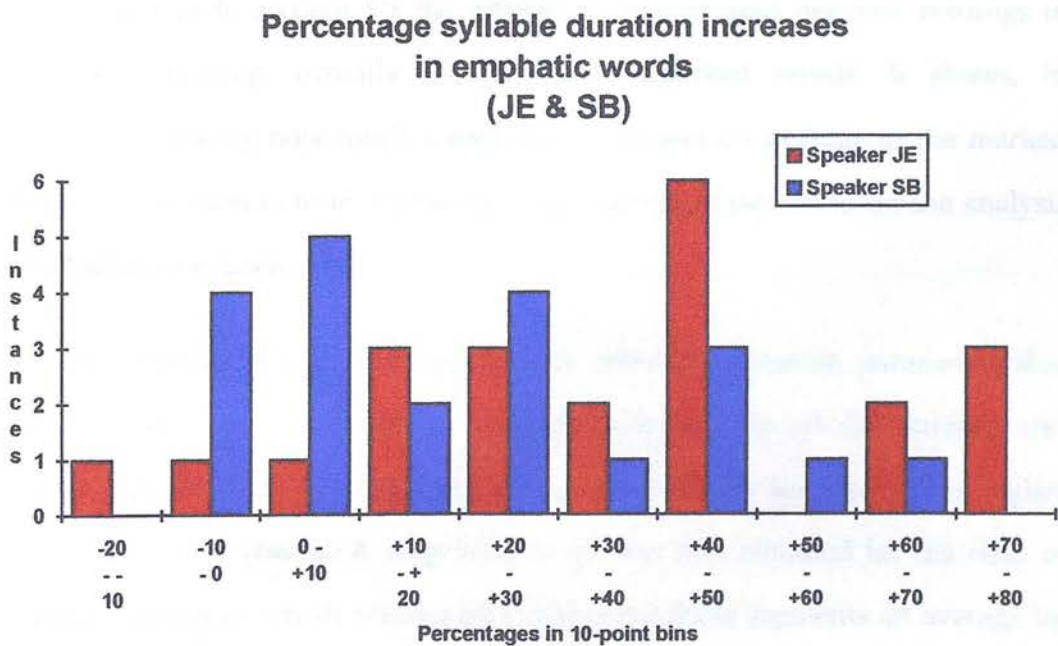
Similarly to the changes in pitch-accent height, the spread of data for syllable duration is wide, for speaker JE mostly between 10% and 40%, with a few considerable lengthenings of around 70%. The lengthenings made by speaker SB are generally more modest ranging from -10%, which is actually a shortening, to around 40%.

For the purpose of creating emphasis in resynthesized utterances it was decided to investigate vowel durations, distinguishing the broad phonological categories of 'short' and 'long' vowels. This was done due to the complex relationship between segments in the lengthening of syllables (Campbell 1992). This also facilitated the implementation of duration changes in the resynthesis of utterances for perception experiments. The changes in vowel duration are quantified in 4.3.4 below.

## 4.3.4  Summary of parameters

Graph 8 summarizes the differences in the parameters which were measured in order to attempt to account for the differences in emphasis between readings of sentences containing textually marked and unmarked words. It shows, in percentage terms, by how much a parameter increases on average in the marked version. Measurements from utterances which served as pilot data for the analysis are included here (Speaker SF).

In particular, Graph 8 shows in more detail the duration parameters that were identified. Vowel duration has been separated out from syllable duration and suggests that the degree of lengthening in phonologically long vowels is higher than that for short vowels. A surprising result was that obtained for the class of continuant consonants with speaker SB lengthening those segments on average by

40% and speaker JE even by 80%, suggesting that the duration of this class of segments is a heavy carrier of emphasis[17]. For all parameters, average percentage differences are listed with their t-values and significance in Appendix F.

**Graph 8:**

**Average percentage increase of salient parameters in emphatic readings**
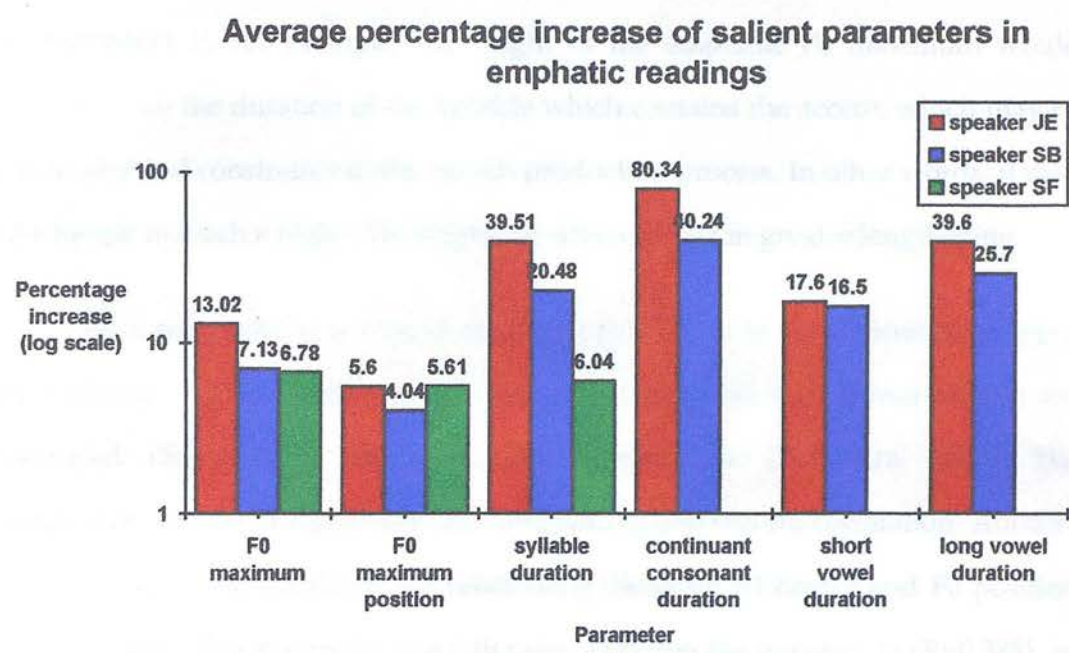


Short and long vowel durations, together with continuant consonant duration and F0 maximum are the parameters which were manipulated in resynthesis and tested for their contribution towards creating emphasis in various perception experiments. Section 4.4 examines whether these parameters alone are sufficient to create

---

[17]These figures are based on the analysis of sparse data since there were not many examples in the analyzed corpus. However, these differences were statistically significant at .01(JE) and .025(SB). Furthermore, values for SB were used subsequently in perception tests (section 4.5) were they were identified as contributing significantly in the generation of percepts of emphasis.

percepts of emphasis and whether different degrees of emphasis based on the knowledge of the range of these parameters can be created.

### 4.3.5 The interaction between parameters of emphasis

This section summarizes the apparent independence of parameters of emphasis by looking at the correlation between parameter values. Parameters could be shown to be dependent if, for example, the height of the emphatic F0 maximum would correlate with the duration of the syllable which contains the accent, which may be due to physical constraints in the speech production process. In other words, it may take longer to reach a higher F0 maximum which results in greater lengthening.

Each parameter was ranked according to the size of the difference between the emphatic reading and the unemphatic reading and rank correlations were calculated (Spearman's rank correlation, Greene & D'Oliveira 1982). The comparison between heightening and lengthening showed no correlation. Another possible correlation involving the relationship between F0 height and F0 position showed a small but insignificant relationship between the parameters (R=0.388), as did the correlation of the parameters duration and F0 position (R=0.266).

In summary, no correlations between the measured parameters of F0 maximum, F0 maximum position and duration could be established from which we may conclude that the parameters are seemingly independent of each other.

### 4.3.6 Inter-speaker differences

The results in sections 4.3.2 to 4.3.4 suggest that speakers differ considerably in the degree to which they use the parameters of emphasis to signal a textually marked

word. In our data, however, there seems to be little evidence that speakers prefer one parameter over the other or that one speaker uses one parameter more prominently at the disadvantage of another. Instead, the differences seem to relate to the overall degree of emphasis provided by each parameter.

Speaker JE uses all the available parameters to a higher degree than Speaker SB with the exception of short vowel lengthening, indicating that he may have been a more expressive reader altogether.

These findings contradict those presented in Godfrey and Brodsky (1986) who suggest that "speakers have different but internally consistent strategies for deploying these acoustic cues" (Godfrey & Brodsky 1986:4). However, in their study, subjects were asked to produce ten self-determined degrees of emphasis which may have led some subjects to 'experiment' with the acoustic cues rather than produce them in a 'natural' context.

### 4.3.7 Acoustic differences between emphasis markers

The small internal survey reported in Chapter 2, which was concerned with differences in the use of textual markers, suggested that there was a perceived difference in function between capitalized words and words surrounded by asterisks, in that capitalized words were thought to indicate loudness, shouting or annoyance, whereas asterisks were thought to signal an exceptional point of emphasis without a change in volume.

In fact, the functional characteristics attributed to asterisks in our survey are rather similar to those attributed to italics in the work of McAteer (1992; 1989), who looks at functional differences in the use of capital letters and italic script in the signalling of emphasis. Again, capital letters are considered to denote size, volume

or loudness, whereas italics indicate more subtle concepts of mood, feeling or emotions.

A further claim made by McAteer is that capital letters are more appropriate for "modulatory"- and italics more for contrastive emphasis. The distinction between modulatory and contrastive emphasis (Dik 1980) is broadly equivalent to the distinction made in this thesis between reinforced emphasis (Examples 6,7&8) and contrastive emphasis (Examples 4&5) shown in section 4.1. Reinforced emphasis emphasizes something which is already prominent to a certain degree, whereas contrastive emphasis concerns previously non-prominent entities. Modulatory emphasis seems to emphasize a word in its own right in order to signal its particular importance, whereas contrastive emphasis emphasizes a word with reference to possible opposites. Although there seems to be a convergence of opinion on the similarities in function between asterisks and italic script as opposed to capital letters, the above examples (6,7&8) show that asterisks are often used for modulatory as well as for contrastive purposes. An examination was carried out in order to establish if suggested differences in function influence the particular set of acoustic parameters examined in this chapter.

Pitch target heights and syllable durations were analyzed separately for words written with capitals and those surrounded by asterisks in order to measure possible acoustic differences between the markers. Although Speaker SB showed a difference in syllable lengthening between the markers, in that for asterisks lengthening was 17.2% and for capital letters it was 24%, the result was not significant. The same applies to a difference in pitch accent heightening which was observed for Speaker JE, which was 14.3% for asterisks but only 11.9% for capital letters.

This may indicate that either a difference in function is not registered

acoustically, at least not in the measured parameters, or that authors used different markers to indicate similar functions. The latter explanation is favoured here, as the examples in the introduction to this chapter show. Although stricter conventions concerning the use of particular emphasis markers may develop over time, it seems the case at this point that a variety of devices are being used simply to indicate that something is *different*, *extraordinary*, or *remarkable* about the way the utterance should be accented, and that the precise function is provided to a large degree by the propositional content of the utterance.

In addition, it should be reiterated that the view presented in this thesis is essentially one of emphasis consisting of gradient parameters which can be scaled up or down to suit particular contexts. Textually annotated emphasis, therefore, is not seen as a rigid system with fixed acoustic representations, although working practices with regard to use and function are undoubtedly developing as the two 'surveys' show. Functional aspects of textually encoded emphasis will be returned to below in section 4.4.3.

Section 4.4 below will show whether the acoustic differences measured in readings with textually marked and textually unmarked words bear any relationship to the perception of emphasis when applied in resynthesized utterances.

## 4.4 The perceptual salience of acoustic correlates of emphasis

Section 4.3 has pointed to some *differences* in fundamental frequency and duration between syllables and segments of textually marked words in comparison to their unmarked counterparts, in readings of sentences where special emphasis markers had been removed. It would be too simplistic, however, to assume that these measured differences amount to the exact quantification of the necessary acoustic parameters for creating perceptually adequate emphasis. In particular, there may be additional salient parameters which were not measured, such as amplitude, energy, or speech rate, and the omission of these may make it impossible to create natural sounding emphasis with the parameters that were extracted. Furthermore, some measured differences in the physical world may not be perceptually salient, in which case some of the measured differences may play no part in the perception of emphasis at all.

The remainder of this chapter is concerned with testing the identified and quantified acoustic parameters of emphasis in perception experiments. It is examined whether these parameters alone are sufficient to create the percept of emphasis (section 4.4.1), and whether it is possible to generate different degrees of emphasis which are adequate for conveying certain functions (section 4.4.3). This is done by using an accent-boost model which contains rules and values regarding the identified acoustic parameters (section 4.4.2). Experiment I investigates whether it is at all possible, given parameter changes in F0 and duration applied to a single syllable, to create stimuli of that syllable which listeners perceive as emphatic. Experiment II examines whether this can be done by the application of rules and values specified in a model, and whether the resultant utterances are considered to have appropriate emphasis for certain contexts that require emphasis.

## 4.4.1 Experiment I: Turning unemphatic into emphatic syllables

### 4.4.1.1 Introduction

This experiment investigates whether it is at all possible to create the percept of emphasis by manipulating the F0 and duration parameters identified above. In particular, the specific case is investigated where the marked syllable in the emphatic reading is the nucleus of the unmarked reading as exemplified by the following sentences (sentences 19 and 20 above):

John kicked the dog versus    John kicked the DOG

Hence, the comparison is between *nuclear* syllables and *emphatic nuclear* syllables. Should subjects be able to differentiate consistently between the two, the experiment would show that specific changes to the fundamental frequency and duration of the nuclear syllable would turn it into an emphatic nuclear syllable.

Subject to comparison were the textually unmarked versions of the utterances and modified versions of the textually unmarked utterances onto which the F0 and duration values from the marked readings were directly transferred.

### 4.4.1.2 Data preparation

The preparation of the data involved a direct transfer of parameter values from the marked reading to the unmarked reading for six base sentences, using a waveform editor (xmg) and an LPC-based re-synthesis program[18] running under the Waves™ environment. This enabled the creation of several resynthesized versions of each unmarked sentence. One version was simply the resynthesized utterance containing

---

[18]xmg was developed at CSTR by Mike Steele, the LPC-based resynthesis program was written by Steve Isard and his students.

the textually unmarked syllable, other versions were based on this version but contained marked values for F0 and duration on the relevant syllable, as extracted from the utterances which contained textually marked syllables.

This allowed for the testing of individual parameters in isolation by creating versions where only one of the parameters was changed. For example, if the marked syllable in the emphatic version had an F0 maximum which was 12 Hz. higher than the corresponding peak in the unmarked version and a vowel duration which was 20 ms. longer, one version would contain these combined 'emphatic' values, others would test individual parameters such as continuant consonant duration (if relevant), and one version would remain as in the unmarked case.

For each base sentence, the following versions were prepared:

A)    a resynthesized version of the textually unmarked sentence

B)    a resynthesized version of the textually unmarked sentence containing the F0 maximum value for the marked version on the relevant syllable

C)    a resynthesized version of the textually unmarked sentence containing the vowel duration of the marked version on the relevant syllable

D)    a resynthesized version of the textually unmarked sentence containing both the F0 maximum value and the vowel duration of the marked version on the relevant syllable

E)    Where possible the continuant consonant duration parameter was used in resynthesized versions of the textually unmarked sentence containing the marked continuant consonant values in isolation as well as in combination with the vowel and the F0 parameters.

This resulted in the preparation of 30 resynthesized utterances constructed from six base sentences which had originally been recorded from text containing no

emphasis markers (see Appendix G:b).

The overall intonation contours of the textually unmarked versions were stylized first, permitting the use of these contours for the resynthesis of the various 'emphatic' versions. Resynthesis was carried out using stylized straight line contours checked for perceptual equivalence with the original with respect to its intonation, following the "close copy" approach developed at IPO (de Pijper 1979; 't Hart 1979). These contours were manipulated locally at the relevant syllable.

### 4.4.1.3 Methodology

Ten subjects were asked to listen to the thirty resynthesized sentences which either contained emphatic values for duration and F0 - in isolation or together - or which contained the original values of the unmarked sentence. The utterances were randomized once and presented in isolation with a gap of six seconds between utterances and preceded by a beep to indicate the start of the next utterance. It was felt that a pairwise comparison of marked and unmarked versions would have been too easy for the subjects and it was considered important to investigate whether subjects could make rapid judgements on a one-off, non-comparative basis.

In the pre-amble to the experiment subjects were introduced to the concept of sentence stress and were told that sentence stress would fall onto the underlined word in each of the sentences on their instruction sheet (see Appendix G). They were then told that speakers and readers often mark out a word for special emphasis or prominence, especially if a word has been specifically marked for emphasis by using italics, bold script or capitalization. Immediately before the main experiment subjects were presented with three pairs of emphatic and unemphatic readings of example sentences for priming. They were then asked to listen to the thirty individual utterances and indicate whether the underlined word had been

read as a result of the reader having encountered a capitalized word or whether it had been read merely with the neutral, nuclear prominence one would expect to fall onto that word.

## 4.4.1.4 Results

Graph 9 shows that for the six resynthesized unemphatic base sentences, four were judged clearly unemphatic, one was undecided and the relevant syllable in one sentence was judged emphatic by seven subjects, although the original reading was intended to be unemphatic. Negative numbers in the graph indicate judgements for [unemphatic], positive numbers represent judgements for [emphatic].

**Graph 9:**

### Perception of six unemphatic base sentences (10 subjects)
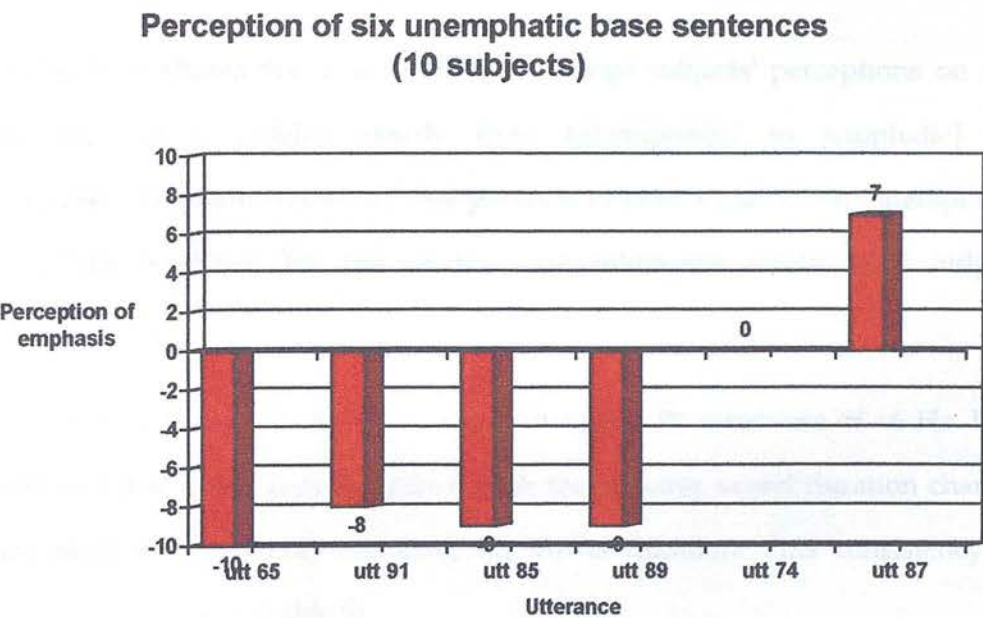


This shows clearly that the nuclei of the unchanged utterances, which are all based on the unemphatic reading, were judged to be unemphatic 'normal' nuclei. For the

following utterance (087) this was not the case, which may be explained by the fact that the auxiliary "should" is in a contrastive position in the utterance.

"For each dialogue, the digits entered are compared with the digits that should have been entered".

This contrastive position becomes more obvious if one accounts for the elided constituent "that were" at the beginning of the utterance.

"For each dialogue, the digits {that were} entered are compared with the digits that should have been entered".

Hence the utterance (087) was unsuitable as an unmarked base utterance since the prominence on "should" in the unmarked reading is not a valid example of regular sentence stress but was contrastive in the first place due to the syntactic construction.

Graph 10 shows that it is possible to change subjects' perceptions on the emphaticness of a syllable clearly from [unemphatic] to [emphatic] by superimposing emphatic values for key parameters onto a previously unemphatic syllable. This is shown for one of the base utterances which were judged unemphatic (065).

Graph 10 further shows that a small change in F0 maximum of +6 Hz. had no effect and that this change combined with the effective vowel duration change had the same effect as only changing the vowel duration. This consistency is examined further below (Table 5).

## The effect of added emphasis parameters on the perception of emphasis
### (Sentence 065)
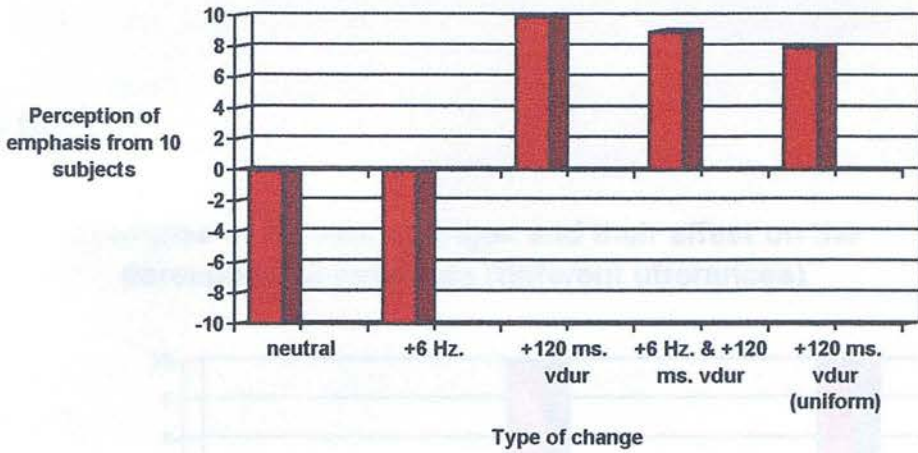


Graph 11 shows the results of manipulating unmarked base sentences by applying changes of differing magnitudes using only the F0 maximum parameter, and further demonstrates the resultant shift in judgements from [unemphatic] to [emphatic]. This shows clearly that F0 maximum is a perceptually salient parameter which can be applied in isolation (see utterances 65; 91; 74).

The results also show that changes of higher magnitudes do not always have the desired effect. For example, a 12 Hz. increase of the F0 maximum in utterance 74 was enough for all subjects to judge the resultant syllable as emphatic. However, the base utterance of 74 was already judged emphatic by half of the subjects and so

---

[19]The vowel duration change indicated in the table under "vdur" was applied to the second part of the diphthong /ei/ in the word "ages" in sentence 65 and elicited [emphatic] judgements from all 10 subjects. However, lengthening the same diphthong uniformly only persuaded 8 subjects of the emphaticness of the syllable. This may indicate that the duration of the second part of the diphthong is more salient in the perception of emphasis.

a small change was enough to render the syllable as unambiguously emphatic. On the other hand, a 24 Hz. increase in the F0 maximum for utterance 089 had no effect at all. This kind of behaviour could suggest a categorical boundary although it should be viewed as strictly anecdotal due to the scarcity of data.

**Graph 11:**

### Examples of F0 max. changes and their effect on the perception of emphasis (different utterances)
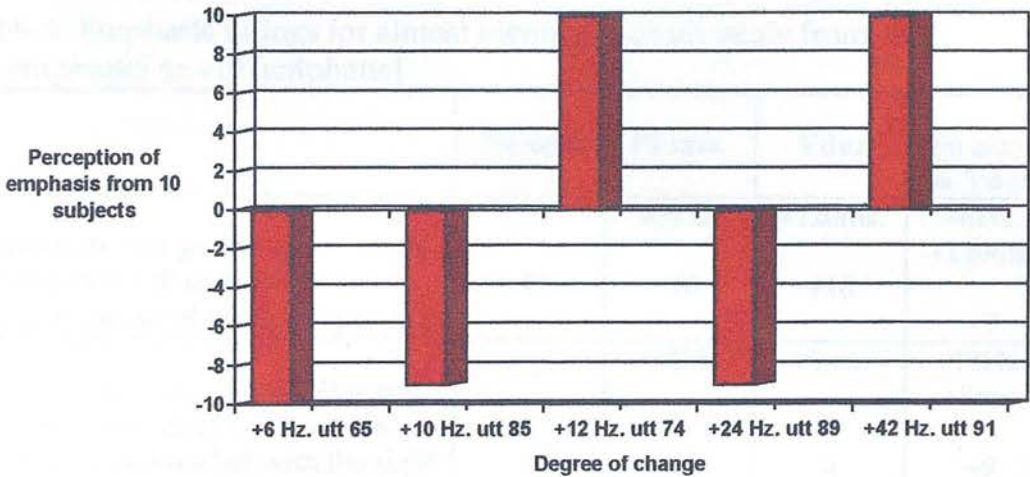


Similar changes in other utterances, involving the parameters F0 maximum, vowel duration and duration of pre-syllabic continuant consonants also managed to change subjects perception from [unemphatic] to [emphatic]. For example, the relevant syllable in sentence 74[20] changed from a split judgement to [emphatic] after the application of an F0 maximum change of +12Hz. on its own, by adding 65

---

[20]Utterance 74: "Passwords in any of the following categories could be broken in MINUTES by a hacker with the right tools".

ms. in continuant consonant duration on /m/ in "MINUTES", and after the application of both these changes together. The combination of these changes with vowel duration changes of insignificant proportions elicited the same responses.

Many versions of utterances in this experiment for which the changes made to the unemphatic base sentences were very small, or versions where similar changes were applied, functioned as control sentences. These show considerable consistency in subjects' responses throughout the experiment. This is shown in Table 5 with the associated utterances.

Table 6: Emphasis ratings for almost identical stimuli (scale from -10 [unemphatic] to +10 [emphatic]

|  | Neutral | F0 max. | Vdur | F0 max. & Vdur |
|---|---|---|---|---|
| Start up rn, and go and get a mug of coffee as it will take AGES to run through the list of newsgroups (065) | -10 | +6Hz. -10 | +120ms. +10 | +6Hz. +120ms. +9 |
| Passwords in any of the following categories could be broken in MINUTES by a hacker with the right tools (074) | 0 | +12Hz. +10 | +8ms. 0 | +12Hz. +8ms. +9 |
| Sorry folks, the BT talk is CANCELLED (091) | -8 | +42Hz. +10 | +21ms. -8 | +42Hz. +21ms. +10 |

For example, for utterance 091 where 8 subjects judged the unmarked version as [unemphatic], an F0 maximum change of 42Hz. resulted in [emphatic] judgements by all 10 subjects, a small vowel duration change of 21ms. had no effect and the F0 change with the vowel change resulted in the same judgements as the F0 change by itself. This suggests that clear and consistent judgements can be made about when a nucleus is a neutral one and when it is an emphatic one.

However, it should also be noted that for one of the unemphatic base sentences (85) substantial changes in vowel duration (+107 ms.), continuant consonant duration (+94 ms.) and both of these combined with a +10 Hz. increase in the F0 maximum, did not result in a significant change of perception from [unemphatic] to [emphatic]. The following two explanations can be offered for the inability to create emphatic perceptions for this utterance.

A)    There was no substantial change in F0 peak height for this utterance as the emphatic F0 height difference between the unmarked and the marked version was only 10 Hz. A bigger change may have resulted in [emphatic] judgements by subjects. It follows that the substantial duration changes that were undertaken did not on their own manage to create a percept of emphasis.

B)    Subjects could have perceived a difference which they did not perceive as correlates of emphasis, or more precisely, did not associate with textual emphasis markers. Hence the manipulations which were undertaken may have sounded odd or 'unnatural'.[21] If this were the case it may be concluded that the manipulations that did work are manipulations which subjects associate with emphasis that sounds appropriate or 'natural'.

### 4.4.1.5 Conclusions

Although this experiment should be seen as exploratory, a number of conclusions can be drawn from the results which provide the necessary background for the main investigation of this chapter, carried out in Experiment II below. The most

---

[21]Unfortunately, original emphatic readings could not be directly compared because they were recorded on different days and could be distinguished by background echo.

important conclusions are the following:

**Conclusion 1:** Evidence presented in Graphs 10 and 11 show that it is possible to change subjects' perceptions from [unemphatic] to [emphatic] using the parameters F0 maximum, vowel duration, and continuant consonant duration in combination and sometimes in isolation.

**Conclusion 2:** Subjects are consistent in their responses when judging on more or less identical stimuli, as shown in Table 5. This indicates that subjects are able to distinguish consistently between nuclear and emphatic nuclear syllables.[22]

As a result, it can be concluded that it is possible to create percepts of emphasis using the identified acoustic parameters discussed in Section 4.4.

4.4.1.6 Extending the scope for creating emphasis

After having identified acoustic parameters of emphasis which can be applied successfully to create the percept of emphasis it is now justified to examine whether it is possible to specify an algorithm for the generation of emphasis which is based on our knowledge of the acoustics of emphatic accents and which allows the generation of emphatic accents at will rather than by the direct approach used in Experiment I. Such an algorithm would then be open to application and testing on accents in various positions in an utterance and may show its general applicability.

A further question which can now be examined relates to whether percepts of emphasis such as the ones created in Experiment I can be judged to be reasonable or 'natural' ones - considering that only a few parameters were being used - or

---

[22]This may suggest that there is a phonologically meaningful category relating to <emphatic accent>.

whether subjects merely equated a difference or even an oddity in the signal with emphasis. Experiment II below is designed to answer both of these questions but the following section first presents an algorithm for generating emphatic accents.

## 4.4.2 The accent-boost model of emphasis

4.4.2.1 General approach

The approach taken in this chapter so far suggests that emphatic accents are simply accents that have been boosted in such a way as to make them acoustically more prominent than their un-boosted or unemphatic but nuclear counterparts. Some of the parameters which may describe this difference between emphatic and unemphatic accents were specified, and it is the difference in acoustic prominence that was presented in 4.3.1. Those differences were subsequently tested in Experiment I above, which attempted to emphasize words by superimposing those emphatic values directly onto unemphatic words. Another feature of the approach taken is that differences are expressed as *local* differences, such as the target-height of a pitch accent or the durations of specific vowels or consonants.

In order to carry out a full evaluation of the approach, it is necessary to translate it into an algorithm so that it can be tested in general application. This is achieved by the Accent-Boost Model, which specifies the changes that are required to make particular accents emphatic by boosting the known duration and F0 parameters of a speaker, or by boosting the parameters of a speech synthesis system.

The Accent-Boost Model is essentially an "amplifier" which takes the parameters of F0 maximum and duration as input and boosts these parameters to create local emphasis. The algorithm is presented below containing values for the

two settings subsequently tested in Experiment II below. One setting generates emphatic accents based around the *Mean* of the parameter values identified for the particular speaker which are laid out in section 4.4 above (first value). Another setting, chosen to give stronger emphasis, is based around the starting point of the *Upper Quartile* of the data spread of the parameter values for that particular speaker (second value). Experiment II tests these two particular settings, but it should be remembered that they are by no means the only possible ones.

## 4.4.2.2 Boosting nuclear accents

Continuing with the description of accents as outlined in section 4.4.1 above, the accents operated on by the following rules are ones which by the most general nucleus placement rule would be nuclear even if not boosted by additional emphasis. Hence, the rules turn nuclear accents to emphatic nuclear accents and are concerned with local changes to the F0 target and local duration changes to certain speech segments. The following changes are specified:

**Rule I:**

I. Change durations as follows:

1. Lengthen the nuclear vowel by (16.5%; 20.2%) if it is phonologically a short vowel

2. Lengthen the nuclear vowel by (25.7%; 40.8%) if it is phonologically a long vowel or a diphthong

3. Lengthen a continuant consonant preceding the nuclear vowel by (40%; 72%)

II. Change F0 as follows:

1. Increase the F0 maximum by (7.1%; 12.7%)

2. Calculate the total amount of inserted material under (I) and

shift the F0 peak to the right by that amount

## 4.4.2.3 Boosting pre-nuclear accents

In contrast to the rule above, the following rule is one which has the effect of shifting the nucleus from its position assigned by the general nucleus placement rule onto another accent. The effect is that an accent is turned from a pre-nuclear accent to an emphatic nuclear accent. As a result, the now post-nuclear accents require to be deaccented.[23] Therefore, the following rules specify accentuation as well as deaccentuation. It can be argued that this has an effect on the overall contour of the utterance, but nevertheless the rules are concerned with local changes at numerous positions in the utterance through the 'promotion' and 'demotion' of certain accents.

For Experiment II it was decided to test the same boost settings as described in 4.4.2.2. This was done in order to test whether the settings specified in Rule I above would generalize across different accent positions, although the absolute F0 values of pre-nuclear accents would be different from those of the nuclear accents to

---

[23]Although, as explained above, this is theoretically not always the case as it depends on context. It was decided to test deaccentuation rules, which would presume some previously established context, which was possible due to the simplicity of the utterances used in the experiment.

which Rule I applies.[24]. In particular, absolute F0 values for the first utterance peak (P1) would have been considerably higher, due to the special status of P1 (Cooper & Sorensen 1981).

**Rule II:**

    I. Repeat the rules in I and II to create an emphatic nuclear accent

      II. Apply deaccentuation rules to the remaining accents as follows:

        1. Keep any boundary tone preceding the boosted accent

        2. Drop the second and all subsequent targets after the boosted

           accent to the average final F0 drop for that speaker (200 Hz.

           speaker SB)

        3. Keep the first target after P1 at 3/4 height between P1 and

           the average final F0 drop for that speaker[25]

        4. Insert a final boundary tone

## 4.4.3 Experiment II: Evaluating an emphasis generation algorithm

The purpose of Experiment II is to show, firstly, that a percept of emphasis can be created by principled changes, that is by application of the accent boost model. It further attempts to show that our knowledge of how to generate emphatic nuclear

---

[24]An alternative technique would have involved normalizing pre-nuclear accent values to make these accents more like nuclear accents prior to applying the accent-boost rules.

[25]Statement 3 and 4 in Rule II model certain aspects of observed emphatic intonation contours for Speaker SB, in particular, statement 3 creates convex falls.

accents can be applied to other positions in an utterance, involving pre-nuclear accents. Another important aim of the experiment is to show that perceptual judgements are based on 'natural' emphasis and not by association to unrelated characteristics in the speech signal.

Experiment II applies two settings of the boost model in various positions in an utterance and provides listeners with a task which indirectly relates to the presence and realization of emphatic accents. The task in Experiment II is to judge utterances with respect to their appropriateness as answers to particular questions.

## 4.4.3.1 Corpus design and collection

The requirement of this corpus is to provide a comprehensive framework for evaluating the performance of the identified emphasis rules and parameters in generating perceptually adequate emphasis in various positions in an utterance. In addition it is aimed at allowing perceptual experiments which provide a more discrete task than that used for Experiment I, by not asking subjects directly about emphasis, but by providing contexts which may require emphasis to fit these contexts, and moreover, may require different degrees of emphasis to be adequate for a given context.

A question and answer scenario seemed the most appropriate way of fulfilling these various criteria. The question would provide the necessary context in terms of which accent position would receive emphasis and which degree of emphasis or emphaticness was required to provide a suitable answer to the particular question. The degree of emphaticness required in the answer was varied by providing three types of questions:

A)       General questions such as:      What did Larry do?

B)       Specific questions such as:      What did Larry run towards?

C)       Contrastive questions such as:      Did Larry run towards the valley?

In all cases the answer was specified as:    "Larry ran towards the mountain".

It was hypothesized that the degree of emphasis on the first syllable in "mountain" would increase through the conditions with a contrastive question provoking the highest degree of emphasis on the nuclear syllable. This hypothesis is supported by a tendency for higher F0 maximum values to occur on the nuclear syllables in the contrastive scenario as compared to the specific scenario.

Ten base sentences were designed which would allow duration rules to operate on short vowels, long vowels and diphthongs, as well as on syllable initial continuant consonants. Furthermore, in each sentence, three accent positions were available for manipulation. The following example shows the positions of the continuant consonants in bold italic script for one of the base sentences:

23      Be*l*inda *f*ollowed the *th*ief

For each base sentence, eight versions were recorded. The first three versions were recorded in isolation and were not as such answers to context-questions. One of the base sentences is used below to show which versions were created. For a listing of all versions for the 10 base utterances see Appendix I.

**Version 1:**    Base sentence recorded in isolation prior to all other recordings:

             For example, "The solicitor arranged the loan"

**Version 2:**    Base sentence with a textual emphasis marker on the 'default' nucleus:

For example, "The solicitor arranged the LOAN"

**Version 3:** Base sentence with a textual emphasis marker on either of the other accent positions:

For example, "The SOLICITOR arranged the loan"      or

"The solicitor ARRANGED the loan"

Versions 4 to 8 were recorded after the first three versions by prompting the recording subject with a question and having the utterance read as if it were a direct answer to that question. These utterances are referred to below as "real answers" and it is mainly these real answer versions which are compared to versions containing rule generated emphasis. Versions 4, 5 and 6 relate to the position of the nuclear accent, Versions 7 and 8 relate to one of the other accent positions, asking specific and contrastive questions (respectively) relating to either the first accent position (P1) or the second accent position (P2). All 10 base utterances were manipulated on the nuclear position, five of those also contained a condition which manipulated P1, the other five also contained a condition which manipulated P2.

**Version 4:** Direct answer to a *general* question such as "What did the solicitor do?"

"The solicitor arranged the loan"

**Version 5:** Direct answer to a *specific* question such as "What did the solicitor arrange?"

"The solicitor arranged the loan"

**Version 6:** Direct answer to a *contrastive* question such as "Did the solicitor arrange the <u>account</u>?"

"The solicitor arranged the loan"

**Version 7:** Direct answer to a *specific* question such as "Who arranged the loan?"

"The solicitor arranged the loan" or

"What did the solicitor do with the loan?"

"The solicitor arranged the loan"

**Version 8:** Direct answer to a *contrastive* question such as "Did the <u>bank manager</u> arrange the loan"?

"The solicitor arranged the loan" or

"Did the solicitor <u>refuse</u> the loan?" "The solicitor arranged the loan"

The central comparisons of the experiment are the comparison between the real question answers for specific questions (weaker emphasis) versus the neutral answer boosted by the mean emphasis rules, and the real question answers for contrastive questions (stronger emphasis) versus the neutral answer boosted by the stronger emphasis rules.

All utterances were recorded by a female subject in a soundproof recording chamber at 10 KHz. using two channels, one channel containing the speech signal, the other a laryngograph signal.

4.4.3.2 Stimulus preparation

All eight versions of the ten base sentences were resynthesized once using the same LPC resynthesis suite as for the data in chapter 3, again running under Waves™. Resynthesis was again carried out using stylized straight line contours checked for perceptual equivalence with the original with respect to its intonation, as indicated above. This resulted in 80 resynthesized utterances based on the original readings which could be compared to utterances with rule generated emphasis.

The utterances containing rule generated emphasis were then resynthesized

as follows. One of the eight versions of each base sentence had been recorded as an answer to a neutral question (Version 4), as the following two examples show:

24    What did Larry do?        Larry ran towards the mountain

25    What did the rhino do?    The rhino amazed the foreigners

These 'neutral' answers with their stylized intonation contours were used as the base utterances which were boosted by the emphasis rules described in sections 4.4.2.2 and 4.4.2.3 above.[26] The following procedure was used to boost the accents on the base utterances:

A)    The utterance was hand segmented using Waves™ facilities in order to determine the durations of the relevant vowels and continuant consonants and to determine the frames in the speech analysis file making up these segments.

B)    The amount of lengthening was calculated for the relevant segments using the boost values contained in Rule I and Rule II, converted into the corresponding number of sample points, and divided by the number of frames making up the segment.

C)    The fixed frame analysis file was then edited by adding the desired number of sample points to each frame[27].

D)    The target specifications of the previously created stylized contour for the base utterance was then changed by manipulating the F0 peak height and peak position. The peak position was shifted right depending on the amount of material inserted by the duration manipulations.

---

[26]For the precise parameter value changes that were made to create emphatic versions, the interested reader is referred to Appendix J.

[27]The segments were therefore uniformly lengthened with the exception that no lengthening was added to the first and the last few frames of each segment were the speech is less stable than in the centre of the segments.

Using two boost settings for mean and strong emphasis, four rule generated emphatic versions were obtained for each of the 10 utterances. One mean and one strongly emphatic original nucleus version (example 26), and one mean and one strongly emphatic originally pre-nuclear accent either in first position (example 27) or in second position (example 28). Strong emphatic versions are marked in brackets and are surrounded by asterisks.

26    Rob annoyed AMANDA (*AMANDA*)

27    ROB (*ROB*) annoyed Amanda

28    Rob ANNOYED (*ANNOYED*) Amanda

In addition, two more versions were created for the first accent position (P1), by taking the first word of the utterance containing rule generated emphasis and splicing it into the 'neutral' answer from which that word had previously been removed. This provided a condition which would only test the accenting rules for this position rather than testing both the accentuation and deaccentuation rules which operated in the fully rule generated version.

A total of 130 resynthesized utterances available for comparison were created. The following section discusses which of these were used to test the performance of the emphasis rules which are aimed at creating perceptually adequate emphasis for various positions in an utterance.

## 4.4.3.3 Methodology

The experiment was carried out in two sessions containing a total of three blocks of pairwise comparisons. The first session consisted of one block of 80 pairwise comparisons involving only the last possible accent position in the utterance. This session was completed by 10 subjects providing a total of 800 judgements. The

second session was split into two blocks, one block of 20 pairwise comparisons concerned with judgements on the first peak in the utterance (P1), and one block of 10 pairwise comparisons concerned with the second peak in the utterance (P2). This session was completed by 10 subjects, eight of which had taken part in the first session, providing a total of 300 judgements.

Thus, each block involved decisions on accents in only one position in the utterance. Separating the experiment into blocks for each accent position would minimize subjects decisions on the correct *placement* of the accents and maximize on judgements concerning the *realization*, i.e. the relative strength or 'emphaticness' of accents. Furthermore, it was felt that a pairwise comparison provided a clear cut task with a clear choice between stimuli, providing the toughest possible test for the emphasis rules.

A sequence of randomized numbers determined the order of presentation of the pairs in each block of the experiment. Each pair was presented once with a gap of six seconds between each pair, which would allow time for subjects to make their judgement. A short beep signalled the start of each new pair.

The following categories of pairs were available for comparison in Session I for each of the 10 base sentences, resulting in 80 pairs. The order of presentation was randomized across the 10 base sentences.

A)    The real question-answers (both for specific and contrastive questions) versus an isolated non-answer recording. This is a preliminary comparison to determine whether subjects would be able to distinguish at all between utterances that are answers to questions and ones which are not.

B)    The real question answers (both for specific and contrastive questions) versus the neutral answer to the general question. This is also preliminary,

although intuitively more difficult than the comparison in A, and would show whether subjects are able to distinguish a question-answer with a nuclear accent on the last accented syllable from contextualized answers with emphatic nuclear accents in that position.

C)    The real question answers for specific questions (weaker emphasis) versus the neutral answer boosted by the mean emphasis rules    and

the real question answers for contrastive questions (stronger emphasis) versus the neutral answer boosted by the stronger emphasis rules. **This is the central comparison in the experiment and would show whether rule generated emphasis is good enough to be confusable with the real answer to the context providing questions.**

D)    The neutral answer boosted by mean emphasis rules versus the neutral answer boosted by strong emphasis rules tested in both specific and contrastive contexts. This would show which emphasis setting subjects prefer and which is more appropriate for the given contexts.

The following categories of pairs were available for comparison in Session II. Five of the 10 base sentences provided for an analysis of P1, the other five tested accents on P2. The order of presentation was randomized across the five base sentences for each peak.

**Block I (P1):**

A)    The real question answers for specific questions (weaker emphasis) versus the neutral answer boosted by the mean emphasis rules    and

The real question answers for contrastive questions (stronger emphasis) versus the neutral answer boosted by the stronger emphasis rules. This tests

for P1 if the emphasis settings create emphasis which is natural enough to be confusable with the real answers. This condition also tests the effectiveness of the deaccentuation rules specified in section 4.4.2.3 above

B)    The real question answers for specific questions (weaker emphasis) versus the emphatic word with mean rule based emphasis spliced into the real answer                                                           and

The real question answers for specific questions (strong emphasis) versus the emphatic word with strong rule based emphasis spliced into the real answer. This condition tests the emphasis rules for P1 but filters out the deaccentuation rules which will provide evidence for the effectiveness of only the emphasis rules for this position

**Block II (P2):**

A)    The real question answers for specific questions (weaker emphasis) versus the neutral answer boosted by the mean emphasis rules        and

The real question answers for contrastive questions (stronger emphasis) versus the neutral answer boosted by the stronger emphasis rules. This condition tests whether it is generally possible to use the boost settings identified above together with deaccentuation rules in this position of the utterance.

In the pre-amble to the experiment subjects were told that the intonation of an utterance often determines whether that utterance is an appropriate one for a given context. They were then given an example of an answer to a specific question and an answer to a contrastive question and were told that the amount of emphasis placed on certain words often determines how appropriate the answer is.

After having listened to three example pairs for priming, subjects were asked to listen to all the utterance pairs in turn and decide *which of the two answers in each pair sounded the more appropriate way of saying the answer to that particular question*. For each pair, the question was listed together with the answer of which they would be hearing two versions. They were asked to indicate their choice in a box adjacent to the question-answer sequence on their instruction sheets. Full instructions for this experiment are listed in Appendix K.

The following summarizes the design of Experiment II:

I. The purpose of the experiment is to test emphasis rules applied to various positions in an utterance for perceptual adequacy. The position of the 'default nucleus' is central to the investigation but rules are also tested on P1 and P2 to see if they can be applied in other positions.

II. Emphasis rules are applied to fit contexts provided by various specific and contrastive questions, the answers to which require different positions and degrees of emphasis.

III. It was thought that the toughest test for the set of emphasis rules would be to compare utterances with rule generated emphasis *directly*, in a pairwise scenario, with the "real answers" to the context providing questions.

4.4.3.4 Results and discussion

Graphs 12, 13 and 14 present the results from Session I of the experiment which compared answer versions which differ in the way the 'default nucleus' is emphasized. This position is shown in utterance 29.

> 29    Rob annoyed aMANda

**Graph 12:**

### Comparison of real answers over other answer versions
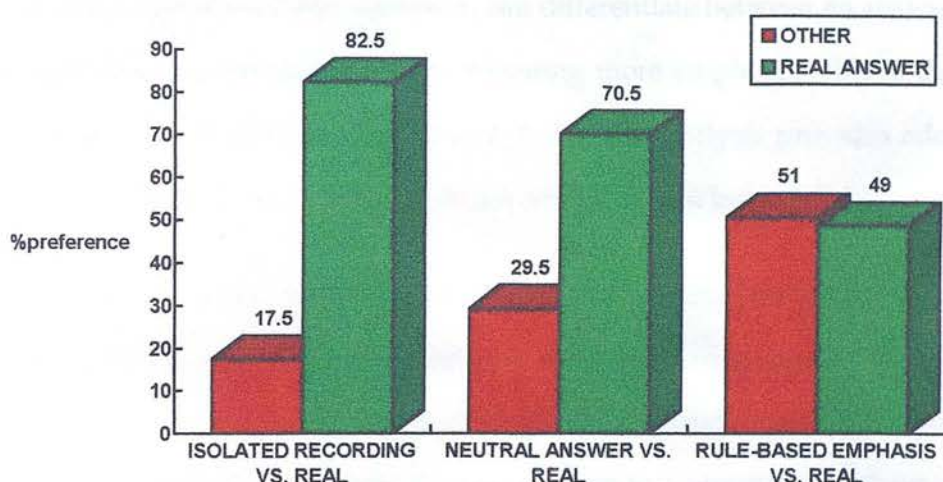


The results show that in a comparison between real answers to a given question and an isolated recording, the real answers are judged overwhelmingly to be more appropriate than the isolated recordings. The real answers were preferred over the isolated recording in 82.5% of all cases (t=8.67, .001). Comparing the real answers to emphasis triggering questions such as "Who did Rob annoy?" with the answer to a more general question such as "What did Rob do?" it can be seen that the real

answers to the emphasis triggering questions are significantly preferred over the answers to the more general question (t=4.12, .001). This holds both for specific (t=3.94, .01) and contrastive contexts (t=2.96, .02). In essence, this shows the difference between the appropriateness of nuclear accents and more emphatic nuclear accents similar to those compared in Experiment I.

More importantly, however, Graph 12 clearly shows that if the answers to the more general question are boosted by emphasis rules on the relevant accent, subjects are unable to distinguish the real answers from the boosted neutral answers with respect to their appropriateness (t=-0.45). Graph 12, therefore, provides evidence for the general perceptual adequacy of the rules boosting nuclear accents (both settings), by establishing that subjects can differentiate between an answer and a non-answer and moreover, can differentiate between an answer and a more appropriate answer for contexts requiring more emphasis. This is the main and most significant result in Experiment II. Further analysis provides additional evidence for the perceptual adequacy of the nuclear accent boost rules.

Graph 13 shows the performance of the two different boost settings by comparing on the one hand, utterances with rules generating mean emphasis with real answers to *specific* questions, and on the other hand, utterances with rules generating stronger emphasis, with the real answers to *contrastive* questions.

The results indicate that listeners tended to prefer the real answers to the ones containing rule generated emphasis for specific contexts in 55% of all cases (t=1.86, not significant), and that they tended to prefer the answers containing rule generated emphasis over the real answers for contrastive contexts in 57% of cases (t=-1.65, not significant). Both of these preferences are not significant providing further evidence for the adequacy of the rule generated emphasis. Those tendencies may also suggest that the mean emphasis rules applied for the specific questions

create emphasis which is considered either too strong or too weak for this context and position in the utterance. Furthermore, the emphasis put onto the real answers for the contrastive cases by the speaker may be either too strong or too weak for this position in the utterance. This provides further evidence for the perceptual adequacy of the emphasis generated by rule, especially for the strong emphasis applied in a contrastive scenario.

**Graph 13:**
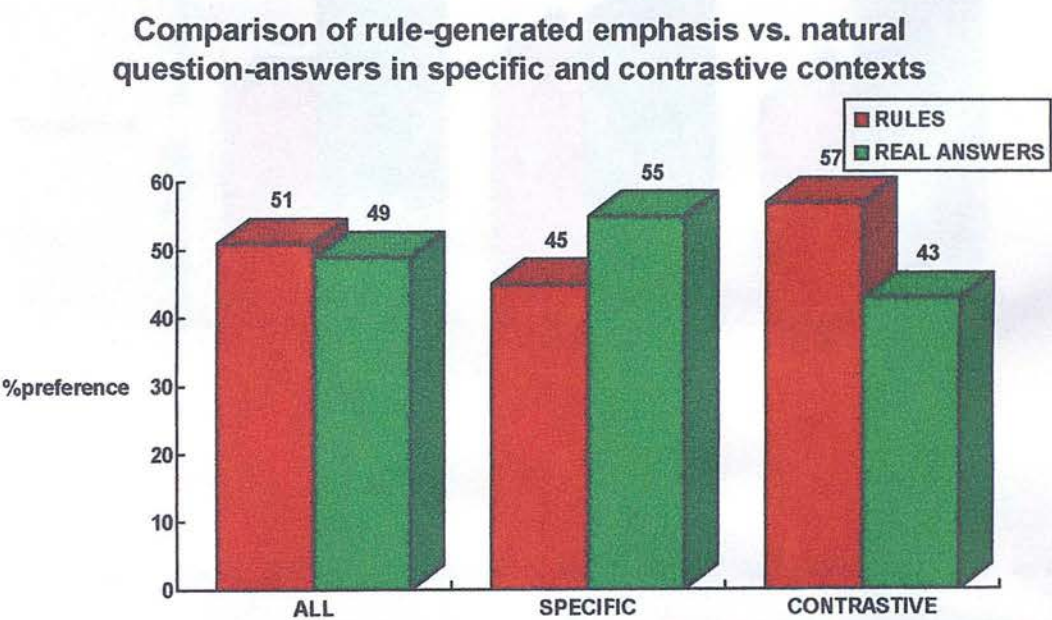


Graph 14 shows the results of the comparison between the neutral answer versions boosted by mean emphasis rules and the same versions boosted by the strong emphasis rules. Hence, this is a direct comparison of the two boost settings in both specific and contrastive contexts.

Although this comparison shows a general preference of the strong boost

settings over the mean boost settings across contexts, the results are not statistically significant (t=1.15).

**Graph 14:**

## Comparison of rule-sets in specific vs. contrastive contexts
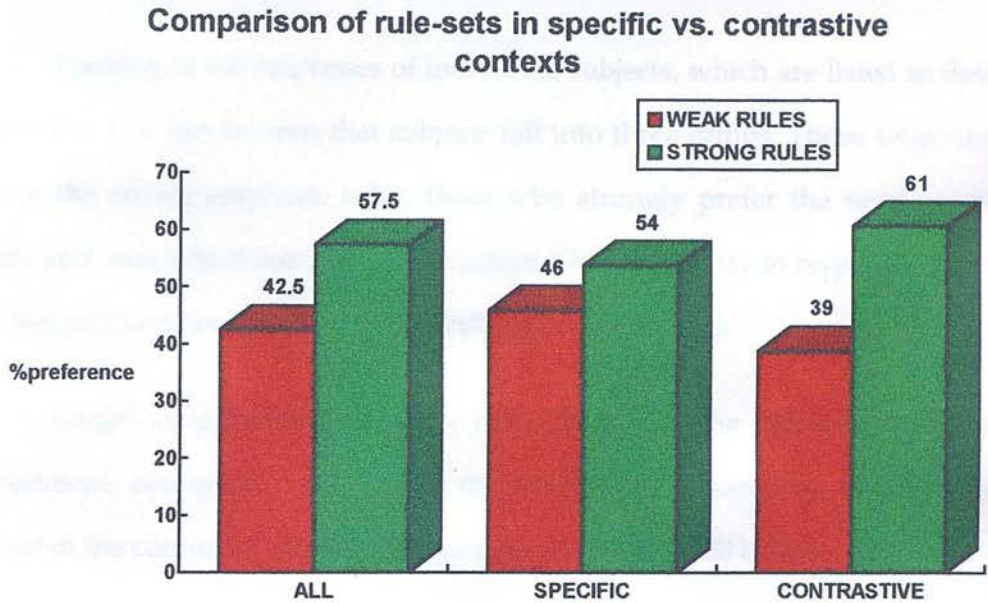


Examining in more detail the differences between the two contexts, there is a stronger preference of the strong rules over the weak rules for contrastive contexts (61% strong rules, 39% weak rules) than for the specific contexts (54% strong rules, 46% weak rules). This shows that with respect to the rules for weaker emphasis, there is a tendency for the strong rules to be judged more appropriate for contrastive contexts than for specific contexts. This tendency however is not statistically significant (t=1.07).

Nevertheless, this may reflect different requirements of the two contexts with respect to different degrees of emphasis and may suggests that a significant

difference between the rule sets for specific contexts in favour of the rules creating strong emphasis would have indicated that the preference of the real answers over the mean boost settings in Graph 13 was due to the rule generated emphasis being too weak. It may even suggest that the strong rules were not strong enough for the contrastive contexts, although these were more than adequate when compared with the real answers (Graph 13)

Looking at the responses of individual subjects, which are listed in detail in Appendix L, it can be seen that subjects fall into three camps. Those who strongly prefer the strong emphasis rules, those who strongly prefer the weak emphasis rules, and ones which seem to prefer neither. This bi-polarity in responses results in the insignificant preference of strong rules over weak rules.

Graph 15 presents the results from Block I of the second session of the experiment, concerned with testing the accent-boost model on the penultimate accent in the corpus utterances. This position is shown in 30 below:

30    The minister aLARmed the layman

In this case a pre-nuclear accent becomes an emphatic nuclear accent after the application of the accent-boost rules. The boost settings that were tested for this position in the utterance are identical to the ones applied above. In addition to the boosting rules, a deaccentuation rule operated on the last accent of the utterance. Hence, this condition tests the adequacy of accentuation as well as deaccentuation rules.

In contrast to the results presented in Graphs 12, 13 and 14, the performance of the individual boost settings is now reversed. For this position in the utterance the rules generating mean emphasis are superior to the rules creating strong emphasis, although the settings which were used were the same as for the final

accent position.

Graph 15:

**Comparison of rule-generated emphasis (including deaccentuation) vs. natural question-answers in specific and contrastive contexts for P2**
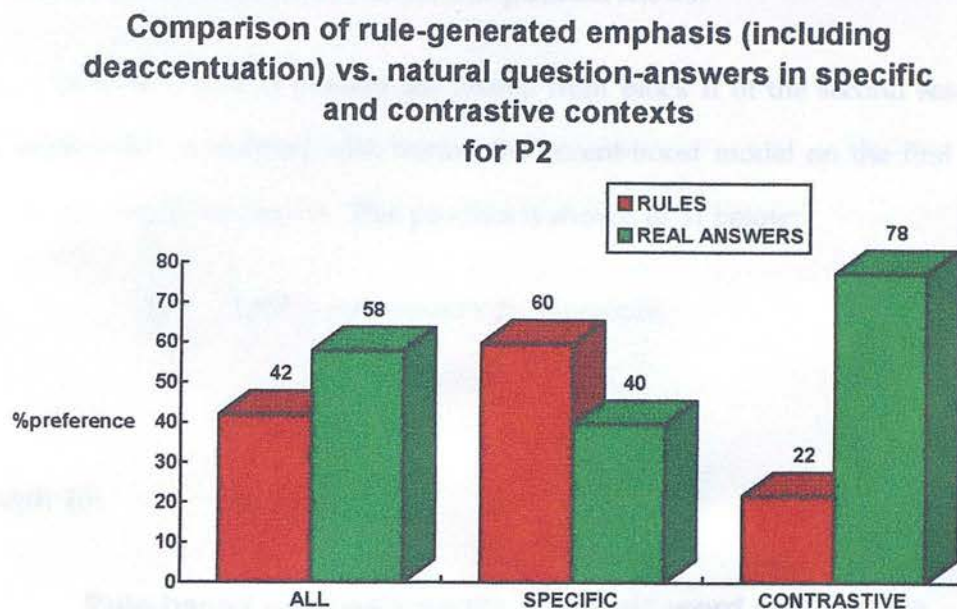


Mean emphasis rules are preferred in 60% of cases over the real answers in a specific context, as provided by a question such as, for example "What did the minister do to the layman?" This preference of the rule generated emphasis over the real answers is not significant (t=-1.93), showing that the mean emphasis rules provide perceptually adequate emphasis for this position in the utterance.

In contrast, the strong emphasis rules this time failed to generate perceptually adequate emphasis and were significantly less preferred (t=6, .01) than the real answers to a contrastive context provided by a question such as "Did the minister reassure the layman?".

The difference in the results between boosted accents in final- and boosted

185

accents in penultimate positions in the test utterances suggests that although rule generated emphasis (both settings) is perceptually adequate or even preferred to the real answers in final position, boost-settings for prefinal accents have to be carefully controlled, as the difference between the success of the specific rules and the failure of the contrastive rules for this position shows.

Graphs 16 and 17 present the results from Block II of the second session of the experiment, concerned with testing the accent-boost model on the first accent (P1) in the corpus utterances. This position is shown in 31 below:

> 31    LARry ran towards the mountain

**Graph 16:**

**Rule-based emphasis on the emphatic word spliced into the real answer vs. natural question-answers in specific and contrastive contexts for P1**
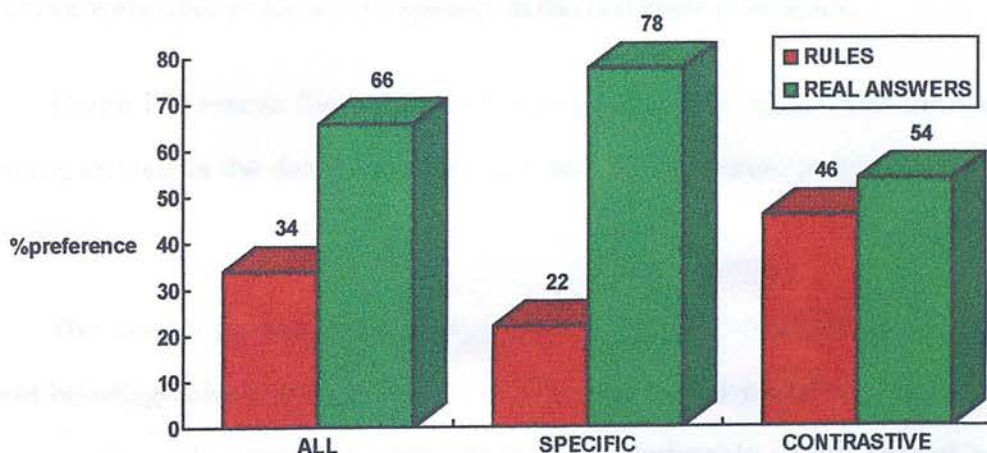


Again, a pre-nuclear accent becomes an emphatic nuclear accent after the application of the accent-boost rules. The boost settings that were tested for this

position in the utterance were again identical to the ones applied above. In addition to the boosting rules, deaccentuation rules operated on the remaining accents of the utterance, as specified in section 4.4.2.3 above. However, in order to provide a condition which tests the accentuation rules independently from the deaccentuation rules the accented word was spliced into the real answer to the question in one condition. The results are presented in Graph 16.

Graph 16 shows clearly that the version with the word containing mean emphasis generated by rule, spliced into the real answer, was not considered perceptually adequate (.02) when compared to the real answer to a specific question such as "Who ran towards the mountain?".

The splicing technique could have been at fault if it were not for the fact that for contrastive contexts set by questions such as "Did <u>Bill</u> run towards the mountain?", spliced versions were judged not significantly inferior to the real answers. A different explanation may be that there was a mismatch in the relationship between the boosted accent and the way the remaining accents in the utterance were deaccented by the speaker in the real answer versions.

Graph 17 presents the results for the condition which tested both the accent-boosting as well as the deaccentuation rules for the first accent position in the test utterances.

The results provide further evidence for the perceptual adequacy of the accent-boosting rules specified as part of the accent-boost model of emphasis. In this condition, rule generated emphasis is again confusable - with respect to its appropriateness - with the real answers to questions setting specific and contrastive contexts. Moreover, both boost-settings seem to be adequate for creating the necessary emphasis for their respective contexts. In addition, the deaccentuation

rules seem to be making the necessary changes to the remaining accents in the utterance.

**Graph 17:**

## Comparison of rule-generated emphasis (including deaccentuation) vs. natural question-answers in specific and contrastive contexts for P1
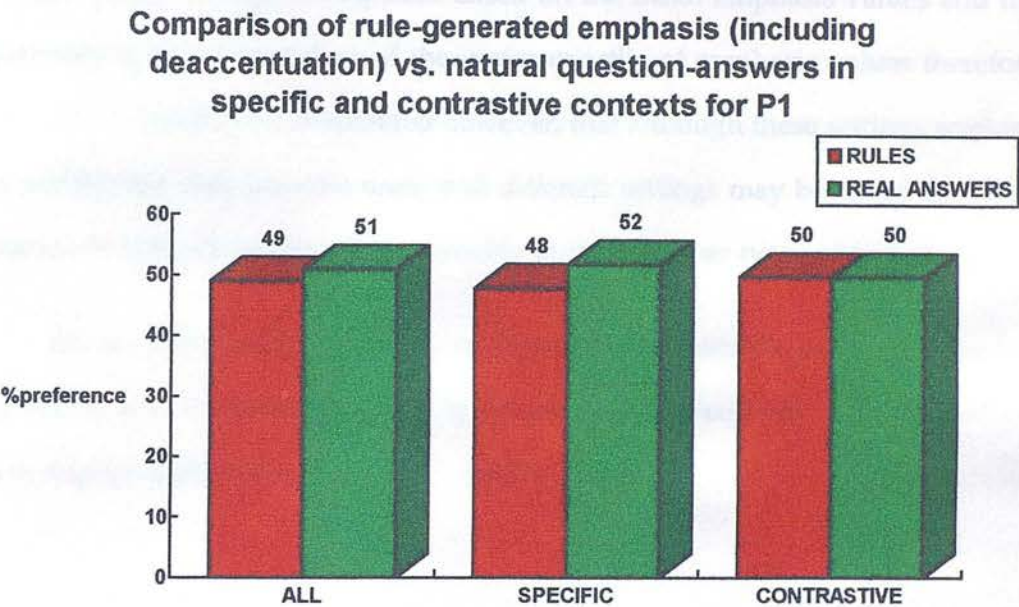


4.4.3.5 Conclusions

Experiment II has shown that perceptually adequate emphatic accents can be created by application of a set of rules manipulating various parameters of emphasis. This was shown by boosting the answers to general questions, which had been shown to be inadequate when compared to real answers. Once boosted, the resultant emphatic versions, however, were confusable with the real answers in terms of their appropriateness as answers to the context setting questions. (Graph 12).

Strong rules were judged to be more appropriate for boosting nuclear accents (Graphs 13; 14) whereas rules creating mean emphasis were judged to be more appropriate for penultimate, pre-nuclear accents (Graph 15). Both settings were equally successful for the first accent in each utterance (Graph 17).

The overall approach summarized in the accent-boost model, together with the selected test settings of emphasis based on the mean emphasis values and the higher setting based on values of the upper quartile of emphasis values therefore seem valid. It should be remembered however, that although these settings worked, they are not the only possible ones, and different settings may be tested for other scenarios of contrast and emphasis, possibly involving other types of accent.

We are therefore in a position to state that it is possible to create adequate emphasis at will, in various positions in an utterance, appropriate even for contexts which require subtle differences in emphasis.

## 4.5 Conclusions

This chapter was concerned with a detailed acoustic analysis of textual prosodic markers which may have various functional characteristics such as signalling contrast, warnings, annoyance or general extraordinarity, and which all seemingly result in some form of acoustic emphasis. A broad approach was taken which would not subdivide between more or less subtle distinctions in function, but would carry out a careful and meticulous analysis of the acoustic parameters which may operate in the generation of acoustic emphasis in readings of sentences containing textual emphasis markers.

The identified acoustic parameters of pitch accent target-height, to some

extent accent target-position, together with various duration parameters, were then applied directly to create emphatic nuclear accents from nuclear accents and tested in a perception experiment. This experiment provided enough evidence to conclude that it was possible to create a percept of emphasis using the mentioned parameters (Experiment I).

A more ambitious task was then set upon, which would test whether an algorithm could be applied which would be able to create emphatic accents at will rather than by direct transfer of parameters which is not easily repeatable by automatic application. It was expected that the most likely success of the algorithm would be its ability to generate emphatic nuclear accents from nuclear accents because the analyzed parameters were largely based on an analysis of emphatic nuclei.

The accent-boost model was specified, based on the general philosophy adopted in this chapter which was to look in detail at local differences between emphatic and non-emphatic accents. Two sets of rules were identified, one set for creating emphatic nuclei on otherwise nuclear accents, and one set for generating emphatic nuclei on otherwise pre-nuclear accents, implying a shift of the nucleus to the left and deaccentuation of the remaining accents. A further aim, therefore, was to investigate whether a percept of emphasis could be generated in various positions in an utterance.

However, the main goal of the second suite of experiments was to show that the generated accents would be perceived as perceptually adequate for certain contexts. This was investigated in a framework which would provide contexts by using the functional notions of *specificity* and *contrastiveness* in a question and answer scenario. These contexts were thought to require differing degrees of emphasis which would allow the testing of two different accent-boost settings.

Experiment II clearly showed that it is possible to create perceptually adequate emphasis at will, by the application of rules specified within the accent-boost model. The experiment further showed that adequate emphatic accents could be generated at various positions in an utterance and further evidence for the validity of the model was provided by the fact that different boost-settings are required for different accent positions within an utterance.

We can conclude that the careful acoustic parameter analysis of readings containing textual emphasis markers has allowed the specification of a model containing a set of rules which are able to generate perceptually adequate emphasis for specific and contrastive contexts for various accent positions in an utterance.

# Chapter 5

# Conclusions

The aim of this thesis was to show that certain punctuation devices, symbolic markers and type-face settings have prosodic correlates in read speech which can be analyzed and quantified, which can be applied in speech production and which are important for the perception and interpretation of speech.

A working hypothesis was adopted by which symbolic markers such as italic- or bold script, capital letter spellings, emoticons and more traditional punctuations such as parenthesis or quotes would be considered *prosodic markers*. Evidence for the applicability of this working hypothesis was gathered by presenting a traditional argument concerning the linguistic status of written- as opposed to spoken language, which was particularly concerned with the question of whether punctuation is a direct 'transcription' of prosodic events or whether there is merely a general correspondence between punctuation and prosody.

The thesis then turned to the examination of a relatively new language genre, which we named *Written Conversation*, where a "transcriptionalist" argument was presented which sees a direct correspondence between certain textual markers and their prosodic form - especially with respect to emphasis markers such as bold print or capital letter spellings. This transcriptionalist stance was supported by establishing firstly that written conversation, which includes electronic written communications such as electronic mail, electronic news- or discussion groups, as well as texts on the World Wide Web, combines forms from both the written and

the spoken language, rendering the genre a hybrid between the two main language forms. Evidence from statistical text analyses carried out by Yates (1992a; 1992b) was used to argue that certain forms of writing are more 'spoken' in linguistic character than others, an observation which applies strongly to these Internet communications.

The strongest argument for the prosodic status of certain punctuation devices, however, came from the recognition of the discrepancy between the medium (writing) and the 'spoken' character of the contents of these writings. It was argued that authors find it necessary at the time of writing to counterbalance the lack of certain devices available in the spoken language for the expression of meaning, such as gesture and intonation, by inventing and inserting textual devices into their texts. The main motivation, it was suggested, lies in the avoidance of misunderstandings brought about by a difference between intention and interpretation.

This transcriptional hypothesis was further underlined by consulting a language production model by Levelt (1989) which makes reference to an internal representation of speech available for monitoring and immediate modification preceding the eventual spoken output. It was argued that this internal representation of speech was also available in the language generation process which produces written rather than spoken output. Levelt's model was adapted into a "blueprint for the conversational writer" which provided a possible framework for referring to prosodic markers and for the "inner voices" or "acoustic images" which transcriptionalists referred to when arguing in favour of the direct mapping between prosody and punctuation.

It was argued that the analysis of textual markers of prosody would be of benefit to certain areas of phonological and phonetic theory which had received

only marginal attention, either because the phenomena were considered to be understood or because they were left aside in favour of the description of the more regular events. It was suggested that the interest in the analysis of certain textual markers lies in their function as indicators of expressive *choice*, which is achieved by pointing to an intended interpretation of an utterance which may deviate from the expected interpretation and is therefore *marked*, or by enforcing or enhancing a particular interpretation even if it corresponds to the expected, *unmarked* interpretation.

A further motivation for determining the acoustic correlates of such markers came from the field of Speech Technology. In particular, knowledge of the acoustic correlates of important information - signalled through emphasis, and less important information - signalled for instance by parenthesis, would be beneficial to the task of key-word spotting, especially in the absence of full automatic speech understanding. The immediate as well as the longer-term benefits for automatic Text-to-Speech conversion were also discussed, by arguing that through the acoustic analysis of textual prosody markers more expressive and more conversational prosody could be achieved for texts which contain these markers. In the longer term, knowledge of the mapping between linguistic concepts and their acoustic form was considered vital, especially when more sophisticated linguistic analysis of unrestricted text becomes available.

The remainder of this thesis (chapters 3 and 4) was then concerned with analysing the acoustic correlates of the two most frequently occurring textual markers in written conversation - parentheticals and emphasis markers. The aim was to show that these markers do have prosodic correlates which speakers manipulate when reading aloud texts which contain such markers. This was not always straightforward as the collected corpus contained real language usage, with

most of the utterances that were analyzed having been written by different authors. As a result, the analysis of the prosodic correlates had to take into account the complex variety of occurrences, both with respect to the types of the markers used as well as the positions in the utterances in which they occurred. The general methodology adopted was to present subjects with texts for reading aloud which contained textual markers, and the same texts from which these markers had been removed, in order to be able to measure the changes that speakers made on encountering the particular devices.

For the investigation of parentheticals a corpus of some 120 utterances was recorded and analyzed in order to provide measurements of their acoustic correlates and furthermore, to investigate the claim that these constructions are somehow "prosodically independent", set aside by pausing and pitch range effects and showing no influence on their surrounding clauses.

Two perception experiments which tested whether subjects could detect if a parenthetical had been removed from within an utterance demonstrated the prosodic independence of some parentheticals but not others. A detailed acoustic analysis of the recordings subsequently revealed factors by which parentheticals can be argued to be prosodically independent and factors by which they are not independent. The former include pre- and post-parenthetical pausing and pitch range compression within the parenthetical clauses. The factors by which some parentheticals were not prosodically independent mainly related to effects on their surrounding main clause fragments. In particular, it was demonstrated that parenthetical clauses constitute an environment for pre-clausal lengthening on the final syllables of the first main clause fragment, and that utterance final parentheticals do not trigger preceding utterance final intonation. All these acoustic

correlates were quantified in detail and can easily be added to existing rules for text to speech conversion.

In chapter 4 the thesis turned to the examination of textual markers of emphasis and their acoustic correlates. For this a corpus of some 150 utterances was recorded mainly for two speakers. Some of the acoustic parameters which were thought to play a role in the generation of acoustic emphasis were analyzed for the scale and range with which the speakers used these parameters to signal emphasis in utterances which contained emphasis markers. Acoustic correlates were quantified for the parameters, F0 maximum, F0 maximum position and the duration parameters of syllable duration, vowel duration and the duration of pre-vocalic continuant consonants.

A first perception experiment then established that these emphasis parameters superimposed on non-emphatic syllables would change subjects' perception from unemphatic to emphatic. This provided enough evidence for specifying an emphasis model which was designed to be able to apply emphasis, in principle, to any word in an utterance. This *Accent-Boost-Model* specified the changes necessary to the F0 and duration parameters for producing emphatic accents based on a statistical analysis of the range with which the readers used each individual parameter.

A second series of perception experiments tested the validity and generative power of the accent-boost model which was supplied with two emphasis settings for all its parameters - one designed to create strong emphatic syllables, the other designed to create more modest emphasis. A question and answer scenario was designed which required subjects to judge on the appropriateness of certain answers to particular questions. The questions were designed as to require different amounts of emphasis on particular syllables. For example, it was thought that the

last syllable in an answer such as "Larry ran towards the mountain" required stronger emphasis to fit a contrastive question such as "Did Larry ran towards the valley?" than to fit a more general question such as "What did Larry do?" or to fit a specific question such as "What did Larry run towards?" Pairwise perception experiments were carried out which directly compared utterances containing emphasis generated by application of the accent-boost model with utterances which had been recorded as direct answers to these questions. This scenario allowed for the investigation of three main questions.

1. Was the statistical characterization of the ranges of the investigated parameters powerful enough to provide answers to emphasis triggering questions which were considered as appropriate as the real answers?

2. Were the implemented emphasis settings appropriate to signal subtle differences in required emphasis?

3. Was it possible to apply the accent-boost-model in various positions in an utterance and, were the implemented emphasis settings appropriate for these different positions?

The central finding was that the subjects found the answers which contained rule-generated emphasis as appropriate as the real answers to the given questions. Furthermore, through a comparison of the two boost-settings applied in the model, it was shown that the strong emphasis rules were judged to be more appropriate for boosting nuclear accents, whereas rules creating weaker emphasis were judged to be more appropriate for penultimate, pre-nuclear accents. Both settings were equally successful for the first accent in each utterance. This also demonstrated that both the weak and the strong boost-settings were able to convey subtle differences in meaning, appropriate for specific and contrastive scenarios respectively.

From this it was possible to conclude that the careful analysis of textual emphasis markers had allowed the specification of a model containing a set of rules which were able to generate perceptually adequate emphasis for specific and contrastive contexts for various positions in an utterance. Again, it would be a straightforward task to recognize emphasis markers in running text by means of a text pre-processor, and applying the necessary changes for creating acoustic emphasis for automatic Text-to-Speech conversion by applying the accent-boost model.

The results for both parenthetical constructions and emphasis markers show that it was possible to achieve a thorough analysis of the acoustic correlates of these textual markers in read speech, even accounting for their complex patterns of occurrence in a corpus of real language usage. Despite having achieved this for the two most frequently occurring markers in *Written Conversation* much work still needs to be carried out.

For example, there are many more textual markers which require thorough investigation such as the quotation mark, which is beginning to take on a new meaning in that it is often used for imaginary but stereotypical quotes with an intended sarcastic undertone. Possibly even traditional punctuation marks such as exclamations or question marks should be re-examined using the methodology adopted in this thesis.

Most importantly however, we were concerned with analyzing the linguistic phenomena of parentheticals and emphasis, but electronic written communications are also full of paralinguistic markers in form of 'smiley-faces' or 'emoticons' which require detailed analysis. Their function of signalling happiness, sadness or sarcasm and their power to signal to a reader that something should not be taken as seriously as it may appear on the surface, are communicative mechanisms which

are vital and essential for the interpretation of these utterances. Although the characterization of vocal emotion has made immense progress (e.g. Scherer 1974, 1980; Murray et al. 1991; Cahn 1989; Williams & Stevens 1972; Fairbanks & Pronovost 1939; Laver 1992; Rice et al. 1987), the study of paralinguistic textual markers and their acoustic correlates may help to further our understanding in this important area.

Furthermore, it is suggested that *Written Conversation* as such is worth serious further study. First of all, it is an almost infinite corpus of language usage, millions of words and sentences flow through the Internet every single day of the year - all in machine readable form. It is a genre which combines linguistic characteristics from the written as well as the spoken language and is therefore suitable for the study of both language forms. In addition, written conversation is still in its infancy, it is a language form with which most of its users are still experimenting. Therefore, it is likely that changes will be observed constantly and that conventions will establish over time. In a way, it is the responsibility of the users of this language form to establish and obey these conventions, otherwise the genre will not reach its full communicative potential.

However, language is a living beast, its constant change and development never stops and it is therefore not inconceivable that *Written Conversation* may develop into a medium which offers the full range of expressive power that language makes available.

# Appendix A (Chapter III)

**Database of utterances containing relative clause parentheticals in their original form, with parentheticals removed (.2) and with parentheticals replaced by commas (.1).**

001 For those whose memory is worse than mine - in 1989, after upsetting "Big Mac", Haarhuis beat someone (I have a vague feeling that it was Sampras!!) but lost to Krickstein (who lost tamely to Becker then, right??). Basically, my forecast is based on my (whimsical) conclusion that the maximum number of matches that Haarhuis can last on his top-class-player beating spree is 2.

001.1 For those whose memory is worse than mine - in 1989, after upsetting "Big Mac", Haarhuis beat someone (I have a vague feeling that it was Sampras!!) but lost to Krickstein, who lost tamely to Becker then, right??. Basically, my forecast is based on my (whimsical) conclusion that the maximum number of matches that Haarhuis can last on his top-class-player beating spree is 2.

001.2 For those whose memory is worse than mine - in 1989, after upsetting "Big Mac", Haarhuis beat someone (I have a vague feeling that it was Sampras!!) but lost to Krickstein. Basically, my forecast is based on my (whimsical) conclusion that the maximum number of matches that Haarhuis can last on his top-class-player beating spree is 2.


002 Is it possible to use Waffle for DOS with a multi-user environment? I am considering using it with something like VM/386 (a serial port multi-user system) or Lantastic.

002.1 Is it possible to use Waffle for DOS with a multi-user environment? I am considering using it with something like VM/386, a serial port multi-user system, or Lantastic.

002.2 Is it possible to use Waffle for DOS with a multi-user environment? I am considering using it with something like VM/386 or Lantastic.


003 I had a pair of the original Verticals from around 1987 (the white ones with the red, orange and blue flashes). I used to take them out for powder and bumps.

003.1 I had a pair of the original Verticals from around 1987, the white ones with the red, orange and blue flashes. I used to take them out for powder and bumps.

003.2 I had a pair of the original Verticals from around 1987. I used to take them out for powder and bumps.

004 I believe what he means is the use of 'that' to introduce an embedded sentence used as a noun (usually a direct object) in the outer sentence.

004.1 I believe what he means is the use of 'that' to introduce an embedded sentence used as a noun, usually a direct object, in the outer sentence.

004.2 I believe what he means is the use of 'that' to introduce an embedded sentence used as a noun in the outer sentence.

005 Is there a general consensus on how one would pronounce the acronym SCSI? I originally heard the term (in Canada) "skuzzy" (rhymes with fuzzy), but have lately been using "skizzy" (rhymes with fizzy).

005.1 Is there a general consensus on how one would pronounce the acronym SCSI? I originally heard the term (in Canada) "skuzzy", rhymes with fuzzy, but have lately been using "skizzy" (rhymes with fizzy).

005.2 Is there a general consensus on how one would pronounce the acronym SCSI? I originally heard the term (in Canada) "skuzzy", but have lately been using "skizzy" (rhymes with fizzy).

005.3 Is there a general consensus on how one would pronounce the acronym SCSI? I originally heard the term (in Canada) "skuzzy" (rhymes with fuzzy), but have lately been using "skizzy", rhymes with fizzy.

005.4 Is there a general consensus on how one would pronounce the acronym SCSI? I originally heard the term (in Canada) "skuzzy" (rhymes with fuzzy), but have lately been using "skizzy".

005.5 Is there a general consensus on how one would pronounce the acronym SCSI? I originally heard the term (in Canada) "skuzzy", but have lately been using "skizzy".

006 A colleague in the Netherlands seems to prefer the term "skoozy" (rhymes with choosy or UZI).

006.1 A colleague in the Netherlands seems to prefer the term "skoozy", rhymes with choosy or UZI.

006.2 A colleague in the Netherlands seems to prefer the term "skoozy".

007 In between, I had inspected one of the files that looked anomalous the first time, and found that from knox it looked wrong but when seen from minto (the file's home machine) it was all right.

007.1 In between, I had inspected one of the files that looked anomalous the first time, and found that from knox it looked wrong but when seen from minto, the file's home machine, it was all right.

007.2 In between, I had inspected one of the files that looked anomalous the first time, and found that from knox it looked wrong but when seen from minto, it was alright.

008 Right after this set of full dumps, I'm going to be starting the ball moving to upgrade all the suns to SunOs 4.1.1 (the latest version of the OS) which (according to the release notes) should fix this and similar problems.

008.1 Right after this set of full dumps, I'm going to be starting the ball moving to upgrade all the suns to SunOs 4.1, the latest version of the OS, which (according to the release notes) should fix this and similar problems.

008.2 Right after this set of full dumps, I'm going to be starting the ball moving to upgrade all the suns to SunOs 4.1, which (according to the release notes) should fix this and similar problems.

009 There are corresponding non-sonorant codes files (which in fact are identical for the two accents) in the same directories (which are under Polyglot).

009.1 There are corresponding non-sonorant codes files which in fact are identical for the two accents, in the same directories (which are under Polyglot).

009.2 There are corresponding non-sonorant codes files in the same directories (which are under Polyglot).

009.3 There are corresponding non-sonorant codes files (which in fact are identical for the two accents) in the same directories, which are under Polyglot.

009.4 There are corresponding non-sonorant codes files (which in fact are identical for the two accents) in the same directories which are under Polyglot.

009.5 There are corresponding non-sonorant codes files (which in fact are identical for the two accents) in the same directories.

009.6 There are corresponding non-sonorant codes files, which in fact are identical for the two accents, in the same directories (which are under Polyglot).

009.7 There are corresponding non-sonorant codes files in the same directories.

010 Here's another problem. In Australia it is very common for neighboring speech forms (a term I will adopt as neutral between "dialect" and "language") to be mutually comprehensible, but for intelligibility to drop off rapidly as distance increases.

010.1 Here's another problem. In Australia it is very common for neighboring speech forms, a term I will adopt as neutral between "dialect" and "language", to be mutually comprehensible, but for intelligibility to drop off rapidly as distance increases.

010.2 Here's another problem. In Australia it is very common for neighboring speech forms to be mutually comprehensible, but for intelligibility to drop off rapidly as distance increases.

011 On the other hand there is equally no sign of it in the one tape (rebroadcast a couple of years back) that I possess.

011.1 On the other hand there is equally no sign of it in the one tape, rebroadcast a couple of years back, that I possess.

011.2 On the other hand there is equally no sign of it in the one tape that I possess.

012 Note that in asserting that statement P is true, one makes an assertion about the propositional content of P (since it occurs in an indirect context of oratio obliqua), rather than P itself.

012.1 Note that in asserting that statement P is true, one makes an assertion about the propositional content of P, since it occurs in an indirectcontext of oratio obliqua, rather than P itself.

012.2 Note that in asserting that statement P is true, one makes an assertion about the propositional content of P, rather than P itself.

013 The same thing is true of the International Morse Code, the flag hoists on seagoing ships, and the Dvorak Simplified Keyboard (refers to typewriters, not Bechstein pianos).

013.1 The same thing is true of the International Morse Code, the flag hoists on seagoing ships, and the Dvorak Simplified Keyboard, refers to typewriters, not Bechstein pianos.

013.2 The same thing is true of the International Morse Code, the flag hoists on seagoing ships, and the Dvorak Simplified Keyboard.

014 The remaining hulk will be 100% inoperable without any disk whatsoever (ie, it will perform marginally worse than it did previously :-) ), and indeed it may well be traded to Linguistics.

014.1 The remaining hulk will be 100% inoperable without any disk whatsoever, ie, it will perform marginally worse than it did previously, and indeed it may well be traded to Linguistics.

014.2 The remaining hulk will be 100% inoperable without any disk whatsoever, and indeed it may well be traded to Linguistics.

# Appendix B (Chapter III)

## Database of utterances containing a variety of parentheticals in their original form for Experiment II[1]

001 Isn't this the same Connors (or Jimbo) who stunned Stefan Edberg (and me) in three glorious sets just two (or is it three??!!) years back at the very same US Open Tournament?

002 My mother, no slouch as an English and Creative Writing teacher, explained that many sentences beginning with the and/or/but class of conjunctions (among her students) ended up as sentence fragments.

003 But developing a new system as exciting as the Macintosh (or even the PC :-)) is not the same thing as writing Cobol.

004 Along the way he tells of the obsession (including his) of the bodybuilding lifestyle.

005 It was the same Agassi who tanked the fourth set (6-0, for those with short memories) against the very same Connors hoping (correctly) that Connors would run out of gas in the fifth.

006 So, you buy an eighth higher than you would have (assuming enough stock available at the price), and an eighth lower on selling.

007 This is *not* intended to start a flame war (and I am not defensive), but it's funny that my parents (shrinks) always warned me about the kids of school teachers, especially of the public school variety :-)

008 This news group barely gets enough traffic as it is (I average around 5 postings a day at my site) so I don't think we have to worry about being inundated with reports from around the world.

009 Of course (and this goes to Lyn's reluctance to post her very personal spiritual experiences here) even if "we" don't speak on the Net, we're out there reading all this, and then "we" are (or at least I feel as I am) a sort of voyeur, an anthropologist.

010 My more important goal right now, though, is to create a separate newsgroup for carrying the many urgent action bulletins and other informational kinds of postings to the NATIVE-L mailing list via Usenet, and to do it in a way that permits as many sites to be reached as possible (which is not a goal that can be easily achieved via the "alt." hierarchy),

---

[1]These are presented as idividual sentences. The subject used for recording had available larger chunks of text providing the necessary context.

and so that the link is fully bidirectional (so that articles posted to the new newsgroup will get sent back to the NATIVE-L moderator for distribution to the mailng list.

011 I (or our organisation) will participate in publicising a world-wide boycott of all HYUNDAI products unless you take heed now.

012 I have several in my collection, and my favorite shows the loyal Indian volounteer (complete with turban) manning an MG-34 against those frightful Brits.

013 Sure, they would've found some crack-pot supporters (Bose wasn't the only one who believed in violence as a means to freedom), but the Gandhi/Nehru leadership of the Congress Party (which led India's freedom struggle) was fairly intelligent and sensible (at that time); I don't think they ever seriously considered exchanging one set of masters for another.

014 A discussion was had with several people down here (myself included), and we came to decide that even if Harold had won, England would have been ripe for plunder from other forces

015 At the same time, or at some later time (depending on how well people cooperate with moving dis-cussion-oriented postings to alt.native.d), I would create a link between the NATIVE-L list and alt.native.

016 If you sent a message saying, "Look I just measured an effect which seems to point to a problem in your model of physics," and you could back it up, and it didn't apparently contradict innumerable other well-documented observations to the contrary ("Hey, I just dropped an apple and it fell up"), I'd treat you with a lot more respect that you seem willing to dole out.

017 If, as is a defensible position, you choose to say that astrology does not employ scientific method (since it started before the scientific method it can't be _based_ on it), then a perfectly good riposte is to point out that the scientific method, in so far as it is understood, is simply one of the latest and most successful forms of human enquiry.

018 On the other hand, my own knowledge of these matters is abysmal (even though I'm the first generation in my family to be born Free), so if anyone feels like they know better, feel free to contradict me.

019 Myself and three others (too many) set up camp in the mountains Friday and came out Monday.

020 My favorite (for sentimental reasons, as I've done no serious study of the subject) would be Papiamentu, the language of Curacao and some nearby islands in the Caribbean, which is said to be a pleasant melange of Portuguese, Dutch and Spanish with some English and French influences.

021 And (putting on my flame-proof suit) there's really no solid reason for treating Norwegian, Danish, and Swedish as three languages. Again, it has to do with culture, history, and the preferences of the speakers.

022 With a touch of spelling reform (to remove some Dutch-isms) and a bit of updated scientific vocabulary, I imagine it would serve as well as most interlanguage proposals.

023 The most probable attack would have come from Scotland, and most likely the Scots would have claimed (or re-claimed depending on your historical viewpoint) Northumbria and York; however, they probably would not have been able to hold most of Northumbria.

024 People here ascribe even the most trivial of life occurences (dialing a wrong number, being late to an appointment) to astrological influences, with unquestioned, deterministic certainty.

025 Besides, psychiatry in the sense of psycho-analysis (which is the form which corresponds most closely to astrology) is merely a form of talking about problems with someone else.

026 If I'm wrong anywhere(and that is likely) I'd be interested to hear from anyone with good sources about a fascinating period in history.

# Appendix C (Chapter III)

**Experimental instructions: Experiment I**

## Instructions

The following two pages contain <u>fourteen </u>sentences or short paragraphs.
Each of these <u>will be read twice</u> in succession by a human female speaker and <u>presented as a pair.</u>

One of the sentences in each pair is an original unedited reading.
The other has been edited;  this editing may affect the natural flow or continuity of the intonation of the utterance.

For each pair, <u>can you tell which reading has been edited?</u>

Please <u>always</u> indicate your choice by putting the numbers 1 or 2 in the adjacent space, depending on whether you perceived the first or the second reading to be the **edited** one.

In case of a short paragraph, concentrate on the underlined sentence.

# Appendix D (Chapter III)

## Experimental instructions: Experiment II

## Instructions

The following three pages contain <u>twenty six</u> sentences.
Each of these <u>will be read twice</u> in succession by a human female speaker and <u>presented as a pair.</u>

One of the sentences in each pair is an original unedited reading.
The other has been edited; **this editing may affect the natural flow or continuity of the intonation** of the utterance.

For each pair, <u>can you tell which reading has been edited?</u>

Please <u>always</u> indicate your choice by putting the numbers 1 or 2 in the
adjacent space, depending on whether you perceived the first or the second reading to be the **edited** one.

For sentences marked with two asterisks please mark the point in the provided text, where you think the edit occurred.

Note:   - All edits have been made somewhere in the middle of the
        sentences.
        - Ignore slight natural hesitations which occur in reading.

# Appendix E (Chapter IV)

## Database of emphatic utterances

## in their original form (including typing errors)

Two recordings of all underlined utterances are available for Speakers SF and JE, one recording with, the other without emphasis markers. Two versions of the sentences numbered in underlined bold script are available for Speakers SF, SB and JE (non-underlined bold script only for SB and JE). Bold numbered utterances were used to compare nuclei and emphatic nuclei for Speakers JE and SB and SF. Sentences 093-102 are the all-voiced sentences recorded for Speaker JE.

All recordings are dual channel, one for the speech signal and one for the laryngograph signal.


001 Usually, when they plug a little transformer into the wall
it's for powering dial lights or other special features of YOUR phones
and has nothing to do with their network.

002 I have also read and heard many things from Eustace
Mullins, and I have *never* found anything that he has
said to be anti-Sematic [sic].

003 If *YOU* want to make this racist, genetic association between
yourself and the Nazis in Germany 1939-45 that's your own neurosis.

004 We review and remind ourselves of the history of these things to
understand that the CURRENT lousy situation is due to that history.

005 But ultimately you're trying to fix the CURRENT situation.

006 That's all you CAN fix.

007 You'd have to *catch* me first.

008 Dont expect to get anny where with the ticket agent, they
have *nothing* to do with issuing passes, except for getting passes for
their own people.

<u>009</u> These are the laminated ID passes members of the
band and their tour crew wear, and are generally IMPOSSIBLE for outsiders
to obtain.

<u>010</u> Not only did I enjoy the show, but you
could honestly tell *they* were enjoying the show as well.

<u>011</u> Well, a staggering response to *that* one, hahahaha!

<u>012</u> I truly want to help others, and I DON'T expect anything.

<u>013</u> Someone might be *your* support for a change.

014  I really don't want to disagree with you, but the problem with being a
NiceGuy is that, at least in my case, I am *always* counted on to be the
strong one, to advise, to console, to comfort.

<u>015</u> Have you ever *tried* being nice to people?

<u>016</u> Stop trying to change yourself into whatever *you* think other people
want, and expecting other people to be responsible for solving all
your problems.

<u>017</u> It's not easy to overcome, and in many cases it can't be done alone -
but I stress that it *can* be done.

<u>018</u> Straighten up and be a *man* for God's sake.

<u>019</u> While I am not of the opinion that people are always to blame for the
state of their lives, it is my opinion that they are *responsible* for
their lives.

<u>020</u> What he needs is to be empowered, to be shown that he
is *not* and has never been at the mercy of the world.

<u>021</u> If and when I lose weight, I will do it because *I* want to, not
because society dictates that I must.

022 Either people accept me as I am, ALL of me, or forget it.

<u>023</u> Oh, as a bartender I saw LOTS of people who did.

<u>024</u> It takes a LOT more than that.

<u>025</u> Ask yourself that question and think about it *really* carefully!

<u>026</u> In many many cases, overweight *can* be dealt with.

<u>027</u> In the last couple of years there have at last been studies done to begin and
understand the *real* physiological basis of overweight.

028 Paid counseling is *much* better at dealing with this
than the ministrations of even a non-fair-weather friend.
029 I haven't read all the responses yet, but those I have seen, I found
to be no-nonsense, but *very* supportive.

030 In order to sustain liquidity in the U.S. economy, Uncle Sam has
shifted from *borrowing* money, to *printing* money, ie, his
total borrowing is lower.

031 Maybe the *real* deficit is around $850 Billion, and he doesn't
anyone to know until he's been funded.

032 If *this* is an example of how you go about it,
I think you are doing your cause more harm than good.

033 I never said that you DID, only that if you engaged in such behavior,
it would make you a homophobe.

034 Therefore, I felt it necessary to once again emphasize WHY I'm
not a homophobe.

035 Wasn't it YOU that brought up the specific points to which I
countered (numerically) above?

036 You may not like the *way* it is being cut, but that's the price you
I pay for George Bush screwing up and doubling our debt in 4 years.

037 Er, well, no Loren; in fact, there is *no* logical connection between these
two statements.

038 The definition of mental disease is bound up *entirely* in whether or not
the diseased individual is responding to society in a statistically
normalized fashion in re the given circumstances.

039 Has *anyone* ever been hurt as a result of these activities?

040 *You* must be emotionally prepared to
realize that some parts of your worldview are inaccurate.

041 I *think* I've now got a complete list of those from CSTR going to
Eurospeech 93.

042 The highlight has been a corncrake of our own - something which is pretty
rare in the south of Scotland (maybe one or two birds in the *whole* of the
Borders region, from Edinburgh south to the border with England).

043 The English use of `Eire' is effectively a refusal to accept that the
Republic of Ireland *is* Ireland (political claims to the six counties
aside).

044 This position can be held by pro-Union people *and* by hardline Republicans.

045 Here's *another* recipe, but one that's fairly explicit, from a cookery writer who's very famous in Britain, and is usually foolproof.

046 Thanks, I was wondering whether they'd bothered giving *anyone* any information, or if they were just hoping that the general uncertainty would deter car owners from bringing their cars in at all!

047 Sorry, maybe I should have listened to him for a bit longer - long enough to get *his* telephone number.

048 Doesn't this sound like an *ad* to you?

049 It's in the class of tape decks just under the high-end home machines and sounds *much* better than the average decks.

050 Yes, this is *the* Moving Sale to catch!

051 Be sure to wash out this prized dirt, however, AFTER cutting the leeks.

052 They are GROSS when crisp.

053 A very good friend of my brother-in-law's proved to my brother-in-law that height has absolutely NOTHING to do with the skill of play.

054 PS - when I went back home (rural Northern New York) for vacation EVERY bar had those damn electronic dart boards.

055 Having a semi-auto is nice, but you don't really *need* one to play paintball or do you?

056 If you *really* want to live, move north to the *real* California!

057 If I'm the one raising them, then their MY children.

058 I avoided it on POPULARITY grounds for months, but finally was in a position where it was the only beer available.

059 Be sure to sterilise *EVERYTHING* which comes into contact with the mead and do not use ferrous materials.

060 It looked really impressive coming back due to it's center of mass and *very* fast rotation.

061 For what it is worth *my* family life was very happy, easy and pleasant.

<u>062</u> It doesn't say that I *will* go there, just that I would be the happiest there.

<u>063</u> Now, I relax, *listen* to the client with an aware ear, and let them describe the chart to me!

<u>064</u> Eusip and hume manage to clear down the lines MOST of the time but you shouldn't rely on it.
<u>065</u> Start up rn, and go a get a mug of coffee as it will take AGES to run through the list of newsgroups.

<u>066</u> The moral is that it's worth checking your codes file if things go wrong, or preferably *before* things go wrong!

<u>067</u> If you can ALWAYS win the game, no matter what the adversary does, the language isn't regular.

<u>068</u> I recently found an interesting article in OUTDOOR magazine about the two guys who are continuing to compete for the world's longets bungee jump.

<u>069</u> Please look over the following and let me know NOW if you foresee any probles with the schedule.

<u>070</u> Hopefuylly those enlightened individuals (like yourself) keep contributing, and this will be a GREAT list, not just a good one.

<u>071</u> EVERYONE has a turn chairing a session.

072 We *do* have a way to reduce problems such as this.

073 I agree with you: the subject of the verb isn't just *one* person.

074 Passwords in any of the following categories could be broken in MINUTES by a hacker with the right tools.

075 Please come along if you're interested, and if you're not then come and *get* interested.

076 If you are interested but not on the list, please let me know IMMEDIATELY.

077 As phoneticians, linguists, engineers and computer scientists working in this field, we believe that all the topics mentioned in the charter *do* belong in a single group.

078 The idea of posting at this stage is not to vote, but to encourage others to vote in favour when the time comes, in 3 weeks I think, to generate *enthusiasm*.

079 Further to Andie's message about security, please do make sure you change your passwords TODAY.

080 This is designed to fit into ANY options class.

081 It's also useful for those of us who don't know programming languages but would be happy to help people out using what we *do* know.

082 I don't care if it DOES take longer.

083 For each dialogue, the digits entered are compared with the digits that SHOULD have been entered.

084 Ah, but *which* C?

085 Could the person who took the mouse form Malcolm's old PC please return it NOW.

086 25000 pounds for *that* ?

087 It was just announced during the 1 o'clock news on BBC1 that apparently the coup in the Soviewt Union is OVER.

088 Even if the numbers on the phones do, when you dial it, a phone with a DIFFERENT number rings.

089 If the customer, account and machine are all represented by *objects*, then the initial task is to decide what attributes and behaviour would be appropriate for each.

090 Seriously though, I just LOVE to use recursion where speed doesn't matter and the chance to use default parameters appealed also.

091 Sorry folks. The BT talk is CANCELLED.

092 Okay folks, this weather is *good*.

# Appendix F (Chapter IV)

**Full tabulated results from the parameter analysis**

**Table Y:** Increase of duration and F0 parameters in emphatic vs. non-emphatic readings for 3 speakers.

| Measure: Mean % UQ % = start of Upper Quartile of data points | syllable dur. | F0 max. | vowel dur. | cont. cons. dur. | F0-max position |
|---|---|---|---|---|---|
| nuclear | JE:+39.51% t = 7.06 (25)<br><br>SB:+20.48% t = 4.27 (21)<br><br>SF: +6.04% (pilot only) | JE:+13.02% t = 6.04 (25)<br><br>SB:+7.13% t = 4.47 (21) UQ: +12.7%<br><br>SF:+6.78% (7) (pilot only) | JE:+28.63%<br><br>Short V:<br>JE: +17.6% t = 2.54 (16)<br>SB: +16.5% t = 2.47 (11)<br>UQ: +20.24%<br>Long V and Dip:<br>JE: +39.6% t = 2 (5)<br>SB: +25.7% t = 3.37 (10)<br>UQ: +40.8% | JE:+80.34% t = 3.67 (6)<br><br>SB:+40.24% t = 3.37 (5) UQ: +72 % | JE:+5.6% t = 1.87 (25)<br><br>As %elapsed vowel (JE) t = 1.7 (22)<br><br>SB:+4.04% t = 1.41 (21)<br><br>SF:+5.61% (pilot only) |
| accented - nuclear | | | SF:+18.37% (11) (pilot only)<br>JE:+35.44% (10) (pilot only) | JE:+137.3% (5) (pilot only) | |

# Appendix G (Chapter IV)

**Experimental instructions: Experiment I**

# Instructions

In British English declarative sentences there are particular syllables which are more prominent than others. These syllables are often referred to as carrying 'sentence stress', or to use Hallidays terms - they are 'nuclear'. This prominence falls onto the stressed syllable of the word which is the recipient of this sentential prominence.

The following resynthesized sentences from one speaker are examples where such a nucleus falls on a syllable in the underlined word in a neutral declarative reading. This results in a certain prominence of the syllable or word in question as described above.

Speakers and readers, however, often exercise the right to mark out a word for special emphasis or prominence if they consider it to be particularly important for the discourse. As for the case of readers it is often a result of reading a word which has been specifically marked for emphasis by the author of the text. This is often done with the help of capitalized or italicized words or by using bold typescript.

After listening to each of the following sentences, can you decide whether the underlined word was read as the result of the reader encountering a capitalized word or whether it was read merely with the neutral, nuclear prominence one would expect?

**In other words, was the underlined word in your text capitalized in the original and hence read with special emphatic prominence or not?**

Please always indicate your choice by putting the numbers 1 (for capitalization) or 0 (for neutral prominence) in the adjacent space.

## A:     Examples:

1. Please come along if you're interested, and if you're not then come and get interested.

2. Please come along if you're interested, and if you're not then come and get interested.

3. Seriously though, I just love to use recursion where speed doesn't matter and the chance to use default parameters appealed also.

4. Seriously though, I just love to use recursion where speed doesn't matter and the chance to use default parameters appealed also.

5. If you are interested but not on the list, please let me know immediately.

6. If you are interested but not on the list, please let me know immediately.


## B:     Base sentences for the construction of thirty re-synthesized versions

1. Passwords in any of the following categories could be broken in minutes by a hacker with the right tools.

2. Sorry folks. The BT talk is cancelled.

3. Could the person who took the mouse from Malcolm's old PC please return it now.

4. For each dialogue, the digits entered are compared with the digits that should have been entered.

5. If the customer, account and machine are all represented by objects, then the initial task is to decide what attributes and behaviour would be appropriate for each.

6. Start up rn, and go and get a mug of coffee as it will take ages to run through the list of newsgroups.

# Appendix H (Chapter IV)

## Full tabulated results from Experiment I

**Table X:** The effect of changes in F0 and duration on the perception of emphatic and non-emphatic nuclei. (Judgements: N=300)

| Type of change / Sentence | None | F0 (Hz.) | | Vdur (msec.) | | Cdur (msec.) | | F0+Vdur (Hz;msec.) | | F0+Cdur (Hz;msec.) | | Vdur+Cdur (msec.) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Perc. | Mag. | Perc. | Mag. | Perc. | Mag. | Perc. | Mag. | Perc. | Mag. | Perc. | Mag. |
| 065 | -10 | -10 | +6 | +10 | +120 | | | | | | | | |
| *uniform length. | | | | +8* | +120 | | | +9 | +6 +120 | | | | |
| 091 | -8 | +10 | +42 | -8 | +21 | | | | +42 +21 | | | | |
| *no peak alignment | | +9* | +42 | | | | | | | | | | |
| 074 | 0 | +10 | +12 | 0 | +8 | +7 | +65 | +9 | +12 +8 | +6 | +12 +65 | +8 | +8 +65 |
| 089 | -9 | -9 | +24 | | | | | | | | | | |
| 085 | -9 All: -6 | -9 | +10 | -7 | +107 | -9 | +94 | -10 | +10 +107 | -8 | +10 +94 | -6 | +107 +94 |
| 087 | +7 | | | | | +8 | +46 | | | | | | |

218

# Appendix I (Chapter IV)

## Sentence design: Experiment II

| 105.0 | Larry ran towards the mountain |
| 105.1 | Larry ran towards the MOUNTAIN |
| 105.2 | LARRY ran towards the mountain |

What did Larry do?
| 105.5 | Larry ran towards the mountain |
What did Larry run towards?
| 105.6 | Larry ran towards the mountain |
Did Larry run towards the valley?
| 105.7 | Larry ran towards the mountain |
Who ran towards the mountain?
| 105.8 | Larry ran towards the mountain |
Did George run towards the mountain?
| 105.9 | Larry ran towards the mountain |

| 106.0 | Neville missed the replay |
| 106.1 | Neville missed the REPLAY |
| 106.2 | NEVILLE missed the replay |

What did Neville do?
| 106.5 | Neville missed the replay |
What did Neville miss?
| 106.6 | Neville missed the replay |
Did Neville miss the match?
| 106.7 | Neville missed the replay |
Who missed the match?
| 106.8 | Neville missed the match |
Did David miss the match?
| 106.9 | Neville missed the match |

| 107.0 | Mary leased the mansion |
| 107.1 | Mary leased the MANSION |
| 107.2 | MARY leased the mansion |

What did Mary do?
| 107.5 | Mary leased the mansion |
What did Mary lease?
| 107.6 | Mary leased the mansion |

Did Mary lease the castle?
107.7          Mary leased the mansion
Who leased the mansion?
107.8          Mary leased the mansion
Did Sarah lease the mansion?
107.9          Mary leased the mansion

---

108.0          Rob annoyed Amanda
108.1          Rob annoyed AMANDA
108.2          ROB annoyed Amanda

What did Rob do?
108.5          Rob annoyed Amanda
Who did Rob annoy?
108.6          Rob annoyed Amanda
Did Rob annoy Fiona?
108.7          Rob annoyed Amanda
Who annoyed Amanda?
108.8          Rob annoyed Amanda
Did Sam annoy Amanda
108.9          Rob annoyed Amanda

---

109.0          The mouse vanished immediately
109.1          The mouse vanished IMMEDIATELY
109.2          The MOUSE vanished immediately

What did the mouse do?
109.5          The mouse vanished immediately
When did the mouse vanish?
109.6          The mouse vanished immediately
Did the mouse hesitate?
109.7          The mouse vanished immediately
Who vanished immediately?
109.8          The mouse vanished immediately
Did the cat vanish immediately?
109.9          The mouse vanished immediately

---

110.0          The solicitor arranged the loan
110.1          The solicitor arranged the LOAN
110.3          The solicitor ARRANGED the loan

What did the solicitor do?
110.5          The solicitor arranged the loan
What did the solicitor arrange?
110.6          The solicitor arranged the loan
Did the solicitor arrange the account?
110.7          The solicitor arranged the loan

What did the solicitor do with the loan?
110.a            The solicitor arranged the loan
Did the solicitor refuse the loan?
110.b            The solicitor arranged the loan

---

111.0            Belinda followed the thief
111.1            Belinda followed the THIEF
111.3            Belinda FOLLOWED the thief

What did Belinda do?
111.5            Belinda followed the thief
Who did Belinda follow?
111.6            Belinda followed the thief
Did Belinda follow the police?
111.7            Belinda followed the thief
What did Belinda do to the thief?
111.a            Belinda followed the thief
Did Belinda run away from the thief?
111.b            Belinda followed the thief

---

112.0            The minister alarmed the layman
112.1            The minister alarmed the LAYMAN
112.3            The minister ALARMED the layman

What did the minister do?
112.5            The minister alarmed the layman
Who did the minister alarm?
112.6            The minister alarmed the layman
Did the minister alarm the preacher?
112.7            The minister alarmed the layman
What did the minister do to the layman?
112.a            The minister alarmed the layman
Did the minister reassure the layman?
112.b            The minister alarmed the layman

---

113.0            The rhino amazed the foreigners
113.1            The rhino amazed the FOREIGNERS
113.3            The rhino AMAZED the foreigners

What did the rhino do?
113.5            The rhino amazed the foreigners
Who did the rhino amaze?
113.6            The rhino amazed the foreigners
Did the rhino amaze the natives?
113.7            The rhino amazed the foreigners
What did the rhino do to the foreigners?
113.a            The rhino amazed the foreigners

Did the rhino bore the foreigners?
113.b            The rhino amazed the foreigners

---

114.0            The recession united the management
114.1            The recession united the MANAGEMENT
114.3            The recession UNITED the management

What did the recession do?
114.5            The recession united the management
Who did the recession unite?
114.6            The recession united the management
Did the recession unite the workforce?
114.7            The recession united the management
What did the recession do to the management?
114.a            The recession united the management
Did the recession split the management?
114.b            The recession united the management

# Appendix J (Chapter IV)

## Table of changes applied to .5 (see Appendix i) sentences for the creation of emphatic nuclei, second peaks (P2) and first peaks (P1)

### 1. Nuclei

| Utt-number | | mean-emphsis F0 increase | mean-emphasis Cdur increase | mean-emphasis Vdur increase | boosted emphasis F0 increase* | boosted emphasis Cdur increase* | boosted emphasis Vdur increase* |
|---|---|---|---|---|---|---|---|
| **Utt-number** | **105 - nuc** | | | | | | |
| F0 peak | 253 Hz. | 18 Hz. | | | 32 Hz. | | |
| Cdur | 69 ms. | | 27.6 ms. | | | 49.7 ms. | |
| Vdur | 136 ms. | | | 34.9 ms. | | | 55.4 ms. |
| Frames | | | 128-132 | 134-145 | | 128-132 | 134-145 |
| s-points/frame | | | 55 | 29 | | 99 | 46 |
| **Utt-number** | **106 - nuc** | | | | | | |
| F0 peak | 265 Hz. | 18.8 Hz. | | | 33.6 Hz. | | |
| Cdur | 44 ms. | | 17.6 ms. | | | 31.7 ms. | |
| Vdur | 112 ms. | | | 28.8 ms. | | | 45.7 ms. |
| Frames | | | 186-190 | 192-200 | | 186-190 | 192-200 |
| s-points/frame | | | 35 | 32 | | 63 | 51 |
| **Utt-number** | **107 - nuc** | | | | | | |
| F0 peak | 276 Hz. | 19.6 Hz. | | | 35 Hz. | | |
| Cdur | 64 ms. | | 25.6 ms. | | | 46.1 ms. | |
| Vdur | 115 ms. | | | 19 ms. | | | 23.2 ms. |
| Frames | | | 76-79 | 83-91 | | 76-79 | 38-91 |
| s-points/frame | | | 64 | 21 | | 115 | 26 |
| **Utt-number** | **108 - nuc** | | | | | | |
| F0 peak | 237 Hz. | 16.8 Hz. | | | 30 Hz. | | |
| Cdur | 61 ms. | | 24.4 ms. | | | 43.9 ms. | |
| Vdur | 133 ms. | | | 21.9 ms. | | | 26.9 ms. |
| Frames | | | 129-133 | 136-146 | | 129-133 | 136-146 |
| s-points/frame | | | 49 | 20 | | 88 | 24 |
| **Utt-number** | **109 - nuc** | | | | | | |
| F0 peak | 256 Hz. | 18.2 Hz. | | | 32.5 Hz. | | |
| Cdur | 51 ms. | | 20.4 ms. | | | 36.7 ms. | |
| Vdur | 104 ms. | | | 26.7 ms. | | | 42.4 ms. |
| Frames | | | 175-177 | 180-187 | | 175-177 | 180-187 |
| s-points/frame | | | 68 | 33 | | 122 | 53 |
| **Utt-number** | **110 - nuc** | | | | | | |
| F0 peak | 271 Hz. | 19.2 Hz. | | | 34.4 Hz. | | |
| Cdur | 35 ms. | | 14 ms. | | | 25.2 ms. | |
| Vdur | 212 ms. | | | 54.5 ms. | | | 86.5 ms. |
| Frames | | | 196-198 | 200-218 | | 196-198 | 200-218 |
| s-points/frame | | | 47 | 30 | | 84 | 10*46;9* 45 |
| **Utt-number** | **111 - nuc** | | | | | | |
| F0 peak | 276 Hz. | 19.6 Hz. | | | 35 Hz. | | |

|  |  | mean-emphsis F0 increase | mean-emphasis Cdur increase | mean-emphasis Vdur increase | boosted emphasis F0 increase* | boosted emphasis Cdur increase* | boosted emphasis Vdur increase* |
|---|---|---|---|---|---|---|---|
| Cdur | 112 ms. | | 44.8 ms. | | | 80.6 ms. | |
| Vdur | 147 ms. | | | 37.8 ms. | | | 60 ms. |
| Frames | | | 114-120 | 126-136 | | 114-120 | 126-136 |
| s-points/frame | | | 64 | 34 | | 115 | 6*55;5*54 |
| **Utt-number** | **112 - nuc** | | | | | | |
| F0 peak | 265 Hz. | 18.8 Hz. | | | 33.7 Hz. | | |
| Cdur | 53 ms. | | 21.2 ms. | | | 38.1 ms. | |
| Vdur | 150 ms. | | | 38.6 ms. | | | 61.2 ms. |
| Frames | | | 225-228 | 230-243 | | 225-228 | 230-243 |
| s-points/frame | | | 53 | 28 | | 95 | 44 |
| **Utt-number** | **113-nuc** | | | | | | |
| F0 peak | 236 Hz. | 16.7 Hz. | | | 30 Hz. | | |
| Cdur | 97 ms. | | 38.8 ms. | | | 69.8 ms. | |
| Vdur | 135 ms. | | | 22.3 ms. | | | 27.3 ms. |
| Frames | | | 242-248 | 254-263 | | 242-248 | 254-263 |
| s-points/frame | | | 55 | 22 | | 100 | 27 |
| **Utt-number** | **114-nuc** | | | | | | |
| F0 peak | 260 Hz. | 18.5 Hz. | | | 33 Hz. | | |
| Cdur | 61 ms. | | 24.4 ms. | | | 43.9 ms. | |
| Vdur | 104 ms. | | | 17.1 ms. | | | 21 ms. |
| Frames | | | 186-190 | 192-200 | | 186-190 | 192-200 |
| s-points/frame | | | 49 | 21 | | 88 | 6*23;3*24 |

*boosted emphasis values are based on the starting point of the Upper Quartile of the analysed data for the particular speaker

## 2. Second Peaks (P2)

| Utt-number |  | mean-emphsis F0 increase | mean-emphasis Cdur increase | mean-emphasis Vdur increase | boosted emphasis F0 increase* | boosted emphasis Cdur increase* | boosted emphasis Vdur increase* |
|---|---|---|---|---|---|---|---|
| **Utt-number** | **110 - P2** | | | | | | |
| F0 peak | 262 Hz. | 18.6 Hz. | | | 33.3 Hz. | | |
| Cdur | 59 ms. | | 23.6 ms. | | | 42.5 ms. | |
| Vdur | 111 ms. | | | 28.5 ms. | | | 45.3 ms. |
| Frames | | | 145-148 | 152-160 | | 145-148 | 152-160 |
| s-points/frame | | | 59 | 32 | | 106 | 6*50;3*51 |
| **Utt-number** | **111 - P2** | | | | | | |
| F0 peak | 256 Hz. | 18.1 Hz. | | | 32.5 Hz. | | |
| Cdur | 111 ms. | | 44.4 ms. | | | 79.9 ms. | |
| Vdur | 96 ms. | | | 15.8 ms. | | | 19.4 ms. |
| Frames | | | 67-73 | 77-84 | | 67-73 | 77-84 |
| s-points/frame | | | 63 | 20 | | 114 | 24 |
| **Utt-number** | **112-P2** | | | | | | |
| F0 peak | 252 Hz. | 17.8 Hz. | | | 32 Hz. | | |
| Cdur | 62 ms. | | 24.8 ms. | | | 44.6 ms. | |
| Vdur | 155 ms. | | | 39.8 ms. | | | 63.2 ms. |
| Frames | | | 182-186 | 188-200 | | 182-186 | 188-200 |
| s-points/frame | | | 50 | 7*30; 6*31 | | 89 | 8*49;5*48 |

| Utt-number | 113-P2 | | | | | | |
|---|---|---|---|---|---|---|---|
| F0 peak | 241 Hz. | 17.1 Hz. | | | 30.6 Hz. | | |
| Cdur | 80 ms. | | 32 ms. | | | 57.6 ms. | |
| Vdur | 171 ms. | | | 43.9 ms. | | | 69.8 ms. |
| Frames | | | 198-204 | 206-220 | | 198-204 | 206-220 |
| s-points/frame | | | 50 | 29 | | 82 | 8*47;7*46 |
| Utt-number | 114-P2 | | | | | | |
| F0 peak | 252 Hz. | 17.8 Hz. | | | 32 Hz. | | |
| Cdur | 64 ms. | | 25.6 ms. | | | 46.1 ms. | |
| Vdur | 151 ms. | | | 38.8 ms. | | | 61.6 ms. |
| Frames | | | 138-143 | 146-152 | | 138-143 | 146-152 |
| s-points/frame | | | 43 | 4*55; 3*56 | | 77 | 103 |

## 3. First Peaks (P1)

| Utt-number | | mean-emphsis F0 increase | mean-emphasis Cdur increase | mean-emphasis Vdur increase | boosted emphasis F0 increase* | boosted emphasis Cdur increase* | boosted emphasis Vdur increase* |
|---|---|---|---|---|---|---|---|
| Utt-number | 105 - P1 | | | | | | |
| F0 peak | 268 Hz. | 19 Hz. | | | 34 Hz. | | |
| Cdur | 77 ms. | | 30.8 ms. | | | 55.4 ms. | |
| Vdur | 82 ms. | | | 13.5 ms. | | | 16.6 ms. |
| Frames | | | 30-34 | 38-41 | | 30-34 | 38-41 |
| s-points/frame | | | 62 | 34 | | 111 | 2*42;2*41 |
| Utt-number | 106 - P1 | | | | | | |
| F0 peak | 285 Hz. | 20.2 Hz. | | | 36.2 Hz. | | |
| Cdur | 45 ms. | | 18 ms. | | | 32.4 ms. | |
| Vdur | 79 ms. | | | 13 ms. | | | 16 ms. |
| Frames | | | 125-128 | 130-136 | | 125-128 | 130-136 |
| s-points/frame | | | 45 | 19 | | 81 | 23 |
| Utt-number | 107 - P1 | | | | | | |
| F0 peak | 283 Hz. | 20 Hz. | | | 35.9 Hz. | | |
| Cdur | 40 ms. | | 16 ms. | | | 28.8 ms. | |
| Vdur | 103 ms. | | | 26.5 ms. | | | 42 ms. |
| Frames | | | 10-13 | 16-22 | | 10-13 | 16-22 |
| s-points/frame | | | 40 | 38 | | 72 | 60 |
| Utt-number | 108 - P1 | | | | | | |
| F0 peak | 283 Hz. | 20 Hz. | | | 35.9 Hz. | | |
| Cdur | 33 ms. | | 13.2 ms. | | | 23.8 ms. | |
| Vdur | 102 ms. | | | 16.8 ms. | | | 20.4 ms. |
| Frames | | | 70-72 | 74-82 | | 70-72 | 74-82 |
| s-points/frame | | | 44 | 19 | | 79 | 6*23;3*22 |
| Utt-number | 109 - P1 | | | | | | |
| F0 peak | 276 Hz. | 19.6 Hz. | | | 35 Hz. | | |
| Cdur | 71 ms. | | 28.4 ms. | | | 51.1 ms. | |
| Vdur | 147 ms. | | | 37.7 ms. | | | 60 ms. |
| Frames | | | 99-104 | 107-117 | | 99-104 | 107-117 |
| s-points/frame | | | 47 | 34 | | 85 | 6*55;5*54 |

# Appendix K (Chapter V)

## Experimental instructions: Experiment II

# Instructions

In spoken communication the intonation of an utterance often determines whether that utterance is an appropriate one for a given context. For example, if you are asked by a friend who else had passed their exams and you know who passed them, you might answer: "JOHN passed the exams".

You may also know that John passed the test rather than the exams, and if asked whether John passed the exams you may answer "John passed the TEST".

*Often, the amount of emphasis that is placed on certain words determines how appropriate the answer is.*

The following pages contain combinations of questions and answers. For each combination you will hear two versions of the answer, presented as a pair, in the form of a human voice which has been resynthesized. Each answer-pair is preceded by a tone.

Your task is to read out quietly each of the questions and **decide which of the answers in each pair sounds the more appropriate way of saying the answer to that particular question.**

Please always indicate your choice by putting the numbers 1 or 2 in the adjacent space, depending on whether you perceived the first or the second answer to be the more appropriate one.

# Appendix L (Chapter IV)

## Full tabulated results from Experiment II (Nuclear Emphasis)

| Nucleus | Isolated recording vs. real answer | Neutral answer vs. real answer | Rule-based emphasis vs. real answer | Rule-based emphasis vs. real answer for specific contexts | Rule-based emphasis vs. real answer for contrastive contexts | Neutral answer vs. real answer for specific contexts | Neutral answer vs. real answer for contrastive contexts |
|---|---|---|---|---|---|---|---|
| F1 | 5 - 15 | 5 - 15 | 8 - 12 | 3 - 7 | 5 - 5 | 3 - 7 | 2 - 8 |
| M1 | 0 - 20 | 4 - 16 | 9 - 11 | 5 - 5 | 4 - 6 | 2 - 8 | 2 - 8 |
| M2 | 6 - 14 | 6 - 14 | 11 - 9 | 4 - 6 | 7 - 3 | 6 - 4 | 0 - 10 |
| F2 | 3 - 17 | 2 - 18 | 10 - 10 | 5 - 5 | 5 - 5 | 1 - 9 | 1 - 9 |
| M3 | 1 - 19 | 9 - 11 | 9 - 11 | 5 - 5 | 4 - 6 | 4 - 6 | 5 - 5 |
| M4 | 3 - 17 | 9 - 11 | 10 - 10 | 4 - 6 | 6 - 4 | 3 - 7 | 6 - 4 |
| M5 | 8 - 12 | 12 - 8 | 11 - 9 | 6 - 4 | 5 - 5 | 5 - 5 | 7 - 3 |
| F3 | 4 - 16 | 4 - 16 | 13 - 7 | 5 - 5 | 8 - 2 | 2 - 8 | 2 - 8 |
| F4 | 3 - 17 | 3 - 17 | 11 - 9 | 4 - 6 | 7 - 3 | 2 - 8 | 1 - 9 |
| F5 | 2 - 18 | 5 - 15 | 10 - 10 | 4 - 6 | 6 - 4 | 3 - 7 | 2 - 8 |
| TOTAL | 35 - 165 | 59 - 141 | 102 - 98 | 45 - 55 | 57 - 43 | 31 - 69 | 28 - 72 |
| Statistics (t - two tailed) | t = 8.67 sig@ .001 | t = 4.12 sig@ .001 | t = - 0.45 not sig | t = 1.86 not sig | t = - 1.65 not sig | t = 3.94 sig@ .01 | t = 2.96 sig@ .02 |

| Nucleus | Rules for weak vs. rules for strong emphasis | Rules for weak vs. rules for strong emphasis for specific contexts | Rules for weak vs. rules for strong emphasis for contrastive contexts | Selection of the first vs. the second answer in the presentation pairs |
|---|---|---|---|---|
| F1 | 10 - 10 | 6 - 4 | 4 - 6 | 44 - 36 |
| M1 | 7 - 13 | 3 - 7 | 4 - 6 | 38 - 42 |
| M2 | 4 - 16 | 4 - 6 | 0 - 10 | 39 - 41 |
| F2 | 8 - 12 | 4 - 6 | 4 - 6 | 37 - 43 |
| M3 | 13 - 7 | 7 - 3 | 6 - 4 | 48 - 32 |
| M4 | 14 - 6 | 7 - 3 | 7 - 3 | 40 - 40 |
| M5 | 12 - 8 | 6 - 4 | 6 - 4 | 47 - 33 |
| F3 | 10 - 10 | 7 - 3 | 3 - 7 | 43 - 37 |
| F4 | 6 - 14 | 2 - 8 | 4 - 6 | 41 - 39 |
| F5 | 1 - 19 | 0 - 10 | 1 - 9 | 36 - 44 |
| TOTAL | 85 - 115 | 46 - 54 | 39 - 61 | 413 - 387 |
| Statistics (t - two tailed) | t = 1.15 not sig | t = 0.52 not sig | t = 1.59 not sig | t = - 1 not sig |

# Appendix M (Chapter IV)

**Full tabulated results from Experiment II (Emphasis on first and second peaks (P1 and P2 respectively)**

| P1 | Rule-based emphasis and de-accentuation rules vs. real answer | Rule-based emphasis and de-accentuation rules vs. real answer for **specific** contexts | Rule-based emphasis and de-accentuation rules vs. real answer for **contrastive** contexts | Rule-based emphasis spliced into the real answer vs. real answer | Rule-based emphasis spliced into the real answer vs. real answer for **specific** contexts | Rule-based emphasis spliced into the real answer vs. real answer for **contrastive** contexts | Selection of the first vs. the second answer in the presentation pairs |
|---|---|---|---|---|---|---|---|
| F1 | 3 - 7 | 3 - 2 | 0 - 5 | 3 - 7 | 1 - 4 | 2 - 3 | 10 - 10 |
| M1 | 4 - 6 | 2 - 3 | 2 - 3 | 3 - 7 | 1 - 4 | 2 - 3 | 9 - 11 |
| M2 | 6 - 4 | 2 - 3 | 4 - 1 | 3 - 7 | 0 - 5 | 3 - 2 | 11 - 9 |
| F2 | 8 - 2 | 3 - 2 | 5 - 0 | 5 - 5 | 1 - 4 | 4 - 1 | 13 - 7 |
| M3 | 1 - 9 | 1 - 4 | 0 - 5 | 1 - 9 | 1 - 4 | 0 - 5 | 12 - 8 |
| M6 | 2 - 8 | 0 - 5 | 2 - 3 | 5 - 5 | 3 - 2 | 2 - 3 | 11 - 9 |
| M5 | 7 - 3 | 4 - 1 | 3 - 2 | 3 - 7 | 0 - 5 | 3 - 2 | 10 - 10 |
| F3 | 2 - 8 | 2 - 3 | 0 - 5 | 1 - 9 | 0 - 5 | 1 - 4 | 9 - 11 |
| F6 | 6 - 4 | 2 - 3 | 4 - 1 | 5 - 5 | 1 - 4 | 4 - 1 | 9 - 11 |
| F5 | 10 - 0 | 5 - 0 | 5 - 0 | 5 - 5 | 3 - 2 | 2 - 3 | 9 - 11 |
| **TOTAL** | **49 - 51** | **24 - 26** | **25 - 25** | **34 - 66** | **11 - 39** | **23 - 27** | **103 - 97** |
| **Statistics (t - two tailed)** | **t = 0.10 not sig** | **t = 0.22 not sig** | **t = 0 not sig** | **t = 3.20 sig@ .02** | **t = 4.02 sig@ .02** | **t = 0.50 not sig** | **t = -0.66 not sig** |

| P2 | Rule-based emphasis and de-accentuation rules vs. real answer | Rule-based emphasis and de-accentuation rules vs. real answer for **specific** contexts | Rule-based emphasis and de-accentuation rules vs. real answer for **contrastive** contexts | Selection of the first vs. the second answer in the presentation pairs |
|---|---|---|---|---|
| F1 | 5 - 5 | 4 - 1 | 1 - 4 | 7 - 3 |
| M1 | 4 - 6 | 3 - 2 | 1 - 4 | 6 - 4 |
| M2 | 6 - 4 | 4 - 1 | 2 - 3 | 6 - 4 |
| F2 | 4 - 6 | 3 - 2 | 1 - 4 | 8 - 2 |
| M3 | 3 - 7 | 3 - 2 | 0 - 5 | 5 - 5 |
| M6 | 4 - 6 | 3 - 2 | 1 - 4 | 6 - 4 |
| M5 | 4 - 6 | 3 - 2 | 1 - 4 | 6 - 4 |
| F3 | 1 - 9 | 1 - 4 | 0 - 5 | 5 - 5 |
| F6 | 5 - 5 | 3 - 2 | 2 - 3 | 7 - 3 |
| F5 | 6 - 4 | 3 - 2 | 2 - 3 | 4 - 6 |
| **TOTAL** | **42 - 58** | **30 - 20** | **11 - 39** | **60 - 40** |
| **Statistics (t - two tailed)** | **t = 1.71 not sig** | **t = -1.93 not sig** | **t = 6 sig@ .01** | **t = -2.73 sig@ .05** |

# Appendix N (Chapter II)

## Example listing of smiley face inventories

The following listing was extracted from the 'Unofficial Smilie Dictionary' by:
Melinda Varian (Princeton University), supplemented by Chris Thomas (UCLA) & Paul
Zarnowski & Larry Chace (Cornell University)

Basic ones:

| | |
|---|---|
| :-) | Your basic smilie. This smilie is used to inflect a sarcastic or joking statement since we can't hear voice inflection via email |
| ;-) | Winky smilie. User just made a flirtatious and/or sarcastic remark. More of a 'don't hit me for what I just said' smilie |
| :-( | Frowning smilie. User did not like that last statement or is upset or depressed about something |
| :-I | Indifferent smilie. Better than a Frowning smilie but not quite as good as a happy smilie |
| :-> | User just made a really biting sarcastic remark. Worse than a :-) |
| >:-> | User just made a really devilish remark |
| >;-> | Winky and devil combined. A very lewd remark was just made |
| :-7 | User just made a wry statement |
| :-& | User is tongue-tied |
| :-? | User has tongue in cheek |

Somewhat less common ones include:

| | | | | |
|---|---|---|---|---|
| (-: | User is left handed | | :'-( | User is crying |
| %-) | User has been staring at screen too long | | :'-) | User is so happy, s/he is crying |
| :*) | User is drunk | | :-@ | User is screaming |
| [:] | User is a robot | | :-# | User wears braces |
| 8-) | User is wearing sunglasses | | :~) | User has a broken nose |
| B:-) | Sunglasses on head | | =:-) | User is a punk rocker |
| ::-) | User wears normal glasses | | =:-( | real punk rockers don't smile |
| B-) | User wears horn-rimmed glasses | | :=) | User has two noses |
| 8:-) | User is a little girl | | +-:-) | User is the Pope or holds some other religious office |
| :-)-8 | User is a big girl | | |-I | User is asleep |
| :-{) | User has a mustache | | |-O | User is yawning/snoring |
| :-{} | User wears lipstick | | :-Q | User is a smoker |
| {:-) | User wears a toupee | | O :-) | User is an angel (at heart, at least) |
| }:-( | Toupee in an updraft | | :-P | Nyahhhh! |
| :-[ | User is a vampire | | :-S | User just made an incoherent statement |
| :-E | Bucktoothed vampire | | | |
| :-F | Bucktoothed vampire with one tooth missing | | :-D | User is laughing (at you!) |
| | | | :-X | User's lips are sealed |
| :-* | User just ate something sour | | :-C | User is really bummed |
| :-)% | User drools | | :-/ | User is skeptical |
| :-%) | User has a cold | | C=:-) | User is a chef |

| | | | |
|---|---|---|---|
| @= | User is pro-nuclear war | ):-) | " |
| *<:-) | User is wearing a Santa Claus Hat | ):-( | unsmiley big-face |
| :-o | Uh oh! | )8-) | scuba smiley big-face |
| *:o) | Bozo the Clown! | =:-) | smiley punk-rocker |
| d8= | User's pet beaver is wearing goggles and a hard hat. | =:-( | real punk rockers don't smile |
| :-9 | User is licking his/her lips | +:-) | smiley priest |
| %-6 | User is braindead | :-q | smiley trying to touch its tongue to its nose |
| [:-) | User is wearing a walkman | :-e | disappointed smiley |
| (:I | User is an egghead | :-t | cross smiley |
| <:-I | User is a dunce | :-i | semi-smiley |
| @:-) | User is wearing a turban | :-o | smiley singing national anthem |
| :-0 | No Yelling!  (Quiet Lab) | :-p | smiley sticking its tongue out (at you!) |
| :-: | Mutant Smilie | :-[ | un-smiley blockhead |
| X-( | User just died | :-] | smiley blockhead |
| | | :-{ | smiley variation on a theme |
| C=}];*()) | | :-} | ditto |
| | Mega-Smilie... A drunk, devilish chef with a toupee in an updraft, a mustache, and a double chin | {:-) | smiley with its hair parted in the middle |
| | | }:-) | above in an updraft |
| | | :-a | lefty smilely touching tongue to nose |

From: page@swan.ulowell.edu (Bob Page)
Organization: University of Lowell,
Computer Science Dept.

| | | | |
|---|---|---|---|
| | | :-s | smiley after a BIZARRE comment |
| :-` | smiley spitting out its chewing tobacco | :-d | lefty smiley razzing you |
| :-1 | smiley bland face | g-) | smiley with ponce-nez glasses |
| :-! | " | :-j | left smiling smilely |
| :-@ | smiley face screaming | :-k | beats me, looks like something, tho. |
| :-#| | smiley face with bushy mustache | :-l | y. a. s. |
| :-$ | smiley face with it's mouth wired shut | :-: | mutant smiley |
| :-% | smiley banker | :-\ | undecided smiley |
| :-6 | smiley after eating something sour | :-| | "have an ordinary day" smiley |
| :^) | smiley with pointy nose (righty) | ;-) | winking smiley |
| :-7 | smiley after a wry statement | :-< | real sad smiley |
| 8-) | smiley swimmer | :-> | y.a.s. |
| :-* | smiley after eating something bitter | :-z | y.a.c.s. |
| :-& | smiley which is tongue-tied | :-x | "my lips are sealed" smiley |
| :-9 | smiley licking it's lips | :-c | bummed out smiley |
| :-0 | smiley orator | :-v | talking head smiley |
| :-( | un-smiley | :v) | left-pointing nose smiley |
| :-) | smiley standard smiley invisible man | :-b | left-pointing tongue smiley |
| | | :-/ | lefty undecided smiley |
| (:-( | unsmiley frowning | :-? | smilely smoking a pipe |
| (:-) | smiley big-face | .-] | one-eyed smilely |
| | | ,-} | wry and winking |
| | | 0-) | smiley cyclops (scuba diver?) |
| | | :-=) | older smiley with mustache |

| | |
|---|---|
| :u) | smiley with funny-looking left nose |
| :n) | smiley with funny-looking right nose |
| :< | midget unsmiley |
| :> | midget smiley |
| }:^#}) | mega-smiley: updrafted bushy-mustached pointy nosed smiley with a double-chin |

# References

Abercrombie, D. (1967): *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Adrianson, L. and Hjelmquist, E. (1985): Small group communication in two media: Face-to-face communication and computer mediated communication. *Göteborg Psychological Reports*,Vol.15, No.1, University of Göteborg, Sweden.

Ainsworth, W. and Holmes, J. (1988): *Speech 88:* Proceedings of the 7th FASE symposium. Edinburgh: IOA.

Allen, J., Hunnicutt, M.S., Klatt, D. (1987): *From Text to Speech: The MITalk system.* Cambridge: Cambridge University Press.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thomson, H. S., Weinert, R. (1991): The HCRC Map Task Corpus. *Language and Speech*, 34(4), pp. 351-366.

Bagshaw, P. C., Hiller, S.M., Jack, M.A. (1993): Enhanced Pitch Tracking and the Processing of F0 Contours for Computer Aided Intonation Teaching. *Proceedings of the 3rd European Conference on Speech Communication and Technology*, Berlin, Germany, pp. 1003-1006.

Berman, A., Szamosi, M. (1972): Observations on Sentential Stress. *Language*, **48**, pp. 304-325.

Bierwisch, M., Schreuder, R. (1992): From Concepts to Lexical Items. *Cognition*, **42**, pp. 23-60.

Bing, J. M. (1979): *Aspects of English Prosody*. Ph.D. dissertation, University of Massachusetts.

Bladen, A. (1990): Evaluating the Prosody of Text-to-Speech synthesizers". *Proceedings, Speech Tech '90*, pp. 215-220.

Bloomfield, L. (1933): *Language*. New York: Henry Holt.

Bolinger, D. (1951): Intonation: Levels versus Configurations. *Word*, 7, pp. 199-210.

Bolinger, D. (1964): Around the Edge of Language: Intonation. In Bolinger, D. (ed.), *Intonation*. Penguin.

Bolinger, D. (1970): Relative Height. In Léon, P.R., Faure, G., Rigault, A. (eds.). *Prosodic Feature Analysis*, 1970, pp.109-125.

Bolinger, D. (1972a): Accent is predictable (if you're a mind reader). *Language,* 48, pp. 633-644.

Bolinger, D. (1972b): *Intonation*. Penguin.

Bolinger, D. (1975): *Aspects of Language*. New York: Harcourt Brace Jovanovich.

Bolinger, D. (1985): Two Views on Accent. *Journal of Linguistics*, 21, pp.79-123.

Bolinger, D. (1987): More Views on 'Two views of Accent'. In Gussenhoven et al., *On Accent,* 1987, pp.124-146.

Bolinger, D. (1989): *Intonation and its Uses*. Stanford: University Press.

Bresnan, J. (1971): Sentence Stress and Syntactic Transformation. *Language,* 47, pp. 257-281.

Bresnan, J. (1972): Stress and Syntax: a Reply. *Language,* 48, pp. 326-342.

Brown, E.K. and Miller, J.E., (1980): *Syntax: A Linguistic Introduction to Sentence Structrure*. London: Hutchinson.

Brown, G., Yule, G. (1983): *Discourse Analysis*. Cambridge: Cambridge University Press.

Bruce, Gösta (1977): *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Cahn, J. E. (1989): Generating Expression in Synthesized Speech. Masters dissertation, MIT, Cambridge (Mass.).

Campbell, W. N. (1992): *Multi-Level Timing in Speech*. Ph.D. dissertation, University of Edinburgh.

Carvalho, P., Lopes, P., Trancoso, I., Oliveria, L. (1994): E-Mail to Voice-Mail Conversion Using a Portuguese Text-to-Speech System. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, pp. 1687-1690.

Chafe, W. (1970): *Meaning and the Structure of Language*. Chicago: Chicago University Press.

Chafe, W. (1979): The Flow of Thought and the Flow of Language. *Syntax and Semantics*, **12**, pp. 159-181.

Chapallaz, M. (1964): Notes on the Intonation of Questions in Italian. In Bolinger, D. (ed.), *Intonation*. Penguin.

Chomsky, N. and Halle, M. (1968): *The Sound Pattern of English*. New York: Harper and Row.

Chomsky, N. (1969): *Deep Structure, Surface Structure, and Semantic Interpretation*. Bloomington: Indiana University Linguistics Club.

Collier, R. (1990): Multi-Lingual Intonation Synthesis: Principles and Applications. *Proceedings of the ESCA Workshop on Speech Synthesis,* Autrans, France, pp. 273-276.

Collier, R., Terken, J. (1987): Intonation by Rule in Text-to-Speech Applications. In Laver & Jack (eds.), **2**, pp.165-168.

Cooper, W.E., Paccia-Cooper, J.P. (1980): *Syntax and Speech*, Cambridge, Mass.: Harvard University Press.

Cooper, W.E., Sorensen, J.M. (1981): *Fundamental Frequency in Sentence Production.* New York: Berlin: Heidelberg: Springer.

Couper-Kuhlen, E. (1986): *An Introduction to English Prosody*. Tübingen: Niemeyer.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Crystal, D. (1969): *Prosodic Systems and Intonation in English*, Cambridge: Cambridge University Press.

Cutler, A. & Ladd, D. R. (1983): *Prosody: Models and Measurements.* Berlin: Springer.

Delattre, P. (1966): The Distinctive Function of Intonation. In Bolinger, D. (ed.), *Intonation.* Penguin.

de Pijper, J. R. (1979): Close-Copy Stylizations of British English Intonation Contours. *IPO Annual Progress Report,* **14**, ppp.66-71.

Dik, S. (1980): On the typology of focus phenomena. In Hoekstra, van der Hulst, and Moortgat (eds.), *Perspectives in Functional grammar*. Dordrecht: Foris.

Docherty, G. J., (1989): *An Experimental Phonetic Study of the Timing of Voicing in English Obstruents.* Ph.D. dissertation, University of Edinburgh.

Downing, B. (1970): *Syntactic Structure and Phonological Phrasing in English*. Ph.D. dissertation, Massachusetts Institute of Technology.

Emonds, J.E. (1976): *A Transformational Approach to English Syntax*. New York: Academic Press.

Emonds, J.E. (1979): Appositive Relatives Have No Properties. *Linguistic Inquiry*, **10**, pp.210-243.

Fairbanks, G., Pronovost, W. (1939): An Experimental Study of the Pitch Characteristics of the Voice During the Expression of Emotions. *Speech Monographs*, **6**, pp.87-104.

Fitzpatrick, E. and Bachenko, J. (1989): Parsing for Prosody: What a Text-to-Speech System Needs from Syntax. *Proceedings of the Annual AI Systems in Government Conference,* Washington DC, USA, pp. 188-194.

Fowler, C.A. and Housum, J. (1987): Talkers signalling of 'new' and 'old' words in speech and listeners perception and use of the distinction. *Journal of Memory and Language,* **26**, pp. 489-504.

Francis, W.N. (1980): A tagged corpus - problems and prospects. In Greenbaum, Leech and Svartvik (eds.). *Studies in English Linguistics - for Randolph Quirk*, pp. 192-209, Longman.

Garreau, J. (1993): Head Turning 'Smileys' Give New Tilt to Computer Chatting. *Washington Post*, August 4th, 1993.

Garrett, M. F. (1988): Processes in Language Production. In Newmeyer, F. J. (ed.), *Linguistics: The Cambridge Survey,* **II**, *Linguistic Theory: Extensions and Implications*, pp. 117-138, Cambridge: Cambridge University Press.

Garro, L. and Parker, F. (1982): Some Suprasegmental Characteristics of Relative Clauses in English. *Journal of Phonetics*, **10**, pp. 149-161.

Garside, R., Leech, G., Sampson, G. (eds.) (1987): *The Computational Analysis of English: A Corpus-Based Approach*, London: Longman

Gartenberg, R. and Hertrich, I. (1988): *Predicting the Timing of F0 Peaks in Partly Controlled Dialogue Situations*. In Ainsworth and Holmes (eds.) 1988, Vol. **3** pp. 997-1004.

Greene, J., D'Oliveira, M. (1982): *Learning to Use Statistical Tests in Psychology*, Milton Keynes: Open University Press.

Grice, M. (1995): Leading Tones and Downstep in English. To appear in *Phonology*, **12**, 1995.

Godfrey, J.J. and Brodsky, J. (1986): *Acoustic Charactersistics of Emphasis*, presented at the 112th Meeting of the Acoustical Society of America, Anaheim, C.A., December 1986.

Guaitella, I. and Santi, S., (1990): Contribution of the Analysis of Punctuation to Improving the Prosody of Speech Synthesis. *Proceedings of the ESCA Workshop on Speech Synthesis*, Autrans, France, pp. 177-180.

Gussenhoven, C. (1983): Focus, mode and the nucleus. *Journal of Language* **19**, pp. 377-417.

Gussenhoven, C. (1985): Two Views on Accent: A Reply. *Journal of Linguistics*, **21**, pp.125-138.

Gussenhoven, C., Bolinger, D., Keijsper, C.E. (1987): *On Accent*. Bloomington, Ind.: Indiana Univeristy Linguistics Club.

Gussenhoven, C. (1987): More Views on 'Two Views on Accent': A Reply. In Gussenhoven et al., *On Accent*, 1987, pp.147-161.

Gussenhoven, C., Rietveld, T. (1988): Fundamental Frequency Declination in Dutch. *Journal of Phonetics*, **16**, pp. 355-369.

Günther, C., Habel, C., Maienborn, C., Schopp, A. (1993): What's up with the printer? Context Relative Presentation of Conceptual Structure and its Prosodic Realization in a Language Production System. In A. Schopp (ed.), *Hamburg Working Papers on Language Production - II.*, GK-Kognitionswissenschaft, AP 13, Universität Hamburg, June 1993.

Günther, C., (1992): Von der abstrakten prosodischen Planung zur phonetischen Realisierung. In M. Herweg (ed.), *Hamburger Arbeitspapiere zur Sprachproduktion - I.*, GK-Kognitionswissenschaft, AP 9, Universität Hamburg.

Halliday, M.A.K. (1967): *Intonation and Grammar in British English*. The Hague: Mouton.

Halliday, M. A. K. (1985): *Spoken and Written Language*. Deakin University Press.

Hayes, B. (1993): "Gesture" in Prosody. In Keating, P. (ed.), *Papers in Laboratory Phonology 3*, Cambridge: Cambridge University Press.

Herweg, M., Maienborn, C. (1992): Konzept, Kontext, Bedeutung - Zur Rolle der Zwei-Ebenen-Semantik in einem Modell der Sprachproduktion. In In M. Herweg (ed.), *Hamburger Arbeitspapiere zur Sprachproduktion - I.*, GK-Kognitionswissenschaft, AP 9, Universität Hamburg.

House, J., Youd, N. (1990): Contextually Appropriate Intonation in Speech Synthesis. *Proceedings of the ESCA Workshop on Speech Synthesis*, Autrans, France, pp.185-188.

House, J., Youd, N. (1991): Stylized Prosody in Telephone Information Services: Implications for Synthesis. *Proceedings of the XIIth International Conference of Phonetic Sciences*, Aix-en-Provence, pp.198-201.

House, J., Youd, N. (1992): Evaluating the Prosody of Synthesized Utterances within a Dialogue System. *Proceedings of the International Conference on Spoken Language Processing*, Banff, Canada, pp.1175-1178.

Hunnicut, S. (1985): Intelligibility Versus Redundancy-Conditions of Dependency. *Language and Speech*, **28**, pp.45-56.

Jackendoff, R. S. (1972): *Semantic Interpretation in Generative Grammar*. Cambridge, Mass: MIT Press.

Johns-Lewis, C., (1986): *Intonation in Discourse*. London & Sydney: Croom Helm.

Kiesler, S. B. (1978): *Interpersonal Processes in Groups and Organizations*. Arlington Harlan Davidson.

Kiesler, S., Siegel, J., McGuire, T. (1984): Social Psychological Aspects of Computer-Mediated Communication. *American Psychologist*, **39**, pp. 1123-1134.

Kiesler, S., Zubrow, D., Moses, A.M., Geller, V. (1985): Affect in Computer Mediated Communication: An experiment in Synchronous Terminal-to-Terminal Discussion. *Human-Computer Interaction*, **1**, pp. 77-104.

Kiesler, S. B., Sproull, L. (eds.) (1987): *Computing and Change on Campus*. Cambridge University Press, Cambridge.

King, R.W., (1991): Layout Processing, User Control and Prosody Insertion in an On-line Synthetic Speech System. *Proceedings of the First European Conference on Speech Communication and Technology*, Paris, France, pp. 121-124.

Kohler, K.J. (1987): The Linguistic Function of F0 peaks. *Proceedings of the 11th International Congress of Phonetic Sciences*, Tallin, pp. 149-152.

Kohler, K.J. (1988): *An Intonation Model for a German Text-to-Speech System*. In Ainsworth and Holmes (eds.) 1988, Vol. **4**, pp. 1241-1247.

Kutik, E.J., Cooper, W.E., Boyce, S. (1983): Declination of Fundamental Frequency in Speakers Production of Parenthetical and Main Clauses. *JASA,* **73**, pp. 1731-1738.

Ladd, D. R. (1980): *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. R. (1983): Phonological Features of Intonational Peaks. *Language*, **59**, pp. 721-759

Ladd, D. R. (1986): Intonational Phrasing: The Case for Recursive Prosodic Structure. *Phonology Yearbook 3*, pp. 311-340.

Ladd, D. R., Scherer, K. R., Silverman, K. (1986): *An Integrated Approach to Studying Intonation and Attitude*. In Johns-Lewis, C. (ed.) 1986, pp. 125-138.

Ladd, D. R. (1993): Constraints on the Gradient Variability of Pitch Range. In Keating, P. (ed.), *Papers in Laboratory Phonology 3*, Cambridge: Cambridge University Press.

Ladefoged, P. (1975): *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

Lakoff, G. (1972): The Global Nature of the Nuclear Stress Rule. *Language*, **48**, pp. 285-303.

Laver, J., Hutcheson, S. (1972): *Communication in Face to Face Interaction*. Harmondsworth: Penguin.

Laver, J., Jack, M.A. (1987): *Proceedings of the First European Conference on Speech Technology*, Vols. **1&2**, Edinburgh: CEP.

Laver, J. (1988): Cognitive Science and Speech - a Framework for Research. *Work in Progress,* 21, pp. 83-114, Department of Linguistics, University of Edinburgh.

Laver, J. (1992): *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lee, W. R. (1956): English Intonation: A New Approach. *Lingua*, **5**, pp. 345-371.

Levelt, W. J. M. (1989): *Speaking: From Intention to Articulation*. Cambridge, Mass: MIT Press.

Levinson, J. (1986): *Punctuation and the Orthographic Sentence: A Linguistic Analysis*. Dissertation, City University of New York.

Liberman, M., Pierrehumbert, J. (1984): Intonational Invariance Under Changes in Pitch Range and Length. In Aronoff, M., and Oehrle, R. (eds.), *Language Sound Structure*, Cambridge, Mass: MIT Press.

Lieberman, P. (1963): Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech,* **6**, pp. 172-187.

Lieberman, P. (1967): *Intonation, Perception and Language*. Cambridge (Mass.): MIT Press.

Lyons, J. (1968): *An Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

Markwardt, A.H. (1942): *Introduction to the English Language*. Oxford: Oxford University Press.

McAteer, E. (1989): *Typeface effects in written language*. Ph.D. dissertation, University of Glasgow.

McAteer, E. (1992): Typeface Emphasis and Information Focus in Written Language. *Applied Cognitive Psychology,* Vol. 6, pp345-359.

McCawley, J. (1982): Parentheticals and Discontinuous Constituent Structure. *Linguistic Inquiry,* **13**, pp. 91-106.

McKeown, K., Swartout, W. R. (1988): Language Generation and Explanation. In Zock, M., Sabah, G. (eds.), *Advances in Natural Language Generation,* **1**, London: Pinter.

McLemore, C. (1991): *The Pragmatic Interpretation of English Intonation: Sorority Speech.* Ph.D. dissertation, University of Pensylvania.

Meyer, C. (1987): *A Linguistic Study of American Punctuation.* NewYork: Peter Lang.

Miller, J.E. (1990): Spoken and Written Language: How Serious is Plato's Problem? *Occasional Paper of the University of Edinburgh Linguistics Department,* University of Edinburgh.

Monaghan, A.I.C. (1988): Generating Intonation in the Absence of Essential Information. In Ainsworth & Holmes (eds.) 1989, **4**, pp.1249-1256.

Monaghan, A.I.C. (1990): A Multi-Phase Parsing Strategy for Unrestricted Text. *Proceedings of the ESCA Workshop on Speech Synthesis,* Autrans, France, pp. 109-112.

Monaghan, A.I.C. (1991): *Intonation in a Text-to-Speech Conversion System.* Ph.D. dissertation, University of Edinburgh.

Monaghan, A.I.C. (1991b): Heuristic Strategies for the Higher-Level Analysis of Unrestricted Text. In Benoit, C. & Bailly, G. (eds.), *Talking Machines.* Amsterdam: Elsevier.

Morris, C. W. (1938): Foundations of the theory of signs. Reprinted in Morris, C. W. (1971): *Writings on the general theory of signs*. The Hague: Mouton.

Murray, I.R., Arnott, J. L., Alm, N., Newell, A. F. A Communication System for the Disabled with Emotional Synthetic Speech Produced by Rule. *Proceedings of the Second European Conference on Speech Communication and Technology*, Genova, Italy, pp. 311-314.

Nunberg, G. (1990): *The Linguistics of Punctuation*. CSLI Lecture Notes, 18.

O'Connor, J. D., Arnold, G. F. (1961): *Intonation of Colloquial English*. London: Longman.

O'Malley, M.H., Resnick H., Caisse, M. (1991): An Anlysis of Strategies for Finding Prosodic Clues in Text. *Proceedings of the Second European Conference on Speech Communication and Technology*, Genova, Italy, pp. 1165-1168.

O'Shaughnessy, D. (1976): *Modelling Fundamental Frequency and its Relationship to Syntax, Semantics, and Phonetics*. Ph.D. dissertation, Massachussetts Institute of Technology.

Panasyuk, A. and Panasyuk, I. (1991): Prosodic Italics: Funtions and Phonetic Realization. *Proceedings of the XIIth International Congress of Phonetic Sciences*. Aix-en-Provence, France, pp. 270-273.

Paris, C. L., Swartout, W. R., Mann, W. C. (1991): *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Boston: Kluwer.

Pasdeloup, V., (1990): Multi-Style Prosodic Model for French Text-to-Speech Synthesis. *Proceedings of the ESCA Workshop on Speech Synthesis*, Autrans, France, pp. 193-196.

Pierrehumbert, J. (1980): *The Phonology and Phonetics of English Intonation.* Ph.D.dissertation, Massachusetts Institute of Technology.

Pierrehumbert, J., Steele, S.A. (1989): Categories of Tonal Alignment in English. *Phonetica*, **46**, pp. 181-196.

Pike, K. L. (1945): *The Intonation of American English.* Ann Arbor: University of Michigan Press.

Pitrelli, J.F., Beckman, M.E., Hirschberg, J. (1994): Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framewrork. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, pp. 123-126.

Postal, P.M. (1964): Constituent Structure, suppl. to *International Journal of American Linguistics*, **30**.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. (`1972): *A Grammar of Contemporary English.* London: Longman

Rice, E.R., Love, G. (1987): Electronic Emotion: Socioemotional Content in a Computer Mediated Network. *Communication Research*, **14 (1)**, pp. 85-108.

Ross, J. R. (1973): In McCawley (1982): Parentheticals and Discontinuous Constituent Structure. *Linguistic Inquiry*, **13**, pp. 91-106.

Sampson, G. (1985): *Writing Systems.* Stanford: Stanford University Press.

Sanderson, D. W. (1993). *Smileys.* ISBN: 1-56592-041-4.

Saussure, F. de. (1974): *A Course in General Linguistics.* Baskin, W., Bally, C., Sechehay, A. and Reidlinger, A. (eds.), Fontana Press.

Scherer, K.R. (1974): Acoustic Concomitants of Emotional Dimensions: Judging Affects from Synthsized Tone Sequences. In Weitz (ed.) *Nonverbal Communication*, pp. 105-111.

Scherer, K.R. (1980): The Functions of Nonverbal Signs in Conversation. In Clair, R. St., Giles, H. (eds.). *The Social and the Psychological Contexts of Language*. Hillsdale, N.J.: Erlbaum, pp. 225-244.

Schmerling, S. F. (1976): *Aspects of English Sentence Stress*. Austin: Texas University Press.

Schmidt, M.S. (1989): Problems in Generating the Pronunciation of Proper Names in Text-to-Speech Systems: An Alternative Approach. M.A. Dissertation, University of Edinburgh.

Schmidt, M.S. (1990): Looking Beyond the Sentence in Text-to-Speech Processing. *Work in Progress*, 23, pp. 74-84, Department of Linguistics, University of Edinburgh.

Schmidt, M.S., Watson, G.S. (1991): The Evaluation and Optimization of Automatic Speech Segmentation. *Proceedings of the Second European Conference on Speech Communication and Technology*, Genova, Italy, pp. 701-704.

Schmidt, M.S., Fitt, S., Scott, C., Jack, M.A. (1993): Phonetic Transcription Standards for European Names (ONOMASTICA). *Proceedings of the 3rd European Conference on Speech Communication and Technology*, Berlin, Germany, pp. 279-282.

Schmidt, M.S, Jack, M.A. (1994): A Multi-Language Pronunciation Dictionary for Proper Names and Place Names: ONOMASTICA. *Proceedings of the Language Engineering Convention*, Paris, France, pp. 125-128.

Selkirk, E.O. (1984): *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge, Mass: MIT Press.

Silverman, K. (1987): *The Structure and Processing of Fundamental Frequency Contours.* Ph.D. dissertation, University of Cambridge.

Silverman, K., Pierrehumbert, J.B. (1991): The Timing of Prenuclear High Accents in English. In Kingston, J. and Beckman, M. (eds.), *Papers in Laboratory Phonology 1,* Cambrigde University Press.

Silverman, K., Blaauw, E., Spitz, J., Pitrelli, J. (1992a): Towards Using Prosody in Speech Recognition/Understanding Systems: Differences between Read and Spontaneous Speech. *Fifth Darpa Workshop on Speech and Natural Language,* New York: Harriman.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. (1992b): TOBI: A Standard for Labeling English Prosody. *Proceedings of the International Conference on Spoken Language Processing,* Banff, Canada, pp. 867-870.

Smith, B. (1987): The Correspondence Continuum. *CSLI-87*-71, Centre for the Study of Language and Information.

Sorensen, E. K. (1991): *Dialogues in Networks: The Dynamic Nature of Asynchronous Linguistic Interaction in Computrer Mediated Communication Systems.* {(Draft copy)}

Spiegel, M. (1975): Probability and Statistics. *Schaum's Outline Series,* New York: McGraw-Hill Book Company.

Spitzer, M. (1986): Writing Style in Computer Conferences. *IEE Transcactions on Professional Communications*, Vol. PC29, No.1.

Sproull, L., Kiesler, S. (1986): Reducing Social Context Cues: Electronic Mail in Organisational Communication. *Management Science*, **32**, pp. 1492-1512.

Sproull, L., Kiesler, S. B. (1991): *Connections: New Ways of Working in the Networked Organization*. Cambridge, Mass: MIT Press.

Svartvik, J. and Quirk, R. (1980): *A corpus of English Conversation*. Lund Studies in English, **56**, Lund: CWK Gleerup.

Terken, J., Lemeer, G. (1988): Effects of Segmental Quality and Intonation on Quality Judgements for Texts and Utterances. *Journal of Phonetics*, **16**, pp.453-457.

Terken, J. (1991): Fundamental Frequency and Perceived Prominence of Accented Syllables. *JASA*, **89**, pp. 1768-1776.

Terken, J., v.d. Hombergh, K. (1992): Judgements of Relative Prominence for Adjacent and Non-Adjacent Accents. *Proceedings of the International Conference on Spoken Language Processing*, Banff, Canada, pp. 735-738.

't Hart, J. (1979): Explorations in Automatic Stylization of F0 Curves. *IPO Annual Progress Report*, **14**, pp.61-65.

Trager, G. L., Smith, H. L. (1951): *An Outline of English Structure*. Washington: American Council of Learned Societies (reprinted 1957).

Vaissière. J. (1983): *Language-Independent Prosodic Features*. In Cutler & Ladd (eds.) 1983, pp. 53-66.

Vonwiller, J. P., King, R. W., Lloyd R. W. T. (1991): Automatic Prosody Assignment for Interactive Synthesized Dialogue Systems. *Proceedings of the Second European Conference on Speech Communication and Technology*, Genova, Italy, pp. 1283-1286.

Waibel, A. (1988): *Prosody and Speech Recognition.* London: Pitman

Waller, R. (1988): *The Typographic Contribution to Language.* Ph.D. dissertation, University of Reading.

Wells, R. (1945): The Pitch Phonemes of English. *Language*, **21**, pp.27-40.

Wells, R. (1947): In McCawley (1982): Parentheticals and Discontinuous Constituent Structure. *Linguistic Inquiry*, **13**, pp. 91-106.

Willemse, R., Boves, L. (1991): Context Free Wild Card Parsing in a Text-to-Speech System. *Proceedings of ICASSP*, 1991, **2**, pp.757-760.

Williams, C. E., Stevens, K. N. (1972): Emotions and Speech: Some Acoustical Correlates. *JASA*, **52**; 4, pp.1238-1250.

Yates, S. (1992a): *Computer Conferencing: A Written or Spoken Environment.* Centre for Information Technology in Education, Report No. 151, The Open University.

Yates, S. (1992b): Conflict and Ambivalence over Oral and Literate Boundries: An Examination of 'Literacy Events' in Computer-mediated Communication. Centre for Language and Communication, *Occasional Papers in communication*, No. 35, The Open University.

Youd, N. (1992): *The Production of Prosodic Focus and Contour in Dialogue.* Ph.D. dissertation, The Open University.