Disfluency and listeners' attention: An investigation of the immediate and lasting effects of hesitations in speech

Philip Collard

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

to

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

2009

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Philip Collard

Abstract

Hesitations in speech marked by pauses, fillers such as *er*, and prolongations of words are remarkably common in most spontaneous speech. Experimental evidence indicates that they affect both the processing of speech and the lasting representation of the spoken material. One theory as to the mechanisms that underlie these effects is that filled pauses heighten listeners' attention to upcoming speech. For example, in the utterance:

(1) She hated the CD, but then she's never liked my taste in *er* music

The hesitation marked by the filler *er* would heighten listeners' attention to the post-disfluent material (*music*) which would then be processed and represented differently to an equivalent stimulus in a passage of fluent speech.

The thesis examines this proposition in the context of an explicit definition of attention. The first half of the work investigates whether hesitations heighten two different aspects of listeners' attention: these are the immediate engagement of attention to post-disfluent stimuli at the point they are encountered, and the continued attention to the representation of stimuli after they are encountered.

In experiment 1, a speech 'oddball' paradigm is used to show that event-related potentials (ERPs) associated with attention (MMN and P3) are affected by a preceding hesitation, indicating an immediate effect of hesitations on listeners overt attention. Experiments 2 and 3 use behavioural responses and eye-movements measures during a change-detection paradigm. These experiments show that there is also an effect on the listeners' attention to the post-disfluent material after the initial presentation of the utterance.

The second half of the thesis concerns itself with the timecourse of the attentional effects. It addresses questions such as: how long-lived is the attentional heightening and what is the attentional heightening trigger? Experiments 4–7 explore the relationship between the filler *er* and periods of silent pause that surround it. Behavioural (exp. 4–6) and ERP (exp. 7) results show that while extending the period of silence after the filler *er* does not affect the immediate engagement of attention, it will affect subsequent attention to the post-disfluent material: constituents that are not immediately preceded by the filler *er* are not attended to in an enhanced way.

Together, these experiments confirm the proposition that hesitations heighten listeners' attention to upcoming speech. The thesis outlines the ways in which the components of this attentional heightening are differentially affected by interaction between the content and timing of the hesitations encountered. Attention has an important role to play in the processing of any stimulus. Using disfluency as a test case, this thesis illuminates its importance in language comprehension.

Acknowledgements

I would like to thank a number of organisations for support during the research and writing of this thesis. The College of Humanities and Social Science, the School of Philosophy, Psychology and Language Science at the University of Edinburgh, and the Overseas Research Students Awards Scheme, all contributed to the funding for this research.

I would also like to express my continued gratitude to a number of individuals. Foremost, I would like to thank Dr. Martin Corley for his supervision and support at all stages of the research and writing. The work was also made possible through the continued support provided by Prof. David Donaldson and the members of his Psychological Imaging Laboratory team at Stirling University.

I am grateful to a number of colleagues at the University of Edinburgh, particularly those within the Edinburgh Disfluency Group and Dr. Patrick Sturt. Special mention goes to Dr. Lucy MacGregor, whose input at many stages in the work proved invaluable. I would like to thank Dr. Manon Jones for providing the voice for Experiments 1 and 7 and Suzy Moat for not only the voice for Experiments 2–6, but the continued advice and support throughout the entire Ph.D. process.

Finally, I would like to thank my family for support from afar and Frances Bentley for support and encouragement close to home, and when I needed it the most.

Contents

Declaration i

Abstract ii

Acknowledgements iv

Chapter 1 Introduction 1

1.1 Thesis overview 2

Chapter 2 Disfluencies 4

2.1 Chapter overview 4

2.2 What are disfluencies? 4

2.2.1 Language production 5

2.2.2 Repetitions 6

2.2.3 Overt repairs 7

2.2.4 Hesitations 8

2.2.5 Summary of disfluency classification 9

2.3 Why hesitations are produced 10

2.3.1 Hesitations due to problems with conceptualisation 11

2.3.2 Hesitations due to problems with formulation 11

2.3.3 Individual differences in disfluency production 13

2.3.4 Intentionality and disfluency production 13

2.4 How hesitations affect a listener 14

2.4.1 Hesitations and markers of disfluency 15

2.4.2 Hesitations and speaker confidence 16

2.4.3 Hesitations and listeners' attention 17

2.4.4 Hesitations and prediction 20

2.4.5 Hesitations and syntactic structure 23

2.4.6 Summary of the effects of hesitations on listeners 24

Chapter 3 Attention 26

3.1 Chapter overview 26

3.2 Defining attention 26

3.3 Directing listeners' attention to speech 28

3.4 Directing listeners' attention during speech 30

3.4.1 Covert attention 31

3.4.2 Good-enough processing and degrees of specification 33

3.4.3 Attention and eye movements during reading 38

3.5 Interim discussion 40

3.6 Neurocognitive approaches to attention 40

3.6.1 Orienting 41

3.6.2 Target detection 42

3.6.3 Alertness 42

3.6.4 Orienting, target detection and alertness 42

3.7 Conclusions 44

Chapter 4 Methodologies 45

4.1 Chapter overview 45

4.2 ERPs and language comprehension 45

4.2.1 N400 47

4.2.2 P600/SPS 49

4.2.3 LPC 51

4.3 ERPs and Attention 52

4.3.1 The MMN 53

4.3.2 The P3 (or P300) 54

4.4 Change detection 57

4.4.1 Change detection for visual scenes 58

4.4.2 Text change detection 60

4.4.3 Same/different matching tasks 63

4.5 Conclusion 64

Chapter 5 Experiment 1 65

5.1 Chapter overview 65

5.2 Experiment 1 65

5.2.1 Method 68

5.2.2 ERP Results 71

5.2.3 Memory Results 77

5.3 Discussion 79

5.4 Conclusion 83

Chapter 6 Hesitations and change detection 85

6.1 Chapter overview 85

6.2 Introduction 85

6.3 Experiment 2 89

6.3.1 Method 90

6.3.2 Results 93

6.3.3 Discussion 95

6.4 Experiment 3 95

6.4.1 Method 96

6.5 Results 99

6.5.1 Behavioural results 99

6.5.2 First-fixation time 100

6.5.3 First-pass time 101

6.5.4 Total time spent in region 102

6.5.5 Number of fixations 103

6.5.6 Probability of regression back into the region 105

6.6 General discussion 107

6.7 Conclusion 109

Chapter 7 Triggers of attentional effects 110

7.1 Chapter overview 110

7.2 Triggers 111

7.3 Experiment 4 114

7.3.1 Method 115

7.3.2 Results 117

7.3.3 Discussion 117

7.4 Experiment 5 119

7.4.1 Method 120

7.4.2 Results 121

7.4.3 Discussion 122

7.5 Experiment 6 124

7.5.1 Method 124

7.5.2 Results 126

7.5.3 Discussion 127

7.6 General discussion 128

Chapter 8 Experiment 7 131

8.1 Chapter overview 131

8.2 Experiment 7 132

8.2.1 Introduction 132

8.2.2 Method 134

8.2.3 ERP Results 136

8.2.4 Memory Results 143

8.2.5 Discussion 144

8.3 Conclusion 148

Chapter 9 Discussion 149

9.1 Chapter Overview 149

9.2 Summary of the findings 149

9.3 Interpretation of the findings 151

9.3.1 Which attentional mechanisms are being affected by hesitations? 152

9.3.2 Which aspects of a hesitations trigger the attentional effects? 155

9.4 Future Research and Conclusions 156

Appendix A Stimuli for Experiments 1 and 7 159

Appendix B Stimuli for Experiments 2–6 167

Appendix C Collard, Corley, MacGregor, and Donaldson (2008) 172

References 180

List of Figures

5.1 Topographic maps illustrating the mean distributions of the deviance effects for fluent and disfluent utterances from Experiment 1 72

5.2 Grand average ERPs showing the deviance effect for fluent utterances from Experiment 1 74

5.3 Grand average ERPs showing the deviance effect for disfluent utterances from Experiment 1 76

5.4 Recognition probabilities for the fluent and disfluent conditions from the memory test in Experiment 1 78

6.1 Mean probability of accuracy for the change-detection task in fluent and disfluent conditions in Experiment 2 94

6.2 Mean probability of accuracy for the change-detection task in fluent and disfluent conditions in Experiment 3 100

6.3 Mean first-fixation time spent in the target region in fluent and disfluent conditions in Experiment 3 101

6.4 Mean first-pass reading time spent in the target region in fluent and disfluent conditions in Experiment 3 102

6.5 Mean total time spent fixating on the target region in fluent and disfluent conditions in Experiment 3 103

6.6 Mean number of fixations in the target region in fluent and disfluent conditions in Experiment 3 104

6.7 Mean proportion of trials where the participant made regressive eye movements back into the target region in fluent and disfluent conditions in Experiment 3 105

7.1 Mean probability of accuracy for the change-detection task in fluent and disfluent conditions in Experiment 4 118

7.2 Mean probability of accuracy for the change-detection task in fluent and disfluent conditions in Experiment 5 122

7.3 Mean probability of accuracy for the change-detection task in fluent and disfluent conditions in Experiment 6 127

8.1 Grand average ERPs showing the deviance effect for fluent utterances from Experiment 7 138

8.2 Grand average ERPs showing the deviance effect for fluent utterances from Experiment 7 139

8.3 Topographic maps illustrating the mean distributions of the deviance effects for fluent and disfluent utterances from Experiment 7 142

8.4 Recognition probabilities for the fluent and disfluent conditions from the memory test in Experiment 7 143

Chapter 1

Introduction

During spontaneous speech, speakers face the difficult task of rapidly deciding what to say and how to say it. One consequence of the speed at which speakers are required to do this is that they are often disfluent, producing phrases that repeat or correct what has already been said, as well as meaningless interjections such as *um* and *er*. Listeners encounter these disfluencies with such regularity that their effects on speech processing are of natural interest to those who research spoken-language comprehension.

This thesis explores the effect of hesitations on the listener. Specifically, it investigates whether hesitations affect listener attention. Previous research has led to suggestions that hesitations "*heighten listeners' attention for upcoming speech*" (Fox Tree, 2001, p. 325), though this claim has yet to be investigated in the context of a clear and defined account of attention. The present thesis takes as its starting point a model of attention, and investigates the effects of disfluency on listeners using neural and behavioural indices of attentional modulation.

1.1 Thesis overview

The aim of this thesis is to not only explore the effect of hesitations on listeners' attention, but to do so in the context of a clear conceptualisation of attention. Chapters 2 and 3 provide a discussion of the research on disfluency and attention. In Chapter 2, a basic taxonomy of disfluencies is provided, and both the production of disfluency and the effects of disfluency on the listener are discussed. Then, the main disfluencies of interest, hesitations, are explored in depth, with specific reference to how these affect a listener. In Chapter 3, attention is discussed with particular emphasis on how attentional processes might operate during speech comprehension.

Chapter 4 explores some of the paradigms that are used in the experimental work reported in this thesis. Event-related potential (ERP) research on both language and attention is discussed, with a particular view to introducing some of the issues and studies that spawned the current research. Then, the use of change-detection paradigms in studies of both language processing and visual attention is discussed.

Chapters 5 and 6 outline the first experiments using these paradigms. In Chapter 5, Experiment 1 addresses the question of what happens at the point at which a disfluency has been encountered. In this experiment, neural indices of attention capture extracted from EEG recorded while participants listen to both fluent and disfluent linguistic stimuli. In Chapter 6, Experiments 2 and 3 use a change-detection paradigm to investigate whether this immediate attentional engagement also results in increased covert attention or an increase in attention to post-disfluent material after it has been mentioned.

Chapters 7 and 8 explore which aspect of disfluency triggers the attentional effects observed in Chapters 5 and 6. In Chapter 7, Experiments 4–6 investigate the fillers in hesitations, and the periods of silent pause that often flank these fillers, exploring how these two components of hesitations interact and how this interaction triggers heightened covert attention. Chapter 8 presents a further ERP experiment which investigates whether immediate attentional enhancement is also sensitive to the intervening time between the filler *er* and the resumption of fluent speech.

Finally, Chapter 9 provides a summary of these findings and an initial interpretation of the results, fitting these into a context of both language processing and attention.

Chapter 2

Disfluencies

2.1 Chapter overview

Having introduced the general themes of this thesis in the previous chapter, the current chapter provides a more in-depth look at disfluencies and aims to give a concise history of the research that has lead to the current state of understanding of these speech phenomena. First, a basic taxonomy of disfluencies is outlined — with each type of disfluency explained in more depth and with specific emphasis on issues related to the experiments outlined in this thesis. Focus is given to hesitation disfluencies which are the disfluencies of primary interest in the current work. After this, research and theory addressing why speaker produce hesitations is outlined. To conclude, the existing research on the major theme of this thesis, 'how do hesitations affect a listener' is discussed.

2.2 What are disfluencies?

The skill of speaking fluently is taught at an early age around much of the world. Even at the level of primary education it is common for students to be taught that when speech is going well, speakers will say only what they need to with no repetition, they will not leave sentences unfinished and will not pause unnecessarily or *um* and *er* while they figure out what to say next. In the simplest sense, disfluencies are phenomena that make speech deviate from this fluent ideal.

One useful definition of disfluencies is that they are parts of an utterance that interrupt the flow of continuous speech and do not add propositional content to the utterance (Fox Tree, 1995). While this definition might not hold strictly true for all things that are referred to as disfluencies, it provides a good basis for the investigation of these speech phenomena outlined in this thesis.

Defined in such terms, disfluencies are common phenomena in normal, day-to-day speech. Even excluding silent pauses, the estimated rate of disfluency production is around six per hundred words (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995; Lickley, 1995; Shriberg, 2001). Below, a classification of disfluency is outlined. While this classification is formulated with a mind to the important issues discussed in this thesis, it is also representative of general consensus among disfluency reseachers. Here, disfluencies are catagoriesed into three groups: repetitions, repairs and hesitations. As previously mentioned, for the purposes of introducing the issues that are of interest in the current thesis, special focus is given to hesitation phenomena. Because this classification of disfluencies is not only reliant on the overt form the disfluency takes, but also the function that they serve during the production process, a brief outline of the production process is given first.

2.2.1 Language production

For the purposes of explanation, the work in this thesis assumes a three stage model of language production (such as the models supplied by Levelt, 1983, 1989). The first stage of the production process involves a speaker planning an utterance. The speaker will conceptualise the intended message in a pre-verbal form. The next stage involves transforming this pre-verbal message into a verbal plan. This includes selecting appropriate word forms and syntactic structures as well as selecting phonological and articulatory forms appropriate to the message. The final stage in this process is the actual articulation of the message.

An important point to note about this three stage model is that the speakers are able to monitor not only the final overt speech, but also their internal speech before it is articulated. This is relevant to the current thesis as some disfluencies are characterised as products of production problems that are 'caught' by the speaker before the erroneous material is articulated. This is discussed further during the classification.

2.2.2 Repetitions

Considering only overt speech acts, repetitions can be viewed simply as repeated syllables, words or phrases, where the repetition does not add to the propositional content of the utterance. In (1) below for example, the speaker might be repeating '*in*' as he or she 'struggles to find the word he or she is looking for'.

(1) I found it difficult in. . . in Berlin because I couldn't speak German.

Investigating this phenomenon further, Heike (1981) provides a sub-classification of repetitions based on the function of the disfluency. Heike (1981) points out that there should be a distinction drawn between prospective and retrospective repetitions. Prospective repetitions are classified as those that are introduced because of perceived upcoming difficulty for the speaker. Retrospective repetitions occur when the speaker has detected that a problem has already occurred. If a speaker realises that an incorrect word has been spoken or that a hesitation has occurred, they can use the repetition of material to re-establish the fluency of their speech. There are empirical findings that support this form of classification. Plauche and

Shriberg (1999) provide evidence that these two types of repetition are distinct in form: the pattern of the fundamental frequency of the speech differs between the two types (see also, Shriberg, 1995).

2.2.3 Overt repairs

Overt repairs can be loosely described as an event where a speaker stops mid speech, and then corrects them self by providing new information intended to replace erroneous or unintended speech that they have already produced (see below, for some discussion of covert repairs). With the three stage model of production in mind, these overt repairs occur when a speaker detects an error after the articulation of the message has commenced.

The entire disfluency is best examined in three constituent parts, the initial speech that the speaker intends to correct, an editing phrase or pause, and then the correct information. Levelt (1983) provides a more precise terminology for these. The larger body— referred to as a repair utterance—is partitioned into three constituent parts: a reparandum; an editing term (or pause); and finally, the repair proper.

(2) I found it difficult in Berlin because I couldn't understand . . . because their accent was so impenetrable.

In (2), the repair utterance starts with the initial 'because'. The phrase 'because I couldn't understand' is the reparandum: that is, the speech that the speaker intends to correct. This is followed by an edit interval. In this case, the pause transcribed by the ellipsis '. . . '. This interval marks the point at which the apparent fluent speech is interrupted. It is sometimes unmarked, though silent and filled pauses may occur here. Following this edit interval is the repair. This is made up of information that replaces the reparandum in the intended utterance (the phrase 'because their accent was so impenetrable', in the example above).

2.2.4 Hesitations

In the context of this thesis the label 'hesitation' is used as an umbrella term for at least three types of disfluency that often co-occur, these are: pauses; non-lexical fillers; and prolongations. Pauses are simply periods of time (or gaps) between the constituent words of an utterance. These gaps might be marked by periods of silence, in conjunction with or in the absence of non-lexical fillers such as *er* or *um* (or any other typological notation of the same sounds such as *uh*, *ah*, *em* or *mm*). Totally silent pauses can present a problem for studies of disfluency as the natural prosody of an utterance might result in an intended, fluent pause between constituents. This issue is discussed at several points throughout this thesis.

Often co-occurring with these filled or silent pauses are prolongations. These hesitations are marked by syllabic lengthening, such as when a speaker pronounces the indefinite article *a* in a long form (such as the vowel that appears in the word *hay*) or the definite article *the* with a long vowel (*thee*). Further discussion of exactly what hesitations are continues throughout this thesis.

One further point that is important to note about hesitations is that there is a history of a further sub-classification of pauses in the literature. For example, O'Shaughnessy (1992) refers two groups of pauses: grammatical (between-clause) pauses and ungrammatical (within-clause) pauses. This classification is based on the assumption that, rather than interrupting fluency, between-clause pauses may be used intentionally by the speaker to convey their utterance in an easy-to-understand way and that within-clause pauses are necessarily unintentional and a result of production difficulty (see also, Ferreira, 2007). Beyond this assumption, there are several reasons to think that these are two at least partially distinct phenomena. For example, Kircher, Brammer, Levelt, Bartels, and McGuire (2004) showed that there are differing patterns of brain activation associated with the production of these two types of pause. Between-clause pauses were associated with more activation in the left temporal cortex, a finding consistent with the idea that these pauses are associated with the process of lexical retrieval — or more specifically, problems with this process.

2.2.5 Summary of disfluency classification

Having reviewed this taxonomy of disfluency, it is evident that this is only one possible classification system of many. For example, *um*s, *er*s and silent pauses might be classified as hesitations because they have the shared characteristic of seeming to 'buy the speaker more time' before he or she is required to continue fluent speech. However, in this scenario, prospective repetitions fulfill a role much the same as the role of hesitations. Despite these complications, the focus here will be on hesitations, with other types of disfluency mentioned only where relevant: either because research being referenced does not explicitly distinguish between hesitations and other types of disfluency or because it examines disfluencies where hesitations co-occur with other types of disfluency.

Having provided this classification, the next obvious area of interest relating to disfluencies is the question of why they are produced. Simply put, repairs occur because some incorrect phonetic plan has already started to be executed. When this incorrect plan is detected by the speaker's self monitoring, he or she will need to back-track and correct the incorrect information. This theory is backed up by experimental evidence from authors such as Postma and Kolk (1992) who showed that interfering with speakers' ability to monitor their own speech using noise masking reduced the rate of self-repairs in their speech. As for the cause of hesitations, there is less consensus and a greater number of alternative accounts. This is due in part to the fact that hesitations could reflect problems at either of the stages of production that precede articulation. The speaker could be struggling to formulate a pre-verbal concept, struggling to specify the linguistic form of a concept or even covertly repairing a message that they have found to be erroneous before it is articulated. It is to these issues that the discussion now turns.

2.3 Why hesitations are produced

Hesitations may well result from phenomena relating to a number of distinct speech processes. One reason that hesitations occur is because the speaker has no correct phonetic plan ready to be executed. Unlike overt repairs, hesitations (like repetitions) might be marking instances of 'covert' repair. That is, the presence of repairs that take place before the speaker has actually executed the incorrect phonetic plan. As previously mentioned, in many models of speech production (e.g. Levelt, 1989) speakers monitor their speech not only by an external medium (how their speech sounds to them) but also by an internal medium (is their speech plan being executed properly). If a speaker finds them self about to produce an incorrect utterance, they might suspend fluent speech until the appropriate speech plan is reformulated. In simple terms, they might just be hesitating because they need time to reformulate an articulatory plan to replace the erroneous one (Levelt, 1983). This explanation does however make certain assumptions about the capabilities of the internal monitoring process. Limits to these capabilities—for example, the monitor not having access to the original speech plan—might influence what type of covert repairs are possible (Levelt, 1983).

Looking one step further than this for an explanation leads to the question, when might a speaker have these problems in accessing an appropriate articulatory speech plan? There are two major reasons why a speaker may have difficulty in forming such a plan for execution at the right time in a speech act. They might not know what they want to say. Or alternatively they might know that, but not how they want to say it. In the formal framework of the three stage model of production like the one presented above, these amount to problems in conceptualisation and formulation.

2.3.1 Hesitations due to problems with conceptualisation

When a speaker is having difficulty formulating the concepts that they wish to express, this can result in disfluency. Evidence for this comes from both observational studies and formal experiments. As an example of the first, speakers rate themselves as less likely to know the answer to a question when they are being disfluent. Smith and Clark (1993) also showed that answers to general knowledge questions were more likely to be incorrect if the speaker was being disfluent. Using the Network task (Levelt, 1983), Schnadt and Corley (2006) provided experimental evidence showing that increasing the difficulty involved in conceptualising utterances leads speakers to be more disfluent. In this task, participants are asked to describe aloud the route a visual marker makes as it moves around a network of pictured objects. Schnadt and Corley show that when conceptualisation is made more difficult by visual blurring of the pictured objects, hesitations (especially prolongations) are produced more frequently.

2.3.2 Hesitations due to problems with formulation

Even if a speaker successfully conceptualises what he or she intends to say, hesitations might also arise from the presence of too many options about the way to formulate these ideas into speech. An example of where too many options can lead to hesitation is provided by Schachter, Christenfeld, Ravina, and Bilous (1991). Schachter et al. (1991), showed that while lecturers in the natural sciences, the social sciences and humanities did not have differing disfluency rates when interviewed on general topics, the rate of fillers used differed when they discussed their own field of expertise. Lecturers in the humanities had the highest rate of filler use, followed by those in the social sciences, who in turn had a higher rate than lecturers in the natural sciences. Schachter et al. (1991) attributed this to the number of linguistic options available in each field, humanities having the highest followed by social sciences and then natural sciences. In a later study, Schachter, Rauscher, Christenfeld, and Crone (1994) reviewed various media from this field and showed that the hierarchy pattern of linguistic options in each field was correct. Taken together, these two studies strongly implicate the idea that when more options for expressing ideas exist, the speakers tended to be more disfluent.

Formulation problems that lead to hesitations have also been shown to arise when speakers are having problems selecting a word to use or assembling a grammatical structure that is appropriate to their intended utterance. At the level of word selection, hesitation rates increase when the speaker is intending to use a low-frequency word (which are typically viewed as harder to access; Levelt, 1983). In addition to this, contextual probability plays a role. Beattie and Butterworth (1979) showed that even when lexical frequency is accounted for, contextual probability still predicts the presence of hesitations. Speakers tended to be more disfluent when attempting to access words that were of low contextual probability.

Evidence suggesting that the burden of constructing complex syntactic structures might result in hesitations is found in a number of studies. Fillers are highly likely to occur at phrase and clause boundaries (Boomer, 1965; Hawkins, 1971). These boundaries are associated with the introduction of new idea units (Butterworth, 1975), and are assumed to require syntactic formulation at the point at which they are introduced. Further evidence for this is provided by Maclay and Osgood (1959) who show that pauses occur more often before complex syntactic structures.

2.3.3 Individual differences in disfluency production

Having assessed the reasons why speakers might make hesitations, it is worth noting that speakers differ in the type and rate of disfluency that they produce. Reviewing data from the switchboard corpus (Godfrey, Holliman, & McDaniel, 1992), Shriberg (1996) coined the term ‘deleters’ and ‘repeaters’ to describe two apparently distinct groups of speakers. Deleters produce a higher rate of complex repairs, where as repeaters produce more repetition disfluencies. Because these two groups differ in speech rate, she suggests that the deleter/repeater distinction reflects differing production strategies. Faster speakers might over estimate their ability to formulate an utterance in time. They would then be likely to have to retrace and repair their speech a lot. Slower speakers might take more time to formulate ideas and this process will be marked by speaker hesitation. When considering this data with regard to the work in the current thesis, it should be noted that that she includes hesitations in her definition of repetitions and also points out that with regards to filled-pause production specifically, the two speaker types do not differ. She also notes individual differences regarding gender where males are more likely to use filled pauses than females – though she suggests social factors such as the desire to ‘hold the floor’ rather than factors concerning speech formulation might account for this.

2.3.4 Intentionality and disfluency production

One issue of contention within the literature on disfluency is about the possible intentionality of their production. Causal support for such a view can be gathered from observations such as the fact that speakers will tend to produce more disfluencies when they are addressing other humans than when they are addressing machines or when they are in monologue situations (Oviatt, 1995). Further to this, factors determining the speaker’s role in a dialogue situation such as whether the speech is descriptive or goal directed will affect their disfluency rate (Bortfeld et al., 2001).

Perhaps the most extreme formulation of the intentional signal view of disfluencies comes from Clark and Fox Tree (Clark, 1994; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). They have advocated that some disfluencies such as repetition and filled-pause hesitations are at times used in a way that intentionally conveys information to a listener. According to this view of disfluency, such phenomena, while not being considered as part of the primary message, should be viewed as part of a collateral message that is providing the listener with useful information about the state of the speaker. A simple example of this would be a situation where a speaker intentionally says *um* in order to let the speaker know that he or she had not finished speaking (i.e. did not want to concede the floor). Further to this, Clark and Fox Tree have also suggested that a speaker will use an *uh* to signal that the pause will be short and an *um* to signal that the pause will be longer. This idea that disfluencies can be signals is not widely accepted by researchers in the area (see O'Connell & Kowal, 2004). The focus of this thesis is the effects hesitations have on listeners and in light of this, the debate about the intentionality of disfluencies is largely inconsequential and is not examined in any further depth throughout this thesis. As is outlined below, hesitations have a number of measurable effects on listeners, and these do not seem dependent upon any intentions of the speaker.

2.4 How hesitations affect a listener

Having examined the production of hesitation disfluencies, this chapter now goes on to examine how these hesitations affect the listener, from their immediate processing of the disfluent speech to their long term representation of the material that they heard. Various groups of researchers have approached the phenomenon from different perspectives, and thus provided different classifications of the disfluencies that they study. For example, research has investigated questions such as: are *er*s treated like words by the listener; how do disfluencies affect a listeners confidence in what the speaker is saying; and what do disfluencies signal to a listener? Despite these distinct approaches, there is clear progress towards a general understanding of how hesitations affect processing.

2.4.1 Hesitations and markers of disfluency

As previously mentioned, one question that has been put forward regarding disfluency production concerns whether or not disfluencies can be considered as words. To paraphrase this question: are disfluencies intentional communicative acts produced by a speaker? With regards to the comprehension of disfluency, this proposition is not important in itself, though it has motivated a number of studies concerned with finding a 'marker' for disfluency. This marker would supposedly be something that will identify disfluencies as entities that stand out from other fluent speech material.

In the area of human speech perception, the search for a marker of disfluency was motivated by the evidence that humans seem to perform poorly when asked to specifically identify disfluency. Lickley (1995) conducted a study where speakers of Dutch were asked to explicitly identify disfluencies in a set of spoken monologues. Participants were asked to follow material which included naturally occurring disfluencies and were provided with the transcription of the passages where the disfluencies had been removed. Their task was to follow the speech and the transcription simultaneously and indicate the points where the two differed. Lickley found that nearly half (45%) of filled pauses went undetected and that filled pauses that occurred within sentences were more likely to be missed than those that occurred between sentences (35% and 49% respectively).

The fact that such a marker would be a valuable tool for artificial speech recognition (ASR) has also motivated studies of the disfluency. ASR parsers are easily disrupted by any input which is not sanitised fluent material. For example, the failure to identify reparanda, whether marked by hesitations or not, could lead to parsing that does not convey the intended meaning of the speech. Furthermore, the failure to identify filled pauses could lead to problems in faithful segmentation of the speech stream into constituent words. To this end, much research has been conducted to identify an acoustic or structural feature common to disfluency that could trigger an editing process. The general conclusion reached by the work of Lickley and colleagues (Lickley, 1995; Lickley & Bard, 1996) is that no such common acoustic or structural feature exists. Consequently, this research provides valuable insight into the ways in which listeners remain unaffected by hesitations.

In addition to the data reported in Lickley (1995), there are many other examples of instances where listeners show an ignorance of hesitations. When asked to explicitly repeat or identify the position of filled or unfilled pauses, listeners perform very poorly (e.g., Duez, 1985; Lindsay & O'Connell, 1995). Such results might lead to the conclusion that listeners are not affected by hesitations at all. However, when more subtle measures of listener susceptibility are considered, it becomes apparent that hesitations affect a listener in some interesting ways. While hesitations might not be breaking into the conscious perception of the listener, there are numerous examples that they affect both the immediate language processing and the subsequent cognitive processes involved in listening to speech.

2.4.2 Hesitations and speaker confidence

When asked to rate a speaker's confidence in what they say, listeners are affected by the presence of filled and silent pauses. Brennan and Williams (1995) asked listeners to explicitly rate how confident they were in the answers that speakers gave to general knowledge questions. These answers were manipulated so that in certain conditions they were preceded by either filled or silent pauses. Intermixed with these were regular fluent answers. Brennan and Williams observed that listeners were less confident in the speakers answer when these were preceded by hesitation and that filled-pauses showed a stronger effect than silent pauses. Similar effects can lead to listeners believing that disfluent speakers are less honest (Fox Tree, 2002) and even less intelligent (Christenfeld, 1995) than when the same speakers are speaking fluently.

While this shows that hesitations can affect a listener and change their 'feeling of another's knowing', further studies have concentrated on phenomena that are arguably more central to the concerns of Psycholinguistics. These have included questions such as how hesitations affect listeners' understanding the structure of what another is saying, and how hesitations might modulate listeners' attention or linguistic predictions. Research addressing these questions is reviewed below.

2.4.3 Hesitations and listeners' attention

As part of the general body of research by Clark and Fox Tree, work regarding the possibility that disfluencies might be an intended act by a speaker in order to help a listener, Fox Tree (2001) reported a study that addressed whether or not comprehension was facilitated by types of filled pauses, *um*s and *uh*s (the filled pauses transcribed in British English as *er*s). If disfluencies are intended to help the speaker, then they should be beneficial to the listener in some way.

Fox Tree asked participants to monitor for target words while they listened to excerpts from a speech corpus (face-to-face stories between students collected by Herb Clark at Stanford University). Intermixed with several filler utterances were critical utterances (like (3) and (4) below) that had originally contained either the filled pause, *um* or *uh* (see the examples below). In a balanced design, each participant heard an equal number of critical utterances with *um*, utterances with *uh*, utterances where an *um* had been excised, and utterances where the *uh* had been excised.

(3) And he said why sure well what kind of *um* **price** range are you looking for?

(4) Then he also tol- sold her on *uh* a couple of *uh* **furniture** items for the ant

For each of the critical utterances, the target word (bold in the examples) was either the word preceded by the critical filled pause or the word that had been preceded by the filled pause before it was excised out. The reaction times for the participants to respond to each of these target words were measured. Fox Tree's results showed that participants were faster to react to the target words when they were preceded by *uh* than when the *uh*s were excised (554ms and 601ms respectively). Participants did not differ significantly in their reaction times to target words preceded by *um* and those where the *um* had been edited out (561ms and 548ms respectively). A second experiment conducted in Dutch reproduced this pattern of results.

Fox Tree's interpretation of these results is that they are due to the fact that listeners use *um* and *uh* as indicators of upcoming delays. As *uh* usually signals a short delay, she suggests that this filled pause "*heightens listeners' attention for upcoming speech*" (p. 325). This is supported by the fact that participants in her experiment identified targets faster after *uh*s then they did in the same sentences where *uh* had been excised. As *um* usually signals a longer delay, she suggests that such heightened attention to upcoming speech is not useful. Again, this is supported by her experimental results.

Additional work showing that pauses might help listeners was conducted by Howell and Young (1991). They showed that repairs preceded by hesitations were rated as easier for listeners to understand than repairs without preceding hesitations. Brennan and Schober (2001) systematically addressed this issue of facilitation in an online task that used fluent and disfluent instructions. They had participants respond to simple instructions in a task where the goal was to identify a single candidate from a set of geometrically-simple shapes. In some cases, the instructions were fluent while in others, they contained filled pauses or silent pauses immediately preceding the speech material important for the task. For example, they played participants instructions such as:

(5) Move to the yel- purple square

Importantly, Brennan and Schober controlled the length of these pauses so that the only systematic difference between the filled and silent pause condition was the phonetic form (the presence of the filler *uh*). Both speed and accuracy of the responses were higher for the utterances that contained pauses than for the fluent utterances and there was no difference between the pause types.

One important point to address with regards to this work is that the pauses were recorded as corrections to utterances that contained previous instructions (e.g. move to the yell - *uh* - purple square). Because of this, the possible set of referents for the correct utterances was smaller for the condition with the pauses than it was for the fluent utterance. Given this, it is difficult to conclude whether the facilitated response to the instructions was specifically due to the hesitation, or whether it was due to the fact that the correction meant that there was one less possible referent (in an already small group of possible referents) for the participant to identify. In light of this confound, it is difficult to say how informative the results are with regards to the filled and silent pauses.

Approaching the topic of facilitation from a different angle Arnold, Tanenhaus, Altmann, and Fagnano (2004) investigated how disfluencies might help listeners predict what is going to be mentioned in up coming discourse. This research is reviewed below.

2.4.4 Hesitations and prediction

Arnold et al. (2004) monitored participants eye movements while they followed instructions about the manipulation of a simple array of objects pictured in front of them. For example:

(6) Put the grapes below the(*e uh*) candle

In a fully crossed design, they manipulated whether the utterance was fluent or disfluent preceding the critical item (candle in the example above) and whether the critical word was given or new to the discourse. They suspected that, because new items are harder for a speaker to access than given entities (Arnold, Wasow, Losongco, & Ginstrom, 2000), listeners would interpret disfluent items as an indication that the speaker intends to mention a new item. Owing to the idea that fixations are affected by lexical access (the participant will look at an object if the speech involves that object) this theory could be measured with eye-movements. If listeners have a stronger expectancy that speakers will mention a discourse new item when they are being disfluent, then this should be reflected in more fixations on discourse new items during disfluent speech.

Their results followed the predicted pattern. When the speaker was referring to a given item, participants were more likely to look at a new competitor in the disfluent condition (.30) than in the fluent condition (.15). When the speaker was referring to a new item, participants were more likely to look at a given competitor in the fluent condition (.34) than in the disfluent condition (.25).

One cautionary point to note when examining the results outlined in Arnold et al. (2004) is that, although they point to a statistically significant interaction between the discourse status (given/new) and fluency (fluent/disfluent), they fail to provide the rigorous statistical examination of these results. As it stands the interaction is unspecified as no pairwise tests are presented. Arnold et al.'s (2004) interpretation of the results supports the hypothesis based on the assumption that this interaction is ordinal. However, the data as it is reported could just as well support an interpretation based on a disordinal interaction. In plain language, an equally valid claim would be that there was facilitation of comprehension of fluent expressions referring to a given item, and that disfluency had no statistically significant facilitation on comprehension of expressions for either old or new items. Despite this, the results—even as far as they are reported—do support the notion that disfluency is affecting listeners' predictions (even if it just means that they are abandoning their predictions).

In a similar vein to Arnold et al. (2004), Arnold, Hudson Kam, and Tanenhaus (2007) tested whether speaker disfluency leads to listeners predicting that the speaker was about to mention something that is difficult to name. In the earlier study, Arnold and colleagues had manipulated the difficulty according to properties exogenous to the referents (whether it was mentioned before in the discourse). Here they manipulated the element of naming difficulty according to properties endogenous to the referents (whether the word referred to an abstract or concrete object; e.g. an ice-cream cone vs. a funny squiggly line). They performed a gating study (predicting what the speaker was to say next) and an eye-tracking study (discussed below) to show that listeners were sensitive to disfluencies indicating difficulty in pronouncing the names of abstract objects.

Arnold et al. played participants spoken instructions such as:

(7) Click on the red ice cream cone

The participants' eye movements were tracked as they carried out the procedure on a computer screen in front of them. The screen contained four images, two familiar objects and two unfamiliar objects ("funny squiggly shape[s]" in the words of the authors; p. 918). For each set of objects, one familiar and one unfamiliar object was presented in one colour, and the remaining two presented in another (examples given are black and red). Half of these critical instructions contained hesitations. These included different duration and pitch pattern of the 'click on' phrase, extending the vowel in the determiner '*thee*' and adding the filled pause *uh* after the determiner to make the instructions sound naturally disfluent.

Eye movements from 200 ms after the onset of the colour phrase were analysed. The major finding was that while participants were equally likely to look at concrete or abstract objects in the fluent condition, the disfluent condition lead to more looks at the abstract objects than at the concrete objects (69% vs. 31% respectively).

Two further experiments show that in some situations inferences about the reasons why the speaker is being disfluent can affect these attentional biases. Listeners looked at familiar and unfamiliar objects equally often in fluent and disfluent conditions if they were told that the speaker had object agnosia, a disorder that would make it difficult to describe familiar objects. However, in situations where the disfluency may have been attributable to the speaker being distracted by background noise (beeps and construction noise added into the recordings by the experimenters) the bias to look at unfamiliar objects in disfluent cases was unaltered. Again, as with Arnold et al. (2004) these results reemphasise that hesitations are having an effect on the predictive processes that participants engage in when listening to speech.

Further to that, they suggest that these processes can be at least partly controlled by top-down inferences at to why a speaker might be being disfluent.

Further to Arnold and colleagues' work, detailed analysis addressing the issue of how hesitations affect listeners' predictions has also been conducted using neural indexes of language processing or event-related potentials (ERPs). The discussion of these studies (e.g. Corley, MacGregor, & Donaldson, 2007) is found later in this thesis, after the introduction of more general ERP work relating to language.

2.4.5 Hesitations and syntactic structure

Bailey and Ferreira (2003) investigated how hesitations might affect the processing of sentences that were syntactically ambiguous. Participants were played recordings of sentences such as:

(8) While the man hunted the deer ran into the woods

These were examples of garden-path sentences: that is, sentences that have a local syntactic ambiguity. In the above example, at the point of hearing the deer the listener does not know whether the noun phrase is part of the initial clause (i.e. the thing that the man hunted) or part of a new clause. The appropriate parsing of the sentence only becomes clear later in the sentence (where the *verb* ran is incompatible with the former option). If a listener does not parse the sentence in this way, then they may think that they have just heard an ungrammatical utterance.

These sentences either appeared in a fluent form, or one of two disfluent forms. These were marked by hesitations either before the ambiguous noun (9) or after it (10). After each of these sentences was presented, the participants were asked to say whether the preceding utterance was grammatical or not.

(9) While the man hunted the *uh uh* deer ran into the woods

(10) While the man hunted the deer *uh uh* ran into the woods

Using this grammaticality judgement as an index of whether the ambiguous region was parsed properly, Bailey and Ferreira (2003) showed that participants rated fluent sentences as grammatically correct 83% of the time. Treating this as a baseline, they then went on to show that hesitations before the noun helped listeners, and hesitations after the noun phrase hindered listeners. Listeners reported that the sentences were grammatical 85% of the time in the first instance and 60% of the time in the second. Bailey and Ferreira (2003) suggest this shows that listeners might be using hesitations as a way to mark out syntactic boundaries. Importantly, this pattern of results was replicated when the fillers in the hesitations were replaced by interruptions not produced by the speaker. For example, environmental noises (such as a door bell ring) that interrupted the continuous flow of speech in the same way as the hesitations, had the same effect. This last point is important for the experiments outlined later in this thesis, as it seems to lead to the suggestion that it is the timing of the hesitation/interruption that is important, rather then the specific form.

2.4.6 Summary of the effects of hesitations on listeners

For the purposes of the current thesis, three themes from the disfluency research should be summarised. First, it appears hesitations affect listeners in a number of different ways, from their overall opinion of the speaker to the way that they process the information in speech. These later changes that relate to language processing might be mediated by a number of different processing mechanisms such as attention, prediction and syntactic assignment. Second, these changes are not apparent at coarse levels of analysis such as when participants are required to explicitly identify hesitation. Rather, they require more subtle measurements that index processes underlying language comprehension in a less invasive way. Third, there are still several areas where hesitation research can be refined. For example, one of the themes mentioned in this chapter is that hesitations might facilitate target word identification by heightening listeners' attention. However, the evidence to support this claim (Fox Tree, 2001) is ambiguous as no explicit measure of attention is recorded. Further, exactly what is meant by listeners' attention is left unspecified. It is this later issue to which the discussion in this thesis now turns.

Chapter 3

Attention

3.1 Chapter overview

In order to address the question of whether hesitations heighten listeners' attention, it is necessary to understand what attention is and how attentional processes might operate during speech comprehension. It is these issues which this chapter addresses, firstly by defining attention with a particular view to speech perception and then by discussing how attention to both speech stimuli and the mental representations that speech invokes might be affected by the characteristics of the speech itself. Finally, the chapter reviews a prominent neuroanotomical model that classifies distinct subsystems of attention.

3.2 Defining attention

Attention is one of the central articles of interest for Cognitive Science and has generated an amount of research that speaks to this importance. Despite this, the term '*attention*' has a notorious reputation, even among those who study it. As a fundamental aspect of the mind, it was being discussed long before any stringency was encouraged in the description of psychological phenomena (Pashler, 1998a). As with many terms given to psychological phenomena, lay-persons have a good grasp of the gist of the concept (just as they have a good grasp of, for example, the concept of memory) and tend to agree on what it means to '*pay attention*' to one thing or another. However, this use of the term by lay-persons seems to be particularly problematic for the study of attention. To complicate matters, at times psychologists will adopt a lay-person's approach to the use of the term *attention* or be ambiguous as to whether the *attention* that they mention is just the lay-person's concept or a strict and critically defined phenomenon. Some confusion also results from the flexibility of the term. *Attention* can refer to the process of attending to something, or to the result of this process, and can be directed at external tasks (activities), stimuli (events) or internal tasks (cognitive processes). Similarly, it can either be voluntarily controlled, or can be involuntarily oriented or 'grabbed' by certain types of stimuli such as those that are particularly odd or intense.

Attention is used to describe a kind of psychological commodity that is spent or assigned at any one time (Pashler, 1998a). If one's attention is spent on or assigned to one particular event, then that event receives a kind of priority for the mind's processing so that other psychological factors such as awareness of, or memory for the event are enhanced. Two factors seem to unify the cognitive sciences' and the lay-person's definition of *attention*: selectivity and capacity limitation (Pashler, 1998b). Both groups seem to appreciate that the mind operates with some constraints, that is it cannot, or does not, fully process all the information available to it at any one time. Both groups also tend to appreciate the consequence of this capacity limitation: some selection of events or activities of interest is necessary.

The aim of this thesis is to explore how hesitations might affect listeners' attention. This chapter will now turn to discussion of attentional research with a particular view to language and where possible, with a view to the type of attentional heightening potentially induced by hesitation. Given this goal, a good operational definition of attention is provided by Spence and Driver (1997, p. 389):

> *Mechanisms of attention allow us to concentrate on events of interest in the environment.*

To summarise this definition, attention will be assumed to be the process of selecting events of interest where this selection has an influence on the cognitive state of the listener.

Fox Tree (2001) suggests that some hesitations heighten listeners' attention for upcoming speech. For example, the *er* in the example:

(1) She hated the CD, but then she's never liked my taste in *er* music

This *er* might be increasing attention to 'music' in several possible ways. The hesitation could be:

1. Increasing the listeners' attention to this particular speech stream, so that they are attending to the speech over and above other environmental stimuli;
2. Changing subsequent activation of the concept of 'music' so that the mental representation of the concept is more strongly activated.

Addressing the first possibility, attentional processes that are involved in language comprehension might involve prioritising one stream of language over another (e.g., paying attention to one conversation and ignoring another), or prioritising language over other environmental noise (paying attention to a speech despite a siren in the background).

3.3 Directing listeners' attention to speech

The issue of directing attention to one stream of speech and ignoring other competing stimuli has received a fair amount of consideration from psychological researchers. Mostly, the research has involved competing speech streams rather than environmental noises or non-auditory stimuli. People are regularly confronted with environments where they will have to attend to one conversation, and block out other distractions. A classic example of this is dubbed the cocktail-party problem (Cherry, 1953). At a cocktail party, one must process and understand the conversation one is having, and attempt to block out other conversations. This 'problem' is easily dealt with in day-to-day life by way of selective attention.

How can studies of selective attention inform us about how hesitations might heighten listeners' attention for upcoming speech? One paradigm that has been particularly fruitful for investigating how people selectively attend to one speech stream over other stimuli involves the dichotic presentation of two auditory streams (most often, two streams of speech). Participants' performance in directing their attention can be measured by asking them to shadow (or repeat back) the target speech stream. Normally, this task is relatively easy and participants perform well. Interestingly, while participants report sound memory and understanding of the messages that they heard in the target stream, the material in the competitor stream is almost entirely forgotten (Cherry, 1953; see also, Moray, 1959).

This method has highlighted the importance of a number of factors which facilitate the ability to selectively focus on one speech stream. For example, auditory characteristics that allow the listener to directly locate the competing streams (or at least form a mental representation of its location) facilitate selective attention (Cherry, 1953). Localisation of any one sound source is determined by, among other factors: the difference in sound intensity between the two ears, the independence of timing compared to other noise, and the difference in the time the signal reaches either ear. Shadowing performance decreases if the target and competitor streams are presented with the same intensity at either ear (Treisman, 1964), or if the two streams are synchronised word-for-word (Treisman & Riley, 1969). Shadowing performance increases if the delay differs between the two streams (see Scharf & Buus, 1986, for a review).

More interesting to the ideas in the current thesis is the issue of which factors lead to the capturing of attention by a competing stream. This attention capture can either be measured by the explicit report of the participant or by an observed decrease in shadowing performance. The general consensus is that auditory 'oddness' will heighten attention to the competing stream. Moreover, this oddness must be associated with the simplest physical aspects of the stream rather than the aspects such as meaning (though see below for one exception to this). For example, when the speech in the competing stream is replaced by a 400-Hz tone, participants will often identify the change. Similarly, when the pitch of the speech in the competing stream is altered due to a change in speaker gender, the participant is likely to notice (Cherry, 1953; see also Underwood & Moray, 1971). However, when the speech in the competing stream is played backwards, participants are unlikely to report the change (Cherry, 1953). Conversely, when the language of the competing stream is changed, participants will have difficulty shadowing the target stream at the point of change (Scharf & Buus, 1986). One result using this paradigm that is often referred is that the listener's own name will cause attention to orient to the competitor stream (Moray, 1959). Despite this exception, most of the aspects that orient listener attention to a competitor stream during these dichotic tasks are related to changes in the physical characteristics of the competitor speech stream.

3.4 Directing listeners' attention during speech

Proceeding on from the issue of competing sources of speech streams, how might attention fluctuate during the course of listening to a single short discourse that is being spoken, and what role can hesitations have in determining this? They might facilitate the processing of the speech stream of interest, or they might involve prioritising one character, object or event over another. To paraphrase this, attention might be directed either at the acoustic speech stream, or at some other aspect of the linguistic proposition.

To explain this last point further, if one listens to a story that contains many characters, objects or events (collectively called constituents), one may need a way to determine which of these are of interest. This selection of constituents of interest might then affect the way one processes this information at the point it is encountered, the way one processes subsequent material about this constituent, or the way one stores these representations in memory. The question that will be addressed below is, how does selective attention interact with language processing to promote important events or referents? At the time of writing, very little research has looked at this selectivity in language and specifically referred to this as attention. There are however, phenomena that are discussed in psycholinguistic research that satisfy the characteristics of attentional processes but are not necessarily referred to as such. These include phenomena such as linguistic focus, which satisfies the characteristics of an attentional process because it helps a listener select and promote the processing of certain constituents. Before discussing focus and similar phenomena, it is necessary to understand two underlying concepts that support this link between attention and these phenomena. These concepts are the referred to as *covert attention* and *good-enough representations*.

3.4.1 Covert attention

The term *covert attention* has been borrowed by some in the field of Psycholinguistics to describe certain characteristics of a state of the cognitive system at any one time (Altmann & Kamide, 1999, 2007; Kamide, Altmann, & Haywood, 2003). To explain this, it is easiest to look at the use of the term in other fields. In fields such as those concerned with visual attention, an important distinction is made between overt and covert attention. Overt attention refers to the process which results in physically directing the body or parts of the body toward a stimulus (e.g., fixating gaze on an object of interest). In this sense, overt attention is directly measurable by means such as eyetracking. In this field, *covert attention* is the term used to describe attention that occurs in the absences of these physical responses (Posner, 1980). For an example of where these two types of attention might be directed toward different stimuli, a driver might be directing his or her overt attention to the road ahead, but might—without looking—be focusing covert attention to the inside of the car where a passenger is adjusting the radio.

The use of the term *covert attention* in Psycholinguistics differs slightly from that given above. The term is used to describe the state of the comprehenders' cognitive system where there are differing levels of activation for the concepts that have been mentioned. To explain this in a context suited to this thesis, a typical view in Psycholinguistics is that while listening to speech, the listener is creating mental representations of the characters, objects and events being discussed. If covert attention is heightened or shifted to one particular representation, this will make the representation more salient or accessible to the cognitive system (Altmann & Kamide, 2007). This might have particular effects on language processing such as making the constituent more accessible during anaphor resolution. To summarise this, heightened covert attention could also be looked at as an increase in the activation for the representation of one constituent or even a small set thereof. In this context, covert attention is operationally defined as attention to mental representations and overt attention is defined as attention to actual stimuli at the time of presentation.

In the context of psycholinguistic research, the most thorough discussion of covert attention has been conducted by Altmann and colleagues (Altmann & Kamide, 1999, 2007; Kamide et al., 2003). The aim of this work was to explore how language mediates shifts in covert attention, and how this then influences changes in overt attention such as visual attention (projected eye movements and fixations) in a scene. Using the visual world paradigm discussed in the last chapter, Altmann and colleagues have reported a number of important findings that have begun to explore the relationship between language, covert attention and overt attention to visual stimuli. For example, they have shown that a speech driven change in covert attention can be measured using overt attention and that this process of covert attentional shift can occur even before a referent is mentioned explicitly, due to featural overlap of preceding verbs and nouns (Altmann & Kamide, 1999). They have shown that this covert attentional shift is sometimes mediated by real world knowledge such as knowledge about the likely actions of particular characters (Kamide et al., 2003) and real world knowledge about an objects affordances (Altmann & Kamide, 2007). While not contributing to the specific understanding of hesitation which is the concern here, this work does support the operational definition of covert attention given above: overt attention operates at the level of incoming stimuli and covert attention operates at the level of the mental representations of constituents that are not being mentioned at the time, either because they have already been mentioned or are yet top be mentioned. The discussion in this chapter now turns from psycholinguistic phenomena that are explicitly referred to as attention, to those which fill the criteria for attentional processes but are nevertheless not referred to as such.

3.4.2 Good-enough processing and degrees of specification

Perhaps one reason why the concept of attention is relatively ignored in Psycholinguistics is that the idea of selection of important constituents was at odds with some of the underlying assumptions of a classical view of sentence processing. With such a view, there is little room to talk about situations where one constituent of an utterance will be attended to more or another attended to less. This is because in a classical view, sentence processing is both compositional and thorough. As will be shown in this section, there is substantial evidence that shows that this is not the case and that as a consequence, sometimes unattended information is not represented entirely authentically.

In the classical view of language processing, each and every individual word is processed and integrated by hierarchical organisation (syntax) into structures that are assigned meaning based on their components. Ferreira, Bailey, and Ferraro (2002) point out that this classical view assumes that compositionality results in "*complete, detailed, and accurate representation of the linguistic input*" (p. 11). In the vocabulary of attention research, the classical view does not take into account that there might be a capacity limit on the processing of each individual constituent in a passage of discourse: if a new constituent is mentioned, then there should and will be enough cognitive resource to process it fully. The consequence of this lack of capacity limitation is that once processing is finished, all parts of a sentence are represented as thoroughly as possible and none are given priority as being events of interest.

Recently, research in sentence processing has highlighted that such thorough, compositional processing is not the typical way that language is processed. Instead, comprehenders of language might just process the linguistic input so that their representation is just good enough for the task at hand (Ferreira et al., 2002). A. J. Sanford and Sturt (2002) describe the same concept in terms of the modulation of degrees of specification. During comprehension, one might specify important information to a greater extent than information deemed unimportant.

To provide experimental evidence for the assertion that people sometimes leave information underspecified during comprehension, A. J. Sanford and Sturt (2002) mention a number of phenomena in human language processing. For example, if specification happens in full, then logical form would have to be fully determined (e.g., the number of women being described in the phrase "*for every man, there is a woman that he loves*"). However, it seems unlikely that one fully specified logical form is adopted, as statements that presuppose one possibility over another do not seem to present a problem for comprehenders (Tunstall, 1998 cited in A. J. Sanford & Sturt, 2002). This is evidence that the processor is sometimes not fully committed to one particular interpretation and is therefore operating with underspecified information.

More importantly for the current thesis is the case where underspecification is not driven by necessity (as with the example above), but merely a consequence of non-complete processing. A good example of this comes from work on the so-called Moses Illusion (Erickson & Matteson, 1981). When asked about the validity of the statement "*Moses put two of each sort of animal on the Ark*" people often respond that it is true because they fail to see that the statement relates to Moses rather than Noah, the protagonist in the biblical story involving the Ark. In a similar vein, people will tend to miss the anomaly inherent in a statement such as "*After an air crash on the border of France and Spain, the authorities were trying to decide where the survivors should be buried*" (paraphrased from Barton & Sanford, 1993).

Much of the research on these good-enough representations has focused on how non-complete representations might shape processing. The discussion in the current thesis is concerned with heightened attention, so it is sensible to explore the alternate side of this issue: that is, how heightened attention might ensure processing is more thorough or complete. The mechanisms that control this attention, or degree of specification, might include phenomena specific to speech such as prosody and, important to the current thesis, disfluency (Ferreira et al., 2002). Since there is relatively little research regarding the way these act on listener attention (though see A. J. S. Sanford, Sanford, Molle, & Emmott, 2006), the current chapter will first look at other non-speech specific controllers of attention that operate during language comprehension. A. J. Sanford and Sturt (2002) argue that focus, subordination, and focalisation can control depth of processing and provide a number of supporting examples taken from studies involving written language comprehension.

Focus and degrees of specification

Changes in focus will alter the semantic meaning of a sentence by changing the way in which information is presented. Rooth (1992, 1995) defines focus along the lines of a set of alternatives. Whereas the two statements "*It was the money for a deposit that Mary loaned John*" and "*It was John who Mary loaned money to for a deposit*" might express the same situation, the first would make sense as the answer to the question "*For what did Mary loan John money?*" and the second for a question like "*To whom did Mary loan money?*". The two statements display the way that speakers highlight characters, objects or events of interest. For effective communication, listeners must appreciate the importance of these constituents in the context of the overall message and ideally, concentrate their cognitive resources on those important constituents.

Given the operational definition of attention—mechanisms that allow us to concentrate on or select constituents of interest—there are two ways that focus could drive this process of selection. This might happen at the point of encountering the constituent (overt attention) or at the level of attention given to the mental representation of the constituent after its initial mention (covert attention).

Focus has been shown to affect processes related to both of these things. Focus will affect how a comprehender treats incoming information. An example of this from a study of written language is given in Birch and Rayner (1997). Here the authors show that readers will spend more time processing words in regions of a sentence involving focused constituents.

There is experimental evidence for the idea that focus also changes covert attention. Birch and Garnsey (1995) show that readers have heightened activation for words denoting focused constituents and for information about those constituents when compared with syntactically de-emphasised constituents (see also, Birch, Albrecht, & Myers, 2000).

It is worth noting that although these measures assess the phenomenon that is defined here as covert attention, it may be that this is being mediated in the first instance by overt attention. As an example of evidence that could relate to either overt or covert attention, Bredart and Modolo (1988) show that if statements such as those used in the example of the Moses Illusion (above) are rephrased to focus on the protagonist (such as in the cleft structure "*It was Moses who put two of every animal on the Ark*") then detection of the anomaly by readers is significantly higher. Further evidence of how focus could affect listener attention is discussed later in Chapter 4.

Subordination and degrees of specification

Subordination seems to affect listener attention in the opposite way to focus. Baker and Wagner (1987) provide evidence that shows that logically flawed statements are harder to detect if they are in subordinate clauses. Baker and Wagner (1987) asked participants to evaluate the truthfulness of statements, some of which had false information. This false information was contained in either the main clause (e.g., "*Emerald City, the home of the Wizard of Oz, was named after the precious red stone*") or in a subordinate clause (e.g., "*Emerald City, named after the precious red stone, was the home of the Wizard of Oz*"). Participants were better at spotting flaws when they were in the main clause (89%) than when they were in the subordinate clause (81%). This could be explained by the fact that listners' attention is oriented to the characters, objects and events in the main clause, and therefore the processing of this information receives priority.

Focalisation and degrees of specification

In stories or speech episodes that involve more than one character, it is usual for a listener to adopt the perspective of the main character (or the narrator where one exists). This is referred to as focalisation (see, Genette, 1980). This process of focalisation establishes a bias that is analogous to increased attention to the main character. Experimental evidence points to this fact. For example, inferences about the main or focalised character are facilitated (A. J. Sanford, Clegg, & Majid, 1998) compared to situations that call for the same inferences to be made about a secondary character. In this case, it seems that heightened covert attention to the main character is serving as an attractor for the inferences.

To summarise the preceding information, components in language such as focus, subordination and focalisation help establish the relative importance of characters, objects and events. Experimental evidence shows that these components lead to changes in a comprehenders' cognitive state that fulfill the definitional criteria for attentional processes.

3.4.3 Attention and eye movements during reading

One final issue that should be covered in this section is the relatively large body of literature that addresses one area that connects written language comprehension and attention. Morrison (1984) attempted to explain eye movements in reading by proposing a model where the eye movement patterns across successive words in a text are determined by the speed of encoding of each word and attentional mechanisms.

Readers do not fixate every word in a text sequentially, instead their pattern of eye movements is complex: their fixations sometimes skip words or jump back and forth. In an attempt to model this behaviour, Morrison suggested that the eye movements from a particular word (word n) to the next point of focus are determined by the speed of encoding of the successive words. More specifically, Morrison's model proposes that a reader will fixate word n, and that this will continue until the word is encoded. Once encoded, attention will move to word *n+1.* Because the eyes have not moved at this stage, this is considered covert attention. Once covert attention shifts, this initiates an eye movement to word *n+1.*

This model not only explains phenomena such as when words are skipped in reading but also when readers fixate on space between words. The model suggests that if word *n+1* is successfully encoded before the eye movement to the position of *n+1* is initiated, then readers might skip *n+1* and jump straight to *n+2* or perhaps fixate a position in between the two.

Models (most notably the E-Z reader models such as those in Reichle, Pollatsek, Fisher, & Rayner, 1998) have extended these ideas, seeking to explain eye movement phenomena with more accuracy and in more detail (e.g., further explaining reading time rather than just probability of fixation). While this work constitutes a large body of work that addresses the relationship between attention and language processing, a full discussion of this literature is beyond the scope of this thesis. This is primarily because the models in this literature are concerned with spatial attention, rather than the attention that is strictly confined to language stimuli that is outlined in this chapter. While it is the case in text, that spatial attention might be highly correlated with attention to language stimuli, this is not the case in speech.

3.5 Interim discussion

The research outlined in this chapter thus far, has shown that attentional processes operate on the language system to prioritise input streams and select items of interest from within those streams. In fact, even if it is not referred to explicitly as such, attention is of fundamental interest to studies of language processing.

This is especially apparent in the case of covert attention, where Psycholinguistics will refer to the phenomenon of directing covert attention to a character, object or event using a number of other terms. These might include saying that it is processed to a '*higher degree of specification*' or that its representation '*receives an increased amount of activation*'. This definition of overt and covert attention gives a framework in which to assess the assertion that hesitations heighten listener attention.

Revisiting Fox Tree's (2001) suggestion that hesitations (marked by *uh*) "*heighten listeners' attention for upcoming speech*", this seems to be a statement about overt attention. Nevertheless, as increases in overt attention seem to be accompanied by increases in covert attention in reading (Birch & Rayner, 1997), the same process might operate during spoken language comprehension. Hesitation-driven increases in attention related to the incoming speech stimuli (overt attention) might be accompanied by increased attention to the mental representations of the constituents (covert attention). The remainder of this chapter reviews a neuroanotomical model that classifies distinct subsystems of attention and discusses which of these subsystems might be involved in the heightening of attention by hesitations.

3.6 Neurocognitive approaches to attention

Much of the current understanding of the attention system has been informed by studies of the neurocognitive aspects of this system. Guided by information from clinical populations, animal models and functional neuroimaging, these approaches have uncovered three partially distinct functional components of attention. These components involve: orienting; detecting targets; and maintaining alert states (Posner & Petersen, 1990). The final section in this chapter will briefly describe these three components and examine the neurocognitive evidence that supports their inclusion as distinct functional components of attention. It will also examine each of these in relation to their possible role during language processing, especially with respect to disfluency.

3.6.1 Orienting

Both overt and covert attention to a particular object in one's environment will enhance the processing efficiency for that object (Posner, 1988). If attention is to be oriented to an object, this will involve disengaging from the previous focus of attention, then shifting the locus of attention to its new position, and finally processing the object at this location. Neurocognitive evidence supports the theory that the mechanisms responsible for this orientation are at least partially distinct from other attentional processes and furthermore, that the process of disengagement, shift, and reengagement are partially distinct from each other. For example, humans with damage to the posterior parietal lobe have been observed to have difficulty disengaging from previously focused locations in order to orient their attention to something new (Posner, Walker, Friedrich, & Rafal, 1984). Conversely, damage to the midbrain or specifically the superior colliculus will not affect the disengagement, but will affect the shift process. People with such damage show the ability to disengage from previous focal points, but will show an increased tendency to refixate on areas of the visual world they have already explored (Posner, 1988). Furthermore, if this attentional shift is covert rather than overt, the lateral pulvinar region of the thalamus is important in "*reading out data*" from the indexed location (Posner & Petersen, 1990, p. 29).

3.6.2 Target detection

Neuroanotomical evidence also points to a distinct subsystem of attention that controls target detection. Target detection involves preferentially selecting input to process, based on existing knowledge about what the important information is for the task at hand. A distinct system involving the anterior cingulate gyrus and supplementary motor area seems to support the process of identifying important objects or events and making them available to conscious processing (see Petersen, Fox, Posner, Mintun, & Raichle, 1988). This system governed by this anterior network is sometimes referred to as *focal attention.* Interestingly, the input to the anterior cingulate supports the idea that the region in involved with selection or integration from both language and visual modalities (see, Goldman-Rakic, 1988).

3.6.3 Alertness

The final major neurocognitive subsystem of attention is one that governs the initiation and maintenance of an alert state. Such a state is important for preparing to process an important signal. This system seems to be heavily reliant on the right hemisphere. Right-cerebral damage can often result in neglect characterised as failure to be vigilant to input from the left visual field (Heilman, Watson, & Valenstein, 1993). Furthermore, split-brain patients often perform poorly at tasks which require vigilance directed to the left visual field (Dimond & Beaumont, 1973).

3.6.4 The role of orienting, target detection and alertness during speech processing

Exactly how important this identification of three anatomically and functionally distinct mechanisms of attention is to the study of speech processing is difficult to judge. To start with, one problem with interpreting the three-stage orienting process is that these mechanisms have been identified and examined in the context of the visual-attention system. Consequently, the descriptions of these mechanisms are largely concerned with orienting attention in space, rather than in time. Given the serial nature of speech, the modulation of attention over time is of primary importance. With visual scene examination, overt attention can be shifted from place to place. During speech perception, overt attention can only be applied to a constituent at the time it is uttered. Any attention to a constituent after it is mentioned is necessarily covert. One consequence of this is that if a similar *attentional shift* occurs in speech processing, then it must necessarily be either a covert shift to a constituent not mentioned at the time, or shift from covert attention on an already mentioned constituent to overt attention directed towards the constituent that is being mentioned.

The processes involved in the target detection sub-system are more easily applied to the attention heightening important during speech processing. A target identification system could even conceivably form the basis for linguistic processes such as focus. To explain this, controllers of attention such as focus or disfluency might be informing the listener as to what the upcoming 'target' information is. One reason why this target system is of interest here is that the initial evidence for attentional heightening supplied by Fox Tree (2001) involved a paradigm that specifically asked participants to monitor for targets while processing speech. It is worth considering whether this particular task had any effect on the type of processing in which the listeners in the experiment were engaging.

Again, the subsystem governing alertness is an interesting candidate for involvement in speech processing. It is particularly relevant to speech processing as the system seems to govern how attention is modulated over time. For example, Posner (1978; cited in, Posner & Petersen, 1990) suggests that in periods of alertness, responses to target stimuli occur more rapidly. The heightening of attention described by

Fox Tree (2001) also improved the rapidness of target-related responses, suggesting that hesitations could be affecting the state of alertness of the listener.

These three subsystems all provide explanations as to the possible mechanisms of attention that could be heightened by hesitations. Hesitations could be: acting to initiate the disengage-shift-reengage posterior subsystem; informing the listener of upcoming targets; or increasing the general level of alertness for the listener.

3.7 Conclusions

The experimental work described in this thesis investigates both the immediate engagement of attention and lasting activation that might be driven by hesitations using two paradigms. First, an ERP paradigm is used to assess the idea that hesitations immediately heighten attention to the input of upcoming speech. Second, a change detection paradigm is used to explore the idea that this is then reflected in a shift in covert attention where the representation of the post-hesitation material in enhanced. Before this experimental work is discussed, Chapter 4 will explore the previous results from the paradigms used in this thesis. This information will be presented with a view to explaining how these paradigms can be used to address the questions posed by the proposition that hesitations heighten listeners' attention. Because the previous findings regarding attention approach the issue from a neurocognitive stand-point, this neurocognitive model is first discussed.

The following chapter will explore previous results from the major paradigms used in this thesis: ERP studies and change-detection studies. This information will be presented with a view to explaining how these paradigms can be used to address the questions posed by the proposition that hesitations heighten listeners' attention.

Chapter 4

Methodologies

4.1 Chapter overview

The experiments outlined in the current thesis investigate the role of attention in the processing of disfluent language. Two paradigms are used to do this; an ERP paradigm is used to investigate the immediate engagement of attention to post-hesitation speech, and a change-detection paradigm is used to investigate the lasting affects of this on the covert attention given to constituents encountered after hesitations. The current chapter aims to introduce the history of the use of these paradigms in language comprehension and attention research.

4.2 ERPs and language comprehension

Event-related potentials or ERPs can provide a useful tool for investigating cognitive events. The temporal resolution of these measures means that neural signatures for very rapid or transient cognitive events can be measured and compared. To understand how these measures can be interpreted, it is important to know how they arise. Some cognitive events are correlated with synchronised activity in localised pools of neurons within the brain. For example, the onset of a sound stimulus causes such synchronous activity within pools of functionally similar neurons within the auditory cortices. When these pools of neurons are of a certain type (e.g., pyramidal cells) and oriented in the same direction, the electrical fields associated with this synchronised activity will summate and become large enough to be measurable on the scalp. These electrical potentials can be recorded by an Electroencephalogram (EEG). EEG is typically recorded using an array of electrodes at various sites on the scalp where data from each of these electrodes represents a continuous recording of the voltage at the electrode site relative to one or more specific reference electrode(s).

Components of these EEG signals are sometimes difficult to identify because of the low signal-to-noise ratio inherent in the EEG. Some components can however, be identified by averaging the EEG from specific epochs, time locked with stimulus properties (e.g., stimulus onset), thus boosting the signal-to-noise ratio. These time-locked averages of the neural responses associated with stimuli are referred to as event-related potentials or ERPs (see Luck, 2005, for an extensive overview).

By comparing ERPs associated with stimuli in different conditions, inferences about the underlying cognitive processes associated with these conditions can be made. The differences between these ERPs are typically described in relation to their polarity (positive or negative), timing (usually in milliseconds post-stimulus for language-related ERPs), and topography (often characterised by the positions on the scalp of the maxima). The process of comparing baseline ERPs to ERPs from experimentally significant events (e.g., anomalies) can be complicated. Therefore, the aim of the following section is to explore some of the findings from ERP studies of language comprehension in order to clarify how new information is garnered from this technique.

4.2.1 N400

The ERP component known as the N400 was first reported by Kutas and Hillyard (1980) as a neural event related to violations of semantic expectancies. Kutas and Hillyard (1980) showed participants sentences with semantically inappropriate final words such as:

(1) He took a sip from the *transmitter*

Compared to sentences that end in appropriate words, these lexical semantic anomalies elicit a negative deflection in the EEG between 250ms and 600ms which peaks at around 400ms. This negativity is predominant over central and posterior scalp sites. Because of its peak latency and polarity, the component is named the N400.

The N400 is present for lexical semantic anomalies in both auditory and visual stimuli though tends to be more biased to the right hemisphere with visual stimuli and bilateral with auditory stimuli. With auditory stimuli, its onset can be considerably earlier than 250ms post-stimulus (as low as 50ms post-stimulus in continuous speech; Holcomb & Neville, 1991).

Not long after its discovery, it became apparent that the N400 is a neural correlate not just of semantic anomalies at the end of otherwise acceptable sentences but of a number of other antecedent conditions. For example, modulation of the N400 is shown during the presentation of word pairs. In such cases, the N400 effect associated with the second word is larger in an unrelated word pair, than in a pair where the words associated with each other (Bentin, McCarthy, & Wood, 1985). Similarly, evidence emerged showing that the N400 is not specific to linguistic stimuli, but rather an effect that could be elicited in with a number of different stimuli (e.g., line drawings) where series of these stimuli could elicit a semantic context (e.g., Federmeier & Kutas, 2001; Ganis, Kutas, & Sereno, 1996). The N400 is now viewed not as a correlate of semantic anomaly, but as a component modulated by the semantic fit of a stimulus into its preceding context, so long as some meaningful content can be extrapolated from this context.

If the N400 is a common correlate of stimulus presentation that is inversely proportional to the degree of fit of the stimulus to its context, then it should be the case that modulation of the component is not only contingent on the word's goodness-of-fit into a context, but also by factors inherent to the context (such as the degree of constraint built up by this context; Kutas & Hillyard, 1984). This appears to be the case. For example, when words from a regular well-formed sentence are presented in a sequence, the N400 amplitude for the first open-class (content) word will be relatively large. The N400 amplitude related to the following open-class words will then decline as the context is established and the range of permissible words is constrained (Van Petten & Kutas, 1990, 1991).

The N400 is not only sensitive to the build-up of sentence level constraints, but also of discourse level constraints (Van Berkum, Zwisterlood, Hagoort, & Brown, 2003) as well as constraints derived from real-world knowledge extrinsic to the language material (Hagoort, Hald, Bastiaansen, & Petersson, 2004). For example, the word *quick* in:

(2) Jane was to wake her sister and her brother at five oclock in the morning. But the sister had already washed herself, and the brother had even got dressed. Jane told the brother that he was exceptionally *slow*.

will elicit an N400 that is larger than the N400 for the word *quick* in the same context, and also larger than the N400 for the word *slow* in the case where the final sentence is presented in isolation (see also, Nieuwland & Van Berkum, 2005;

Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). Further work specifically relating to disfluency processing and the N400 is discussed in Chapter 5.

4.2.2 P600/SPS

Where the N400 indexes processes associated with meaning, other ERPs have been shown to index processes associated with structure. The P600 or Syntactic positive shift (hence, simply P600) is one such component. For example, a violation of subject verb number agreement such as:

(3) The talented musician *perform* the solo with ease

will elicit such a positivity. When ERPs time locked to the verb *perform* are compared with ERPs from the syntactically permissible verb *performs*, there is an increased positivity in the ERP waveform initiating at about 500ms after the onset of the verb and extending to 800ms or beyond. The distribution of this positive difference is typically bilateral and widespread, with a centroparietal maximum. This P600 effect has been shown to be elicited by a variety of syntactic violations. It is known to be associated with subject-verb agreement of both number (Coulson, King, & Kutas, 1998) and gender (Osterhout & Mobley, 1995); violations of subcategorisation (Ainsworth-Darnell, Shulman, & Boland, 1998); violation of argument structure (Osterhout & Holcomb, 1992); and violations of subjacency (Neville, Nicol, Barss, Forster, & Garrett, 1991).

Unlike the N400, the component can be elicited even in the context of otherwise senseless sentences. It appears as a reaction to violations of structure in sentences comprised of normal function words connecting non-related content words (Hahne & Jescheniak, 2001). It also appears in a reduced form for structural violations in sentences of psudowords connected by normal function words such *that* in (5) compared with *that* in (4) below (though see Münte, Matske, & Johannes, 1997 for a similar example where the P600 is not elicited).

(4) Minno can kogg the mibe with *that* nove

(5) Minno can kogg the mibe with her *that* nove

As with the N400, the effect is not limited to outright rule violations, but also violations of expectancies (Kaan & Swaab, 2003). For example, in the absence of any outright structural violation, the presence of a P600 on the *is* below is presumably because of the preference to assign the noun phrase *the garage* to the the initial clause.

(6) The man is painting the house and the garage *is* already finished

The functional interpretation of the P600 is disputed. In the context of language research, it has been interpreted generally as a correlate of processes associated with: parsing difficulty (Münte, Heinze, Matzke, Wieringa, & Johannes, 1998), or specifically the inability to assign a preferred structure to an input (Hagoort, Brown, & Groothusen, 1993); the difficulty of integrating syntactic information (Kaan, Harris, Gibson, & Holcomb, 2000); or as a correlate of not difficulty but the controlled process of syntactic re-analysis (see, Kutas, Van Petten, & Kluender, 2006 for an overview).

Because similar components are seen in ERPs taken from non-linguistic studies, the P600 has also been interpreted as a neural response to violations within contexts where structure is important (e.g., music; Patel, Gibson, Ratner, Besson, & Holcomb, 1998). In addition to this, the P600 has been interpreted by some researchers as part of the P300 family of components (e.g., Coulson et al., 1998; see below), suggesting it is a subtype of a family of domain-general reactions to unexpected events. An exhaustive overview of these particular debates is beyond the scope of this thesis.

4.2.3 LPC

Another electrophysiological component of interest to psycholinguists is a late positive deflection in the ERP associated with certain attributes of semantic processing (here referred to as a late positive complex or LPC). A number of studies have identified components that arise around 500ms after stimulus onset and have a frontal focus and occasional left-hemisphere bias.

While there is some disagreement on how similar individual instances of the LPC are, and how applicable a single label for these different instances of frontal components is, the antecedent conditions that have been observed to elicit these effects all seem to share the same property: these are situations involving controlled memory retrieval. LPCs have been observed: in relation to unexpected sentence ending words in highly constrained sentences when compared with predictable sentence ending words in less constrained sentences (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007); as a reaction to probe words that are unrelated to previously presented jokes when these are compared with probe words that were related (Coulson & Wu, 2005); and in a study of Spanish-English bilinguals in reaction to idiom type sentences where the final word either involves an abrupt shift from English to Spanish, or were this final word is incorrect, but related to the normal idiom ending (e.g., The truck hit me like a ton of *stones*).

The 'memory process' interpretation of the LPC is supported by evidence that it is attenuated for repeated words within a sentence (Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991) and is topographically similar to other effects associated with effortful retrieval (Ranganath & Paller, 1999; Rugg, Allan, & Birch, 2000).

4.3 ERPs and Attention

Two attention-related ERPs are of particular importance in the context of this thesis. These are the mismatch negativity (MMN) and the P3 components. These ERPs are typically elicited during an experimental paradigm that has become known as the *oddball* paradigm. Because discussions of the research related to these components is best understood in the context of this paradigm, an outline of this is first presented below.

The oddball paradigm is an experimental procedure that has become important for studying the electrophysiological aspects of attention and related cognitive processes. A typical auditory oddball study might proceed as follows. While EEG is recorded, participants are presented with a continuous stream of stimuli at a standard rate of presentation. These stimuli differ in one or more specific characteristic, for example, participants might be presented with tone pips of two distinguishable frequencies. Importantly, in this paradigm these are not equally likely to occur: the presentation is biased so that one token is likely, and the other is unlikely (a deviant or oddball stimulus). This ratio of standard to deviant tokens could be anywhere upwards of 2:1. An important point to highlight here is that this deviance is exogenous to the stimuli: these deviant tokens are only odd insofar as they are unlikely to occur within the context of the experimental stimuli.

When ERPs time-locked to the onset of these deviant sounds are compared with those ERPs taken from the standard stimuli, large deflections associated with the oddballs can be seen. Depending of the nature of the paradigm (especially the task demands) these deflections might include an early negative difference (the MMN) and a later positive complex (the P3). These two ERP effects are discussed below.

4.3.1 The MMN

Even in auditory oddball paradigms where the deviant sounds have no direct relevance to the task, a mismatch negativity (MMN) will be elicited by those sounds. This MMN is a negative component that can be seen when the ERP associated with the normal or frequent sounds is subtracted from the ERP associated with the deviant sounds. First distinguished from the broader N2 component family by Näätänen, Gaillard, and Mantysalo (1978), this difference can initiate as early as 50ms after the deviant stimulus onset, though it usually occurs later at approximatly 100ms. Most commonly, it offsets around 250ms after stimulus onset. The MMN is usually distributed over the central or centrofrontal regions of the scalp, though as with all ERPs this is dependent on the reference point. When referenced to the nose tip (e.g., Alho, Paavilainen, Reinikainen, Sams, & Näätänen, 1986), the effect inverts at the mastoids, suggesting that the neural generators for this mismatch system are contained in the primary auditory cortex or adjacent areas (Scherg, Vajsar, & Picton, 1989).

The MMN is sensitive to the magnitude of deviation of the oddball stimuli. For example, in the environment of frequent 800Hz tone pips, a 840Hz tone pip will produce a larger component than an 820Hz tone pip. It is also sensitive to the intensity (amplitude) and duration of the stimuli. Where larger deviations result in MMN components of a larger magnitude and earlier onset latency (Scherg et al., 1989). The onset latency decrease tends to correlate with the stimulus discriminability better than the magnitude increase (Näätänen & Gaillard, 1983). Even when the stimuli differ only at the behavioural-discrimination threshold level, an MMN can still be detected (Näätänen, 1992).

Importantly for the studies that will be later discussed in this thesis, the MMN is not only sensitive to simple deviance such as the frequency of tones, but also to complex deviation. For example, Aaltonen, Niemi, Nyrke, and Tuhkanen (1987) reported a study on speakers of Finnish using the vowels /i/ and /y/. When the presentation of one of the vowels was more probably, the improbably (deviant) vowel would elicit an MMN (and P3 when the stimuli were attended). These two vowels were similar in all but the frequency of the second formant (F2 = 1800Hz and 2500Hz for /i/ and /y/ respectively).

4.3.2 The P3 (or P300)

The variants of the P3 components (the P3a or Novelty-P3 and P3b or P300) have a long history in the field of electroencephalography. P3 research represents some of the earliest work using ERPs. Sutton, Braren, Zubin, and John (1965) presented participants with stimuli in visual (light flashes) and auditory (clicks) modalities. They showed that when participants were uncertain of the sensory input they were about to receive, there was a positive deflection in the ERP that peaked at 300ms post stimulus. Importantly, they showed that this positive deflection was sensitive to the probability of the stimulus. When stimuli in one medium were less probable, the deflection at 300ms was larger.

From this initial observation, the P3 — as it would later be come to be known — has spawned more research than any other cognition-related ERP component to date. During the history of this research, focus has shifted away from the certain/uncertain distinction made by Sutton et al. (1965) onto specific internal (task requirement) and external (stimulus probability and classification) factors that correlate with the P3 effect size (Pritchard, 1981).

Task requirements can influence the amplitude of the P3. If a task requires more effort, then it will elicit a larger P3. The amplitude of the effect increases with decreasing probability of the target (for an extensive review of this phenomenon see, Duncan-Johnson & Donchin, 1977). Even a 2:1 standard-to-novel stimulus ratio is high enough to induce such an effect. This sensitivity is consistent across the different manipulations used to induce novelty.

More interesting to note, is that the sensitivity to stimulus probability is not contingent upon the strict numerical probability of a stimulus but the probability based on a task-defined classification. For example, in simple design where letters are presented to a participant using repeated serial presentation, if a target letter is frequent relative to the stimulus stream (appearing say, 10% of the time, and more frequently than any other single letter), but infrequent relative to the task demands (where the participant has to respond using one button for target and another button for all other letter stimuli), a P3 will be elicited by the target stimuli (Vogel, Luck, & Shapiro, 1998). Moving even further beyond the strict physical differences that distinguish the stimuli in earlier work, a P3 can be observed in reaction to stimuli that are only infrequent relative to high-level categorisation (e.g., infrequent female names in lists of predominantly male names, Kutas, McCarthy, & Donchin, 1977).

At this point, it is important to highlight how these findings lead to an important distinction between the MMN and the P3. As with the MMN, the P3 components arise in reaction to novel events such as those outlined in the auditory oddball paradigm. However, unlike the MMN the P3 is not modality dependent and has been studied in the context of visual processing as well as auditory processing.

Another point of distinction between these two components is specifically related to attention. Unlike the P3, the presence of an MMN is not contingent on the participant attending to the stimuli stream. In fact, many studies will involve asking the participants not to attend to the stimuli, instead having them perform a secondary task, or even a non-experimental activity such as reading a book (though see, Woldorff, Hackley, & Hillyard, 1991). For reliable P3 elicitation, the participants must at least be attending to the stimuli stream, even if their task does not require them to react differently to the novel stimuli. Finally, while the MMN is a component that is associated with early sensory processing, studies such that by Kutas et al. (1977) show that the P3 is an ERP that is sensitive to manipulations that occur at the post-perceptual level of processing.

A further point of distinction between the MMN and P3 is that the latter is often used to describe a 'family' of electrophysiological components. At least two subclassifications of the P3 have been studied in great detail. These are the early anterior P3a associated with the stimulus properties of novelty, and the later centro-posterior P3b associated with the task in which the participant is required to engage.

The P3a and P3b

The P3a and P3b are essentially distinguished by scalp distribution and task demands. The frontally maximal P3a can be seen in paradigms such as the auditory oddball paradigm outlined previously. When there are additional demands that render the oddballs relevant to the task — such as when a participant is asked to count the number of oddballs in a particular experimental block — the parietally maximal P3b will be elicited in addition to, or in the absence of the P3a.

Of the two components, the classification of the P3a has been more controversal. At the same time that Squires, Squires, and Hillyard (1975) identified the P3a in a typical two-stimulus oddball paradigm such as the one in the example above, Courchesne, Hillyard, and Galambos (1975), identified a component that they called the *Novelty-P3*. Courchesne et al.'s P3 was identified in a three-stimulus paradigm where there were: standard stimuli; target stimuli; and task-irrelevant and infrequent novel stimuli. The Novelty-P3 was associated with the last of these.

To complicate matters, the terminology for these P3 components has been inconsistently applied. For example, the term *P3a* has been used in the context of a three-stimulus oddball paradigm (e.g., Comerchero & Polich, 2000). In addition to this, the P3a effect as it was observed in the Squires et al. paradigm has proven difficult to replicate. There is evidence that this controversy may stem from a false distinction. For example, a comparison of the principle components for P3a and Novelty-P3 effects after factor analysis provides little justification for a distinction (R. F. Simons, Graham, Miles, & Chen, 2001). This controversy is beyond the scope of this thesis, though with this in mind, an explicit clarification of the terminology is necessary. The term *P3a* will hence be applied to the frontal component of the positivity that is elicited by deviant stimuli in an oddball paradigm. The term *P3b* will be applied to the centro-posterior component the deviant stimuli and the term *P300* will be used as an umbrella term covering both components.

4.4 Change detection

As outlined at the beginning of this chapter, some experiments in this thesis use a change-detection paradigm to investigate heightened covert attention given to constituents encountered after hesitations. While not widely used in investigations of language, it is used extensively in research on visual-scene processing. With this in mind, the following section briefly outlines the history of the use of this paradigm in visual scene research first, then goes on to describe the how the paradigm has been used withing psycholinguistic research.

4.4.1 Change detection for visual scenes

In visual-cognition research, change-detection paradigms involve showing a participant a scene either in one extended presentation or over successive presentations. During this, the participant is required to decide whether objects within the scene remain constant or change. For example, in visual scenes made up of a simple array of elements such as dots scattered over a computer screen, there is a relatively high detection rate when dots appear, disappear or move about so long as these events happen at random times while the scene is being examined. This detection rate falls noticeably when changes occur in a scene during a saccade (see Bridgeman, Heijden, & Velichkovsky, 1994 for an overview). Similar results are reported by Grimes (1996) where he shows that in natural scenes, such as those depicted in photographs, even seemingly large changes, such as when people within the picture exchanged heads, will be poorly detected if the change occurs during a saccade (see Grimes, 1996).

More recently, similar results have been shown in the absence of changes that are timed to coincide with eye-movements. Much of this work has been done with natural scenes such as photographs of common arrays of objects (e.g., a scene of pots, pans and appliances in a kitchen). When there is a change in the scene such as one of the appliances disappearing, participants are relatively good at reporting the change. This detection rate falls when there is a brief gap or visual discontinuation (such as a blank screen) in between the original and changed scenes (e.g., Rensink, O'Regan, & Clark, 1997). Similar 'blindness' can be induced with changes that occur during other masking processes such as: blinks (e.g., O'Regan, Deubel, Clark, & Rensink, 2000); when the object is temporarily occluded (e.g., Vaughan & Yantis, 1999); and when brief distractors suddenly appear in the scene (e.g., Rensink, O'Regan, & Clark, 2000; see Rensink, 2002a for a summary). This striking failure to detect changes extends beyond the controlled environment of the lab. D. J. Simons and Levin (1998) showed that even when ones interlocutor is substituted for another individual mid conversation (though after a brief visual occlusion), many people fail to notice that the individual has switched and will carry on the conversation as if nothing odd has happened.

The common theme in all these techniques which succeed in 'blinding' a participant to the change in a scene is that the techniques all diminish the participants' ability to make use of the short-lived information available from the visual cues of motion or luminance. If such changes occur without the masking of these intervening manipulations, then the motion or luminance change can focus the observers' attention toward the changed object. This increases awareness of the change and subsequently the detection rate.

Change detection and focused attention

What information can these visual scene change detection studies tell us about attention? Attention seems to affect the probability of being aware of or detecting a change. If an object is being more attended to due to being semantically central, or due to some other attention-capturing feature (e.g., luminance change), a change to that object will more probably be detected. The flip-side of this logic is that the probability of change detection can inform researchers about the locus of attention during scene examination. To paraphrase this idea, if certain manipulations make changes more easy to detect, this could be because those manipulations serve to bring that object into the locus of participants' attention.

This technique has lead to a number of interesting insights into visual attention. For example, while several items can be attended to at once (Rensink, 2002a) only one change can be seen at any one time (Rensink, 2002b), leading to the suggestion that up to five items are pooled into a single nexus of attention where changes to this nexus, rather than changes to the items might be detected (see, Rensink, 2000; Wheeler & Treisman, 2002). The investigation of this phenomenon has also lead to the suggestion that during the formation of this nexus, differing features on the same dimension for the items compete for capacity while features on differing dimensions can be represented in parallel (Wheeler & Treisman, 2002).

Further to this, changes to semantically central elements of a scene are detected more rapidly (see Rensink et al., 1997). Focused attention to an object enhances detection of changes to that object, even in contexts where the physical salience of the change is constant (Kelley, Chun, & Chua, 2003). Focused attention is however, not always sufficient to enhance change-detection rates. For example, Triesch, Ballard, Hayhoe, and Sullivan (2003) show that at least during some tasks, it is not attention to an object that is important for change detection, but the task relevance of the object's changed attributes that will enhance detection (see also Williams & Simons, 2000). This insight has led to the suggestion that rather than attention binding features into a complete representation, object perception just might be specified to the level that is required at any one moment for the task at hand.

Given the significance of the findings from change detection in the visual world, how might the techniques be used to tell us about how attention operates during language processing? The following section explores this issue.

4.4.2 Text change detection

The use of change detection techniques for investigating language processes has a surprisingly long history. Sachs (1967) played participants short discourse passages that contained target sentences. These target sentences could be presented to participants a second time either straight after their first presentation or at various points after (80 or 160 syllables of related material after). In this second presentation, the target sentences could be identical to the first presentation targets or changed in one of two ways, semantically changed, or changed in structural ways that did not greatly affect the meaning (e.g., switching from passive to active voice, active to passive voice, or formal changes such as information ordering).

Detection rates for the two types of change did not differ greatly when no interpolated material separated the first and second presentation of the target material (averages ≈82–92%). When interpolated material was present between the two presentations, change detection rates fell slightly to between ≈76% and ≈81% depending on the amount of interpolated material. This was in stark contrast to the detection of structural changes that fell abruptly to levels just higher than chance (≈62%) when any interpolating material was present. Sachs concluded that while the meaning of sentences is represented in memory relatively well, "the original form of the sentence is stored only for the short time necessary for comprehension to occur" (p. 437).

Change-detection paradigms have also been used to investigate language processing in much the same way as the anomaly-detection paradigm outlined in the previous chapter. The detection of an anomaly such as the one in "*After an air crash on the border of France and Spain, the authorities were trying to decide where the survivors should be buried*" suggests that the anomaly falls within the participants' locus of attention. Where anomaly detection is assumed to increase if the anomalous constituent is attended to, or more deeply processed (Barton & Sanford, 1993), similar logic can be applied to non-anomalous constituents that are changed from one presentation to the next.

Sturt, Sanford, Stewart, and Dawydiak (2004) showed participants passages in two successive presentations. These included manipulations of focus using either cleft constructions (7) or prior context (8).

(7) Everyone had a good time at the pub. A group of friends had met up there for a stag night. What Jamie really liked was the *cider*, apparently.

(8) Everybody was wondering which man got into trouble. In fact, the man with the *hat* was arrested.

The target word appeared in either a focused position as above, on in a non-focused condition (below).

(9) Everyone had a good time at the pub. A group of friends had met up there for a stag night. It was Jamie who really liked the *cider*, apparently.

(10) Everybody was wondering what was going on that night. In fact, the man with the *hat* was arrested.

In the second presentation, target words were sometimes changed to a semantically related word (e.g., beer → cider; hat → cap), or a semantically unrelated word (e.g., beer → music; hat → dog).

Sturt et al. (2004) found that sematically unrelated changes were detected frequently (≈90%) regardless of whether the target word was in focused or not, but that semantically unrelated changes were detected well when the target was focused (≈90% and ≈68% in the cleft construction and contextual focus passages respectively) and poorly when the target was not focused (≈84% and ≈45% in the cleft construction and contextual focus passages respectively). From the results of their change detection task, Sturt et al. (2004) conclude that "focus influences the specificity of the meaning representations of content words" (p. 887). Extensions of this technique are discussed further in Chapter 6.

4.4.3 Same/different matching tasks

One paradigm that bears much resemblance to change detection and therefore warrants mention here is the sentence matching task employed in studies such as those of Freedman and Forster (1985) and Murray and Rowan (1998). In these studies, participants are shown two written sentences on a screen and are asked to decide whether they are the same or different. Rather than comparing the physical similarity of the sentences with regard to visual information, it seems that the fastest way to do the comparison is to produce something akin to a sentence-level representation (i.e., a representation that includes semantic and syntactic information; see Murray & Rowan, 1998 for a brief review).

This paradigm has been used to inform on several aspects of language comprehension. For example, besides the evidence that shows that the passages are often processed to the sentence level of representations (i.e., where ungrammatical strings are slower to compare), the paradigm has also been used to show that pragmatic processes occur early and automatically (i.e., where pragmatically implausible passages again show slower processing; Murray & Rowan, 1998); as well as to establish basis of the psychological difference between different subsets of ungrammaticality (e.g., where sentences with subjacency violations seem to be processed just as fast as grammatical sentences within the paradigm; Freedman & Forster, 1985).

While these studies do not provide any information regarding the issue of attention during language processing, and especially not attention during speech processing, they nevertheless do show a history or using repeated presentations to assess linguistic phenomena even where those phenomena exist, in everyday reality, as non-repeated stimuli.

4.5 Conclusion

This chapter has outlined the use of ERP and change-detection paradigms in language comprehension and attention research. These two paradigms provide complementary techniques that will be used in the research outlined in this thesis. The two paradigms will be used to explore the role of attention during the processing of disfluent speech, specifically addressing the question of whether hesitations heighten listener attention. In Experiments 1 and 7, an ERP paradigm is used to measure the effects of hesitations on the immediate engagement of attention. The presence of MMN and P3 components is assessed for ERPs recorded from oddball stimuli presented during a sentence comprehension experiment. In experiments 2–6 a change-detection paradigm is used to measure the the lasting effects of this heightened attention. Discourse passages are presented to participants initially in speech and then again in text. The experimental manipulation involves changes to the constituents over the successive presentation and the probability of detecting these changes is used as a measure of the covert attention, or increased activation paid or attributed to these constituents.

Chapter 5

Experiment 1

5.1 Chapter overview

In Chapter 2, the argument was put forward that disfluency can affect the way in which an utterance is understood. Hesitations in speech affect the confidence that listeners have in speakers' knowledge (Brennan & Williams, 1995), and disfluent corrections of a message may leave a lingering representation of the original content (Ferreira, Lau, & Bailey, 2004). The current chapter outlines the first experiment in this thesis which addresses the question: what happens at the point at which a disfluency has been encountered? The experiment described here is reported in Collard, Corley, MacGregor, and Donaldson (2008), a transcript of which is provided in Appendix C.

5.2 Experiment 1

Corley et al. (2007) used Event-Related Potentials (ERPs) to demonstrate an immediate effect of hesitations while listening to spoken utterances such as (1) and (2).

(1) Everyone's got bad habits and mine is biting my [er] nails.

(2) Everyone's got bad habits and mine is biting my [er] tongue.

Corley et al. were interested in the possibility that hesitation disfluency changes the predictions a listener will make about upcoming speech. This hypothesis is based on the finding that disfluency often precedes less predictable linguistic material (Beattie & Butterworth, 1979; Levelt, 1983; Schnadt & Corley, 2006). If listeners are sensitive to this pattern, then it may be that they "interpret hesitation as a signal that the following words may not follow from the preceding context" (Corley et al., 2007, p. 660).

Evidence supporting a similar hypothesis is provided by Arnold et al. (2004). To review this briefly, listeners make online predictions during language comprehension (e.g., Altmann & Kamide, 1999; Van Berkum et al., 2005) and eyetracking evidence suggests that hesitations marked by prolongations such as *thee* and filled pauses such as *uh* lead listeners to update their predictions about upcoming words (this work is outlined in Chapter 2). Specifically, Arnold et al. (2004) showed that following hesitation, listeners were more likely to predict the upcoming mention of a discourse-new object.

Following this logic, Corley et al. (2007) used ERP measures to assess whether hesitations changed the process of prediction and integration of words in a situation where these processes were not limited to a small set of immediately visually accessible referents (as was the case with Arnold et al.'s design). Using the N400 effect as an index of integration difficulty, they compared listeners' responses to unpredictable (difficult to integrate) words (2) against predictable words (1) in fluent contexts, and in disfluent contexts where the critical words were preceded by hesitations. The magnitude of the N400 (predictability) effect was significantly reduced for disfluent utterances, showing a clear effect of hesitations on listeners' language processing. Interestingly, the N400 differences were associated with representational differences: listeners were more likely to remember words which had been preceded by a hesitation in a forced-choice recognition task.

It is clear that disfluency can affect linguistic processes, such as prediction, but such processing differences may in turn be predicated on other mechanisms, such as attention. Compared to the most likely continuation of an utterance (fluent production of the next word), disfluency introduces novelty. Such novelty might occupy attention and hence limit the processing of the following part of the utterance. Alternatively, the novelty could enhance attention to, and facilitate the processing of, subsequent words.

Given these suggestions, the aim of the present experiment is to investigate directly the immediate engagement of attention material preceded by hesitations, using a 'speech oddball' ERP paradigm.

The ERP measures that are likely to occur during such a paradigm are outlined in Chapter 4. To review this briefly, in oddball experiments listeners are occasionally presented with stimuli that are physically deviant from more frequent standard stimuli, for example with respect to pitch or amplitude. The deviant stimuli elicit a cascade of neural events related to their detection and the orientation of attention towards them. The ERP effects commonly elicited by such oddball stimuli are the Mismatch Negativity (MMN) and members of the P300 family of components, such as the P3a and P3b. The MMN, an early (100–250ms post stimulus) centro-frontal negative difference wave (Schröger, 1997) appears to index neural processes involved in identification of deviance in the acoustic environment and can be modulated by highly focused attentional states (Alho, 1995). Occurring after the MMN at around 300ms post stimulus, the frontally maximal P3a and the subsequent parietally maximal P3b are positive components typically associated with identification of, and attentional orientation to, deviant stimuli, and with the subsequent induced memory updating (Polich, 2004, 2007). Modulation of these attention related ERP components following hesitations would provide strong evidence that the hesitations alter the attentional state of listeners.

In the current experiment, participants listened to recorded utterances containing infrequent changes to the auditory characteristics of single words. Half of the time, the manipulated words followed hesitations. These were marked by natural changes to the speech, such as elongations of words within the hesitation (e.g., *thee*), and the filled pause *er*. The acoustic changes were designed such that the manipulated words would be physically deviant from the acoustic regularities set up by the preceding speech, but did not alter the linguistic content of the utterances. Because the deviant words were infrequent and therefore novel with respect to their contexts, they would be expected to induce equivalent attention-related ERPs in both fluent and disfluent conditions, unless — as we predicted — the attentional state of listeners was affected by preceding disfluency. If hesitations result in changes to the processing of subsequent words (indexed by alterations to the ERP signal) then longer lasting changes to the representation of these words are expected. Following this reasoning, and the protocol given in Corley et al. (2007), we assessed these long-term effects using a surprise recognition-memory test at the end of the experiment.

5.2.1 Method

Participants

Twelve native English speakers participated in the experiment (7 male; mean age 23 years; range 17–36). All were right handed and reported no known neurological impairment. Informed consent was obtained in accordance with the University of Stirling Psychology Ethics Committee guidelines. Participants were given financial compensation and course credit where applicable.

Materials

The stimuli consisted of 160 pairs of recorded utterances taken from Corley et al. (2007; an example is given in 1 above; see appendix A for the full list) which ended with a highly predictable target word (mean cloze probability 0.84). Fluent and disfluent versions of utterances were recorded by a native British English speaker who was instructed to produce the utterances as naturally as possible. Disfluent versions incorporated a hesitation before the utterance-final word which included signs of disfluency that were natural to the speaker, such as prolongations to preceding words (e.g., the prolonged definite article *thee*) and culminated in a filled-pause *er*. Utterances were recorded with a pseudo-target 'pen' so that there were no acoustic cues to the upcoming word. Targets were recorded in separate carrier sentences and spliced onto the fluent and disfluent utterances, resulting in acoustically identical targets across the fluent and disfluent contexts. An additional 80 unrelated filler utterances were recorded. These were of a similar length to the experimental utterances. Half contained various types of disfluency, including hesitations marked by filled pauses, and disfluent repairs at varying positions within the utterances. Using the 320 experimental recordings, 320 additional stimuli were created by manipulating the target words to make them acoustically deviant. To do this, we applied an equalisation pattern that was biased to the mid-range frequencies from the target word onset until the end of the utterance and resulted in an amplification of 2.8dB across all frequencies except for the 125–1000Hz range. In this range we applied a bell curve-like pattern which ranged from 2.8dB to 18dB and peaked at 500Hz. The salient effect of the manipulation was to make the speech sound momentarily compressed, not unlike speech over a poor telephone line.

Four versions of the experiment were created, each containing 40 fluent normal, 40 disfluent normal, 40 fluent manipulated, and 40 disfluent manipulated recordings. Each target word occurred only once in each version of the experiment. Two copies of each of the 80 fillers were added to each set, resulting in a total of 320 recordings of which 80 ended in deviant target words. Thus the overall deviant to normal utterance ratio was 1 in 4, ensuring that manipulated stimuli remained relatively novel ‘oddballs’ throughout the experiment.

Procedure

The experiment comprised two sections. In the first, participants listened to the 320 experimental utterances and fillers. Materials were presented in a random order via computer loudspeakers in two blocks lasting around 20 minutes each, and separated by a break of a few minutes. Participants were instructed to listen to the recordings as if they were part of a normal conversation, but were not given any other task. They were not told specifically about the presence of the disfluencies or acoustically manipulated words, but were told that occasionally, the sound editing quality would drop, which they should ignore. Electroencephalogram (EEG) was recorded from 61 silver/silver-chloride electrodes embedded in an elasticized cap at standard 10–20 locations (Jasper, 1958), using a left-mastoid reference. Electro-oculargrams (EOGs) were collected to monitor for eye-movements. EEG and EOG were amplified (bandpass filtered online, 0.01–40 Hz) and continuously digitized (16 bit) at 200Hz. Electrode impedances were kept below 5KΩ. Epochs were created from the EEG (150ms before the onset of the target words to 800ms after the onset) and these data were re-referenced offline to the average of the left and right mastoid electrodes, baseline corrected (relative to the average over the pre-stimulus interval) and smoothed over 5 points. Before averaging into ERPs, individual epochs were screened for drift of $\pm$ 75μV over 500ms (amplitude difference between first and last data point of each epoch), and for artifacts of $\pm$ 75μV. The screening process resulted in the loss of 10.47% of epochs, with no significant variation in rejections between conditions [$F(3, 33) = 1.756$]. Average ERPs were formed, time locked to the onset of target words for each participant (minimum of 16 artefact free trials were required for inclusion).

In the second section of the experiment participants performed a surprise recognition memory test for the material that they had heard. The 160 utterance-final (previously heard) target words were presented visually interspersed with 160 frequency-matched foil words, which had not been uttered at any previous point during the experiment. After a 500ms fixation cross, each word was presented for 750ms, followed by a blank screen for 1750ms. Participants were instructed to decide whether each word had occurred at any previous point during the experiment and respond ‘old’ or ‘new’ via a button-box placed in front of them. Responses which took longer than 2500ms were discarded.

5.2.2 ERP Results

ERPs associated with the onsets of deviant target words were compared to ERPs to non-manipulated standard controls for fluent and disfluent conditions. Because pre-stimulus baselines in fluent and disfluent utterances were different (including an *er* for disfluent cases), effects related to the acoustic manipulations were analysed separately for fluent and disfluent conditions.

Figure 5.1 shows the distribution of the oddball effects over 100–400ms. In fluent utterances, deviant words elicit an early negativity with an initial left hemisphere bias (100–150ms) which spreads laterally into a very typical MMN distribution (150–200ms). A large positive difference appears fronto-centrally at the midline (250–300ms) and develops into a widespread centroparietally maximal positivity (300–400ms). This pattern represents a typical P300 complex.

In disfluent utterances, effects are much smaller and less widespread. There is some indication of early negativity at the midline fronto-centrally (100–150ms) which

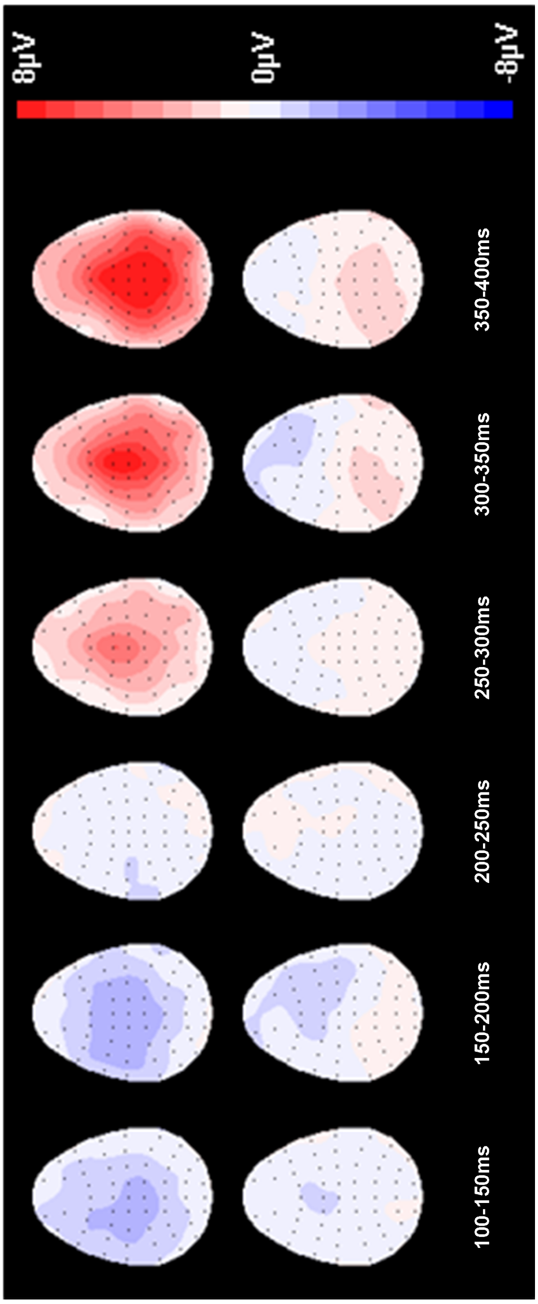

Figure 5.1: Experiment 1: Topographic maps (anterior up; electrodes shown as black dots) illustrating the mean distributions of the deviance effects (deviant minus standard event-related potentials) over 100-400 ms (in 50-ms time windows) for fluent (top) and disfluent (bottom) utterances.

becomes lateralised with a right hemisphere bias (150–200ms). No frontocentral positivity is apparent although a less focal and greatly diminished centroparietal positivity can be seen later (300–400ms).

Figures 5.2 and 5.3 show the waveforms of the MMN and P300 effects at electrodes used in the statistical analyses (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4), for fluent and disfluent utterances respectively. In fluent utterances (figure 5.2), deviant stimuli give rise to midline dominant MMN and P300 effects. There is clear indication of a P3a-like early frontal component (as with the topographic depiction of the data; figure 5.1). Data from disfluent utterances are presented on the same scale (figure 5.3) and show oddball effects which are much smaller in magnitude.

ERPs were quantified by measuring the mean voltages for deviant and standard targets over two time windows, consistent with the MMN (100–200ms) and the P300 (250–400ms), for fluent and disfluent utterances separately. Greenhouse-Geisser corrections to degrees of freedom were applied and corrected F and p values are reported where appropriate.

Analyses used three-way ANOVAs with factors of deviance (infrequent deviant, standard), location (electrodes F, C and P) and laterality (electrodes 3, z and 4). For the fluent conditions, in the MMN time window, results showed a significant main effect of deviance [$F(1, 11) = 13.152$, $\eta_p^2 = .545$, $p = .004$], indicating that deviant stimuli elicited a widespread negativity across the scalp (mean voltages of -1.701μV and -.118μV for deviant and standard stimuli respectively). No other effects involving the factor of deviance reached significance [Fs < 2.170].

In the P300 time window there was a significant effect of deviance [$F(1, 11) = 51.080$, $\eta_p^2 = .823$, $p < .001$] reflecting a positivity associated with deviant words that was widespread across the scalp (mean voltages of 4.390μV and .325μV, for deviant and standard stimuli respectively). Significant deviance by laterality

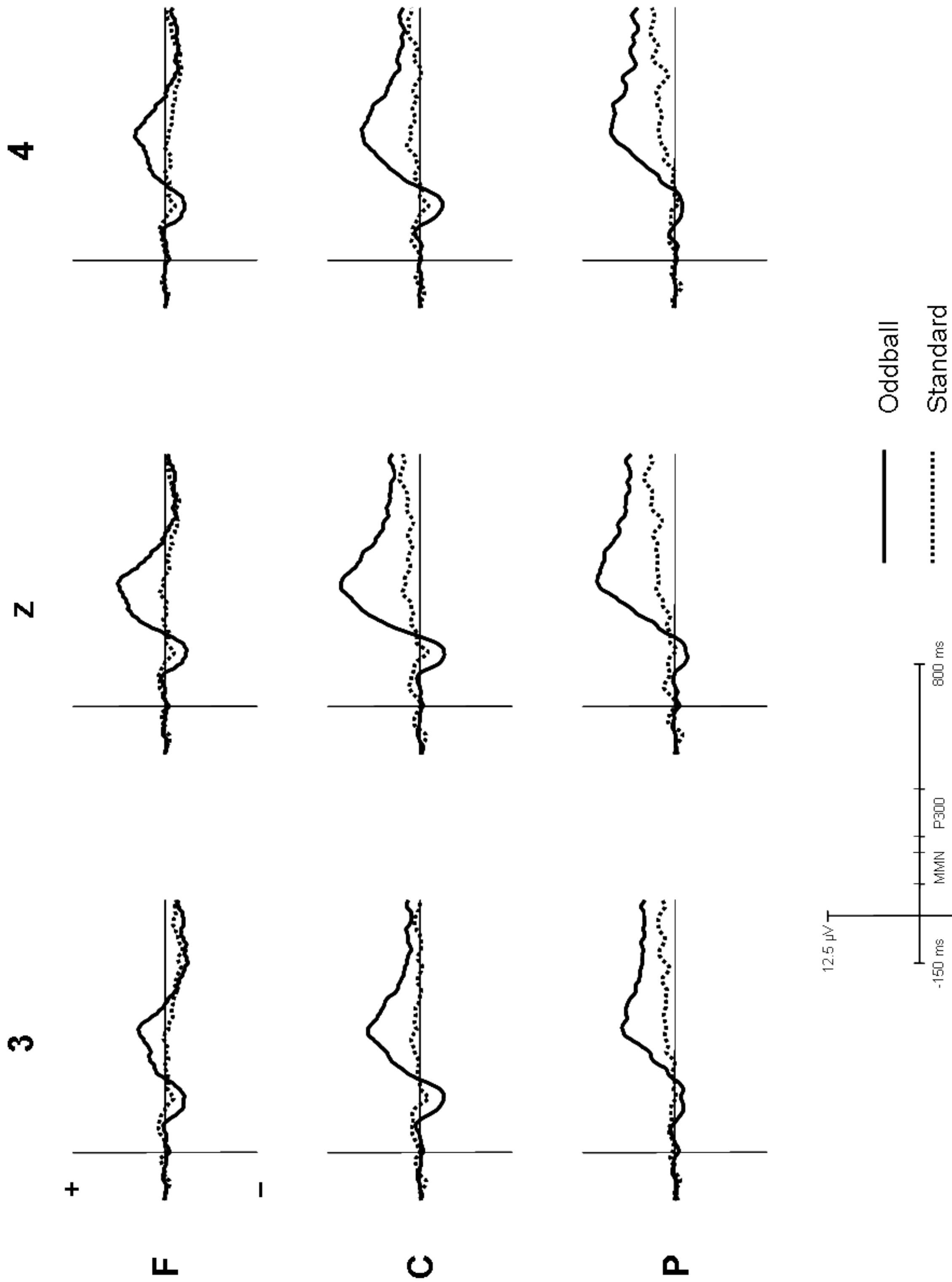

Figure 5.2: Experiment 1: Grand average event-related potentials for deviant (continuous lines) relative to standard (dotted lines) target words in fluent utterances (positive up). Waveforms show data from left, midline, and right electrodes at frontal, central, and parietal sites labeled according to the 10-20 system (from left to right and top to bottom: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4).

$[F(2,22) = 10.045$, $\eta_p^2 = .477$, $p = .001]$ and deviance by location by laterality $[F(4,44) = 7.920$, $\eta_p^2 = .419$, $p < .001]$ interactions indicate that the deviance effect was larger over midline sites, and that this midline bias was largest at frontal and posterior sites. No other effects involving the factor of deviance reached significance $[F\text{s} < 1.668]$.

For the disfluent conditions, in the MMN time window, there was a significant deviance by location interaction $[F(2,22) = 4.950$, $\eta_p^2 = .310$, $p = .017]$, indicating negativity associated with deviant words that was confined to frontal and central sites (mean voltages of .776μV, .100μV, -.864μV for frontal, central and posterior sites respectively for the standard stimuli and -.266μV, -.908μV, -.859μV for the deviant stimuli). No other effects involving deviance reached significance $[F\text{s} < 2.092]$.

In the P300 time window there was a significant deviance by location interaction $[F(2,22) = 6.033$, $\epsilon = .553$, $\eta_p^2 = .354$, $p = .028]$, indicating positivity associated with deviant words that was confined to posterior sites (mean voltages of 1.619μV, .606μV, -.171μV for frontal, central and posterior sites respectively for the standard stimuli and .836μV, 1.084μV, .974μV for the deviant stimuli). No other effect involving deviance reached significance $[F\text{s} < 2.024]$.

These analyses demonstrate robust and typical MMN and P300 effects for acoustically deviant words in fluent stimuli. In disfluent contexts, the early negativity and later positivity are much weaker and less widespread, and there are some distributional differences between fluent and disfluent ERPs. However, the antecedents and gross topographies of the effects support an interpretation of MMN followed by P300 complex in each case. We therefore conducted a further analysis to compare effect sizes across fluent and disfluent conditions. Because the disfluent condition gave rise to interactions between deviance and location in both the MMN and P300

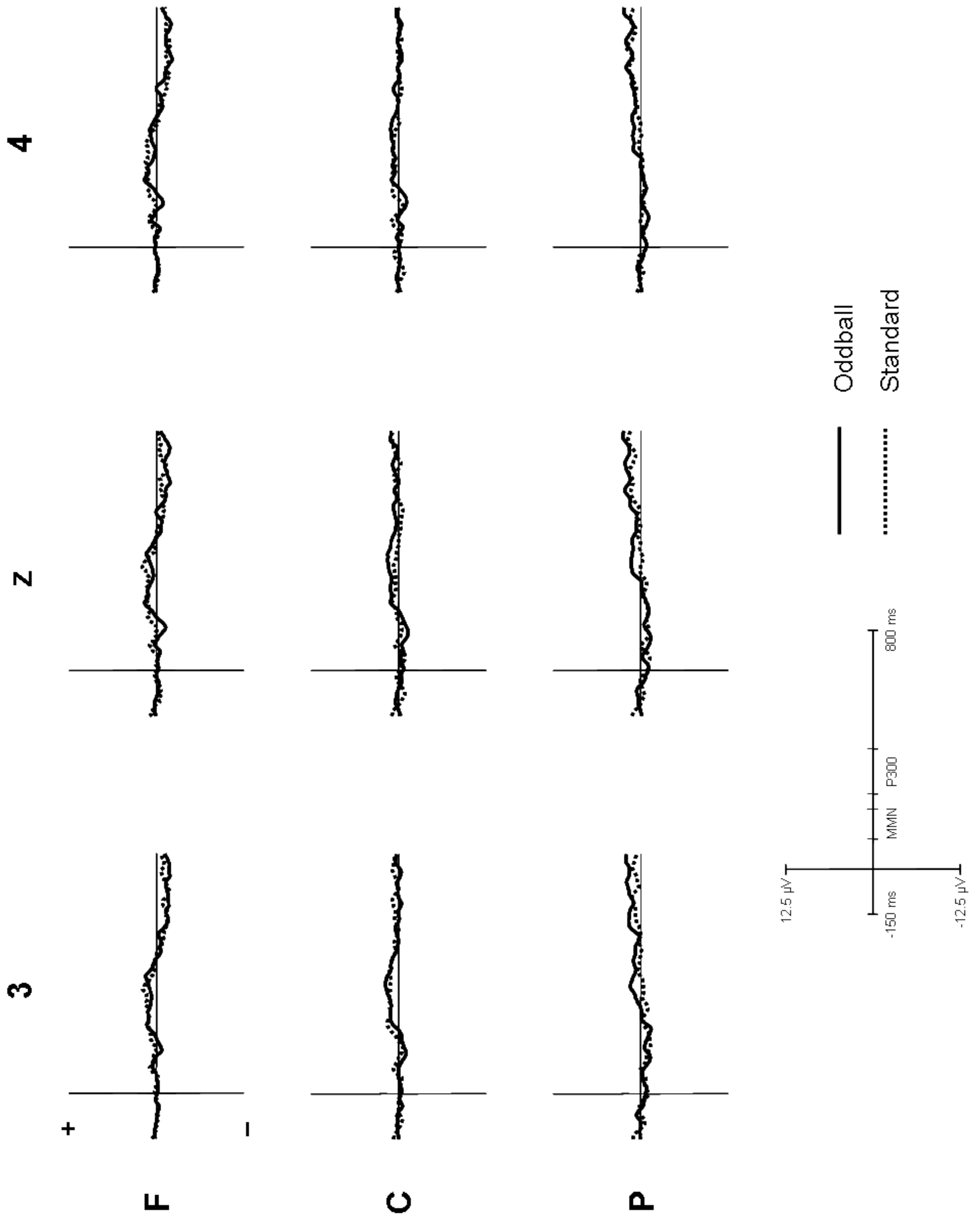

Figure 5.3: Experiment 1: Grand average event-related potentials for deviant (continuous lines) relative to standard (dotted lines) target words in disfluent utterances (positive up). Waveforms show data from left, midline, and right electrodes at frontal, central, and parietal sites labeled according to the 10-20 system (from left to right and top to bottom: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4).

windows, location was also included as a factor in these comparisons. Each analysis was conducted on the deviance effect (ERPs to deviant items minus standard ERPs) using the factors of fluency and location, with the same electrode set as the previous analyses, collapsed across laterality. In the MMN time window, there were no significant effects involving fluency [Fs < 2.804]. This is perhaps surprising in light of Figure 5.1, which corresponds to a mean difference between conditions of .898μV across electrodes. In the P300 time window, a large difference between the fluent and disfluent conditions (mean of 4.434 μV and .187 μV for fluent and disfluent respectively) was confirmed [$F(1, 11) = 32.484$, $\eta_p^2 = .747$, $p < .001$]. The interaction of fluency and location was not significant [$F(2, 22) = 1.476$].

A final consideration was addressed using an additional analysis which examined the responses to disfluent items over time. By comparing responses during the first and second halves of the experiment, we were able to establish that the responses to deviant items following a hesitation did not differ over the course of the experiment, either for the MMN (Fs < 1.433 for all effects involving half) or the P300 (deviance by half: $F(1, 11) = 2.187$; other Fs < 1.035).

5.2.3 Memory Results

A second analysis focused on performance in the recognition task. As in Corley et al. (2007), the probability of correctly identifying words heard in the comprehension block of the experiment was quantified with stimulus identity treated as a random factor. Traditional adjustments for individual error-rates, such as d', are inappropriate, since the properties of 'old' stimuli are determined by their context of occurrence and hence there are no comparable categories of 'new' stimuli. Using stimulus identity as a random factor ensures that per-participant biases to respond 'old' or 'new' are controlled for across the experiment.

Overall, 57% of the previously-heard words were correctly recognized (false alarm rate 18%). Figure 5.4 shows the recognition probability of utterance-final words by fluency and deviance.

A 2-way ANOVA with factors of fluency and deviance revealed a significant interaction between the two factors [$F(1,147) = 5.382$, $\eta_p^2 = .035$, $p = .022$]. For standard stimuli, a pairwise comparison of recognition probabilities for words which had been heard in fluent or disfluent contexts showed a significant difference [$t(147) = 2.114$, $\eta_p^2 = .030$, $p = .036$], suggesting that acoustically normal words were more likely to be recognized following disfluency. Conversely, there was no difference in the recognition probabilities for deviant words [$t(147) = 1.083$].

Twelve target words were inadvertently repeated in the experiment, resulting in 148 distinct targets. Removing data from the repeated targets did not affect the

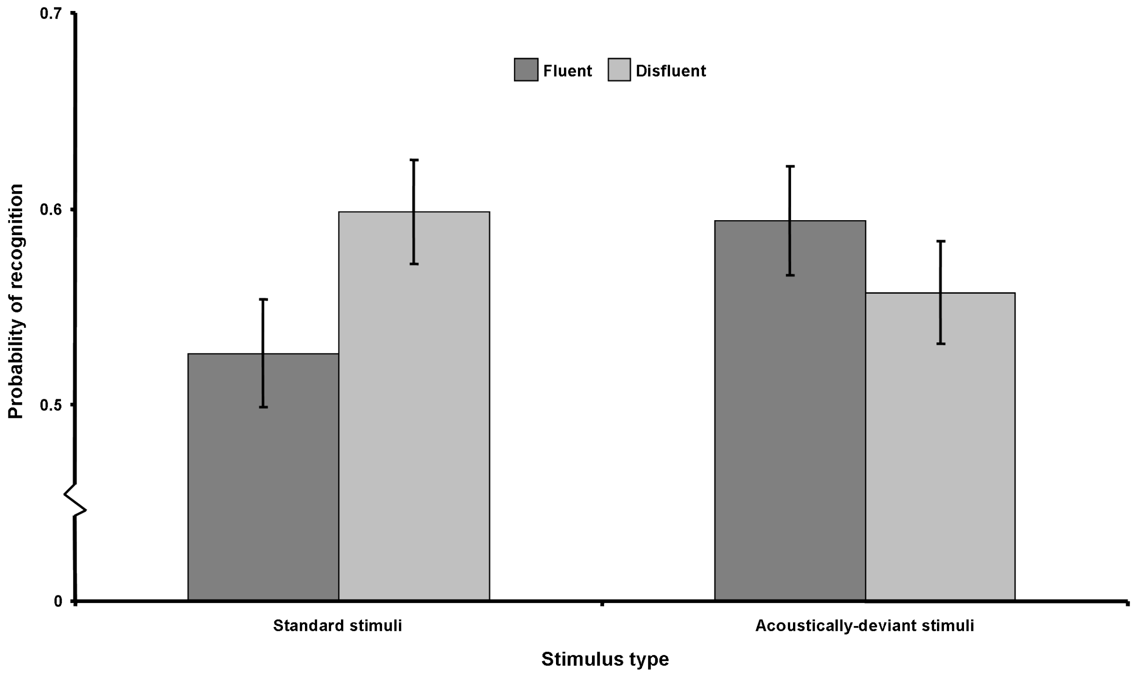

Figure 5.4: Experiment 1: Recognition probabilities for utterance-final words that were originally presented as standard or acoustically deviant stimuli and fluent or disfluent contexts (bars represent ±1 std. error of the mean).

outcome of the ANOVA, but the fluency effect for standard stimuli became marginal [$t(135) = 1.993$, $\eta_p^2 = .027$, $p = .055$].

5.3 Discussion

Experiment 1 aimed to address the question: what happens at the point at which a disfluency has been encountered? ERP measurements were taken from standard speech and infrequently-occurring acoustically deviant words (oddballs) both in fluent contexts and in contexts immediately preceding hesitations. Large deflections in the ERPs were observed when participants encountered deviant words in standard fluent speech. Given their polarities, distributions, timings and antecedent conditions, it is clear that these ERP deflections correspond to the typical neural signatures of attention capture and orientation, the MMN and P300. When the same deviant words were encountered following a hesitation, there was some evidence for MMN and P300-like effects in the appropriate time windows. However, compared to the fluent case, amplitudes were greatly reduced, and distributions were less widespread.

Information regarding the MMN and P300 components can be found in Chapter 4. Polich (2004) provides a model of ERPs elicited by auditory deviance. In his model, the MMN is associated with the detection of deviance by attentional systems. The P300 is driven by the novelty of the stimulus, and is associated with orientation of attention towards deviant stimuli (frontal P3a component) and subsequent memory-updating processes (parietal P3b component). The reduction of the observed ERP effects following disfluency in the present study provides prima facie evidence that hesitation affects the listener's attentional system. Moreover, the reduced response to novelty suggests that when the acoustically deviant words were encountered, attention was already oriented towards the speech, consistent with previous claims that hesitations heighten attention.

At first glance, these findings are reminiscent of results from attentional-blink experiments (e.g., Raymond, Shapiro, & Arnell, 1992). In attentional-blink experiments participants are less likely to detect a second target stimulus after a first, to which attention has presumably been oriented; this is accompanied by a reduced P300 to the second target (Vogel et al., 1998). Because of this similarity, a section giving an overview of the attentional blink is provided below.

The attentional blink

A paradigm known as rapid visual serial presentation (RSVP) has been used extensively to investigate attention. The paradigm involves presenting participants with stream of visual stimuli (letters, pictures, words etc.) at a rapid rate (around 10 items per second). One major finding from studies using the RSVP paradigm that is relevant to the in the context of this thesis, is the attentional blink effect.

When participants are asked to identify two targets within a rapidly presented visual series, they show a decreased ability in identifying the second of two targets (T2) at certain latencies after the presentation of the first target (T1). The decrease in performance for T2's identification peaks when the lag between T1 and T2 is about 3 items (or about 300ms given the most cited presentation rate of 10 items per second; e.g., Raymond et al., 1992). Importantly for this thesis, these attentional blinks have also been shown to occur with auditory stimuli (Arnell & Jolicoeur, 1999; but see also Chun & Potter, 2001).

The explanation of these results is that there is interference between the post-perceptual processes associated with T1 and T2. Processes such as attention and memory-updating are occupied with processing information from T1 and the processing of T2 by the same resources is, as a result, degraded (hence the name *attentional blink*). There are other accounts of this effect, but they are beyond the scope of this thesis.

An explanation similar to this would account for our recognition test data from the manipulated stimuli.

As far as the comparison of the ERPs results go, the disfluency paradigm here as well as the attentional-blink paradigm show an analogous attenuation of the P300. In the attentional blink paradigm, P300 components relating to T2 are attenuated at a lag of 3 items after T1. Vogel et al. (1998) recorded ERPs from participants performing two tasks during an RSVP paradigm. During the presentation of some of the series, there were either one target or two targets. Where there were two, the first was a number occurring in a series that was otherwise letters (T1). Participants were asked to monitor for these types of stimuli. The second was a letter presented in white typeface in a series that was otherwise black typeface (T2). Again, participants were asked to monitor for these stimuli. These two targets were separated by a lag of either 1, 3 or 7 based on the number of intervening items. Participants were required to respond as to whether the T1 item (a number) was odd or even and whether T2 (the letter in white) was the letter E or not. The letter E was infrequently occurring (15% of trials), and therefore expected to elicit a P300 component in the ERP record. A clear P300 effect was seen in response to the infrequent E in the single-task trials as well as at lags of 1 and 7 in the dual-task trials. However, the P300 was absent when the T2 E was presented at a lag of 3.

This result is reminiscent of the ERP results in the current experiment, in which attenuation of the P300 effect is observed after the supposed attention orienting effect of the hesitation.

Experiment 1 and the attentional blink effect

Although there is enough similarity between the results from the current experiment and those from attentional blink studies to warrant mention here, we stress that most of the evidence points to a clear difference between the cognitive events underlying the two experimental phenomena. There are three reasons to suggest that the present findings cannot be accounted for in terms of an attentional blink. First, Corley et al. (2007) have demonstrated that the N400 effect related to low cloze probability words is attenuated following hesitations. There is no equivalent N400 attenuation in the attentional-blink paradigm (Vogel et al., 1998). Second, in attentional-blink paradigms, the attentional attenuation tends to be maximal about 300ms after the onset of the initial orienting target (at a lag of 3 items, 100ms/item; Vogel et al., 1998). In the present study, the mean delay between the onset of the *er* and that of the target word was 598ms (SD 103ms; this is a low estimate for the time between events because signs of disfluency such as word prolongations sometimes occurred before the *er*).

The third and most important reason for rejecting an attentional blink account comes from the recognition task. Hesitations cause subsequent (acoustically normal) target words to be more likely to be later recognised, in direct contrast to what would be predicted if hesitation induced an attentional blink. This increase replicates the finding of Corley et al. (2007) that differences in the processing of fluent and disfluent utterances lead to long-term differences in the representations of those utterances, and further suggests that despite the acoustic manipulations necessary for the purposes of the present study, participants were engaged in comparable language processing. Salient (here, deviant) items were recognised equally often whether they had originally been encountered in fluent or disfluent utterances. This was possibly ascribable to a ceiling effect, given the numbers of stimuli and time between encoding and recognition of up to 55 minutes. Taken together, the results of the present study suggest that hesitations orient listeners' attention to the ongoing utterance. In contrast to attentional blink studies, attention is not 'occupied' by hesitation, rather it is heightened so that listeners specifically attend to (and subsequently recognise) the words which follow. If the subsequent word is acoustically deviant, the standard MMN and P300 responses to deviance are attenuated, because attention is already oriented to the disfluent utterance. This provides a straightforward account for the increased likelihood of recognition following hesitations, as well as for the facilitated reaction times for targets that have been found in earlier studies (Brennan & Schober, 2001; Fox Tree, 2001).

The interaction of attentional and linguistic processes

Previous accounts of disfluency processing have either focused on changes in attention (Fox Tree, 2001) or changes to linguistic mechanisms (Arnold et al., 2004) that occur when hesitations marked by filled pauses are encountered. However, these accounts are not mutually exclusive. Hesitations may induce a low-level response that heightens listeners' attention, and this may in turn affect linguistic processes which alter the linguistic availability of subsequent material. Clearly, such an account would require elaboration: for example, it is presently unclear whether the effects on the attentional system precede changes to prediction. Such issues remain questions for future research and will not be explored in this thesis.

5.4 Conclusion

The importance of Experiment 1 is that it provides clear evidence that attention is immediately affected by hesitation in an utterance. This is in agreement with other finding such as those described by Fox Tree (2001), who showed that hesitations speed-up identification of target words during a speech monitoring task. However, in this case, the experiment uses ERP measures that are directly implicated in the attentional-orienting process. Thus, the work is an important initial step in integrating theories of disfluency processing into a clear and defined model of attention. The next chapter builds on this finding by assessing whether the immediate attentional effects of hesitations might also be accompanied by what could be described as a modulation of *covert attention* or the increase in activation for constituents that are encountered in this disfluent context.

Chapter 6

Hesitations and change detection

6.1 Chapter overview

In Chapter 5, Experiment 1 provided evidence that supported the hypothesis that hesitation disfluency heightens listener attention. The immediate engagement of attention to speech was measured by comparing the attention-related ERP components and showing that these are attenuated after hesitation. In Chapters 3 and 4 it was suggested that the attentional effects of hesitations might not only manifest in immediate attentional engagement, but might also reflect an effect that could be described as a modulation of *covert attention* or the increase in activation for constituents that are encountered in this disfluent context. The current chapter presents two experiments addressing this possibility.

6.2 Introduction

Attention has been defined as the mechanisms that allow people to concentrate their cognitive resources on important events. If hesitations heighten listener attention, this might manifest in an immediate attentional effect, which changes the way incoming information is treated during initial processing. A second possibility is that hesitations could cause a listener to temporarily mark out material encountered in these contexts as more important than other information. This could be described as increasing the listeners' covert attention to the constituent encountered after the hesitation or the increase in activation that the mental representation of this constituent receives (see Chapters 3 and 4 for an overview of this). The experiments outlined in this chapter explore this possibility using a change-detection paradigm. In these experiments, participants are presented with a passage of speech which they are to listen to and remember. Following this, the same passage is then presented to them as text, with or without changes to constituents in the passage. In Experiment 2, the change/no-change judgment accuracies are compared to see if changes to constituents are more easily noticed when encountered in the context of hesitations. An increase in the probability of change detection for constituents in disfluent contexts is indicative of increased covert attention to these constituents. In Experiment 3, eyetracking measures are used while the participants engage in the same task. Here, changes to the reading patterns (or eye-movement data) relative to the post-hesitation constituents might be indicative of increased covert attention to the post-hesitation constituents.

Sturt et al. (2004) showed that this sort of change-detection paradigm could be used to investigate how linguistic focus affects what could be described, using the vocabulary of the current thesis, as the locus of the comprehenders' attention during language processing. Sturt et al. (2004) showed participants passages in two successive presentations. These included manipulations of focus, where specific target words were either brought into, or shifted out of focus. In the second presentation, target words were sometimes changed to a semantically related word (a close-semantic change), or a semantically unrelated word (a distant semantic change). Sturt et al. (2004) found that semantically unrelated changes were detected frequently regardless of whether the target word was in focused or not, but that semantically unrelated changes were detected well when the target was focused and poorly when the target was not focused (see Chapter 4 for a more in-depth description). Since this study, the same technique has be used to investigate phenomena similar to focus and explore how these operate during language comprehension.

A. J. S. Sanford et al. (2006) investigated whether depth of processing is mediated through phenomena that they term *Attention Capturers*. A. J. S. Sanford et al. (2006) identify a number of candidates for such a role. They suggest that both text-specific examples (e.g., changes in typography style) and speech-specific examples (e.g., spoken emphasis) of these attention capturers exist, as well as capturers that are independent of the input medium (e.g., cleft structures). To investigate these, they used a paradigm similar to Sturt et al. (2004) and similar to the experiments outlined in this chapter. Their first experiment involved a change-detection paradigm where the first and second presentations were both passages of text. Here they showed that typographical novelty can act as an attention controller. Participants were more likely to notice both close and distant changes if the changed words are presented in italics. As with the focus manipulation in Sturt et al. (2004), the effect of the italics was largest in the close-change condition.

Following this, they showed a similar effect for spoken material. Using broad and narrow focus conditions, focus was manipulated both through the contextual focus inherent in the material and through differing stress on focused words, where longer durations and larger minimum-to-maximum pitch ranges marked out the focused constituents. In this case, both presentations of the passages were auditory, rather than visual (or auditory followed by visual as is the case for the current experiments). Here the results patterned with their previous experiment, participants were more likely to notice both close and distant changes if the changed words were in focus. Again, this effect was largest in the close-change condition.

Following similar logic, A. J. S. Sanford, Sanford, Filik, and Molle (2005) used a change-detection paradigm to test whether sentential load affects the depth of processing. In a series of experiments they manipulated the processing load prior to a changed word. Here again, the change could be either close or distant. In the vocabulary of the current thesis, this research was investigating whether high processing loads result in less attentional allocation to subsequent material. Processing load was manipulated via syntactic manipulations (object extraction in the high-load condition and subject extraction in the low-load condition) and referential load (full noun phrases in the high-load condition and pronouns in the low-load condition). Even though comprehension did not break down when the processing load was high, participants were less likely to detect changes. However, unlike the results of Sturt et al. (2004), this effect was similar for both close and distant semantic changes. In Sturt et al. (2004), the attention capturing effect of focus could be said to affect the level of granularity with which the attended constituent is processed. However, since there was no interaction between the sentential-load manipulation and the change distance in A. J. S. Sanford et al.'s (2005) study, it seems that focus and sentential load affect change-detection (and by extension, what is described here as attention) via mechanisms that are at least partially distinct.

Ward and Sturt (2007) investigated the eye-movement behaviour of people participating in a change-detection experiment. They had participants engage in an experiment where they read over the passages from Sturt et al. (2004). While participants were doing this task, their eye-movements were being recorded. One notable difference between this study and Sturt et al. (2004), is that in Ward and Sturt (2007) there was no distant-change condition. The eye-movement measures were taken from both the first and second presentations of the change-detection material. While first presentation behaviour is not discussed at length here, one important observation is that in the first presentation, there was no effect of focus on eye-movement measures regarding the target (to be changed) region. In the second presentation, Ward and Sturt (2007) noted a number of effects of focus and change on the participants' eye-movement behaviour. The average duration of the first fixation of the target words, the gaze duration (first-pass time) and the total time spent in the target region were longer for targets in the change condition than in the no-change condition. There was also an increased number of fixations on changed targets compared to non-changed targets. An interaction between the focus condition and the change condition was observed in both the total time and the number of fixation measures. In both cases, the interaction was driven by the fact that changed words were fixated more often, and for longer times, when they occurred in the focus condition than when they occurred in the non-focus condition.

These studies show that change-detection can be a useful paradigm for investigating attentional allocation during language processing. In the experiments reported in this chapter, a change-detection paradigm is used to investigate how the attentional engagement that is induced by a hesitation disfluency might affect the covert attention directed towards the post-hesitation constituents.

6.3 Experiment 2

In Experiment 2, participants were presented with recordings of passages adapted from Ward and Sturt (2007). These were either fluent utterances or disfluent utterances marked by filled-pause hesitations.

(1) The doctor checked to see how much longer he had to work. He saw that the patient with the [er] virus was at the front of the queue. A kind but strict-looking nurse brought the boy in.

After listening to the passage, they were asked to read a second presentation in text that was either an identical passage, or a passage with a one word substitution (e.g. virus → infection, in a close semantic change condition or virus → magazine in a distant semantic change condition). The position of these critical words was such that they were always out of focus, embedded in the second sentence in a prepositional phrase associated with the subject. To avoid the participants becoming accustomed to the positions of the changes, the positions of the disfluencies and the co-occurrence of the two, obvious (salient) disfluencies and obvious changes were inserted into filler passages in positions that differed to those in the target materials.

One expected result is a replication of the effects seen in Sturt et al. (2004): there might be an effect of change distance on disfluency, where distant changes are detected more frequently. If hesitations enhance the covert attention to post-hesitation constituents, then there should also be an effect of disfluency: where changes are detected more in the disfluent condition than in the fluent condition. If the effects of hesitations manifest in a similar way to those of focus, then the effects of fluency and change distance will interact: distant changes should be picked up nearly all the time and close changes picked up more when they appear in the context of disfluency.

6.3.1 Method

Participants

Twenty-four members of the University of Edinburgh student community were recruited to take part in the study. They were each paid £2 compensation for their participation. All procedures regarding participation complied with the BPS ethical guidelines.

Materials

Using a similar approach to Sturt et al. (2004: experiment 2), the experiment involved the manipulation of two factors; *fluency* (fluent and disfluent) and *change* (no change, close change, distant change). Twenty-four passage frames were adapted from the non-focus condition in Ward and Sturt (2007) and a further 12 frames of the same format were created in order to satisfy an additional level included in the change factor (see example (1) or appendix B for the full list). The second sentence of each frame contained a critical noun in non-focused position in a prepositional phrase. During the second presentation of each frame in the experiment, this noun either remained the same as the one in the first passage (no-change), or was changed to one of two equally plausible constituents. This was either a word which was semantically related or a near synonym to the noun in the first presentation (close-change), or a word which was not closely related to he noun in the first presentation (distant-change). Frequency information from both the spoken and written sections of the British National Corpus (1995) was used to assess whether the target words differed across conditions. No significant differences in log transformed frequency (per million words) was found between the groups of target words (means 3.113–3.219; $F[2, 105] < 1$). Similarly, no difference in the number of characters in each word was found (means 6.194–6.417; $F[2, 105] < 1$).

Each version of the passage frames was digitally recorded 6 times (16-bit, 44000Hz mono). The actor who produced these passages was a native speaker of British English and was instructed to speak as clearly and naturally as possible, and in the case of disfluent sentences, to make the disfluency sound as natural as possible. This resulted in 216 recorded passages which would satisfy the 2x3 design (fluent and disfluent condition versions of 36 target frames, each containing one of three possible nouns in the change location).

In order to minimise the difference between the different versions of each frame, the second (target) sentence the recordings were spliced between the first and last sentence example of each frame. The volume level of the non-target sentences was modified manually by the experimenter in order to fit naturally into these new frames. This resulted in a new set of 6 audio passages for each frame, where each differed only with regards to the target sentence.

The 36 target passages were divided using a Latin-square method, so that each participant heard 18 fluent targets and 18 disfluent targets with hesitations. Within these groups, each participant heard: 6 no-change targets, 6 close-change targets and 6 distant-change targets.

Procedure

The experiment lasted approximately 20 minutes and was run with E-prime software (Schneider, Eschmann, & Zuccolotto, 2002). The 54 trials (including 18 fluent and disfluent filler trials) were presented in a random order. Participants were told that during each trial, they would listen to a recorded presentation of the passage via headphones. During this, a fixation cross was presented in the centre of the screen. This audio passage was followed by a brief delay (500ms), then a text version of the same passage would be presented on screen. This was presented one sentence per line in 12 point Garamond font, as white text on a black background to reduce visual fatigue. The text was aligned to a left-hand margin offset roughly 5cm from the left edge of the screen and 10cm from the top edge. Participants were asked to read through this passage and indicate (with a button press on a keyboard in front of them) when they were finished. Once this was done, a question presented in the centre of the screen in the same font asked whether the participant had detected a change. Participants responded via the keyboard, and if they responded yes, they were asked to type the word that had been presented in the first (speech) presentation of the passage that had changed in the second (text) version. After either typing this word or indicating 'no-change', the participants were presented with a written message on screen asking them to indicate when they were ready to start the next trial. Two long breaks where participants were told to 'relax and have a stretch' were included in the procedure, equally spaced throughout the experiment.

Data Analysis

Initial comparisons for the data involved the simple measure of accuracy of the response for the change/no change question. Because the measures of accuracy (expressed as a proportion) were close to the maximum value (1) in some conditions, all the statistical analysis was performed on the arcsine transform of the data. A pairwise t-test comparison between the accuracy of the responses in the fluent and disfluent no-change conditions was run to ensure that fluency was not affecting the general processing of the non-changed passages (analogous to a false-alarm in the change detection paradigm). Next, an ANOVA with the factors of fluency (fluent and disfluent) and change-type (close change and distant change) was performed. Finally, to see if the disfluency was acting as an attention orienting device similar to linguistic focus (affecting close changes primarily), a critical planned comparison between fluent and disfluent response accuracy was run on the data from the close-change condition.

6.3.2 Results

The mean probability of correctly responding 'no change' to passages where no change had occurred was .931 (S.D. = .120) in the fluent condition and .951 (S.D. = .092) in the disfluent condition (see figure 6.1). This difference was not found to be statistically significant by either participants [$t(23) < 1$] or items [$t(35) < 1$].

The mean probability of correctly responding to a change that had occurred to passages with close changes was lower in the fluent condition (mean = .667; S.D. = .220) than in the disfluent condition (mean = .778; S.D. = .201). There was no mean difference at all between the fluent and disfluent conditions for the probability of

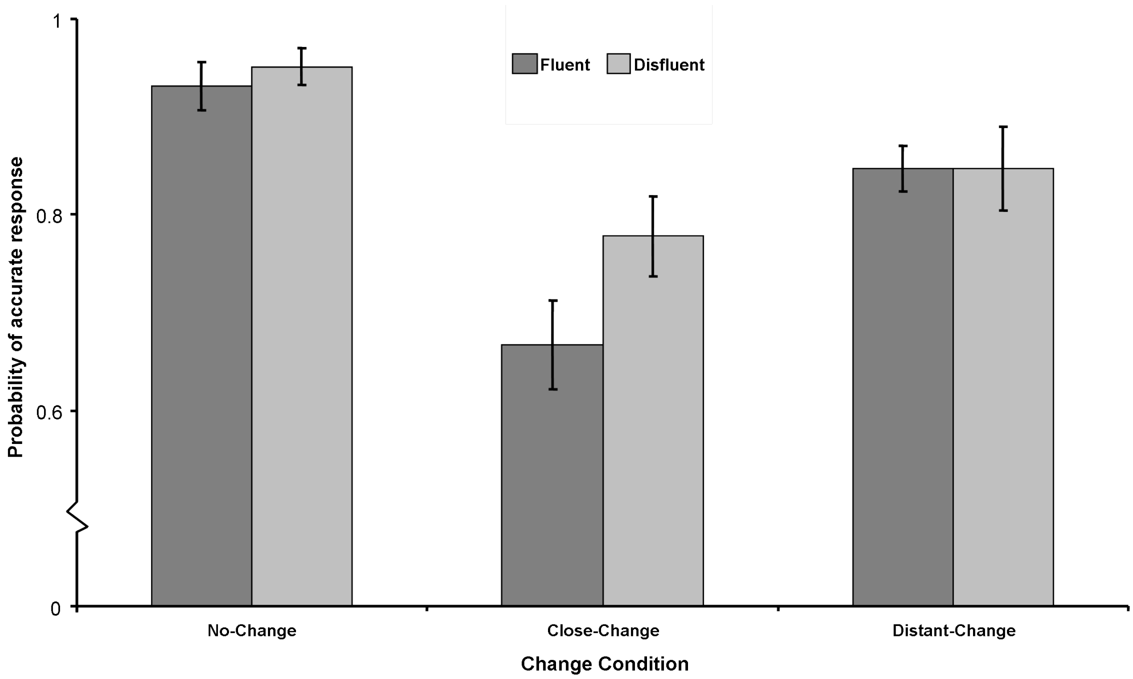

Figure 6.1: Experiment 2: Mean probability of accuracy in the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

correctly responding to a change that had occurred in passages with distant changes (mean = .847 in both cases; S.D. = .115 and .208 respectively). The ANOVA showed a main effect of disfluency [$F(1, 23) = 5.600$, $\eta_p^2 = .196$, $p = .027$] by participants, but no such effect by items [$F(1, 35) = 2.454$, $\eta_p^2 = .066$] and a main effect of change type by both participants [$F(1, 23) = 8.055$, $\eta_p^2 = .259$, $p = .009$] and items [$F(1, 35) = 15.941$, $\eta_p^2 = .313$, $p < .001$]. The interaction of the two effects was marginally significant by participants [$F(1, 23) = 3.826$, $\eta_p^2 = .143$, $p = .063$] but not so by items [$F(1, 35) = 1.874$, $\eta_p^2 = .051$, $p = .180$].

The final pairwise comparison was run on the close-change data only. This showed that the probability for detecting a close change was significantly higher in the disfluent condition than in the fluent condition in the by participants analysis [$t(23) = 3.336$, $\eta_p^2 = .326$, $p = .003$]. This significance of the difference was marginal by items [$t(35) = 1.802$, $\eta_p^2 = .085$, $p = .080$].

6.3.3 Discussion

Experiment 2 aimed to explore the attentional effects of hesitations on the representation of constituents encountered after the disfluency. A change-detection paradigm was used to assess whether the hesitations had induced an effect on the covert attention directed towards these constituents. If covert attention towards these constituents had indeed increased, then it is more likely that the participants would detect changes to them. The results from this experiment provide some support for this hypothesis. Participants were more likely to detect subtle (close) changes to constituents if those constituents were preceded by hesitation (though this effect was marginal by items). This experimental effect resembles the enhanced representation for focused constituents reported in Sturt et al. (2004). In the following section an extension of this study is reported. The same experimental paradigm is used, although here the participants' eye-movements are measured during the second presentation of the passages.

6.4 Experiment 3

Experiment 3 sought to investigate the influence of hesitations on the representation of spoken material using more subtle measures than the explicit-mention measure provided in Experiment 2. There are three motivation for this. First as Chapter 2 shows, previous work on disfluency has highlighted that measures involving explicit mention can sometimes mask effects that are occurring at a preconscious level. Past examples of these explicit mention tasks ofter required participants to specifically identify disfluency (e.g. Lickley, 1995). Even though this is not the case in Experiment 2, the extra burden involved with the additional task of monitoring the language and making explicit mention regarding the constituents could be masking other subtler changes in processing. Eyetracking provides one alternative method with which to explore this possibility. Second, the eyetracking data might help to explore exactly how far the analogy between the attentional effects of focus and those of disfluency can be taken. A. J. S. Sanford et al. (2006) list both disfluency and focus as *attention capturers*. However, in the absence of a formal definition of attention, it is difficult to say exactly how far and in what way there is a commonality between the two phenomena. Comparing the eyetracking data from the change-detection paradigm here with the eyetracking data of change-detection studies involving a focus manipulation (e.g., Ward & Sturt, 2007), will help to explore this commonality further. Ward and Sturt (2007) showed that measures of early processing (first-fixation times and first-pass times) were not modulated by focus, but instead the effects of focus were seen in broader measures indexing global processing (total time and number of fixations). If hesitations are affecting attention in a similar way to focus, then modulations of these measures of global processing are expected. Last, if the eyetracking data appears to be more sensitive to the disfluency manipulation, then this experiment will provide a good basis for the further investigation of the attention-heightening effects of hesitations.

6.4.1 Method

Participants

Twenty-four members of the University of Edinburgh student community were recruited to take part in the study. They were each paid £5 compensation for their participation. All procedures regarding participation complied with the BPS ethical guidelines.

Materials

The stimuli used for Experiment 3 were those used in Experiment 2. The experimental manipulation again involved two factors: *fluency* (fluent and disfluent) and *change* (no-change, close change, distant change). The 36 target passages were divided using a Latin-square method, so that each participant heard 18 fluent targets and 18 disfluent targets with hesitations. Within these groups, each participant heard: 6 no-change targets, 6 close-change targets and 6 distant-change targets.

Procedure

Eye movements were recorded with an SR EyeLink II video-based head-mounted eyetracking system. Viewing was binocular, though only the data from the participants self-reported dominant eye was sampled (500 Hz). After positioning the equipment on the participants head, a calibration was performed using a 9-point grid spanning the entire presentation screen. Drift correction was performed before the start of each experimental trial.

The experiment lasted approximately 30 minutes and was run on Eye Track software (Stracuzzi & Kinsey, 2003). The 54 trials (including 18 fillers) were presented in a random order. During each trial participants listened to the recorded version of a presentation of the passage via speakers within the testing chamber. During this, a fixation cross was displayed in the position of the first character in the subsequent text passage. Five-hundred milliseconds after the sound file finished, a text version of the same passage was presented on screen. This was presented one sentence per line in 14 point Courier New font, as white text on a black background to reduce visual fatigue. The text was aligned to a left-hand margin with a 20 point offset from the left edge of the screen and a 200 point offset from the upper edge. The passage remained on screen until the participant indicated they were ready to move on (with a button press on a control pad in front of them). Once this was done, a question presented in the centre of the screen in the same font asked whether the participant had detected a change.

If the participant responded *yes* to the change/no-change question, they were asked to speak out loud, the word that had been presented in the first, (speech) presentation of the passage that had change in the second (text) version. After this, participants were presented with a written message on screen asking them to indicate when they were ready to start the next trial.

Data analysis

The region of interest for the analysis was the critical word (e.g., *virus* in example (1) above). Five measures of eye-tracking were analysed. For the sake of comparison with Ward and Sturt (2007), these included the four measurements used in their study of focus: first-fixation time, first-pass time, total time, number of fixations, and probability of regression. One further analysis was conducted to determine whether participants were more likely to 'check' the region that had changed. Checking might be reflected in refixations of the word once the participants gaze had had moved to the right. To explore this possibility, regressions into the critical region were also analysed.

The first-fixation time was a measure of the duration of the initial fixation in the region of interest. The first-pass measure was calculated for each trial by addition of the duration of all of the fixations from the first time the target region is entered (from the left) until it is first exited (to the left or right). The total-time measure was calculated from the addition of the durations of all of the fixations on the target region irrespective of the direction of entry into the region and regardless of whether the participant had already fixated the region previously. The number of fixations measure was calculated from the number of times the region was fixated regardless of the origin of the eye movement into the region. The probability-of-regression measure was calculated as the average proportion of trials within each condition where the participant fixated the target region, exited it to the right and then subsequently made regressive eye movements back resulting in a fixation of the target region again. These measures were submitted to the statistical tests outlined in the data analysis section in Experiment 2. The analyses were run on the data from all trials regardless of the participants' responses (in line with the analyses reported in Ward & Sturt, 2007).

6.5 Results

6.5.1 Behavioural results

The mean probability of correctly responding 'no change' to passages where no change had occurred was .944 (S.D. = .106) in the fluent condition and .965 (S.D. = .085) in the disfluent condition (see figure 6.2). This difference was not found to be statistically significant by either participants [$t(23) < 1$] or items [$t(35) < 1$].

The mean probability of correctly responding to a change that had occurred to passages with close changes was lower in the fluent condition (mean = .472; S.D. = .244) than in the disfluent condition (mean = .667; S.D. = .177). The mean probability of correctly responding to a change that had occurred to passages with distant changes was lower in the fluent condition (mean = .688; S.D. = .227) than in the disfluent condition (mean = .756; S.D. = .170). The ANOVA showed a main effect of fluency [$F(1,23) = 23.928$, $\eta_p^2 = .510$, $p < .001$] by participants, and by items [$F(1,35) = 8.892$, $\eta_p^2 = .203$, $p = .005$] and a main effect of change type by both participants [$F(1,23) = 14.803$, $\eta_p^2 = .392$, $p = .001$] and items [$F(1,35) = 14.679$, $\eta_p^2 = .295$, $p = .001$]. The interaction of the two effects was not significant by participants [$F(1,23) = 2.760$, $\eta_p^2 = .107$, $p = .107$] or by items [$F(1,35) = 2.059$, $\eta_p^2 = .056$ $p = .160$].

A final pairwise comparison was run on the close-change data only. This showed that the probability for detecting a close change was significantly higher in the disfluent

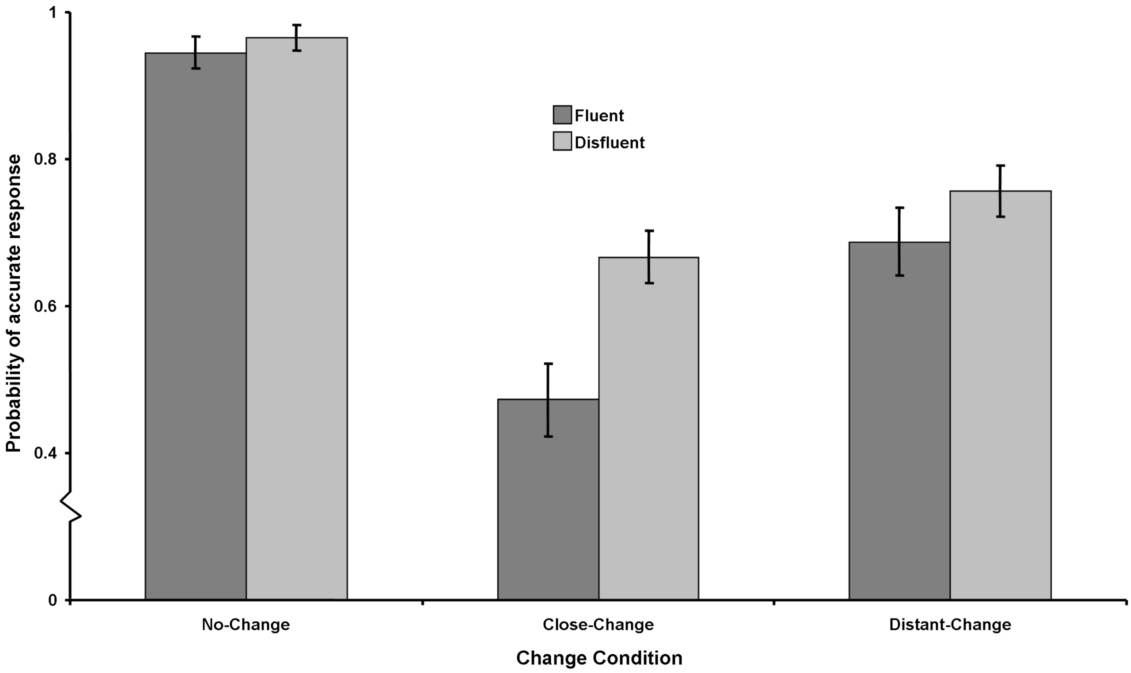

Figure 6.2: Experiment 3: Mean probability of accuracy in the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

condition than in the fluent condition in the by participants analysis [$t(23) = 4.071$, $\eta_p^2 = .419$, $p < .001$]. and by items [$t(35) = 2.844$, $\eta_p^2 = .188$, $p = .007$].

6.5.2 First-fixation time

In the no-change conditions, the mean first-fixation time spent in the target region was 237ms (S.D. = 49) in the fluent condition and 234ms (S.D. = 49) in the disfluent condition (see figure 6.3). This difference was not found to be statistically significant by either participants [$t(23) < 1$] or items [$t(35) < 1$].

In the change conditions, the mean first-fixation time spent in the target region was lower in the fluent condition (mean = 299ms; S.D. = 91) than in the disfluent condition (mean = 305ms; S.D. = 71). The mean mean first-fixation time spent in the target region was lower in the close-change condition (mean = 279ms; S.D. = 74) than in the distant-change condition (mean = 325; S.D. = 88). An ANOVA showed that there was no main effect of disfluency by participants [$F(1, 23) < 1$] or by

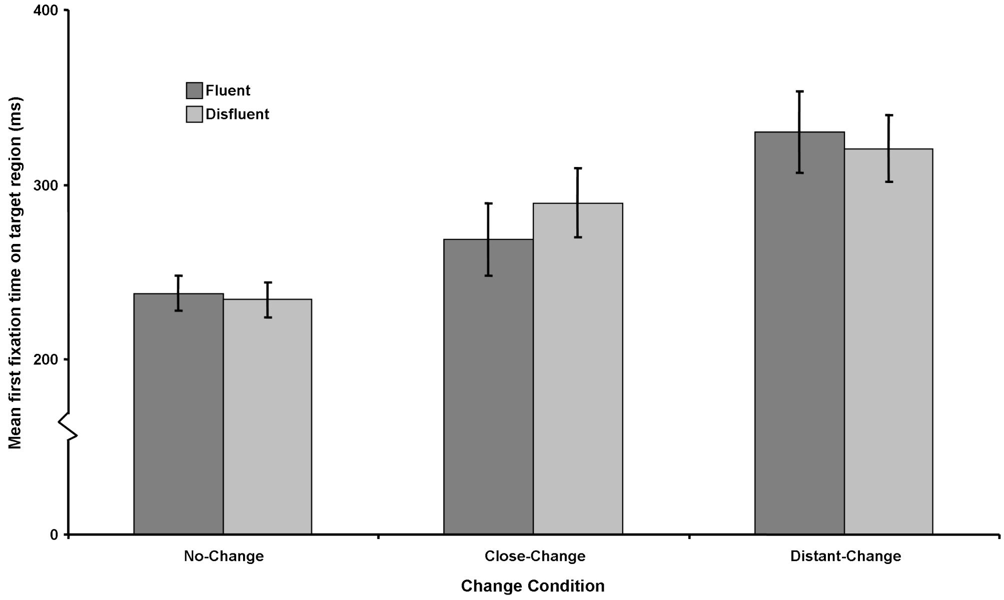

Figure 6.3: Experiment 3: Mean first-fixation time spent in the target region for the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

items [$F(1, 35) < 1$]. There was a main effect of change type by both participants [$F(1, 23) = 6.033$, $\eta_p^2 = .208$, $p = .022$] and items [$F(1, 35) = 4.415$, $\eta_p^2 = .112$, $p = .043$]. The interaction of the two effects was non-significant by participants [$F(1, 23) < 1$] and items [$F(1, 35) < 1$].

6.5.3 *First-pass time*

In the no-change conditions, the mean first-pass reading time spent in the target region was 279ms (S.D. = 98) in the fluent condition and 259ms (S.D. = 81) in the disfluent condition (see figure 6.4). This difference was not found to be statistically significant by either participants [$t(23) = 1.176$] or items [$t(35) < 1$].

In the change conditions, the mean first-pass reading time spent in the target region was lower in the fluent condition (mean = 385ms; S.D. = 117) than in the disfluent condition (mean = 397ms; S.D. = 265). The mean first-pass reading time spent in the target region was lower in the close-change condition (mean = 350ms; S.D. =

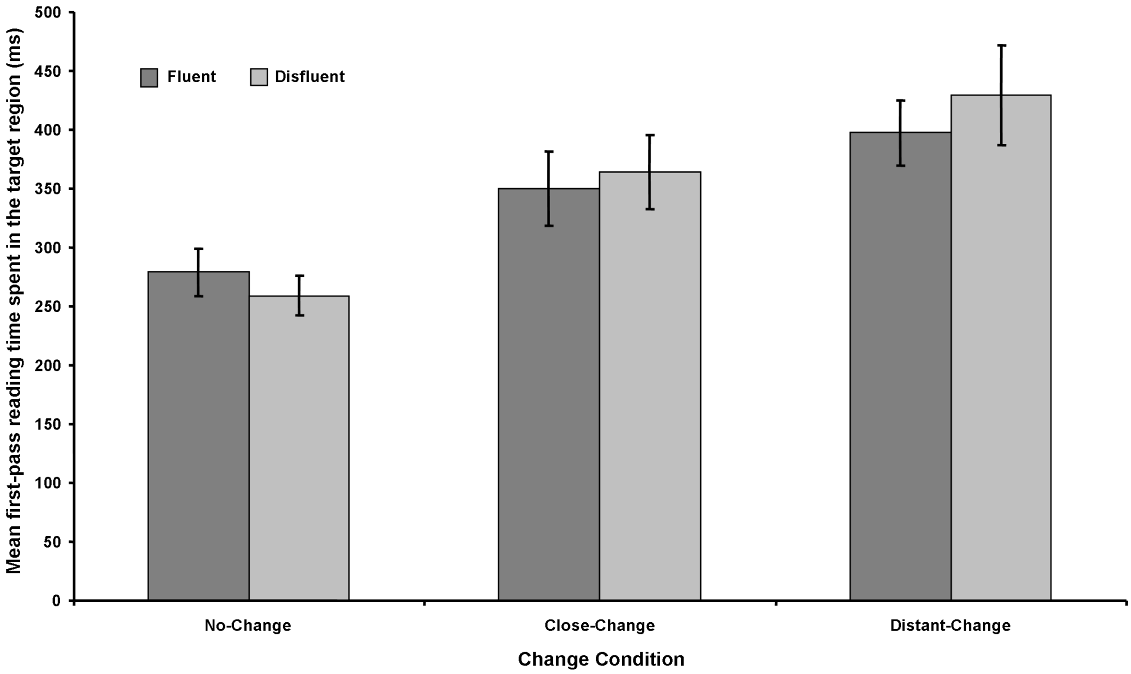

Figure 6.4: Experiment 3: Mean first-pass reading time spent in the target region for the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

154) than in the distant-change condition (mean = 413; S.D. = 156). An ANOVA showed that there was no main effect of disfluency by participants [$F(1, 23) < 1$] or by items [$F(1, 35) < 1$]. There was no main effect of change type by either participants [F(1,23) = 3.071] or items [$F(1, 35) < 1$]. The interaction of the two effects was non-significant by participants [$F(1, 23) < 1$] and items [$F(1, 35) < 1$].

6.5.4 Total time spent in region

In the no-change conditions, the mean total time spent fixating on the target region was 288ms (S.D. = 112) in the fluent condition and 285ms (S.D. = 123) in the disfluent condition (see figure 6.5). This difference was not found to be statistically significant by either participants [$t(23) < 1$] or items [$t(35) = 1.012$].

In the change conditions, the mean total time spent fixating on the target region was lower in the fluent condition (mean = 720ms; S.D. = 285) than in the disfluent condition (mean = 808ms; S.D. = 291). The mean total time spend fixating on the target region was lower in the close-change condition (mean = 680ms; S.D. = 250) than in the distant-change condition (mean = 831ms; S.D. = 316). The data for mean total time spent fixating on the target were then run in an ANOVA (see description above). The ANOVA showed that there was no main effect of disfluency by participants [$F(1, 23) = 2.020$] or by items [$F(1, 35) < 1$]. There was a main effect of change type by both participants [$F(1, 23) = 5.102$, $\eta_p^2 = .182$, $p = .034$] and items [$F(1, 35) = 7.967$, $\eta_p^2 = .185$, $p = .008$], indicating that total time spent in the target region was significantly higher in the distant-change condition. The interaction of the two effects was non-significant by participants [$F(1, 23) < 1$] and items [$F(1, 35) < 1$].

6.5.5 *Number of fixations*

In the no-change conditions, the number of fixations in the target region was 1.279 (S.D. = .450) in the fluent condition and 1.321 (S.D. = .601) in the disfluent con-

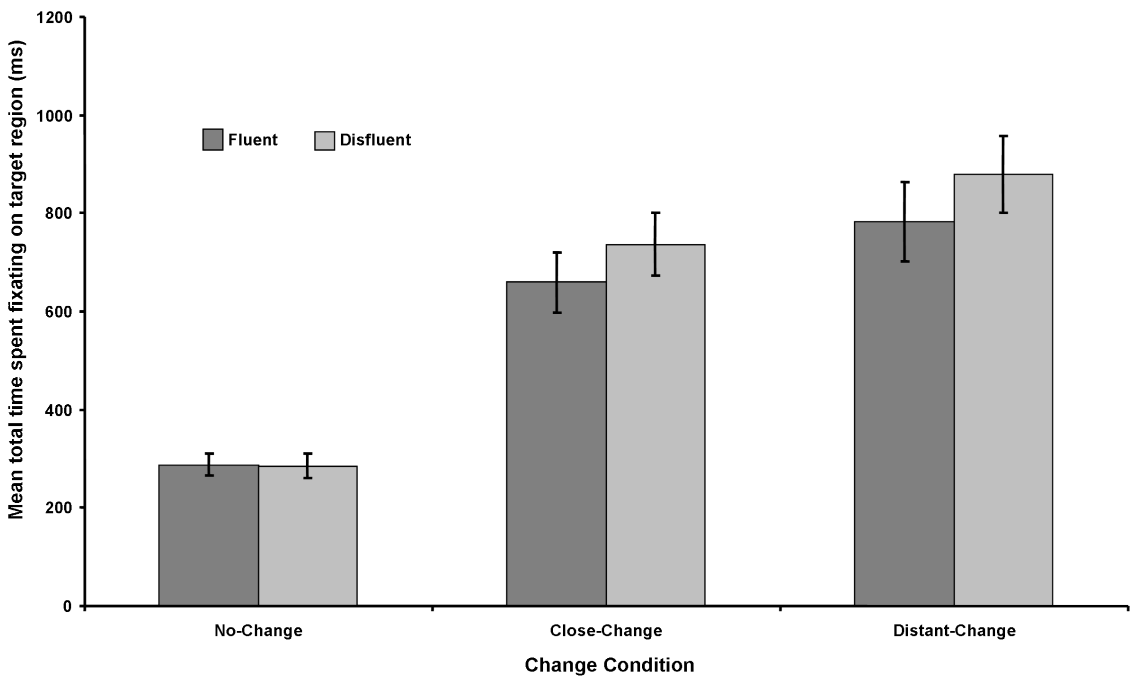

Figure 6.5: Experiment 3: Mean total time spent fixating on the target region for the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

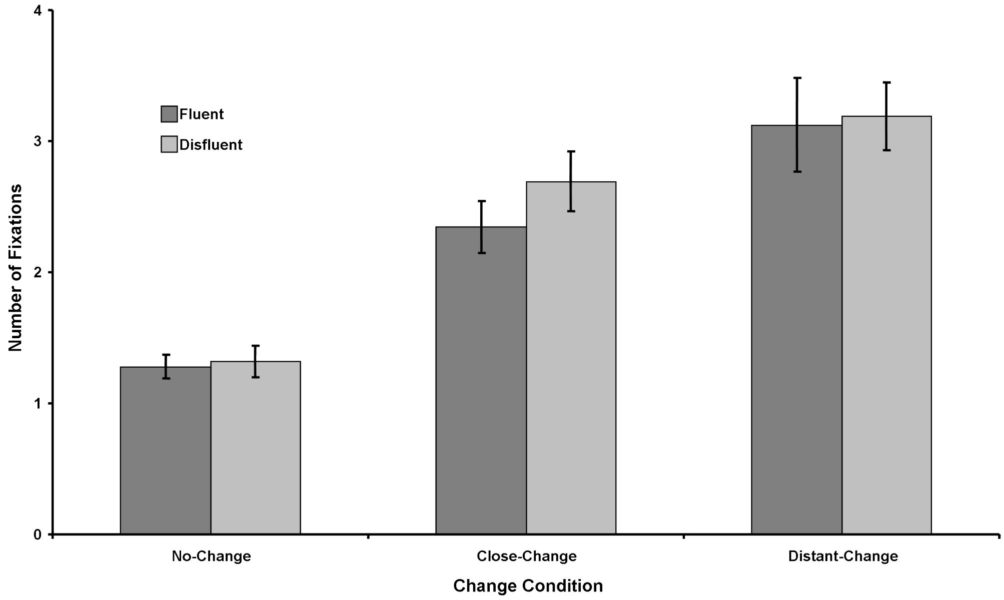

Figure 6.6: Experiment 3: Mean number of fixations in the target region for the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

dition (see figure 6.6). This difference was not found to be statistically significant by either participants [$t(23) < 1$] or items [$t(35) < 1$].

In the change conditions, the number of fixations in the target region was lower in the fluent condition (mean = 3.902; S.D. = 1.397) than in the disfluent condition (mean = 4.285; S.D. = 1.492). The number of fixations in the target region was lower in the close-change condition (mean = 3.688; S.D. = 1.213) than in the distant-change condition (mean = 4.715; S.D. = 2.016). An ANOVA showed that there was no main effect of disfluency by participants [$F(1, 23) < 1$] or by items [$F(1, 35) < 1$]. There was a main effect of change type by both participants [$F(1, 23) = 9.302$, $\eta_p^2 = .288$, $p = .006$] and items [$F(1, 35) = 4.7875$, $\eta_p^2 = .120$, $p = .035$]. The interaction of the two effects was non-significant by participants [$F(1, 23) < 1$] and items [$F(1, 35) < 1$].

6.5.6 Probability of regression back into the region

In the no-change conditions, the mean proportion of trials where the participant made regressive eye-movements back into the target region was .185 (S.D. = .193) in the fluent condition and .260 (S.D. = .208) in the disfluent condition (see figure 6.7). This difference was not found to be statistically significant by either participants [$t(23) = 1.580$] or items [$t(35) < 1$].

In the change conditions, the mean proportion of trials where the participant made regressive eye-movements back into the target region was lower in the fluent condition (mean = .484; S.D. = .208) than in the disfluent condition (mean = .568; S.D. = .147). The mean proportion of trials where the participant made regressive eye-movements back into the target region was lower in the close-change condition (mean = .483; S.D. = .188) than in the distant-change condition (mean =

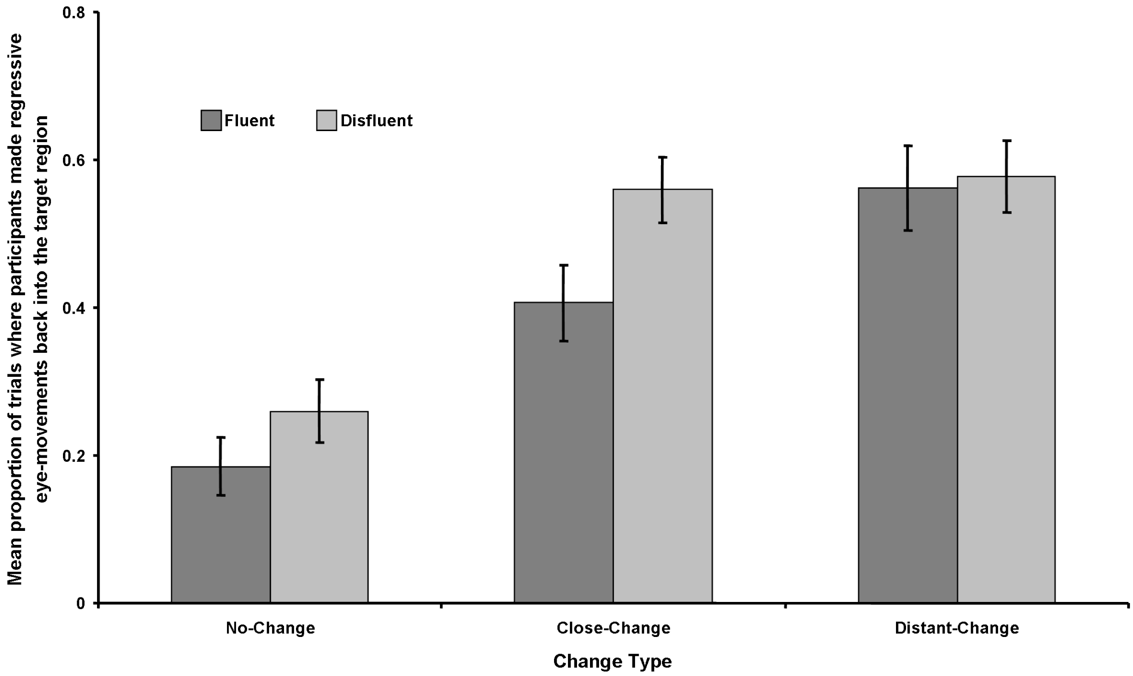

Figure 6.7: Experiment 3: Mean proportion of trials where the participant made regressive eye-movements back into the target region for the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

.570; S.D. = .219). An ANOVA showed that there was a marginal effect of disfluency by participants [$F(1, 23) = 3.715$, $\eta_p^2 = .1.39$, $p = .066$] and by items [$F(1, 35) = 3.307$, $\eta_p^2 = .086$, $p = .078$]. There was no main effect of change type by participants [$F(1, 23) = 2.131$] though there was a marginally significant effect by items [$F(1, 35) = 3.170$, $\eta_p^2 = .083$, $p = .084$]. The interaction of the two effects was marginally significant by participants [$F(1, 23) = 3.237$, $\eta_p^2 = .123$, $p = .085$] but not by items [$F(1, 35) = 2.526$]. Because this data appears to pattern with the results of Experiment 2, the final planned comparison between the probability of regression in the fluent and disfluent close-change conditions was performed. This revealed a significant effect of fluency where the proportion of regressive eye-movement was significantly higher in disfluent contexts (mean = .406, S.D. = .251) than in the fluent context (mean = .559, S.D. = .220) in both participants [$t(23) = 2.611$, $\eta_p^2 = .229$, $p = .016$] and items analyses [$t(35) = 2.228$, $\eta_p^2 = .124$, $p = .032$].

Discussion

Experiment 3 aimed to explore the attentional effects of hesitations on the representation of constituents encountered after the disfluency. Eyetracking measures taken during a change-detection paradigm were used to assess whether the hesitations had induced an effect on the covert attention directed towards these constituents. If covert attention towards these constituents had indeed increased, then modulation of the measures of eye movements might be expected, these could include looking at the changed constituents longer or being more likely to regress back into the region of the change. The results from the experiment supported this hypotheses. Participants were more likely to regress back into the region of text that changed if the changed constituent was first encountered in a passage where it was preceded by hesitation. Measures from the first-fixation times, first-pass times, total time spent in the target region and number of fixation on the target region did not show effects that were sensitive to the presence of hesitation. One further point of interest is that despite these regressive eye movements the data from the number of fixations and the total-time measures remained statistically insensitive to the added attention to the changed area. Interestingly, these results mirror Raney and Rayner (1995) who found that during a re-reading task (where each passage to be read was presented to the participant twice visually), participants did not alter their eye-movement behaviour for words that had changed from near synonyms presented in the first presentation (though a non-significant trend for longer total fixation times on near synonyms was observed).

6.6 General discussion

Experiments 2 and 3 support the hypothesis that hesitations affect the covert attention directed towards constituents that are immediately preceded by hesitations. Experiment 2 showed this using the data recorded from participants overt responses to the question 'did a change occur?' and Experiment 3 showed this using data from regressive eye-movements made by participants as they read over a second presentation of the passage.

Further experiments outlined in this thesis will explore more subtle distinctions in hesitations to investigate the cause of this attentional heightening. In light of this, one purpose for conducting Experiment 3 was to see whether eyetracking might provide a more useful measure of any attentional modulation. The only eyetracking measure that showed sensitivity to the fluency manipulation was the proportion of regressions made back into the changed region. The usefulness of this was assessed in two ways; the type of measurement, and the effect size. As far as the type of measurement is concerned, data from regressive eye-movements does not represent an advance over the accuracy measures from Experiment 2. If the total-time or first-pass time showed a similar sensitivity, the fact that these are continuous measures might represent some sort of advantage over the accuracy scores as the data could be used in more sophisticated and powerful statistical techniques such as modeling the effect using regression. The regressive eye-movement data however is binary, regressive eye movements were either made or not made on a trial by trial basis. While this does does not rule out the use of techniques such as logistic regression, the eye movement data provides nothing over-and-above what the behavioural data provide. As far as the effect size is concerned, again the eyetracking data represents no improvement upon the accuracy data and in fact the effect size for the critical comparison (the effect of fluency in the close-change condition) was higher in the accuracy score data ($\eta_p^2 = .419$) than it was in the eye-movement data($\eta_p^2 = .229$).

While the eye-tracking measures do not seem to provide a useful avenue for the further work on disfluency that will be outlined in this thesis, one use for the current data is that it allows us to explore whether hesitations affect the readers eye-movements in a similar way to linguistic focus. It seems that the attention capturing effects of hesitations lead to subtly different effects when compared to processes involved with focus. Ward and Sturt (2007) found an interaction trend between the effects of focus and change distance when looking at global measures of processing, such as total-time and number of fixations. In Experiment 3, these measures of global processing did not show any sensitivity to the fluency manipulation. While design differences between the current experiment and Ward and Sturt (2007) must be considered, this difference does suggest the two processes may have at least partially distinct mechanisms of action.

One advantage of Experiment 3 is that the data mitigate the paradigm against one potential criticism: that the participants were using the occurrence of a disfluency to guide their answering. If this were the case, then there should be a difference between the fluent and disfluent eye-movement behaviour for the no-change conditions. This does not appear to be the case: there was no significant difference between the two fluency conditions in any of the measures (all $ts < 1.580$).

6.7 Conclusion

The importance of Experiment 2 and 3 are that they provide evidence of covert attentional heightening to constituents encountered after hesitations. Further work in this thesis will go on to look at the specific aspects of the hesitation that trigger this attentional heightening. The paradigm presented in Experiment 2 provides a good platform from which to start this research. In the next chapter, the same technique is used to explore whether attentional heightening occurring after disfluency is a result of the gap between propositional content that a disfluency introduces, or whether specific auditory aspects of the filler are acting to heighten attention.

Chapter 7

The role of the filler and the surrounding silence

7.1 Chapter overview

In Chapters 3 and 4 it was suggested that the attentional effects of hesitations might not only manifest in immediate attentional engagement, but might also induce changes that could be said to modulate *covert attention*. The current chapter presents three experiments exploring the possible mechanisms that trigger these effects. In Chapter 6, Experiment 2 provides a method for testing the effect on covert attention that is induced by hesitation disfluency. After their initial mention, constituents that are encountered after hesitations are represented differently from equivalent fluent material and changes to these constituents are noticed more frequently. The following chapter uses this technique to investigate disfluency further. Specifically, it investigates the fillers in hesitations, and the periods of silent pause that often flank these fillers. The experiments outlined here explore how processes associated with these two components of hesitations interact and how this interaction triggers heightened attention.

7.2 Triggers

Hesitations often include a number of physically distinct components: prolongations, pauses and fillers (see section 2.2.4). Research outlined in Chapter 2 along with the previous experiments reported in this thesis show that there is strong evidence to suggest that hesitations affect listener attention. Furthermore, there is evidence to suggest that this effect is concomitant with changes to the processing of the spoken material (e.g., changes to prediction processes; Corley et al., 2007) and changes to the lasting representation of the information (e.g., the memory results in Chapter 5). One issue surrounding these effects regards their underlying trigger. Disfluent pauses change speech in at least two ways; they change the sound of the speech (introducing an uncommon phonetic interruption), and they change the timing of the speech (increasing the delays between the propositional elements). Mechanisms related to either of these changes might be driving these effects.

For an example of how auditory characteristics of the disfluency might drive the prediction and representation effects, the non-propositional gap in the disfluency associated with a break in sound input could be inducing attentional heightening through its novelty. As is outlined in Chapter 3, novel stimuli heighten attention both to a speech stream and to stimuli within a speech stream (e.g., Cherry, 1953; A. J. S. Sanford et al., 2006). Such novelty could be provided by the pitch pattern in hesitations. Fillers are often associated with flat pitch patterns or with rapid drops in fundamental frequency and then by abrupt rises occurring prior to the resumption of fluent speech. Furthermore, the fundamental frequency during the pause is remarkably low (in the bottom 15% of the speakers range; see O'Shaughnessy, 1992).

Alternatively, the effects could be driven by the delay introduced by the disfluency. The gap in propositional content is, for the listener, essentially a gap in any important incoming material. This gap could potentially be exploited by the processing system, allowing the listener to clear out any backlog of processing tasks and interpret the post-hesitation material under conditions of lower load. Such an effect could operate at the level of the language processor, or even in more domain general perceptual and post-perceptual processing systems.

These issues are systematically addressed by only a small set of studies, though various ideas have been proposed in light of other research that explores topics at the fringe of this issue. This research is covered in Chapter 2. A brief elaboration, outlining specifically how this research relates to the issues explored in the current chapter, is provided below.

In Fox Tree (2001), participants monitored speech for target words whilst listening to utterances with naturally occurring disfluencies. Participants were quicker to identify these when they were preceded by the filled pause *uh* than in equivalent fluent speech. This was not the case for the filled pause *um*. Importantly, in this experiment both the average duration and the phonetic characteristics changed between the conditions (*uh* and *um*). The *um*s not only sounded different from the *uh*s, but produced longer periods of silence before the speaker resumed. This is reflected in Fox Tree's statement:

> *The brief delays signalled by uh heightens listeners' attention for upcoming speech. The longer delays after um do not appear to alter listeners' attention in the same way.* (p. 325)

In this work (with regard to the English stimuli in Experiment 1), the *uh*s were of a shorter duration that the *um*s (327ms and 384ms, respectively). Notably, the periods of silence surrounding the fillers also varied. The pauses around the *uh*s were of a shorter duration before the filler (349ms and 362ms) whereas the opposite pattern occurred in the pauses after the filler (355ms and 334ms respectively).

Overall, the hesitations marked by *uh* were of shorter duration (1031ms) than those marked by *um* (1080ms).

Bailey and Ferreira (2003) showed that filled pauses changed the way listeners syntactically represented ambiguous sentences. They also found that environmental noises (such as bell rings of an equal duration to the fillers) had the same effect. Though the authors do not fully commit to any ideas about the nature of the trigger of these effects, they strongly favour an explanation indicating that they think it is the delay or gap in propositional content that is driving the effects.

Brennan and Schober (2001) come closer to systematically addressing the issue of an added gap in time and how this affects the listener. Participants responded to simple instructions in a task where the goal was to identify a single candidate from a set of geometrically simple shapes. Both speed and accuracy of the responses were higher for the utterances that contained pauses before the critical information than for fluent utterances. The effects were no different for the filled or unfilled pauses which were controlled for length. This result seems to support an explanation of the effects of hesitations that implicates timing, or the gap in propositional content (henceforth, a 'timing-gap' explanation). However, because of a design flaw (a smaller set of possible referents in the disfluent conditions compared to the fluent condition) it is difficult to say how informative the results are with regard to the filled and silent pauses.

Finally, MacGregor (2008) has shown that the 'timing-gap' explanation on its own will not explain some experimental results. Using the N400 effect as an index of integration difficulty, she compared listeners' responses to unpredictable (difficult to integrate) words against predictable words in fluent contexts, and in disfluent contexts where the critical words were preceded by hesitations. In an initial experiment, the magnitude of the N400 effect was significantly reduced for disfluent utterances with an *er*, showing a clear effect of hesitations on listeners' language processing (this work is also reported in Corley et al., 2007). A second experiment used the same paradigm and stimulus set, but the *er*s were replaced by silent pauses of an equivalent length. In this study, the magnitude of the N400 effect was not significantly reduced for disfluent utterances. This suggests that the phonetic form of the filler is important over and above the timing gap provided by the hesitation.

In light of these conflicting results, the current research seeks to explore the issues of the interaction of the sound of the filler (its phonetic form) and the periods of silence that surround it. Specifically, it will investigate whether the gaps in timing before and after the filler affect the attentional allocation, both in terms of covert attention (the current chapter) and immediate attentional engagement (Chapter 8). Experiments 4–6 use a change-detection paradigm to investigate the 'timing-gap' and 'phonetic-form' accounts of the attentional effects of hesitations. The length of the silent pauses before and after the filler are manipulated. If the 'timing-gap' account is correct, then the effects of the hesitation should be contingent upon the addition of the silent pause, rather than its position. If the 'phonetic-form' account is correct, then any experimental effects might be sensitive to the position of added silent pause.

7.3 Experiment 4

If the 'timing-gap' explanation is accurate, then hesitaions that include longer gaps between the propositional material might be expected to produce larger attentional effects than those where the gap is relatively short. Experiment 4 aimed to test this prediction. Effectively, it is a replication of Experiment 2 with an additional level in the fluency variable. In this experiment, participants carrying out a change-detection task heard both fluent and disfluent passages. Unlike Experiment 2, the hesitation were either standard hesitations (unaltered recordings such as those in

Experiment 2), or hesitations that had been digitally altered to extend the length of the silent pause that followed the filler.

7.3.1 Method

Participants

Twenty-four members of the University of Edinburgh student community were recruited to take part in the study. They were each paid £2 compensation for their participation. All procedures regarding participation complied with the BPS ethical guidelines.

Materials

The materials for Experiment 4 were created from a modified set of those used in Experiments 2 and 3. The same 36 passage frames were used with the exception of those in the distant change condition. An extra level was added to the factor of fluency. Rather than fluent and disfluent versions of the utterances, there were now versions that were fluent, versions that included standard hesitation or the new condition of versions that included extended hesitations. For this condition, all the stimuli within the set had an additional silence added between the filler *er* and the onset of the post-hesitation word. This silence was added by digitally splicing 350ms of silence into the existing silence that followed the filler. It is important to note that although a non-varying 350ms of silence was added to each of the passages in the extended pause condition, the non-systematic nature of this pause length was still kept constant by the additional periods of silence that were already present in the pre-manipulated stimuli (those that appeared in the standard hesitation condition). This manipulation allowed both hesitation conditions to have equally varied lengths of pauses following the filler. The change condition was reduced to two levels (no-change and close-change, as in Ward & Sturt, 2007). The 36 target passages were divided using a Latin-square method, so that each participant heard 18 fluent targets and 18 disfluent targets with hesitations. Within these groups, each participant heard: 9 no-change targets, 9 close-change targets, and 18 fluent and disfluent filler passages.

Procedure

The experiment followed the procedure outlined for Experiment 2 (see section 6.3.1). The experiment lasted approximately 20 minutes. The 54 trials were presented to the participant in a random order. During each trial participants listened to a recorded presentation of the passage and then read over a text version of the same passage. The participants were then asked to indicate whether a change to the material had occurred in the second presentation, and to identify the word that had changed.

Data Analysis

Experiment 2 had a 2x3 design where the factor of fluency had two levels and the factor of change type had three levels. Here however, the 2x3 design was switched so that fluency had three levels and the factor of change type had two levels. This required an altered approach to the data analysis. In Experiment 2, ‘change’ and ‘no-change’ conditions were analysed separately as the two processes of answering represent differing cognitive operations. The type of mistake that an inaccurate answer represents differed between the two: the false identification of a change in the first is a mistake that is functionally distinct from the missing of a change in the second. Here, the same approach was taken. However, this required the use of two one-way ANOVAs, one for the fluent conditions and one for the disfluent conditions. Each of these had one factor of fluency with three levels (fluent, standard hesitation, and extended hesitation). After this, planned pairwise comparisons between each level of fluency were carried out, though only on the data from the change condition.

Again, all the statistical analyses were performed on the arcsine transform of the data. In cases where the assumption of sphericity was not met, Greenhouse-Geisser corrections to degrees of freedom were applied and the corrections for F and p values are reported.

7.3.2 Results

The mean probability of correctly responding 'no change' to passages where no change had occurred was .875 (S.D. = .123) in the fluent condition, .882 (S.D. = .134) in the standard-hesitation condition and .847 (S.D. = .162) in the extended-hesitation condition (see figure 7.1). This difference in means was not found to be statistically significant by either participants [$F(23) < 1$] or items [$F(35) < 1$].

The mean probability of correctly responding 'change' to passages where a change had occurred was .681 (S.D. = .235) in the fluent condition, .715 (S.D. = .243) in the standard-hesitation condition and .729 (S.D. = .202) in the extended-hesitation condition. This difference in means was not found to be statistically significant by either participants [$F(23) < 1$] or items [$F(35) = 1.029$]. As these analyses produced non-significant results, none of the planned comparisons were performed.

7.3.3 Discussion

Experiments 4 aimed to explore whether the attentional effects of hesitations were modulated by the gap in time between the propositional contents of an utterance. If attention is heightened because the listener is afforded more time to clear the backlog of processing, then the attentional effect should be larger, or at least as large, when these hesitations are longer. Here, a change-detection paradigm was used to assess whether participants were better at noticing subtle change after either short or long hesitations. The results from this experiment do not support this hypothesis. Participants were no more likely to detect changes to constituents after long hesitations than they they were after short hesitations. Additionally, unlike Experiment 2, participants were no more likely to detect changes after any sort of hesitation than they were after standard fluent speech. This suggests that covert attention was not modulated by the hesitations in this experiment. In contrast to the result in Experiment 2 and the result from the behavioural data in Experiment 3, the null result here could be due to a number of reasons. It could be that extending the hesitations actually counteracted the attentional heightening that occurred in disfluent contexts in Experiments 2 and 3. This however, does not explain why even standard (non-extended) hesitations did not produce these effects here. Alternatively, the null effects could be a result of the difference between this paradigm and the paradigms used in Experiments 2 and 3. Here disfluent passages were more frequent than they were in Experiment 2. They occurred at a ratio of 2:1 (disfluent:fluent) within the experimental materials (as well as in the fillers). This

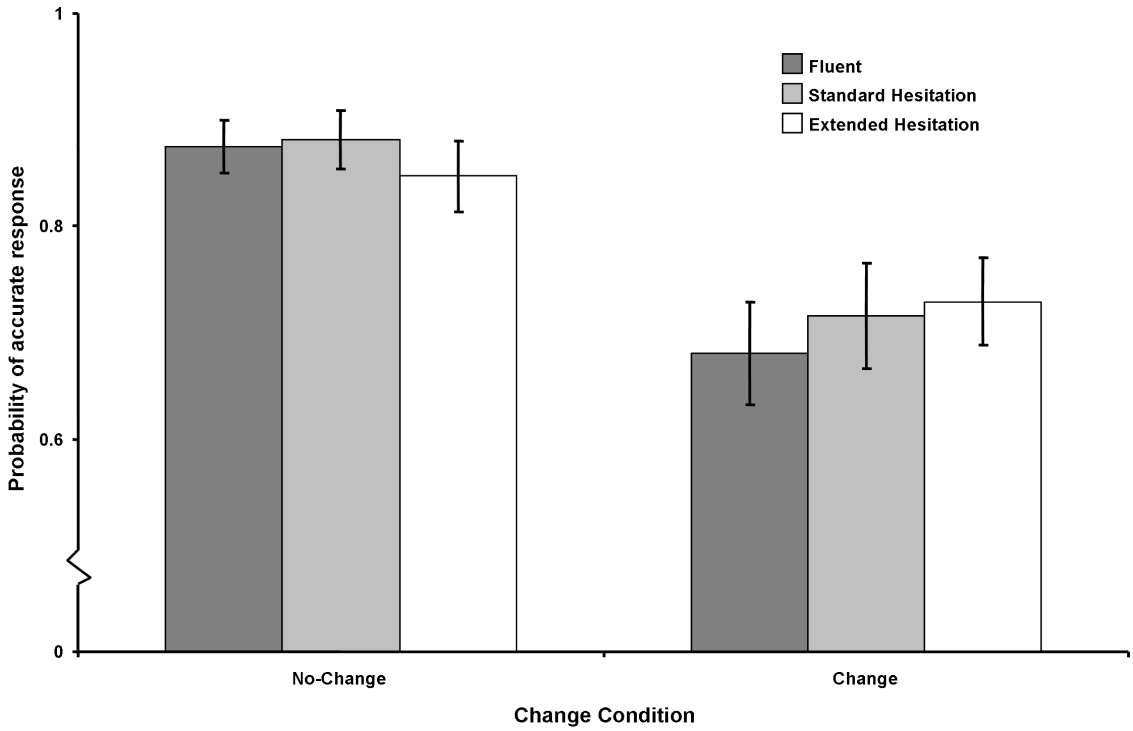

Figure 7.1: Experiment 4: Mean probability of accuracy in the no-change and change conditions for fluent, standard-hesitation and extended-hesitation passages (bars represent ±1 std. error of the mean).

high proportion of disfluencies could be altering participants' approach to the task. Also, the participants' approach to the task might have also been altered by the lack of a distant-change condition. These possibilities are experimentally explored in Experiment 5 below.

7.4 Experiment 5

Given the issues with Experiment 4, Experiment 5 aimed to address the same 'timing-gap' issue in a paradigm that more closely resembled Experiment 2. As with Experiment 4, the predicted result is that if the 'timing-gap' hypothesis is correct, then the effect of attention (more frequent noticing of close changes when they occur after disfluency) should be larger or at least as large as the effect in the context of shorter hesitations (the effect seen in Experiment 2).

To do this, Experiment 5 reverted to the exact paradigm outlined in Experiment 2. This offers three advantages: first, participants might be sensitive to the very high disfluent-to-fluent passage ratio that Experiment 4 involved. Unlike Experiment 4, here (as with Experiment 2) there was a 1:1 ratio of disfluent and non-disfluent passages. Second, when participants are only exposed to semantically close changes, then they might increase their vigilance, resulting in abnormal language processing that nullifies any of the effects of interest. In the current experiment, the change condition included semantically distant changes. Third, the known effect size and power associated with this paradigm makes a null result easier to interpret. Knowing these factors help mitigate against the the possibility of interpreting a null result simply as a type II error (Mosteller, 1948). The effect size of the critical comparison ($\eta_p^2 = .326$) provided in Experiment 2 suggests the same effect or one of a greater magnitude would have a high observed power ($\phi = .891$) with 24 participants (and an alpha level of .050). Thus, if this paradigm also shows that there is no effect for extended hesitations, it is extremely unlikely that this is due to a type II error.

7.4.1 Method

Participants

Twenty-four members of the University of Edinburgh student community were recruited to take part in the study. They were each paid £2 compensation for their participation. All procedures regarding participation complied with the BPS ethical guidelines.

Materials

The materials for Experiment 5 were the same as those used in Experiment 4, though the standard hesitation length passages were excluded. Also, the distant change passages from Experiment 2 were included. The hesitation length for these was extended as is outlined in the materials section of Experiment 4. The 36 target passages were divided as with Experiment 2 so that each participant heard 18 fluent targets and 18 disfluent targets with extended hesitations. Within these groups, each participant heard: 6 no-change targets, 6 close-change targets and 6 distant-change targets, and 18 fluent and disfluent filler passages.

Procedure

The experiment followed the procedure outlined in Experiment 2 (see section 6.3.1). The experiment lasted approximately 20 minutes. The 54 trials were presented to the participant in a random order. During each trial participants listened to a recorded presentation of the passage and then read over a text version of the same passage. The participants were then asked to indicate whether a change to the material had occurred in the second presentation, and to identify the word that had changed.

Data Analysis

The data analysis followed the procedure outlined for Experiment 2. Comparisons for the data involved the measure of accuracy of the response for the change/no-change question and all the statistical analysis was performed on the arcsine transform of this data. A pairwise t-test comparison of the accuracy of the responses in the fluent and disfluent no-change conditions was run to ensure that fluency was not affecting the general processing of the non-changed passages (analogous to a false-alarm in a signal-detection paradigm). Next, an ANOVA with the factors of fluency (fluent and disfluent) and change-type (close change and distant change) was performed. Finally, to see if the disfluency was acting as an attention orienting device similar to linguistic focus, the critical planed comparison between the fluent and disfluent response accuracy was run on the data from the close-change condition.

7.4.2 Results

The mean probability of correctly responding 'no change' to passages where no change had occurred was .938 (S.D. = .096) in the fluent condition and .951 (S.D. = .092) in the disfluent condition (see figure 7.2). This difference was not found to be statistically significant by either participants [$t(23) < 1$] or items [$t(35) < 1$].

The mean probability of correctly responding to a change that had occurred to passages with close changes was higher in the fluent condition (mean = .736; S.D. = .273) than in the disfluent condition (mean = .695; S.D. = .218). The mean probability of correctly responding to a change that had occurred to passages with distant changes was higher in the fluent condition (mean = .854; S.D. = .199) than in the disfluent condition (mean = .847; S.D. = .202). The ANOVA showed no main effect of disfluency by either participants [$F(1, 23) < 1$] or by items [$F(1, 35) < 1$]. A main effect of change type was found by both participants [$F(1, 23) = 12.298$,

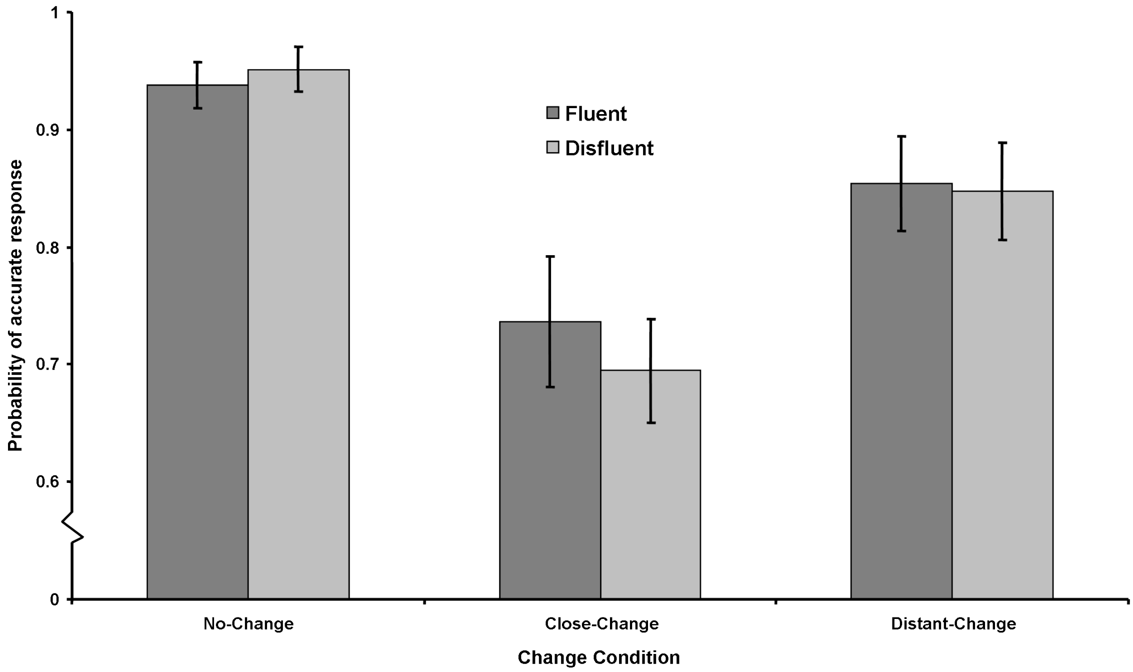

Figure 7.2: Experiment 5: Mean probability of accuracy in the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

$\eta_p^2 = .348$, $p = .002$] and items [$F(1,35) = 8626$, $\eta_p^2 = .198$, $p = .006$]. The interaction of the two effects was not significant by participants [$F(1,23) < 1$] or by items [$F(1,35) < 1$].

For the sake of comparison with the power analysis above (see section 7.4), the critical planned comparison of the fluent and disfluent response accuracy was run on the data from the close-change condition. This showed that the probability of detecting a close change not significantly different in the fluent condition compared to the disfluent condition in either the by participants analysis [$t(23) < 1$] or the by items analysis [$t(35) < 1$].

7.4.3 Discussion

Experiment 5 aimed to explore the possibility that the attentional effects of hesitations were modulated by the gap in time between the propositional contents of an utterance. If attention is heightened because the listener is afforded more time to clear the backlog of processing, then the attentional effect should be larger, or at least as large, when these hesitations are longer. Here, a change-detection paradigm was used to assess whether participants were better at noticing subtle change after long hesitations. The results from this experiment do not support this hypothesis. Participants were no more likely to detect changes to constituents after long hesitations than they they were after equivalent fluent utterances.

Unlike the null effect in Experiment 4, the null effect here can not be accounted for by the high disfluency frequency within the paradigm and is unlikely to be due to a type II error. Experiment 2 showed a positive effect of attentional heightening after hesitations using the same paradigm. The only difference here is that hesitations were extended by adding 350ms of silence after the filler and before the resumption of fluent speech.

To summarise, not only does the added gap between the filler and the post-disfluent material not increase the effect of the hesitation, but it nullifies the effect seen in an experiment conducted with non-extended hesitations (Experiment 2). This seems strong evidence against the 'time-gap' account of the effect of hesitations. If hesitations are heightening listener attention because they provide the listener with a break between propositional content that allows them to wrap-up a backlog of processing, then increasing this gap should enhance the effect or at least leave it unchanged. Here such a gap simply nullified the effect.

If the 'timing-gap' account for the effects of hesitations is not sufficient to account for the data seen in Experiment 2, another likely explanation is that the effect is related to the phonetic form of the hesitation. As outlined above, the fillers in hesitations are marked by the novel occurrence of flat or abruptly falling fundamental frequency of the speech stream (O'Shaughnessy, 1992). If this novelty is the trigger for increased listener attention, this provides a possible account of why no effect was seen in Experiment 5. If the novelty of the filler triggers attentional heightening, then increasing the gap between this filler and the post-disfluent material would simply provide more time for this attentional effect to decay before fluent speech is resumed. Experiment 6 tests this proposition by again including extended hesitations in the paradigms set-up in Experiment 2, thought here the manipulation involves extending the length of the pre-filler pause rather than the post-filler pause.

7.5 Experiment 6

If the phonetic form of the filler is inducing the heightened attention in listeners post-disfluency, then lengthening the intervening period between the filler and the post-disfluent material should reduce the extent to which attention is heightened for the processing of the post-disfluent material. Results fitting this prediction are given in Experiment 5. Alternatively, this reduction in effect could have been a consequence of the extended hesitation length. If longer hesitations simply do not induce attentional heightening, then a similar pattern of results is expected. Experiment 6 aims to address this possibility. Effectively, Experiment 6 is a replication of Experiment 5. However, here the extended silence that is added to the hesitation occurs before the filler rather than after. If the null results in Experiment 5 are due to the fact that extended hesitations do not induce attentional heightening, then the null effect of hesitations seen in Experiment 5 should occur in the current experiment as well. If however, the null effect in Experiment 5 is due to the decay (over the post-filler silence) of heightened attention that was induced by the filler, then the results should pattern with those in Experiment 2.

7.5.1 Method

Participants

Twenty-four members of the University of Edinburgh student community were recruited to take part in the study. They were each paid £2 compensation for their participation. All procedures regarding participation complied with the BPS ethical guidelines.

Materials

The materials for Experiment 6 were the same as those used in Experiment 2, though the hesitation length was again extended. For the disfluent condition, all the stimuli within the set had an additional silence added between the pre-filler word and the filler *er*. This silence was added by digitally splicing 350ms of silence into the existing silence that preceded the filler. Other than the position of the additional silence the extension procedure was identical to the one carried out in Experiment 4. Each participant heard: 6 no-change targets, 6 close-change targets and 6 distant-change targets, and 18 fluent and disfluent filler passages.

Procedure

The experiment followed the procedure outlined for Experiment 2 (see section 6.3.1). The experiment lasted approximately 20 minutes. The 54 trials were presented to the participant in a random order. During each trial participants listened to a recorded presentation of the passage and then read over a text version of the same passage. The participants were then asked to indicate whether a change to the material had occurred in the second presentation, and to identify the word that had changed.

Data Analysis

The data analysis followed the procedure outlined for Experiment 2. Comparisons for the data involved the measure of accuracy of the response for the change/no-change question and all the statistical analyses were performed on the arcsine transform of the data. A pairwise t-test comparison between the accuracy for the response in the fluent and disfluent no-change condition was run to ensure that fluency was not affecting the general processing of the non-changed passages. Next, an ANOVA with the factors of fluency (fluent and disfluent) and change-type (close change and distant change) was performed. Finally, a critical planed comparison between the fluent and disfluent response accuracy was run on the data from the close-change condition.

7.5.2 Results

The mean probability of correctly responding 'no change' to passages where no change had occurred was .938 (S.D. = .108) in the fluent condition and .931 (S.D. = .120) in the disfluent condition (see figure 7.3). This difference was not found to be statistically significant by either participants [$t(23) < 1$] or items [$t(35) < 1$].

The mean probability of correctly responding to a change that had occurred to passages with close changes was lower in the fluent condition (mean = .660; S.D. = .267) than in the disfluent condition (mean = .806; S.D. = .201). The mean probability of correctly responding to a change that had occurred to passages with distant changes was lower in the fluent condition (mean = .813; S.D. = .192) than in the disfluent condition (mean = .931; S.D. = .098). The ANOVA showed a main effect of disfluency by both participants [$F(1,23) = 12.675$, $\eta_p^2 = .355$, $p = .002$] and items [$F(1,35) = 14.989$, $\eta_p^2 = .300$, $p < .001$]. There was a main effect of change type by both participants [$F(1,23) = 17.169$, $\eta_p^2 = .427$, $p < .001$] and items [$F(1,35) = 10.545$, $\eta_p^2 = .232$, $p = .003$]. The interaction of the two effects was not significant by participants [$F(1,23) < 1$] or by items [$F(1,35) < 1$].

The critical planned comparison between the fluent and disfluent response accuracy was run on the data from the close-change condition. This showed that the probability for detecting a close change was significantly higher in the disfluent condition than in the fluent condition in the by participants analysis [$t(23) = 2.454$, $\eta_p^2 = .207$,

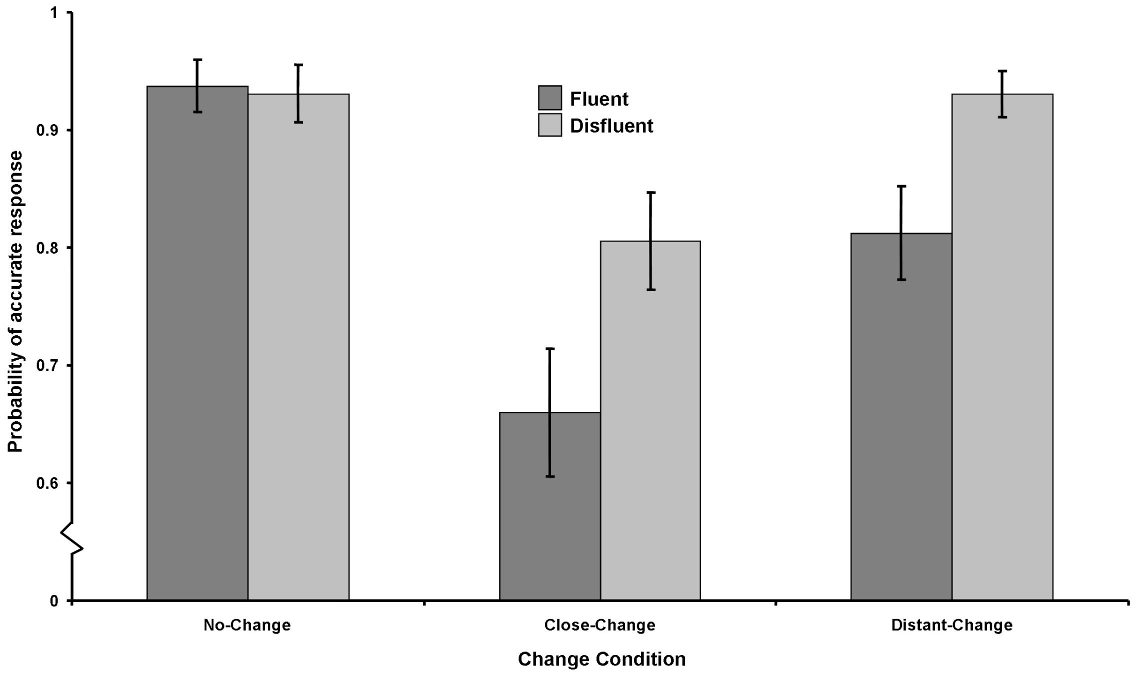

Figure 7.3: Experiment 6: Mean probability of accuracy in the no-change, close-change and distant-change conditions for fluent and disfluent passages (bars represent ±1 std. error of the mean).

$p = .022$] and in the by items analysis [$t(35) = 3.477$, $\eta_p^2 = .257$, $p < .001$]. Finally, because of the pattern of two significant main effects within the data, an unplanned comparison was run between the fluent and disfluent response accuracy from the distant-change condition. This showed that the probability for detecting a distant change was significantly higher in the disfluent condition than in the fluent condition in the by participants analysis [$t(23) = 2.207$, $\eta_p^2 = .175$, $p = .038$] and in the by items analysis [$t(35) = 2.786$, $\eta_p^2 = .181$, $p = .009$].

7.5.3 Discussion

Experiment 6 aimed to explore whether the attentional effects of hesitations could be attributed to the phonetic form of the filler. A change-detection paradigm was used to assess whether the hesitations that included extended pauses before the filler *er* had induced an effect on the covert attention directed towards these constituents. If covert attention towards these constituents had indeed increased, then it is more likely that the participants would detect changes to them. The results from this experiment supported this hypothesis. Participants were more likely to detect changes to constituents if those constituents were preceded by these extended hesitations. The data support the hypothesis that the null effect seen in Experiment 5 is due to the period of silence intervening between the filler and the post-disfluent word and not due simply to the duration of the hesitation.

Unlike Experiment 2, the effect of disfluency was not limited to the data for the close-change condition. In Experiment 6, the mean detection rates were higher in the disfluent condition for both close and distant changes.

7.6 General discussion

Experiments 5 and 6 provide evidence that the attentional heightening effect of hesitations in speech is not a result of the extended gap in timing that the disfluency provides, but instead might be triggered by phonetic characteristics of the filler. How does this information relate back to other studies of the effects of disfluency?

The interpretation of Fox Tree's (2001) work suggests that characteristics of the filler are important. Notably though, the distinction she makes between hesitations that induce heightened attention and those that do not is based on the differing phonemic characteristics of the filler *uh* (transcribed in this thesis as *er*) and the filler *um*. The only details that are provided about these disfluencies besides their phonemic form is information about their duration and the duration of the surrounding silences. While *um*s were longer and were preceded by longer silent pauses, they occurred with a slightly shorter post-filler pause. This would seem to indicate that the mechanism suggested to account of the null effect in Experiment 5 here (the rapid decay of attentional heightening) was not operating in Fox Tree's experiment. One possible alternative to account for her data is that the *um*s could possibly be associated with less auditory novelty than *uh*s.

Similarly, Bailey and Ferreira (2003) showed that filled pauses changed the way listeners syntactically represented ambiguous sentences. While they strongly favour a 'timing-gap' account of their findings, two aspects associated with this idea are worth note. First, unlike much of the other research on hesitations, Bailey and Ferreira (2003) were interested in how disfluencies affect syntactic processing. It could be that the 'timing-gap' explanations are more applicable in the context of syntactic assignment than they are in the context of linguistic prediction or attention. Second, it is worth noting that the equivalent effects that they observe in the context of environmental noises confound the 'novelty' aspect of the phonetic form account with the added time that these noises add between propositional content.

While the data from the silent pause effect in Brennan and Schober (2001) favour a 'timing-gap' account over a 'phonetic form' account, the experimental flaw outlined above make this result difficult to interpret reliably (see section 7.1). Finally, the data from MacGregor (2008) seem to pattern with the data from the experiments outlined in this chapter, suggesting that aspects of the phonetic form of the filler are at least partially responsible for the effects of hesitations.

Finally, the data from studies being carried out concurrent to the present research seem to support the idea of an effect on attention that decays rapidly. A. J. S. Sanford and Molle (2006) used a paradigm similar to the change-detection paradigm used here. They showed that filled pauses induced similar *attention capturing* effects in words that immediately followed the filler. Changes to these words were detected more frequently, though changes to the words one position on from these in the passages were not noticed more frequently. However, this data does not fully support the phonetic characteristic explanation of the attentional effects of hesitations as silent pauses gave rise to similar effects.

In the next chapter, the ERP paradigm for Experiment 1 is again used to explore issues raised by the research outlined in this chapter. The work in the current chapter suggests that effects triggered by the phonetic characteristics of the filler induce attentional heightening, but that this decays at a rapid rate. Experiment 7 explores this theory using measurements of immediate attentional engagement.

Chapter 8

Experiment 7

8.1 Chapter overview

In Chapter 5, Experiment 1 provided evidence supporting the hypothesis that hesitation disfluency heightens listener attention. The immediate engagement of attention to speech was assessed by measuring attention-related ERP components. The results showed that these are attenuated after hesitations. In Chapter 6, Experiment 2 provides a method for testing the effect on covert attention that is induced by hesitation disfluency. Constituents that are encountered after hesitations are represented more strongly and changes to these constituents are noticed more frequently. In Chapter 7, Experiments 5 and 6 provided evidence suggesting that these effects on covert attention are sensitive to the intervening time between the filler *er* and the resumption of fluent speech: the increase in covert attention to post-disfluent constituents is only induced if there is a short gap in time between these two events. The current chapter explores the issue of whether immediate attentional engagement after hesitations is also sensitive to this gap.

8.2 Experiment 7

8.2.1 Introduction

In Chapter 7, various research was discussed that explored the underlying trigger for the effects of hesitations on processing. The two most prominent candidates for this triggering role are the timing gap that the hesitation provides and the phonological form of the filler. Evidence suggesting that listeners will not only use hesitations to inform them about the syntactic structure of the utterance that they are hearing, but also interruptions induced by environmental noises (such as sneezes or bells) indicate that it may be the timing gap that is important for effects of hesitations on syntactic assignment (Bailey & Ferreira, 2003). Similarly, evidence suggesting that listeners respond to instructions faster after hesitations marked by either filled or unfilled pause (Brennan & Schober, 2001) supports the timing-gap explanation.

However, there are results that indicate the phonetic form of the filler is important. Fox Tree (2001) showed that fillers that differ in phonological form have deferential effects on how quickly post-disfluent target words are identified. Importantly for the work outlined in the current chapter, MacGregor (2008) showed that phonetic form is important for immediate processing. She compared listeners' responses to unpredictable (difficult to integrate) words against predictable words in fluent contexts, and in disfluent contexts where the critical words were preceded by hesitations with no fillers but extended silent pauses (Experiment 3). Whereas hesitations that include the filler *er* affect how the unpredictable words are integrated into their preceding context (indexed by the N400; Corley et al., 2007), the integration of the unpredictable words is not affected by hesitations marked by silent pauses.

Although this lack of N400 attenuation for silent pauses in MacGregor (2008) indicates that the phonetic form of the filler is influencing immediate processing, there were other effects that were influenced by silent pause hesitations. Notably, participants were more likely to recognise words that were encountered in disfluent contexts. This applied exclusively to predictable words, just as in Experiment 1 here the memory effect for disfluency applied only to acoustically regular words. These results suggest that the silent pauses affect the way post-disfluent words are stored even if they do not affect the way these words are integrated during processing (as indexed by the N400).

In addition to this effect on memory, the silent pauses also affected processing as the utterances were being heard. While the N400 remained unaffected by the silent pauses, MacGregor (2008) noted what appeared to be an ERP effect resembling the late positive complex (LPC). This was present for difficult to integrate (unpredictable) words in the disfluent (silent-pause) condition. These unpredictable words elicited an increased positivity at frontal and left parietal sites, between 600 and 900ms. In line with with other LPC interpretations (e.g., Federmeier et al., 2007), MacGregor (2008) tentatively interpreted this as indication that the increased timing gap provided by the silent pauses might affect memory retrieval and control processes that occur during speech comprehension.

The results of Corley et al. (2007), Fox Tree (2001), and MacGregor (2008) as well as those in Experiments 5 and 6 here, suggest that some effects of hesitations are a result of more than just the increased timing gap that hesitations provide. Important cues from the phonetic form of the filler might be necessary to fully explain results showing that hesitations affect target identification and integration. The question addressed in Experiment 7 is whether the phonetic form of a hesitation is important in triggering the immediate engagement of attention induced by hesitations.

In the current experiment, EEG was recorded while participants listened to recorded utterances containing infrequent changes to the auditory characteristics of single words. As with Experiment 1, half of the time the manipulated words followed hesitations. These were marked by natural changes to the speech, such as elongations to words within the hesitation (e.g., *thee*), periods of silent pause, and the filler *er*. In contrast to Experiment 1, the periods of silence after the filler were digitally extended by splicing in 350ms of additional silence before the resumption of fluent speech.

Because the deviant (manipulated) words were infrequent and therefore novel with respect to their contexts, they would usually be expected to induce attention-related ERPs (MMN and P300) in both fluent and disfluent conditions. If the the attentional state of listeners was affected by preceding disfluency, then this change might be reflected in an effect of fluency on these ERPs such as was seen in Experiment 1. However, if the attentional heightening induced by the filler decays before the resumption of fluent speech (as the results from Experiment 5 indicate), then this effect on the ERPs may not manifest here. Thus, the Experiment provides a way to test the timing-gap and phonetic-form hypotheses with regards to the immediate attentional heightening that hesitations have been shown to produce. As with Experiment 1, the long-term effects of these disfluencies was also assessed using a surprise recognition-memory test at the end of the experiment.

8.2.2 Method

Participants

Twelve native English speakers participated in the experiment (4 male; mean age 21 years; range 19–25). All were right handed and reported no known neurological impairment. Informed consent was obtained in accordance the University of Stirling Psychology Ethics Committee guidelines. Participants were given financial compensation and course credit where applicable.

Materials

The materials for Experiment 7 were created from a modified set of those used in Experiments 1. The same 640 recorded passages were used. Here however, the hesitations in disfluent versions were digitally extended by adding 350ms of silence between the filler *er* and the onset of the post-hesitation word. With this manipulation aside, in all other ways the materials were identical to those used in Experiment 1, with the oddball manipulation (amplification and 125–1000Hz range manipulation), filler utterances (160) and standard to deviant ratio (3:1) remaining the same.

Procedure

The experiment comprised two sections. In the first, participants listened to the 320 experimental utterances and fillers. Materials were presented in a random order via computer loudspeakers in two blocks lasting around 20 minutes each, and separated by a break of a few minutes. Participants were instructed to listen to the recordings as if they were part of a normal conversation, but were not given any other task. They were not told specifically about the presence of the disfluencies or acoustically manipulated words, but were told that occasionally, the sound editing quality would drop, which they should ignore. Electroencephalograms (EEG) were recorded from 61 silver/silver-chloride electrodes embedded in an elasticized cap at standard 10–20 locations (Jasper, 1958), using a left-mastoid reference. Electro-oculargrams (EOGs) were collected to monitor for eye-movements. EEG and EOG were amplified (bandpass filtered online, 0.01–40 Hz) and continuously digitized (16 bit) at 200Hz. Electrode impedances were kept below 5KΩ. Epochs were created from the EEG (150ms before the onset of the target words to 800ms after the onset) and these data were re-referenced offline to the average of the left and right mastoid electrodes, baseline corrected (relative to the average over the pre-stimulus interval) and smoothed over 5 points. Before averaging into ERPs, individual epochs were screened for drift of ± 75μV over 500ms (amplitude difference between first and last data point of each epoch), and for artifacts of ± 75μV. The screening process resulted in the loss of 3.86% of epochs, with no significant variation in rejections between conditions [$F(3, 33) = 1.416$]. Average ERPs were formed time locked to the onset of target words for each participant (minimum of 16 artefact free trials were required for inclusion).

8.2.3 ERP Results

ERPs associated with the onsets of deviant target words were compared to ERPs to non-manipulated standard controls for fluent and disfluent conditions. Because pre-stimulus baselines in fluent and disfluent utterances were different (including an *er* for disfluent cases), effects related to the acoustic manipulations were analysed separately for fluent and disfluent conditions.

Figures 8.1 and 8.2 show the waveforms of the MMN and P300 effects at electrodes used in the statistical analyses (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4), for fluent and disfluent utterances respectively. As with Experiment 1, ERPs were quantified by measuring the mean voltages for deviant and standard targets over two time windows consistent with the MMN (100–200ms) and the P300 (250–400ms), for fluent and disfluent utterances separately. Greenhouse-Geisser corrections to degrees of freedom were applied and the corrections for F and p values are reported where appropriate. Analyses used three-way ANOVAs with factors of deviance (infrequent deviant, standard), location (electrodes F, C and P) and laterality (electrodes 3, z and 4).

For the fluent conditions in the MMN time window, results showed a significant main effect of deviance [$F(1, 11) = 23.000$, $\eta_p^2 = .676$, $p < .001$], indicating a widespread negativity associated with deviant stimuli (mean voltages of .808μV and

-1.455μV, for standard and deviant stimuli respectively) and a significant deviance by laterality interaction [$F(2, 22) = 6.342$, $\eta_p^2 = .366$, $p = .007$], indicating a larger effect over midline sites (mean voltages of .693μV, 1.096μV, .635μV for left hemisphere, midline and right hemisphere sites respectively for the standard stimuli and -1.521μV, -1.574μV, -1.272μV for the deviant stimuli) No other effects involving the factor of deviance reached significance [Fs < 2.037].

The results for the fluent conditions in the P300 time window showed a significant deviance by location interaction [$F(2, 22) = 12.055$, $\eta_p^2 = .523$, $p < .001$] indicating a positivity associated with deviant stimuli that was largest over central and posterior sites (mean voltages of .423μV, 1.638μV, 1.495μV at frontal, central and posterior sites respectively for the standard stimuli and 1.474μV, 4.960μV, 4.968μV for the deviant stimuli). Results also showed a significant deviance by laterality interaction [$F(2, 22) = 5.296$, $\eta_p^2 = .325$, $p = .013$], indicating that the positivity was largest over midline sites (mean voltages of .973μV, 1.789μV, .794μV at left hemisphere, midline and right hemisphere sites respectively for the standard stimuli and 3.185μV, 5.368μV, 2.850μV for the deviant stimuli). A deviance by location by laterality interaction was also observed [$F(4, 44) = 2.834$, $\eta_p^2 = .205$, $p = .036$], indicating that the midline bias for the deviance effect was largest at central and posterior sites. The main effect of deviance was marginally significant [$F(2, 22) = 4.495$, $\eta_p^2 = .290$, $p = .058$].

For the disfluent conditions in the MMN time window, there was a significant deviance by location interaction [$F(2, 22) = 5.373$, $\epsilon = .586$, $\eta_p^2 = .309$, $p = .032$] indicating negativity at frontal sites and positivity at central and posterior sites associated with deviance (mean voltages of .874μV, -.189μV, -1.456μV at frontal, central and posterior sites respectively for the standard stimuli and .146μV, .049μV, -.433μV for the deviant stimuli). No other effects involving the factor of deviance reached significance [Fs < 1].

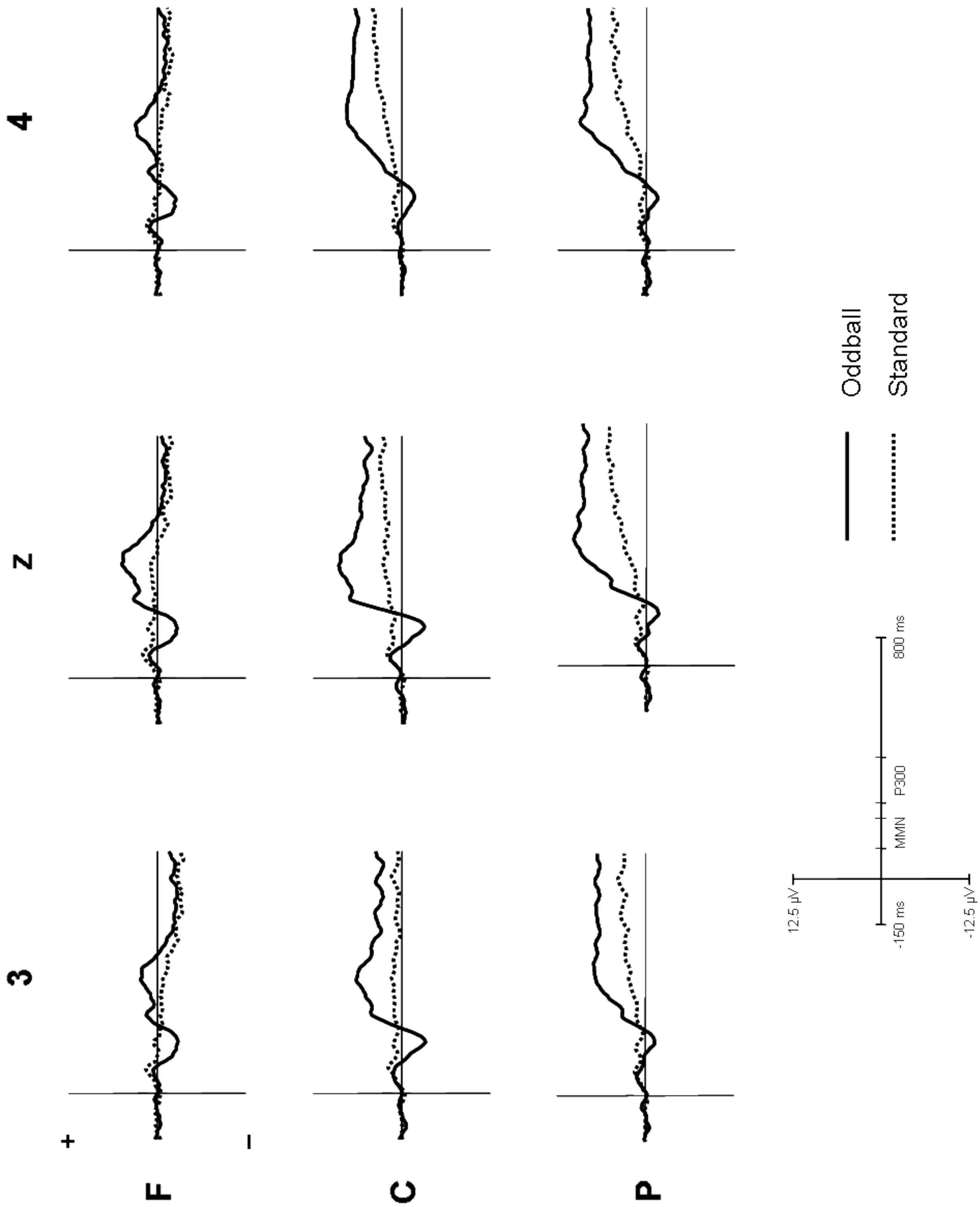

Figure 8.1: Experiment 7: Grand average event-related potentials for deviant (continuous lines) relative to standard (dotted lines) target words in fluent utterances (positive up). Waveforms show data from left, midline, and right electrodes at frontal, central, and parietal sites labeled according to the 10-20 system (from left to right and top to bottom: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4).

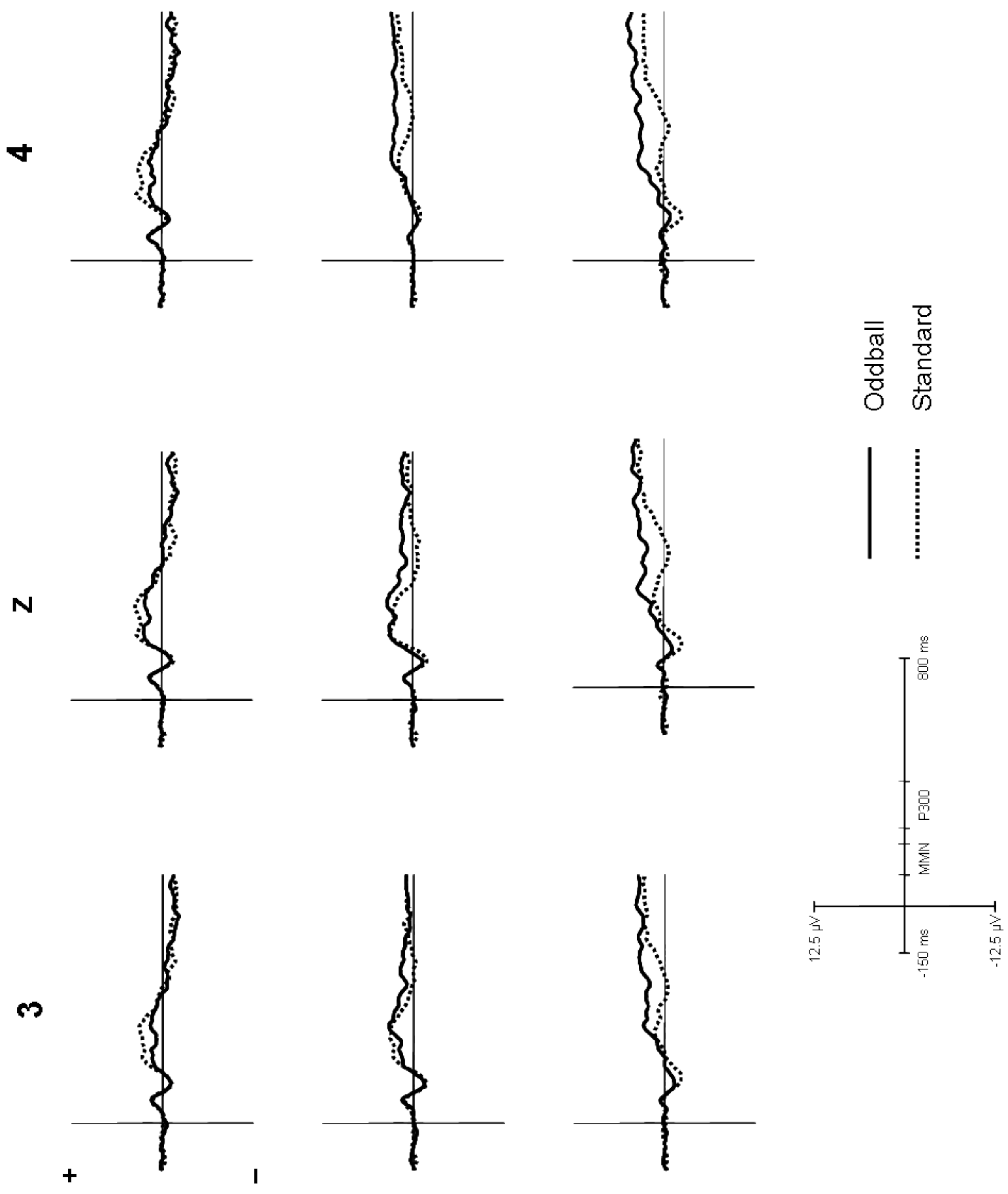

Figure 8.2: Experiment 7: Grand average event-related potentials for deviant (continuous lines) relative to standard (dotted lines) target words in disfluent utterances (positive up). Waveforms show data from left, midline, and right electrodes at frontal, central, and parietal sites labeled according to the 10-20 system (from left to right and top to bottom: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4).

For the disfluent conditions in the P300 time window, there was a significant deviance by location interaction [$F(2,22) = 9.783$, $\epsilon = .603$, $\eta_p^2 = .449$, $p = .005$] again indicating negativity at frontal sites and positivity at central and posterior sites associated with deviance (mean voltages of 2.482μV, 1.959μV, .656μV at frontal, central and posterior sites respectively for the standard stimuli and 1.084μV, 2.230μV, 2.285μV for the deviant stimuli). No other effects involving the factor of deviance reached significance [Fs < 1.828].

These analyses demonstrate typical MMN and P300 effects for acoustically deviant words in fluent contexts. This finding is in line with the results of Experiment 1. As with Experiment 1, data from the effects for acoustically deviant words in the disfluent condition differ from those in the fluent condition.

In Experiment 1, the effects in the disfluent condition were interpreted as a reduced MMN and P300 effects. The data in Figure 8.3 suggest that this may not be the case. The scalp distribution of the effects here show an early negativity followed by a later positivity, just as was observed in Experiment 1 (see Figure 5.1). These appear to be of larger magnitude in the current experiment. The small magnitude early negativity and late positivity for the disfluent condition in Experiment 1 were interpreted as reduced MMN and P300 effects. However, the larger magnitude early negativity and late positivity for the disfluent condition in the current experiment do not appear to have the correct scalp distribution for this interpretation.

Here the early negativity appears to be more frontally focused than the MMN and spread out in a bilateral pattern that does not extend to the dorsal region of the scalp. The MMN by contrast, typically has a dorsal fronto-central maximum.

The later positivity appears to be more posterior than the P300, focusing at parietal and occipital regions on the scalp. The following section outlines analyses that aim to see if the disfluent effects in the current experiment resemble those in Experiment 1.

Comparison of the oddball effects in the disfluent conditions of Experiment 1 and Experiment 7

To determine whether there were distributional differences between the oddball effects (acoustic-deviance ERPs minus non-deviance ERPs) in the disfluent conditions of Experiment 1 and Experiment 7 the data were rescaled using the technique provided by McCarthy and Wood (1985). Following the rescaling, the data were submitted to an ANOVA involving the factors of location (electrodes F, C and P) and laterality (electrodes 3, z and 4) with a between participants factor of experiment (Experiment 1 and 7). If the disfluent context oddball effects are similar, all statistical effects involving experiment should be non-significant. In the early time-window (100–200ms), no effects involving the factor of experiment reached significance [Fs < 2.456]. In the later time-window (350–400ms), again no effects involving the factor of experiment reached significance [Fs < 1.653]. This result indicates that there is no evidence to suggest that the distributions of the oddball effects in the disfluent condition for Experiments 1 and 7 are different from each other in the time-windows analysed.

To determine whether there were magnitude differences between these effects, similar analyses were run on the raw voltage data. Here however, the analyses were confined to the sites where the effects were maximal (frontal electrodes for the early effects and parietal electrodes for the late effects). The analysis of the early frontal effect revealed no difference between the mean voltages for the effects in Experiment 1 and Experiment 7 [Fs < 2.146]. The analysis of the late posterior effect revealed a significant interaction between experiment and laterality [$F(2, 44) = 5.105$, $\eta_p^2 = .188$, $p = .010$] indicating a greater magnitude for the positive effect centrally

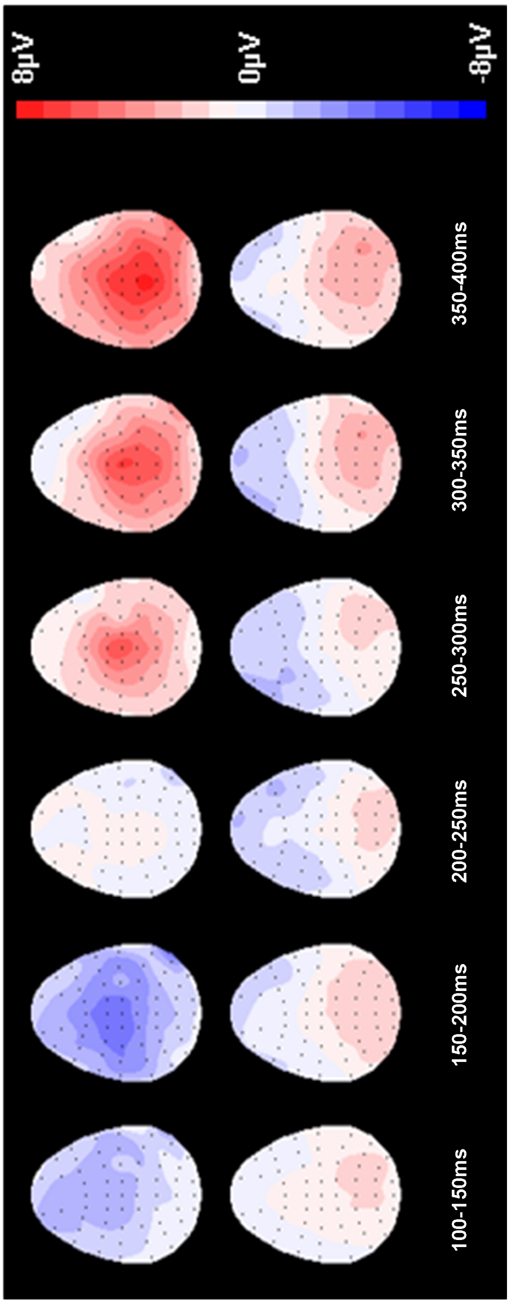

Figure 8.3: Experiment 7: Topographic maps (anterior up; electrodes shown as black dots) illustrating the mean distributions of the deviance effects (deviant minus standard event-related potentials) over 100-400 ms (in 50-ms time windows) for fluent (top) and disfluent (bottom) utterances.

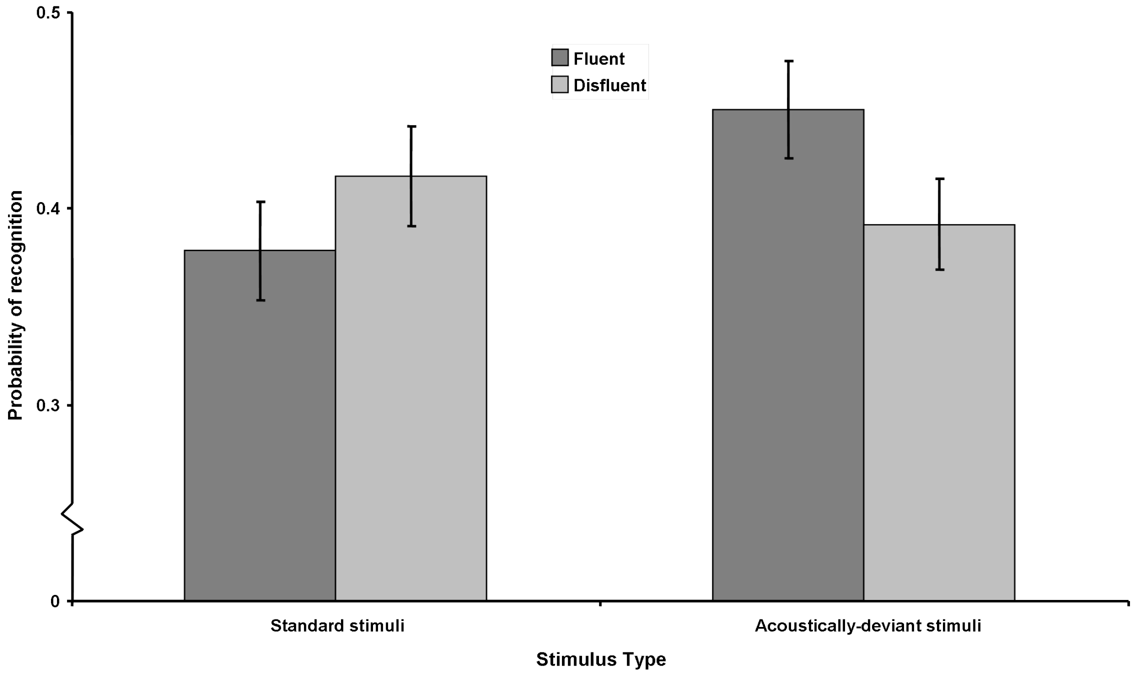

Figure 8.4: Experiment 7: Recognition probabilities for utterance-final words that were originally presented as standard or acoustically deviant stimuli and fluent or disfluent contexts (bars represent ±1 std. error of the mean).

and in the right hemisphere in Experiment 7. Because this interaction indicates a distributional difference, this last analysis was also performed on the rescaled data. When global scalp magnitude differences were partialed out of this local analysis, it was confirmed that the distribution of these two effects was not different even when the analysis is confined to posterior regions [Fs < 2.244].

8.2.4 Memory Results

A second analysis focused on performance in the recognition task. As in Corley et al. (2007) and Experiment 1 here, the probability of correctly identifying words heard in the comprehension block of the experiment was quantified with stimulus identity treated as a random factor. Overall, 40% of the previously-heard words were correctly recognized (false alarm rate 16%). Figure 8.4 shows the recognition probability of utterance-final words by fluency and deviance.

A 2-way ANOVA with factors of fluency and deviance revealed a significant interaction between the two factors [$F(1, 147) = 4.059$, $\eta_p^2 = .027$, $p = .027$]. For standard stimuli, a pairwise comparison of recognition probabilities for words which had been heard in fluent or disfluent contexts showed no significant difference [ts < 1.817], suggesting that acoustically normal words encountered after disfluency were no more likely to be recognized than those encountered in fluent contexts.

As with Experiment 1, twelve target words were inadvertently repeated in the experiment, resulting in 148 distinct targets. Removing data from the repeated targets did not affect the outcomes of the ANOVA the or pairwise comparison.

8.2.5 Discussion

Experiments 5 and 6 of this thesis provide evidence to suggest that the effects of hesitation on attention are not due solely to the timing gap that the hesitation provides and instead that the filler (here *er*) has an important role to play in the heightening of attention. Furthermore, increasing the gap between this filler can nullify attentional effects that would otherwise occur. Experiment 7 explored whether the immediate attentional engagement after hesitations (seen in Experiment 1) is sensitive to this timing gap (silent pause) between a filler and the resumption of normal fluent speech.

In Experiment 7, ERP measurements were taken from standard speech and from infrequently occurring acoustically-deviant words (oddballs) both in fluent contexts and in contexts immediately following hesitations. In contrast to Experiment 1, these hesitations were digitally altered so that there were extended periods of silence between the filler (*er*) and the post disfluent material. For fluent utterances, large deflections in the ERPs were observed when participants encountered deviant words. The polarities, distributions, timings and antecedent conditions indicate that these ERP deflections correspond to the typical neural signatures of attention capture and orientation, the MMN and P300. When these deviant words were encountered following a hesitation, these effects were not seen. However, there was evidence of a large posterior positivity in relation to deviant words in this extended hesitation condition.

In Experiment 5, an increased gap similar to the one in the current experiment nullified the effects of the hesitation. This result was tentatively interpreted as indicating that the attentional engagement that the filler induced rapidly decayed, so that its effect was reduced by increasing the timing gap between the filler and the post-disfluent material. If the reduction of the MMN and P300 effects after hesitations are interpreted as indicative of heightened attention (the interpretation given in Experiment 1) then the similar effects here and in Experiment 1 suggest that the rapid-decay-of-attentional-heightening explanation provided previously is not supported by these ERP results.

However, the ERP results from Experiment 7 stand in contrast to the memory results. Considering only the standard stimuli, in Experiment 1 words that were preceded by (short) hesitations were more likely to be recognised subsequently than word that were encountered in fluent contexts. Here, with the hesitations that included long periods of post-filler silence, this was not the case. This suggests that the extended hesitations are not affecting the immediate heightening of attention, but are affecting subsequent processes involved in speech comprehension, and moreover, that these effects on processing are changing the way that the post-disfluent material is being represented. In order to explore this last point more systematically, the direct comparison of the results from Experiments 1 and 7 is discussed below.

Comparison of the memory effects in Experiments 1 and 7

The lack of a fluency effect for standard (non-manipulated) words here is in contrasts with the memory results from Experiment 1. It is also in contrast with MacGregor (2008; Experiment 3), where standard (predictable) words were better remembered if they were preceded by silent-pause hesitations. The effects are however, consistent with what was found in Experiment 5 here — where extending the length of silence after the filler was seen to nullify attentional effects seen in reaction to non-manipulated hesitations.

One difference between the experiments in this thesis that is worth exploring is that only some of these experiments (or more specifically the tasks they involved) required overt responses. In Chapter 2, one point that was discussed was how the experimental effects of disfluency might differ depending on the task a participant is performing and it was suggested that the effects related to hesitations might be reduced when a task requires more conscious access to information (see section 2.4.1). In Experiments 1, 5 and 7, the varying tasks differ with regard to the level of conscious access to the material that the tasks require. In Experiment 5, conscious access was required for the task: participants had to recall and report changed words. In that experiment, no effect of the extended disfluency was observed. In the ERP phases of both Experiments 1 and 7, no conscious access was required: the participant were asked to listen to the material as if they were listening to a normal conversation. For these ERP phases, there was an effect of both extended and non-extended hesitations (indexed by the absence of the MMN and P300). In the memory phase of Experiments 1 and 7, participants were required to recognise, though not recall, words from the preceding material. Thus, it was possible, but not necessarily the case, that participants were consciously accessing the material. One possible explanation of this is that in line with other studies (e.g., Duez, 1985; Lickley, 1995; Lindsay & O'Connell, 1995), the effects of hesitations seem to reduce with added emphasis on consciously controlled processing at the point where the task requires a response. If the already small effect of hesitations on memory (seen in Experiment 1) is further reduced by extending the hesitations (as in Experiment 7) then this might be enough to mask the effect altogether in the case of a measure where consciously controlled processing is required. This interpretation is consistent with the memory results from MacGregor (2008). She shows that a more robust effect on memory for material encountered after *er* was reduced to a smaller and only marginally significant effect when the *er* was replaced by a silent pause.

Comparison of the ERP effects in Experiments 1 and 7

For the disfluent conditions in Experiment 1, the small early negative components and later positive components were interpreted as reduced MMN and P300. Though Experiment 7 shows similar effects, they (especially the positivity) are of greater magnitude. In Experiment 7, the effects seemed to obviously deviate from the MMN and P300, and so were not interpreted as such. However, reinterpreting the results from Experiment 1 in light of those in the later study suggest that the earlier interpretation of effects as reduced MMN and P300 might have been erroneous. With the later positivity in particular, a review of the results of Experiment 1 suggests that the positivity is more posteriorly focused that a typical P300.

A topographic analysis of the late positive effects suggested that the effects of deviant sounds in disfluent contexts were the same on Experiments 1 and 7. To test this, a topographic analysis was performed on rescaled data (data in which the magnitude differences between the effects were partitioned out). This showed that there was no reason to suspect that the effects in Experiments 1 and 7 were the result of different neural generators (and therefore functionally different effects). Although this rescaling technique is not without its critics (see, Urbach & Kutas, 2002) the criticisms tend to focus on the use of the technique to provide evidence of a difference, rather than of a similarity. Therefore it seems reasonable to put forward an interpretation suggesting that the deviance effects after hesitations in Experiments 1 and 7 are the same.

Further to this topographic analysis, an effect size comparison was conducted between the late-posterior deviance effects in Experiments 1 and 7. The magnitude of this effect was found to be significantly larger in Experiment 7 than in Experiment 1. Though this analysis is post-hoc in nature, and should be considered with this in mind, it is worth noting that this increase in posterior effect negatively correlates with the effect of disfluency on memory. Furthermore, it is similar to a late effect (LPC) seen in MacGregor (2008) in relation to unpredictable words encountered after silent pause hesitations. MacGregor (2008) interprets this LPC as an index of controlled memory processes that are engaged in subsequent to an unpredictable (here, acoustically deviant) stimulus. A similar account could potentially explain the ERP results here.

8.3 Conclusion

The ERP results from Experiment 7 indicate that an increased silent pause between a filler and the resumption of fluent speech does not affect the way attention is immediately engaged by the hesitation. However, this extended pause does seem to affect the covert attention to the post-disfluent constituent (Experiment 5) and the subsequent memory for this constituent (Experiment 7, memory results). While this pattern of results could potentially be explained by task differences in the various measurements involved, it is also the case that here, as with previous work (MacGregor, 2008), hesitations seem to change the immediate processing of speech even in the absence of a filler (as indicated by the late posterior positivity), and the way this change is manifest differs depending on whether a filler is present (MacGregor, 2008) and how soon fluent speech is resumed (Experiment 7).

Chapter 9

Discussion

9.1 Chapter Overview

The experimental work in this thesis has investigated common hesitations in speech. The effect that these hesitations have on listeners has been measured using both neural and behavioural indices, with a view to exploring how these hesitations might modulate listener attention. This chapter provides a summary of these findings and an initial interpretation of the results, fitting these into a context of both language processing and attention. Finally, the potential applications of the theory and methods developed in this thesis for future work is considered.

9.2 Summary of the findings

The experimental work outlined in this thesis tested the hypothesis that hesitations heighten listeners' attention for upcoming speech. The work was carried out and analysed with a view to fitting this hypothesis into a clear and defined account of attention. The research primarily concentrated on two measures of attention; the immediate engagement of attention and increased covert attention.

Experiment 1 used ERP indices of attentional engagement to test whether oddball stimuli that would usually engage attention were doing so after hesitations. These indices were absent after hesitations, supporting the theory that hesitations had already engaged listeners' attention. This demonstration of immediate attentional engagement was also accompanied by an increased memory for constituents encountered after hesitations, suggesting that the attentional heightening results in, or is concomitant to other changes, in processing that induce better memory for constituents that immediately follow hesitations.

Experiments 2 and 3 used a change-detection paradigm to investigate whether this immediate attentional engagement also resulted in increased covert attention to the constituent (that is, more attention was paid to the constituent even after it had been mentioned). Experiment 2 showed that subtle changes to constituents in that occurred between two presentations of the material were noticed more often if they were preceded by hesitations. Experiment 3 aimed to explore whether this change was accompanied by eye-movement behaviour when reading over the material on the second presentation. While differing eye-movement behaviour was observed related to whether a word changed or not and how subtle this change was, there was little evidence that eye movement data during the reading of the second presentation was sensitive to the presence of a hesitation in the first presentation. The overt answers provided when there was a change were sensitive to the presence of hesitations. When participants were asked to respond to the question 'did a word change', they were more likely to identify subtle changes when the words in the first (spoken) presentation were preceded by a hesitation. They were also more likely to look back at this word, prior to answering the question.

Experiments 4–6 used the change-detection paradigm to explore the mechanisms responsible for the heightening of covert attention by hesitations. The experiments explored issues such as whether the timing gap caused by the hesitation heighten attention because the listener uses the gap to free up processing resources. The combined results of these experiments indicate that the increased timing gap is not the only important feature for the heightening of covert attention, and that the phonetic features of the hesitation might be inducing this attentional effect. The results also indicated the mechanisms behind the covert attentional heightening are sensitive to the length of gap between the filler in the hesitation (here *er*) and the resumption of fluent material. Covert attention to post-disfluent constituents only occurred if those constituents were encountered immediately subsequent to the filler.

Experiment 7 tested whether the decay of the covert attentional effect was mirrored in the data regarding immediate attentional engagement. The ERP indices used in Experiment 1 were again explored using an oddball paradigm, only here, the hesitations included an extended silent period before the resumption of fluent material. The ERP indices associated with attention (MMN and P300) were sensitive to these hesitation in the same way as they were to non-extended hesitations in Experiment 1. This indicates that the immediate attentional engagement did not decay rapidly. There was however, a late posterior positivity associated with these extended hesitations, indicating that some aspect of the immediate processing was affected by the increase in gap between the filler and the following constituent.

9.3 Interpretation of the findings

The following sections explore the experiments in this thesis together, and discuss three major aspects of the reported research. First, the locus of the attentional effects is discussed. In the following section the issue of the signal of disfluency is revisited. The final section considers what can be concluded from the current work, and introduces some suggestions for future research.

9.3.1 Which attentional mechanisms are being affected by hesitations?

One of the major goals of the research in this thesis was to investigate disfluency in the context of a clear and explicit account of attention. Below, the processes likely to be engaged in covert (Experiments 2–6) and overt (1 and 7) attention are discussed.

The findings concerning covert attention in this thesis are consistent with previous findings in the literature, showing that hesitations affect the representation of post-disfluent material both the material is encountered (e.g., in change-detection paradigms; A. J. S. Sanford & Molle, 2006) and following extended periods of time after the material is presented (e.g., the memory results for Corley et al., 2007; MacGregor, 2008). In the change-detection Experiments here, covert attention was also shown to be sensitive to the presence of hesitations, so long as the gap between the filler and the post-disfluent material was not too long. While not all Psycholinguists would call such effects attentional (preferring terms such as *heightened activation*), the evidence presented in Chapter 3 along with that from Experiments 2–6 show that the processes underlying these effects fully fit the definition of attention as comprising "*mechanisms... that allow us to concentrate on events of interest in the environment*" (Spence & Driver, 1997, p. 389).

The findings concerning overt attention in this thesis are unique in that they allow us to say unambiguously that hesitations heighten listeners' immediate attention to upcoming speech. The ERP results of Experiments 1 and 7 show that hesitations do affect neural mechanisms that select important information from the environment. The MMN and P300 are ERPs that index these processes. Experimental results indicating that the mechanisms underlying these ERPs are affected by hesitations, is powerful evidence that the different degrees to which language appears to be processed following disfluency, is a direct consequence of the engagement of attentional processes.

Can they inform us further with regard to which subsystems of attention are being affected by hesitations? There are two ways to look at this issue: which attentional indices are being affected by hesitations and what are the effect of the hesitations on the processing and representation of the post-disfluent material?

In Experiments 1 and 7, hesitations were shown to affect the P300. This is an ERP component associated with attention, therefore the disruptive influence of hesitations might indicate which subsystems of attention are being affected. This line of reasoning is complicated however, by ambiguity regarding which attentional mechanisms are responsible for the P300. As specified in one of the most current models of the P300 (Polich, 2004, 2007), the three attentional subsystems mentioned in Chapter 3 (those systems underlying orienting, target detection and alertness) could all be playing a role.

One general conclusion reached in P300 research is that the stimulus driven (here stimulus novelty driven) frontal orienting system seems to play a part in the elicitation of the P3a and the posterior system seems to play a part in the task-driven target identification that is indexed by the P3b. These separate effects are only easily distinguishable in a three-stimulus oddball paradigm (see section 4.3.2). Since the P300s in Experiments 1 and 7 initiated frontally and then shifted to a parietal maximum later, the effect is likely to be a summation of the P3a and P3b. Whether both of these attentional systems are being affected by hesitations, or whether the effect is primarily driven by one system remains difficult to say given the current data. To complicate matters further, both of these ERP subcomponents are likely to be sensitive to the overall level of attentional arousal (alertness), in a way where heightened arousal (such as that potentially induced by hesitations) will reduce the size of the effects (see, Kok, 2001; Polich, 2007). Thus, the interpretation based on the ERP evidence looks somewhat ambiguous. However, looking at some of the other effects of the hesitations leads to some clearer interpretations.

The heightening of alertness speeds up reactions, so the fact that hesitations speed-up reaction times in Fox Tree (2001) could suggest that hesitations are influencing the alertness subsystem. However, it is unlikely that hesitations are doing so, as heightened alertness also leads to poor selection of responses. Alertness facilitates response speed at a cost to the quality of incoming information (Posner, 1978). Since no increase in error rate was seen in Fox Tree's (2001) study and post hesitation words were seen to be remembered better in Experiment 1 here, it is unlikely that hesitations are affecting alertness.

Hesitations may be affecting the target selection subsystem in a way that provides it with information about upcoming targets. Here this would be the case for the post-disfluent words. An explanation such as this does account for the increased memory performance (Experiment 1) and the increase in change detection performance (Experiments 2, 3, & 6). One reason to suspect that the attentional heightening induced by hesitations does not operate through the target identification subsystem is because again, this attentional heightening comes at a massive cognitive cost, this time to the processing of information subsequent to the target. The cognitive cost involved with target identification means that the subsystem is unlikely to play a major part during normal speech comprehension. Target identification and processing can interfere with the processing of non-target input both in space (e.g., Duncan, 1980) and in time (e.g., Raymond et al., 1992). Given that hesitations enhance the processing for important constituents, but do not seem to inhibit or disrupt the processing of information around this (at least not to the extent seen with other target detection) it is unlikely that any target detection system is working during language processing as it works elsewhere.

Taking all the factors outlined here into account, the orienting subsystem probably provides the most likely candidate for the attention heightening effects of hesitations. As outlined in Chapter 3, one problem with this suggestion is that these mechanisms have been identified and examined in the context of the visual-attention system. As a result of this, the descriptions of these mechanisms are largely concerned with orienting attention in space, rather than in time. Thus, it is difficult to say how accurately hesitations resemble other stimuli that result in orienting. Despite this, the suggestion that hesitations orient listener attention to the speech stream and that this attentional focus is located at a particular point in time does provide an elegant explanation of many of the experimental results here. The system seems to be largely exogenously controlled: that is, sensitive to stimulus properties. As has been explained, hesitations provide auditory novelty that could be a candidate trigger for an orienting response. The focal aspect of orienting induced attention would also explain the attentional decay seen in Experiment 5 and the memory data in Experiment 7. If the focusing of attention manifests in a temporal attention maximum, then the longer period between the focusing event (hesitation) and the event of interest (target/changed word) would result in a reduced attentional effect on this event. Lastly, a very tentative mention should be made of proximity of the late posterior positivity in relation to hesitations (seen in Experiments 1 and 7) and the neural mechanism underlying the orientation process (parietal cortices).

9.3.2 Which aspects of a hesitations trigger the attentional effects?

The findings concerning the characteristics of hesitations that trigger the attentional effects highlight the importance of not only the added time that hesitations allows a listener, but also the orienting effects of the filler. The relationship between these two signals within the hesitation seems to be complex, though the work here represents important intitial progress in understanding it.

Although covert attention appeared to be attenuated when a silent pause followed the filler *er*, resulting in an inability to detect changes and poorer recognition of the word which followed the disfluency, the immediate post-hesitation ERP effects observed in Experiment 1 were not seen in Experiment 7 when a silence followed the *er*. Whereas these findings suggest that the trigger for hesitation effects is not exclusively time-based, they implicate a subtle interaction between the phonetic form of the hesitation, and the time taken to hesitate. Despite this, even in the presence of ERP evidence suggesting that hesitations with extended post-filler pauses still engage listeners immediate attention, representational changes akin to the changes in covert attention were apparent. In Experiment 7, listeners were just as likely to recognise post-disfluent material as they were fluent material, showing that the memory effect (seen in Experiment 1) was affected by this increased gap between the filler and the resumption of fluent speech. As discussed in Chapter 8, the lack of ERP effect for the extended hesitations could be due to task demands. Interestingly, MacGregor (2008) showed the reverse pattern when looking at the effect of hesitations on predictive processes. The silent pauses in her study seemed to elicit no modulation of the N400 (unlike the filled pauses) and seemed to induce the same memory effects (better memory for disfluent material). Here, the extended pause condition produced the same immediate processing difference as the filled pauses (attenuation of the P300) but differing memory effects (though in a context absent of unpredictable sentence ending stimuli). This serves to highlight the fact that the signals which trigger the effects of disfluency may differ depending on which effects are the focus of experimental manipulation. The research in this thesis has provided evidence which speaks to the importance of the phonetic form of the hesitation with regard to the heightening of listener attention.

9.4 Future Research and Conclusions

The present thesis provides a sound framework for future work to build on, both with the methodology and theory it provides. Building on the methodology for investigating attention and disfluency, the 'spoken oddball' paradigm developed here could be used to address a potential problem with much disfluency research. This is the fact that most of the work has been done on scripted disfluency (e.g., Arnold et al., 2004; Bailey & Ferreira, 2003; Brennan & Schober, 2001; Corley et al., 2007). Work on non-scripted disfluency has often required participants to perform a potentially disruptive task not normal to language processing (e.g., monitoring for target words; Fox Tree, 2001). One valid concern regarding these approaches is that the findings might lack ecological validity. In particular, the work in this thesis has shown, the phonetic form of a hesitation is important in determining whether it will induce heightened attention, emphasising the potential effects of non-natural recordings. An obvious area of future research would be to attempt to extend the ERP experiments in this thesis in the context of spontaneously produced disfluency.

The paradigm developed for Experiments 1 and 7 provides a unique and ideally suited method for this. Many other paradigms, even where there is no secondary task, require the manipulation of the probability of the target word (e.g., Arnold et al., 2004; Corley et al., 2007). Since speakers tend not to spontaneously utter improbable sequences, fully natural materials may be hard to come by. However, the 'oddball' manipulation used here can be directly applied to spontaneously produced disfluent language, since utterances do not need to differ in plausibility. This affords a unique opportunity to investigate the direct effects of natural disfluency without involving a secondary task, such as the word monitoring task in Fox Tree (2001). By extending the classification of the hesitations to include exact measures of the silent pauses before and after the filler as well as phonetic characteristics such as the pitch pattern of the hesitation (e.g., the fundamental frequency drop during a filler or rise upon resumption of fluent speech), the paradigm might even be used in a regression approach to explore the interaction of the phonetic form and timing gap more systematically. Though clearly ambitious, the groundwork for such an investigation is provided here by the 'spoken oddball' paradigm.

Building on both the methodology and the theory provided here, an approach grounded in the concepts and paradigms provided in this thesis could be extended to other psychologically interesting linguistic phenomena: for example, linguistic focus. Such an approach would provide a useful means to investigate this in the absence of linguistic anomalies or repeated presentations and could even (as is the case with the research proposed in the previous paragraph) be applied to naturally occurring linguistic stimuli.

The present thesis has clearly implicated attention at the centre of disfluency processing, basing itself on a well-defined attentional model, and demonstrating that attention in the face of a disfluency is modulated by both the temporal and phonetic characteristics of the speaker's hesitation. By specifically siting attention within the processing of language, this work will have repercussions for other investigation of language processing. It shows that attention can be considered a core aspect of language processing, where the production of various (here disfluency) linguistic phenomena might act on attentional systems and modulate the way this material is processed by listeners.

Appendix A

Stimuli for Experiments 1 and 7

The stimuli for Experiments 1 and 7 were edited from the following 160 utterances taken from Corley et al. (2007) and MacGregor (2008). The targets were the utterance-final words which were either standard or acoustically manipulated. Utterances were either fluent or contained a hesitation (*er*), before the target.

1. My sister had a skiing accident and she broke her leg.
2. She said she wouldn't cheat on him but she broke her promise.
3. The sitting room is really cold so I think you should light a fire.
4. We're in a smoking area so now you can light a cigarette.
5. I've got a deadline to meet for tomorrow so I can't afford to waste time.
6. I reuse old envelopes because I hate to waste paper.
7. On the front of the card was a lucky black cat.
8. The invite says that the dress code is black tie.
9. She's a bit dippy and thinks there's fairies at the bottom of the garden.
10. She messed up in her exams and went to the bottom of the class.
11. I love it when I wake up and the birds are singing.
12. It's distracting in lectures when people are talking.
13. Ben's so stubborn he can never admit when he's wrong.
14. Ben's only had a few beers, but I think that he's drunk.
15. I can't post the letter because it hasn't got a stamp.

16. It's only got an address and telephone number but it hasn't got a name.
17. I stood up and started to speak but then my mind went totally blank.
18. The list of names had no order to it and appeared to be totally random.
19. She hated the CD but then she's never liked my taste in music.
20. She hated the jumper but then she's never liked my taste in clothes.
21. I'm really thirsty. Lets go to the pub for a drink.
22. I'm really hungry. Lets go to McDonald's for a burger.
23. We had a great day playing at the seaside. It ended perfectly when we sat outside eating fish and chips.
24. We had a great time playing games at the party. It ended perfectly when we sat outside eating jelly and ice cream.
25. I think it's true that a lot of people stay at University because they appreciate the student lifestyle.
26. You know, if you take your University card you'll get a ten percent student discount.
27. We've not booked and so the first thing to do when we get there is to find somewhere to stay.
28. I've brought the football and so I think we should find somewhere to play.
29. I want to visit South America but to get the most out of it I should really learn to speak Spanish.
30. My little sister puts on a silly voice and so my mum is always telling her to speak properly.
31. It's been really busy in the office today, and the phones haven't stopped ringing.
32. I thought the computer had been fixed, but now it's stopped working.
33. Everyone's got bad habits, and mine is biting my nails.
34. That drink was too hot; I've just burnt my tongue.
35. She's been taking driving lessons all year, but she's still nervous about taking the test.
36. It's hard for the teachers to keep control when there are so many children in the class.
37. I want to travel the world, see new places, and meet new people.

38. They're going to be extending the zoo and are planning to get in some new animals.
39. I'm going to fill up the kettle for a hot water bottle. Then I'm going to bed.
40. I think I've just got time to grab a coffee. Then I'm going to work.
41. I think it's brewed enough by now so pass over a mug and I'll pour you some tea.
42. I filled up my bottle from the stream we passed earlier so pass over yours and I'll pour you some water.
43. I can hear someone knocking: please can you answer the door.
44. I can hear something ringing: please can you answer the phone.door.
45. I've got to remember to take that book I borrowed when I go to the library.
46. I've got to remember to take those money-off coupons I collected when I go to the supermarket.
47. Budweiser is a famous name associated with the King of beers.
48. Louis is a famous name associated with the King of France.
49. Jimmy's parents have been told that he can't hear and they're going to send him to a school for the deaf.
50. Jimmy's parents have been told that he can't see and they're going to send him to a school for the blind.
51. I'm really tired and so I'm going to have an early night.
52. I'm going to bed at 9 because tomorrow I've got an early start.
53. Daniel would have finished his essay but he ran out of time.
54. Daniel would have bought the whole set but he ran out of money.
55. I got to my front door and went to open it, but I couldn't find my keys.
56. I went to text her, but I couldn't find my phone.
57. If you hear the firealarm, move to the nearest exit.
58. After work on Fridays, we always meet in the nearest pub.
59. Because the plane was so delayed, we had to spend the whole night at the airport.
60. We're right by the coast so when the weather's good we tend to spend the whole day at the beach.
61. Marion started crying about her haircut as soon as she left the hairdresser.
62. Marion started crying about her dog's health as soon as she left the vet.

63. Carrots are great for helping to see in the dark.
64. Those goggles are the best for helping to see under the water.
65. The meeting was boring: Rob kept looking at his watch.
66. The new haircut was great; Rob kept looking in his mirror.
67. Terry's bored because there's nothing to watch on the TV.
68. Terry's bored: he's been staring at the picture on the wall.
69. If we're going to the cinema we should book in advance because it's a popular film.
70. I'll have to recall it from the library because it's a popular book.
71. I was caught by the police for speeding in my car.
72. The lake's big but we can cross it in my boat.
73. It's raining so I'm planning to stay indoors. I really don't want to go outside.
74. John suggested going today, but it would be better for me to go tomorrow.
75. My grandma died and I still need to get a black dress for the funeral.
76. I'm looking forward to my sister's birthday but I still need to get a dress for the party.
77. It was really hot yesterday and so I wore a hat so I wouldn't get burnt.
78. It was really rainy yesterday and I wore a hat so I wouldn't get wet.
79. I'm really looking forward to my holiday to Crete and I keep thinking about relaxing and sitting on the beach.
80. When there were no more chairs in the lecture room I thought about sitting on the floor.
81. I don't have time to paint the walls so we should call in the decorator.
82. I think next door have been burgled: we should call in the police..
83. Traditionally, Christians use Sundays for going to church.
84. He's a murderer: he'll be going to jail.
85. I'm so sweaty because I've just come from the gym. I'm just going to have a shower.
86. You know what they say: have a break, have a kit-kat.
87. Watching the bit in the film when Bambi's mother died always makes Caroline cry.
88. Watching the antics of the cat and mouse in Tom and Jerry cartoons always makes Caroline laugh.

89. My grandmother always drinks her tea from a fine porcelain cup.
90. My grandfather always mixes wallpaper paste in a big metal bucket.
91. If everyone has their coffee white, we'll have to buy more milk.
92. If everyone wants sandwiches, we'll have to buy more bread.
93. After the car crash, the three of us spent the night in hospital.
94. After winning the holiday competition, three of us spent the night in a five-star hotel.
95. We could tell it was a formal occasion because Andrew was wearing a smart suit and a tie.
96. We could tell it was going to be cold because Andrew was wearing a woolly hat and a scarf.
97. If you haven't got much time it takes only a few minutes to heat the soup in a microwave.
98. It will take at least an hour for the wine to chill in the fridge.
99. If you want to know the time, why don't you look at the clock.
100. If you want to know the date, why don't you look at the calendar.
101. Jack's very short-sighted, and he spent all his birthday money on a new pair of glasses.
102. Jack's a very keen runner and he spent all his birthday money on a new pair of trainers.
103. When my teacher tells us off he shouts so loudly and always wags his finger.
104. When my dog greets me, he barks so happily and always wags his tail.
105. One of the things I like about this restaurant is that there are always fresh flowers on every table.
106. One of the things I like about this churchyard is that there are always fresh flowers on every grave.
107. If we put more cheese in the trap we'll catch another mouse.
108. If we put more bait on the hook we'll catch another fish.
109. In the summer, he just sits under the trees in the local park.
110. In the summer, he just sits with a beer in the local pub.

111. Watching the cooking programme has made me feel hungry.
112. Watching her leave has made me feel sad.
113. She was easily the quickest for her age and broke loads of records.
114. She was the clumsiest waitress at the restaurant and broke loads of plates.
115. At the end of breaktime we should ring the bell.
116. That smoke is not from a bonfire: we should ring the firebrigade.
117. I pulled a muscle playing hockey and so I'm going home for a long hot bath.
118. It's been a beautifully warm spring and so we're hoping for a long hot summer.
119. I've finished cooking for Christmas. I've just got to put the icing on top of the cake.
120. I've finished decorating for Christmas. I've just got to put the fairy on top of the tree.
121. Management decided to close the pit and the miners have voted to go on strike.
122. Term finishes next week and so I'm planning to go on holiday.
123. There's a housing shortage in rural areas because so many people have second homes.
124. A BA or a BSc doesn't count for much these days because so many people have second degrees.
125. They've been living together for ages so I wasn't surprised to hear they're going to get married.
126. They've been separated for ages so I wasn't surprised to hear they're going to get divorced.
127. I want my birthday present to be a surprise. I wish my sisters would stop dropping hints.
128. The street where I live is really scruffy: I wish people would stop dropping litter.
129. I'm meeting the estate agent after work. He's going to show me round the house.
130. I'm meeting the headteacher after work: he's going to show me round the school.
131. He's going bald and he's going to go on a special diet because he doesn't want to lose any more hair.
132. She's too thin and she's going to go a special diet because she doesn't want to lose any more weight.

133. I wish I was better at foreign languages. It was difficult in Paris because I couldn't speak French.
134. I wish I was better at foreign languages. It was difficult in Berlin because I couldn't speak German.
135. The holiday's all arranged. I've just got to go the the travel agents to collect the tickets.
136. We've signed the tenancy agreement on the flat. The landlord will come on Fridays to collect the rent.
137. I don't like pastry and I loathe sausage rolls.
138. I don't like poodles and I loathe sausage dogs.
139. You can retire when you want, but you have to be 60 to get a pension.
140. Today there are so many graduates looking for work that it's hard to get a job.
141. We need to pay; please can you ask the waiter for the bill.
142. We need to order; please can you ask the waiter for the menu.
143. Susie's mum makes the best chocolate cake.
144. Susie's mum makes the best apple pie.
145. If you want to be safe when you're walking home late then stay in a group and keep together.
146. If you want to keep in shape and feel good then go to the gym and keep fit.
147. I don't think he remembers who I am because he keeps calling me by the wrong name.
148. I tried to call him but I kept getting the wrong number.
149. It was too noisy to have a conversation on the dance floor; we had to shout to be heard.
150. It was too crowded to catch his attention, we had to jump and wave to be seen.
151. She wasn't very involved in the business: she was a sleeping partner.
152. I was cold when we went camping: I forgot to take my sleeping bag.
153. Urban land is expensive so the council have started building sky scrapers.
154. We get pretty good views from the attic because of the sky lights.
155. It was a formal dinner and Ellie was surprised to find herself at the top table.

156. The earth was barren because the rain had washed away the top soil.
157. I need to buy a new coat but I want to get a discount so I'm going to wait until the January sales.
158. I need to buy a new coat but I'm going to wait and see the fashions for the new season.
159. I can't pay for the holiday because I don't have any money.
160. He's unemployed at the moment. In fact I can't remember the last time he did any work.

Appendix B

Stimuli for Experiments 2–6

The stimuli for Experiments 2–6 were edited from the following 32 passages adapted from Ward and Sturt (2007). The second sentence of each frame contained target noun in non-focused position in a prepositional phrase. The first auditory presentation contained one of the three targets (e.g., virus, infection, or tissue in the first example). The second visual presentation contained only the first of these words (e.g., virus). The change condition was induced by playing different versions of the first presentation to correspond with the no-change (virus), close-change (infection), or distant-change (tissue) condition. In disfluent conditions, this target word was preceded by a hesitation including prolongations to preceding words (e.g., the prolonged definite article *thee*) and filled-pause *er*.

1. The doctor checked to see how much longer he had to work. He saw that the patient with the virus/ infection/ tissue was at the front of the queue. A kind but strict-looking nurse brought the boy in.
2. We all wondered where the new employee was going. It was obvious the woman carrying the rucksack/ backpack/ briefcase was a bit lost. In such a big complex it's so easy to lose your way.
3. Tony heard all about the celebrities at the Oscar ceremony. Apparently the film about the aliens/ martians/ dinosaurs had been universally praised. Everybody thought it had been a wonderful ceremony.

4. Simon really needed to decide what to do with his life. He said that the job advertised in the magazine/ newspaper/ church had looked interesting. He really wanted something that would challenge him.
5. The police still didn't know how to proceed with investigations. They thought the boy caught with the lighter/ matches/ gun was a likely suspect. The witnesses had not been very helpful at all.
6. We found out what the neighbours had been up to. The tree that had blocked the street/ road/ view had been cut down. It should make a real difference to their garden.
7. The journalist wasn't sure what he should be doing. He knew that the story about the burglary/ robbery/ budget was long overdue. But his editor would be needing the front page picture.
8. The lawyer wondered how he could construct a solid case. Obviously the document for the building/ property/ judges would be useful. He couldn't afford to let the partners down.
9. The taxi driver didn't know where he was supposed to be. Somehow the apartments with the truck/ lorry/ fence in front seemed familiar. If he didn't find his way soon he would lose the customer.
10. The secretary checked to see what had to be done next. The letter to the client/ customer/ board was on the boss's desk. All the office chores had to be finished by five o'clock.
11. The theatre critic was certain about his latest recommendation. He thought the play about the two policemen/ detectives/ pilots would run for months. He knew the theatre business and was usually right.
12. The air traffic controller checked that everything was running smoothly. The plane carrying the important packages/ parcels/ delegates was approaching the runway. It could be quite a stressful job.
13. The advertising executive explained how to reach the target audience. He said the poster featuring the kitten/ puppy/ model was a safe bet. He had a lot of experience in the advertising industry.

14. The ramblers thought they were getting near to the village. It seemed that the path beside the canal/ stream/ forest was going in the right direction. But without a detailed map there was no way to be certain.
15. It became clear how attitudes in the city had started to change. The reports of the recent killings/ murders/ crimes had made the community more vigilant. But a heavy police presence would still be necessary.
16. The fireman asked us how the incident had started. We pointed out the woman wearing the sweater/ jumper/ scarf who had dialled 999. They wanted to get the full story.
17. The crime squad guessed the criminal was somewhere in the local area. Soon the area behind the pond/ lake/ warehouse was completely surrounded. But he was not found and the search continued for days.
18. I couldn't decide whether I liked the new cinema layout. I hoped the seat by the exit/ door/ aisle would give me a good view. It turned out to be a wonderful evening's entertainment.
19. He asked me if I had ever had a supernatural experience. I told him about the ghost in the graveyard/ cemetery/ mansion that had scared me. I don't think he believed me.
20. The vet wondered what all the noise was about. The dog with the injured legs/ paws/ mouth would not stop barking. The owner was getting quite embarrassed.
21. The student asked the professor for advice about the course. He said that the book on ancient rituals/ ceremonies/ battles would be essential. The student needed all the advice she could get.
22. The museum owner wanted to know about the preparations for the exhibit. It turned out the box containing the old drawing/ painting/ vase was still in the van. There would be terrible trouble if anything went missing.
23. The student would have to choose very carefully this year. The course containing chemistry/ biology/ computers would probably have to be avoided. It was important to have a timetable with no clashes.

24. The zookeeper knew he had some cleaning to do. He had noticed that the cage for the tigers/ lions/ eagles was beginning to smell. It was a big job and would probably take all day.
25. We found out what the commotion was about. The window of the house/ flat/ car had been broken. The act of vandalism was to be discussed at the next community meeting.
26. Everyone at the book launch wondered what had caused the delay. It turned out the bag belonging to the author/ writer/ reporter had been checked thoroughly. Security at events like this was always tight.
27. The girl wondered how easy her homework would be. It was in the bag lying in front of the couch/ sofa/ table in the living room. She hated doing homework for school.
28. The sailor was enjoying being on dry land again. The equipment for his boat/ ship/ mast would take a while to fix. He had a number of friends that he was planning on visiting while he could.
29. The girl was searching all over her room for the tickets. She thought she had left the envelope inside her closet/ wardrobe/ handbag along with the present and card. If she didn't find them soon, she would be very late.
30. The editor had sighed as she pulled into the driveway. The villa which sat beside the coast/ shore/ mountain was always a welcome sight. She had been very busy for the past month and was looking forward to a relaxing weekend.
31. The brewer was always experimenting with new concoctions. The barrel with the wheat/ grain/ berries had started to ferment. He was planning on selling the drink at the local market.
32. The firemen were busy searching through the remains. The old cottage in the woods/ forest/ hills had been abandoned for years. Almost everything had been destroyed in the fire.
33. The farmer had organised his finances more carefully this year. He was already planning for the storms/ rain/ droughts that often happened late in the year. A good harvest would mean he would be debt free by the end of the season.

34. The museum had previously been considered to be very secure. The footprints on the lawn/ grass/ roof showed where the thief had entered. The sculptures had been insured but would be impossible to replace.
35. The athlete was struggling to contain all his emotions. The crowed that had gathered at the stadium/ arena/ airport was like nothing he had experienced. Despite feeling very nervous, he was expecting to enjoy the competition.
36. The two generals met in private for a frank discussion. The conditions of their agreement/ arrangement/ surrender would still have to be negotiated. It seemed obvious to both of them that all sides were hoping for a quick end to the war.

Appendix C

Collard, Corley, MacGregor, and Donaldson (2008)

Journal of Experimental Psychology:
Learning, Memory, and Cognition
2008, Vol. 34, No. 3, 696–702

0278-7393/08/$12.00 DOI: 10.1037/0278-7393.34.3.696

Attention Orienting Effects of Hesitations in Speech: Evidence From ERPs

Philip Collard, Martin Corley, and
Lucy J. MacGregor
University of Edinburgh

David I. Donaldson
University of Stirling

Filled-pause disfluencies such as *um* and *er* affect listeners' comprehension, possibly mediated by attentional mechanisms (J. E. Fox Tree, 2001). However, there is little direct evidence that hesitations affect attention. The current study used an acoustic manipulation of continuous speech to induce event-related potential components associated with attention (mismatch negativity [MMN] and P300) during the comprehension of fluent and disfluent utterances. In fluent cases, infrequently occurring acoustically manipulated target words gave rise to typical MMN and P300 components when compared to nonmanipulated controls. In disfluent cases, where targets were preceded by natural sounding hesitations culminating in the filled pause *er*, an MMN (reflecting a detection of deviance) was still apparent for manipulated words, but there was little evidence of a subsequent P300. This suggests that attention was not reoriented to deviant words in disfluent cases. A subsequent recognition test showed that nonmanipulated words were more likely to be remembered if they had been preceded by a hesitation. Taken together, these results strongly implicate attention in an account of disfluency processing: Hesitations orient listeners' attention, with consequences for the immediate processing and later representation of an utterance.

Keywords: disfluency, attention, speech, P300, ERPs

Disfluency is common in spontaneous speech (Fox Tree, 1995). Listeners encounter disfluency with such regularity that its effects on speech processing are of natural interest to those who research spoken-language comprehension. Converging lines of evidence show that disfluency can affect the way in which an utterance is understood. For example, hesitations in speech affect the confidence that listeners have in speakers' knowledge (Brennan & Williams, 1995), and disfluent corrections of a message may leave a lingering representation of the original content (Ferreira, Lau, & Bailey, 2004). Hesitations also affect syntactic representation, marking breaks in syntactic structure at phrase boundaries (Bailey & Ferreira, 2003).

But what happens at the point at which a disfluency has been encountered? Research addressing this question has tended to focus on hesitation-type disfluencies because these are often associated with local markers, such as elongations to words (e.g., *thee*) and filled pauses (e.g., *um, uh* or, in British English, *er*). Recently, Corley, MacGregor, and Donaldson (2007) used event-related potentials (ERPs) to demonstrate an immediate effect of hesitations while listening to spoken utterances such as (1) and (2).

(1) Everyone's got bad habits and mine is biting my [er] nails.

(2) Everyone's got bad habits and mine is biting my [er] tongue.

Using the N400 effect as an index of integration difficulty, they compared listeners' responses to unpredictable (difficult to integrate) words (2) against predictable words (1) in fluent contexts, and in disfluent contexts where the critical words were preceded by hesitations. The magnitude of the N400 (predictability) effect was significantly reduced for disfluent utterances, showing a clear effect of hesitations on listeners' language processing. Importantly, the N400 differences were associated with representational differences: Listeners were more likely to remember words that had been preceded by a hesitation in a forced-choice recognition task. One account of these findings is based on linguistic prediction, or expectancy. There is increasing evidence that listeners make online predictions during language comprehension (e.g., Altmann & Kamide, 1999; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). Furthermore, eyetracking evidence suggests that hesitations marked by prolongations such as *thee* and filled pauses such as *uh* may lead listeners to update their predictions about upcoming words (Arnold, Tanenhaus, Altmann, & Fagnano, 2004). Specifically, Arnold et al. (2004) showed that following hesitation, listeners were more likely to predict the upcoming mention of a discourse-new object, albeit from a limited set of candidate referents. In the absence of sufficient information from the environment regarding possible speech referents, hesita-

Philip Collard, Martin Corley, and Lucy J. MacGregor, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom; David I. Donaldson, Department of Psychology, University of Stirling, Stirling, Scotland, United Kingdom.

This research was partially supported by the College of Humanities and Social Science at the University of Edinburgh (Philip Collard), a Staff Research Development Grant (Martin Corley), the Economic and Social Research Council (Lucy J. MacGregor), and the Biotechnology and Biological Sciences Research Council (David I. Donaldson). We thank the members of the Psychological Imaging Laboratory at the University of Stirling and the University of Edinburgh Disfluency Group for support throughout the experiment.

Correspondence concerning this article should be addressed to Philip Collard, Psychology, University of Edinburgh, Edinburgh, Scotland EH8 9JZ, United Kingdom. E-mail: P.Collard@sms.ed.ac.uk

tion may cause a reduction in the extent to which specific predictions are made, leading to the N400 attenuation observed by Corley et al. (2007).

It is clear that disfluency can affect linguistic processes, such as prediction, but such processing differences may in turn be predicated on other mechanisms, such as attention. Compared to the most likely continuation of an utterance (fluent production of the next word), disfluency introduces novelty. Such novelty might occupy attention and hence limit the processing of the following part of the utterance. Alternatively, the novelty could enhance attention to, and facilitate the processing of, subsequent words. The existing data seem to support the latter alternative. In a word monitoring task, for example, Fox Tree (2001) found that participants identified targets more quickly following a hesitation including *uh*, which she attributed to heightened attention. An effect of disfluency on attentional processes might also account for findings that listeners respond more quickly to disfluent instructions (Brennan & Schober, 2001) and are more likely to remember words that follow a disfluent hesitation (Corley et al., 2007). Importantly, once attention is directed at an utterance, standard predictions as to what may follow may be affected.

Given these suggestions, the aim of the present study is to investigate directly the contention that attention for subsequent material is affected by disfluent hesitation in speech, using an "oddball" ERP paradigm. In such experiments, listeners are occasionally presented with stimuli that are physically deviant from more frequent standard stimuli, for example with respect to pitch or amplitude. The deviant stimuli elicit a cascade of neural events related to their detection and the orientation of attention toward them. The ERP effects commonly elicited by such oddball stimuli are the mismatch negativity (MMN) and members of the P300 family of components, such as the P3a and P3b. The MMN, an early (100- to 250-ms poststimulus) fronto-central negative difference wave (Schröger, 1997), appears to index neural processes involved in identification of deviance in the acoustic environment and can be modulated by highly focused attentional states (Alho, 1995). Occurring after the MMN at around 300 ms poststimulus, the frontally maximal P3a and the subsequent parietally maximal P3b are positive components typically associated with identification of, and attentional orientation to, deviant stimuli, and with the subsequent induced memory updating (Polich, 2004). Modulation of these attention-related ERP components following hesitations would provide strong evidence that the hesitations alter the attentional state of listeners.

In the current study, participants listened to recorded utterances containing infrequent changes to the auditory characteristics of single words. Half of the time, the manipulated words followed hesitations. These were marked by natural changes to the speech, such as elongations to words within the hesitation (e.g., *thee*), and the filled pause *er*. The acoustic changes were designed such that the manipulated words would be physically deviant from the acoustic regularities set up by the preceding speech but did not alter the linguistic content of the utterances. Because the deviant words were infrequent and therefore novel with respect to their contexts, they would be expected to induce equivalent attention-related ERPs in both fluent and disfluent conditions, unless, as we predicted, the attentional state of listeners was affected by preceding disfluency. If hesitations result in changes to the processing of subsequent words (indexed by alterations to the ERP signal), then we might expect some longer lasting changes to the representation of these words. Following Corley et al. (2007), we assessed this in a surprise recognition memory test.

Method

Participants

Twelve native English speakers participated in the experiment (7 men; mean age = 23 years; range = 17–36). All were right-handed and reported no known neurological impairment. Informed consent was obtained in accordance with the University of Stirling Psychology Ethics Committee guidelines. Participants were given financial compensation and course credit where applicable.

Materials

The stimuli consisted of 160 pairs of recorded utterances taken from Corley et al. (2007; an example is given in [1] above), which ended with a highly predictable target word (mean cloze probability = 0.84). Fluent and disfluent versions of utterances were recorded by a native British English speaker who was instructed to produce the utterances as naturally as possible. Disfluent versions incorporated a hesitation before the utterance-final word, which included signs of disfluency that were natural to the speaker, such as prolongations to preceding words (e.g., the prolonged definite article *thee*) and culminated in a filled-pause *er*. Utterances were recorded with a pseudotarget "pen" so that there were no acoustic cues to the upcoming word. Targets were recorded in separate carrier sentences and spliced onto the fluent and disfluent utterances, resulting in acoustically identical targets across the fluent and disfluent contexts. An additional 80 unrelated filler utterances were recorded. These were of a similar length to the experimental utterances. Half contained various types of disfluency, including hesitations marked by filled pauses, and disfluent repairs at varying positions within the utterances. Using the 320 experimental recordings, we created 320 additional stimuli by manipulating the target words to make them acoustically deviant. To do this, we applied an equalization pattern that was biased to the midrange frequencies from the target word onset until the end of the utterance. This resulted in an amplification of 2.8 dB across all frequencies except for the 125–1000 Hz range. In this range we applied a bell curvelike pattern that ranged from 2.8 dB to 18 dB and peaked at 500 Hz. The salient effect of the manipulation was to make the speech sound momentarily compressed, not unlike speech over a poor telephone line.

Four versions of the experiment were created, each containing 40 fluent normal, 40 disfluent normal, 40 fluent manipulated, and 40 disfluent manipulated recordings. Each target word occurred only once in each version of the experiment. Two copies of each of the 80 fillers were added to each set, resulting in a total of 320 recordings of which 80 ended in deviant target words. Thus, the overall deviant to normal utterance ratio was 1 in 4, ensuring that manipulated stimuli remained relatively novel oddballs throughout the experiment.

Procedure

The experiment comprised two sections. In the first, participants listened to the 320 experimental utterances and fillers. Materials were presented in a random order via computer loudspeakers in two blocks lasting around 20 min each, and separated by a break of a few minutes. Participants were instructed to listen to the recordings as if they were part of a normal conversation but were not given any other task. They were not told specifically about the presence of the disfluencies or acoustically manipulated words but were told that occasionally the sound recording quality would drop, which they should ignore.

Electroencephalogram (EEG) was recorded from 61 silver/silver-chloride electrodes embedded in an elasticized cap at standard 10–20 locations (Jasper, 1958), using a left-mastoid reference. Electro-oculargrams were collected to monitor for eye movements. EEGs and electro-oculargrams were amplified (band-pass filtered online, 0.01–40 Hz) and continuously digitized (16 bit) at 200 Hz. Electrode impedances were kept below 5 KΩ. Epochs were created from the EEG (150 ms before the onset of the target words to 800 ms after the onset), and these data were rereferenced offline to the average of the left- and right-mastoid electrodes, baseline corrected (relative to the average over the prestimulus interval), and smoothed over 5 points. Before averaging into ERPs, individual epochs were screened for drift of ±75 μV over 500 ms (amplitude difference between first and last data point of each epoch), and for artefacts of ±75 μV. The screening process resulted in the loss of 10.47% of epochs, with no significant variation in rejections between conditions, $F(3, 33) = 1.756$. Average ERPs were formed time locked to the onset of target words for each participant (minimum of 16 artefact free trials were required for inclusion).

In the second section of the experiment, participants performed a surprise recognition memory test for the material that they had heard. The 160 utterance-final (previously heard) target words were presented visually, interspersed with 160 frequency-matched foil words, which had not been uttered at any previous point during the experiment. After a 500-ms fixation cross, each word was presented for 750 ms, followed by a blank screen for 1,750 ms. Participants were instructed to decide whether each word had occurred at any previous point during the experiment and respond "old" or "new" via a button box placed in front of them. Responses that took longer than 2,500 ms were discarded.

Results

ERPs associated with the onsets of deviant target words were compared to ERPs to nonmanipulated standard controls for fluent and disfluent conditions. Because prestimulus baselines in fluent and disfluent utterances were different (including an *er* for disfluent cases), effects related to the acoustic manipulations were analyzed separately for fluent and disfluent conditions.

Figure 1 shows the distribution of the oddball effects over 100–400 ms. In fluent utterances, deviant words elicit an early negativity with an initial left hemisphere bias (100–150 ms) that spreads laterally into a very typical MMN distribution (150–200 ms). A large positive difference appears fronto-centrally at the midline (250–300 ms) and develops into a widespread centroparietally maximal positivity (300–400 ms). This pattern represents a typical P300 complex.

In disfluent utterances, effects are much smaller and less widespread. There is some indication of early negativity at the midline fronto-centrally (100–150 ms), which becomes lateralized with a right hemisphere bias (150–200 ms). No fronto-central positivity is apparent, although a less focal and greatly diminished centroparietal positivity can be seen later (300–400 ms).

Figures 2 and 3 show the waveforms of the MMN and P300 effects at electrodes used in the statistical analyses (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4) for fluent and disfluent utterances, respectively. In fluent utterances (see Figure 2), deviant stimuli give rise to midline dominant MMN and P300 effects. There is clear indication of a P3a-like early frontal component (as with the topographic depiction of the data; see Figure 1). Data from disfluent utterances are presented on the same scale (see Figure 3) and show oddball effects that are much smaller in magnitude.

ERPs were quantified by measuring the mean voltages for deviant and standard targets over two time windows, consistent with the MMN (100–200 ms) and the P300 (250–400 ms), for fluent and disfluent utterances separately. Greenhouse–Geisser corrections to degrees of freedom were applied, and corrected *F* and *p* values are reported where appropriate.

ERP effects were analyzed using three-way analyses of variance with factors of deviance (infrequent deviant, standard), location (electrodes F, C, and P) and laterality (electrodes 3, z, and 4).

For the fluent conditions, in the MMN time window, results showed a significant main effect of deviance, $F(1, 11) = 13.152$, $\eta_p^2 = .545$, $p = .004$, indicating that deviant stimuli elicited a widespread negativity across the scalp (mean voltages of –1.701 μV and –.118 μV for deviant and standard stimuli, respectively). No other effects involving the factor of deviance reached significance (Fs < 2.170).

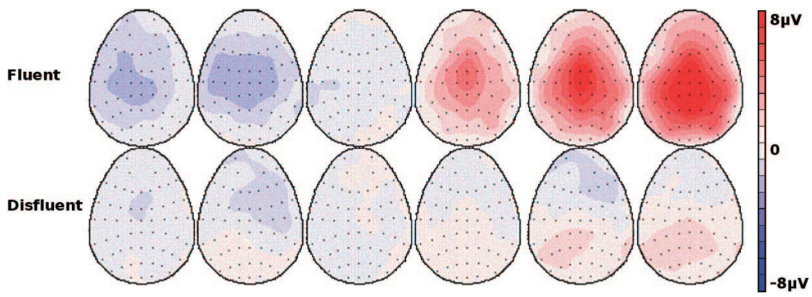

Figure 1. Topographic maps (anterior up; electrodes shown as black dots) illustrating the mean distributions of the deviance effects (deviant minus standard event-related potentials) over 100–400 ms (in 50-ms time windows) for fluent (top) and disfluent (bottom) utterances.

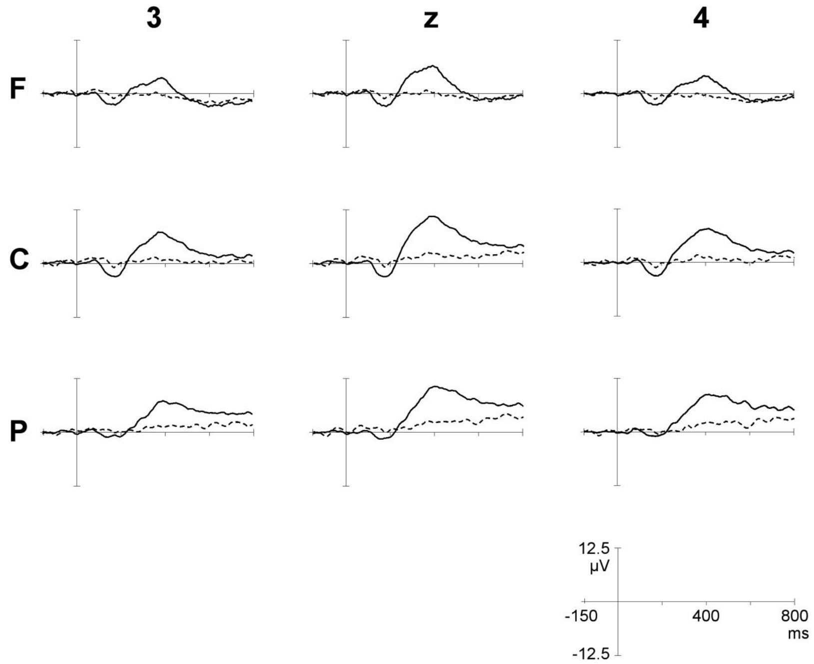

Figure 2. Grand average event-related potentials for deviant (continuous lines) relative to standard (dotted lines) target words in fluent utterances (positive up). Waveforms show data from left, midline, and right electrodes at frontal, central, and parietal sites (from left to right and top to bottom: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4).

In the P300 time window, there was a significant effect of deviance, $F(1, 11) = 51.080$, $\eta_p^2 = .823$, $p < .001$, reflecting a positivity associated with deviant words that was widespread across the scalp (mean voltages of 4.390 μV and .325 μV for deviant and standard stimuli, respectively). Significant Deviance × Laterality, $F(2, 22) = 10.045$, $\eta_p^2 = .477$, $p = .001$, and Deviance × Location × Laterality, $F(4, 44) = 7.920$, $\eta_p^2 = .419$, $p < .001$, interactions indicate that the deviance effect was larger over midline sites and that this midline bias was largest at frontal and posterior sites. No other effects involving the factor of deviance reached significance (Fs < 1.668).

For the disfluent conditions, in the MMN time window, there was a significant Deviance × Location interaction, $F(2, 22) = 4.950$, $\eta_p^2 = .310$, $p = .017$, indicating negativity associated with deviant words that was confined to frontal and central sites (mean voltages of .776 μV, .100 μV, and –.864μV for frontal, central, and posterior sites, respectively, for the standard stimuli and –.266 μV, –.908 μV, and –.859 μV for the deviant stimuli). No other effects involving deviance reached significance (Fs < 2.092).

In the P300 time window, there was a significant Deviance × Location interaction, $F(2, 22) = 6.033$, $\varepsilon = .553$, $\eta_p^2 = .354$, $p = .028$, indicating positivity associated with deviant words that was confined to posterior sites (mean voltages of 1.619 μV, .606 μV, and –.171 μV for frontal, central, and posterior sites, respectively, for the standard stimuli and .836 μV, 1.084 μV, and .974 μV for the deviant stimuli). No other effect involving deviance reached significance (Fs < 2.024).

These analyses demonstrate robust and typical MMN and P300 effects for acoustically deviant words in fluent stimuli. In disfluent contexts, the early negativity and later positivity are much weaker and less widespread, and there are some distributional differences between fluent and disfluent ERPs. However, the antecedents and gross topographies of the effects support an interpretation of MMN followed by P300 complex in each case. We therefore conducted a further analysis to compare effect sizes across fluent and disfluent conditions. Because the disfluent condition gave rise to interactions between deviance and location in both the MMN and P300 windows, location was also included as a factor in these comparisons. Each analysis was conducted on the deviance effect (ERPs to deviant items minus standard ERPs) using the factors of fluency and location, with the same electrode set as the previous analyses, collapsed across laterality. In the MMN time window, there were no significant effects involving fluency (Fs < 2.804). This is perhaps surprising in light of Figure 1, which corresponds to a mean difference between conditions of .898 μV across electrodes. In the P300 time window, a large difference between the fluent and disfluent conditions (mean of 4.434 μV and .187 μV for fluent and disfluent, respectively) was confirmed, $F(1, 11) = 32.484$, $\eta_p^2 = .747$, $p < .001$). The interaction of fluency and location was not significant, $F(2, 22) = 1.476$.

A final consideration was addressed using an additional analysis that examined the responses to disfluent items over time. By comparing responses during the first and second halves of the experiment, we were able to establish that the responses to deviant items following a hesitation did not differ over the course of the experiment, either for the MMN (Fs < 1.433 for all effects involving half) or the P300 (deviance by half: $F[1, 11] = 2.187$; other Fs < 1.035).

The second analysis focused on performance in the recognition task. As in Corley et al. (2007), the probability of correctly

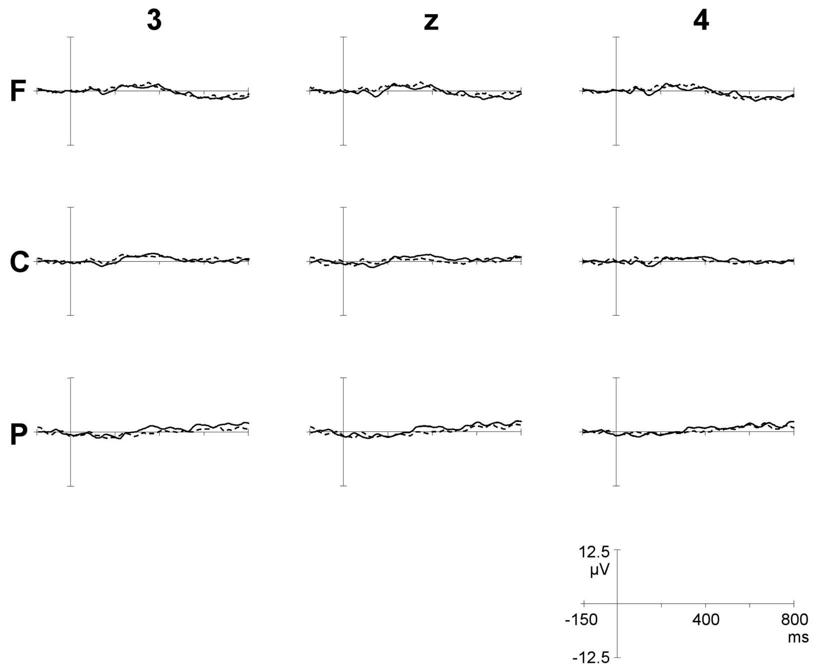

Figure 3. Grand average event-related potentials for deviant (continuous lines) relative to standard (dotted lines) target words in disfluent utterances (positive up). Waveforms show data from left, midline, and right electrodes at frontal, central, and parietal sites (from left to right and top to bottom: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4).

identifying words heard in the comprehension block of the experiment was quantified with stimulus identity treated as a random factor.[1] Overall, 57% of the previously heard words were correctly recognized (false alarm rate = 18%). Figure 4 shows the recognition probability of utterance-final words by fluency and deviance.

A two-way analysis of variance, with factors of fluency and deviance, revealed a significant interaction between the two factors, $F(1, 147) = 5.382$, $\eta_p^2 = .035$, $p = .022$. For standard stimuli, a pairwise comparison of recognition probabilities for words that had been heard in fluent or disfluent contexts showed a significant difference, $t(147) = 2.114$, $\eta_p^2 = .030$, $p = .036$, suggesting that acoustically normal words were more likely to be recognized following disfluency. Conversely, there was no difference in the recognition probabilities for deviant words, $t(147) = 1.083$.

Discussion

Large deflections in the ERPs were observed when participants encountered infrequently occurring acoustically deviant words in standard fluent speech. Given their polarities, distributions, timings, and antecedent conditions, it is clear that the ERP deflections correspond to the typical neural signatures of attention capture and orientation, the MMN and P300. When the same deviant words were encountered following a hesitation, there was some evidence for MMN and P300-like effects in the appropriate time windows. However, compared to the fluent case, amplitudes were greatly reduced, and distributions were less widespread.

Polich (2004) provided a model of ERPs elicited by auditory deviance. In his model, the MMN is associated with the detection of deviance by attentional systems. The P300 is driven by the novelty of the stimulus and is associated with orientation of attention toward deviant stimuli (frontal P3a component) and subsequent memory-updating processes (parietal P3b component). The reduction of the observed ERP effects following disfluency in the present study provides prima faciae evidence that hesitation affects the listener's attentional system. Moreover, the reduced response to novelty suggests that when the acoustically deviant words were encountered, attention was already oriented toward the speech, consistent with previous claims that hesitations heighten attention.

At first glance, these findings are reminiscent of results from attentional blink paradigms (e.g., Raymond, Shapiro, & Arnell, 1992). In attentional blink experiments, participants are less likely to detect a second target stimulus after a first, to which attention has presumably been oriented; this is accompanied by a reduced P300 to the second target (Vogel, Luck, & Shapiro, 1998). However, there are three reasons to suggest that the present findings

[1] Traditional adjustments for individual error rates, such as d', are inappropriate, since the properties of old stimuli are determined by their context of occurrence and hence there are no comparable categories of new stimuli. Using stimulus identity as a random factor ensures that per-participant biases to respond old or new are controlled for across the experiment.

Twelve target words were inadvertently repeated in the experiment, resulting in 148 distinct targets. Removing data from the repeated targets did not affect the outcome of the analysis of variance, but the fluency effect for standard stimuli became marginal, $t(135) = 1.993$, $\eta_p^2 = .027$ $p = .055$.

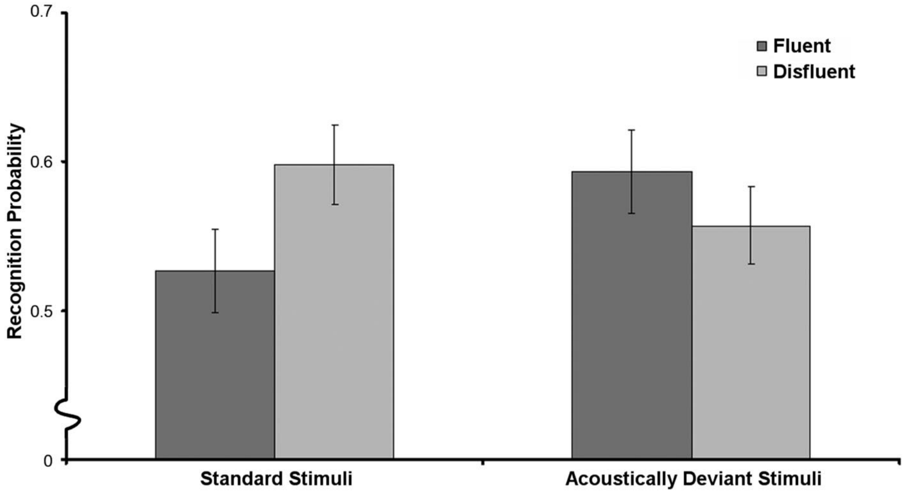

Figure 4. Recognition probabilities for utterance-final words that were originally presented as acoustically deviant or standard stimuli, in fluent or disfluent contexts (error bars represent one standard error of the mean).

cannot be accounted for in terms of an attentional blink. First, Corley et al. (2007) have demonstrated that the N400 effect related to low cloze probability words is attenuated following hesitations. There is no equivalent N400 attenuation in the attentional blink paradigm (Vogel et al., 1998). Second, in attentional blink paradigms, the attentional attenuation tends to be maximal about 300 ms after the onset of the initial orienting target (at a lag of three items, 100 ms/item; Vogel et al., 1998). In the present study, the mean delay between the onset of the *er* and that of the target word was 598 ms ($SD = 103$ ms; this is a low estimate for the time between events because signs of disfluency such as word prolongations sometimes occurred before the *er*).

The third and most important reason for rejecting an attentional blink account comes from the recognition task. Hesitations cause subsequent (acoustically normal) target words to be more likely to be later recognized, in direct contrast to what would be predicted if hesitation induced an attentional blink. This increase replicates the finding of Corley et al. (2007) that differences in the processing of fluent and disfluent utterances lead to long-term differences in the representations of those utterances, and further suggests that despite the acoustic manipulations necessary for the purposes of the present study, participants were engaged in comparable language processing. Salient (here, deviant) items were recognized equally often whether they had originally been encountered in fluent or disfluent utterances, possibly ascribable to a ceiling effect, given the numbers of stimuli and time between encoding and recognition of up to 55 min. Taken together, the results of the present study suggest that hesitations orient listeners' attention to the ongoing utterance. In contrast to attentional blink studies, attention is not occupied by hesitation; rather it is heightened so that listeners specifically attend to (and subsequently recognize) the words that follow. If the subsequent word is acoustically deviant, the standard MMN and P300 responses to deviance are attenuated, because attention is already oriented to the disfluent utterance. This provides a straightforward account for the increased likelihood of recognition following hesitations, as well as for the facilitated reaction times for targets that have been found in earlier studies (Brennan & Schober, 2001; Fox Tree, 2001).

Previous accounts of disfluency processing have either focused on changes in attention (Fox Tree, 2001) or changes to linguistic mechanisms (Arnold et al., 2004) that occur when hesitations marked by filled pauses are encountered. However, these accounts are not mutually exclusive. Hesitations may induce a low-level response that heightens listeners' attention, and this may in turn affect linguistic processes that alter the linguistic availability of subsequent material. Clearly, such an account would require elaboration: For example, it is presently unclear whether listeners' heightened attention is speech specific or represents a more general state of arousal. Such issues remain questions for future research. The importance of the present study is that it provides clear evidence that attention is affected by hesitation in an utterance, either concomitantly with, or as a precursor to, linguistic processes.

References

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing, 16,* 38–51.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247–264.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15,* 578–582.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language, 49,* 183–200.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44,* 274–296.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34,* 383–398.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you say it: Disfluency in speech affects the comprehension process. *Cognition, 105,* 658–668.

Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, parsing, and tree-adjoining grammars. *Cognitive Science, 28,* 721–749.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34,* 709–738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29,* 320–326.

Jasper, H. H. (1958). Report to the committee on methods of clinical examination in electroencephalography. Appendix: The ten-twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10,* 370–375.

Polich, J. (2004). Neuropsychology of P3a and P3b: A theoretical over view. In N. C. Moore & K. Arikan (Eds.), *Brainwaves and mind: Recent developments* (pp. 15–29). Wheaton, IL: Kjellberg.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 849–860.

Schröger, E. (1997). On the detection of auditory deviants: A pre-attentive activation model. *Psychophysiology, 34,* 245–257.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 443–467.

Vogel, K. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1656–1674.

Received April 14, 2007
Revision received January 3, 2008
Accepted January 4, 2008 ■

References

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*, 197-207.

Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. *Journal of Memory and Language*, *38*, 112-130.

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing*, *16*, 38-51.

Alho, K., Paavilainen, P., Reinikainen, K., Sams, M., & Näätänen, R. (1986). Separability of different negative components of the event-related potential associated with auditory stimulus processing. *Psychophysiology*, *23*, 613-623.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by langauge and world knowledge: Linking anticipatory (and other) eye movements to linguistic process. *Journal of Memory and Language*, *57*, 502-518.

Arnell, K. M., & Jolicoeur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental*

Psychology: Human Perception and Performance, *25*, 630-648.

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference resolution. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 914-930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new. disfluency and reference resolution. *Psychological Science*, *15*, 578-582.

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language*, *76*, 28-55.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*, 183-200.

Baker, L., & Wagner, J. L. (1987). Evaluating information for truthfulness: The effects of logical subordination. *Memory and Cognition*, *15*, 247-255.

Barton, S. B., & Sanford, A. J. (1993). A case-study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory and Cognition*, *21*, 477-487.

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*, 201-211.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 343-355.

Birch, S. L., Albrecht, J. E., & Myers, J. L. (2000). Syntactic focusing structures influence discourse processing. *Discourse Processes*, *30*, 285-304.

Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, *34*, 232-267.

Birch, S. L., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory and Cognition*, *25*, 653-660.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*, 148-158.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*, 123-147.

Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effects on a semantic illusion. *Acta Psychologica*, *67*, 135-144.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274-296.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive state of speakers. *Journal of Memory and Language*, *34*, 383-398.

Bridgeman, B., Heijden, A. H. C. Van der, & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences*, *17*, 247–247.

British National Corpus. (1995). Available at www.natcorp.ox.ac.uk.

Butterworth, B. L. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *1*, 75-78.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975-979.

Christenfeld, N. (1995). Does it hurt to say um? *Journal of nonverbal behaviour*, *19*, 171-186.

Chun, M. M., & Potter, M. C. (2001). The limits of attention: Temporal constraints on human information processing. In K. Shapiro (Ed.), (p. 20-35). Oxford University Press.

Clark, H. H. (1994). Managing problems in speaking. *Speech communication*, *15*, 243-250.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73-111.

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 696-702.

Comerchero, M., & Polich, J. (2000). P3a, perceptual distinctiveness, and stimulus modality. *Journal of Cognitive Brain Research*, *7*, 41-48.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*, 658-668.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21-58.

Coulson, S., & Wu, Y. C. (2005). Right hemisphere activation of joke-related information: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *17*, 494-506.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *39*, 131-143.

Dimond, S. J., & Beaumont, J. G. (1973). Difference in the vigilance performance of the right and left hemispheres. *Cortex*, *9*, 259-65.

Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and Speech*, *28*, 377-389.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*, 272-300.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456-467.

Erickson, T. A., & Matteson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*, 540-552.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 202-224.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75-84.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, *22*, 1151-1177.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11-15.

Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, *28*, 721-749.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*, 709-738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition*, *29*, 320-326.

Fox Tree, J. E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes*, *34*, 37-55.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*, 151-167.

Freedman, S. E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition*, *19*, 101-131.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*, 89-106.

Genette, G. (1980). *Narrative discourse: An essay on method* (T. J. E. Lewin, Trans.). Ithaca, NY: Cornell University Press.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: telephone speech corpus for research and development. *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, *1*, 517-520.

Goldman-Rakic, P. S. (1988). Topography of cognition: Parallel distributed networks in primate association cortex. *Annual Reviews in Neuroscience*, *11*, 137-156.

Grimes, J. (1996). Perception. Vancover Studies in Cognitive Science. In K. Akins (Ed.), (p. 5: 89-109). New York: Oxford University Press.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP-measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439-483.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438-441.

Hahne, A., & Jescheniak, J. D. (2001). What's left if the jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, *11*, 199-212.

Hawkins, R. R. (1971). The syntactic location of hesitation pauses. *Language and Speech*, *14*, 277-288.

Heike, A. E. (1981). A content-processing view of hesitation phenomena. *Language and Speech*, *24*, 147-160.

Heilman, K. M., Watson, R. T., & Valenstein, E. (1993). Clinical neuropsychology. In K. M. Heilman & E. Valenstein (Eds.), (Vol. 3, pp. 269–346). Oxford University Press.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related potentials. *Psychobiology*, *19*, 286-300.

Howell, P., & Young, K. (1991). The use of prosody in highlighting alterations in repairs from unrestricted speech. *Quarterly Journal of Experimental Psychology*, *43*, 733-758.

Jasper, H. H. (1958). Report to the committee on methods of clinical examination in electroencephalography. Appendix: The ten-twenty system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 370-375.

Kaan, E., Harris, T., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159-201.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*, 98-110.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133-156.

Kelley, T. A., Chun, M. M., & Chua, K. P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision*, *3*, 1-5.

Kircher, T. J., Brammer, M. J., Levelt, W. J. M., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: Engagement of left temporal cortext during pauses in speech. *NeuroImage*, *21*, 84-90.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557-577.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161-163.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, *197*, 792-795.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994-2005). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., p. 83-143). San Diego, CA: Elsevier Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Lickley, R. J. (1995). Missing disfluencies. In *Proceedings of the international congress of phonetic sciences* (Vol. 4, p. 192-195). Stockholm, Sweden.

Lickley, R. J., & Bard, E. G. (1996, October). On not recognising disfluencies in dialogue. In *Proceedings of the international conference on spoken language processing* (Vol. 3, p. 1876-1879). Philadelphia, USA.

Lindsay, J., & O'Connell, D. C. (1995). How do transcribers deal with audio recordings of spoken discourse? *Journal of Psycholinguistic Research*, *24*, 101-115.

Luck, S. J. (2005). *An introduction to the event-related potential technique.* Cambridge, Massachusetts: The MIT Press.

MacGregor, L. J. (2008). *Disfluencies affect language comprehension: evidence from event-related potentials and recognition memory.* Unpublished doctoral dissertation, University of Edinburgh.

Maclay, H., & Osgood, C. E. (1959). Heistation phenomena in English speech. *Word*, *15*, 19-44.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203-208.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Expeirmental Psychology*, *11*, 56-60.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of experimental psychology: Human perception and performance*, *10*, 667-682.

Mosteller, F. (1948). A k-sample slippage test for an extreme population. *The Annals of Mathematical Statistics*, *19*, 58-65.

Münte, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, *36*, 217-226.

Münte, T. F., Matske, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, *9*, 318-329.

Murray, W., & Rowan, M. (1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research*, *27*, 1-22.

Näätänen, R. (1992). *Attention and brain function.* Lawrence Erlbaum Associates.

Näätänen, R., & Gaillard, A. W. K. (1983). EEG correlates of information processing: Theoretical issues. In A. W. K. Gaillard & W. Ritter (Eds.), (p. 169-177). Amsterdam: Elsevier.

Näätänen, R., Gaillard, A. W. K., & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologia*, *42*, 313-329.

Neville, H., Nicol, J., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*, 151-165.

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary change deafness in discourse comprehension. *Cognitive Brain Research*, *24*, 691-701.

O'Connell, D. C., & Kowal, S. (2004). The history of research on the filed pause as evidence of the written language bias in linguistics. *Journal of Psycholinguistic Research*, *33*, 459-474.

O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, *7*, 191–212.

O'Shaughnessy, D. (1992). Recognition of hesitations in spontaneous speech. In *Proceedings of the ICASSP* (Vol. 1, pp. 521–524).

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

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*, 739-773.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, *9*, 19-35.

Pashler, H. E. (Ed.). (1998a). *Attention.* Psychology Press.

Pashler, H. E. (1998b). *The psychology of attention.* MIT Press.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*, 717-733.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585-589.

Plauche, M. C., & Shriberg, E. E. (1999, July). Data-driven subclassification of disfluent repetitions based on prosodic features. In *Proceedings of the international congress of phonetic sciences* (Vol. 3, p. 1513-1516). Berkley, USA.

Polich, J. (2004). Neuropsychology of P3a and P3b: A theoretical overview. In N. C. Moore & K. Arikan (Eds.), *Brainwaves and mind: Recent developments* (p. 15-29). Wheaton, IL: Kjellberg Inc.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128-2148.

Posner, M. I. (1978). *Chronometric explorations of mind.* L. Erlbaum Associates.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Expeirmental Psychology*, *32*, 3-25.

Posner, M. I. (1988). Masters lectures in clinical neuropsychology. In T. Boll & B. Bryant (Eds.), (p. 173-202). American Psychological Association.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Reviews in Neuroscience*, *13*, 25–42.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience*, *4*, 1863-1874.

Postma, A., & Kolk, H. H. J. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research*, *35*, 537-544.

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, *89*, 506-540.

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian journal of experimental psychology*, *49*, 151.

Ranganath, C., & Paller, K. A. (1999). Frontal brain activity during episodic and semantic retrieval: Insights from event-related potentials. *Journal of Cognitive*

Neuroscience, *11*, 598-609.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 849-860.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125-157.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17-42.

Rensink, R. A. (2002a). Change detection. *Annual Review of Psychology*, *53*, 245-277.

Rensink, R. A. (2002b). Failure to see more than one change at a time. *Journal of Vision*, *2*, 245.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). The Need for Attention to Perceive Changes in Scenes. *Psychological Science*, *8*, 368–373.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, *7*, 127-145.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75-116.

Rooth, M. (1995). Handbook of contemporary semantic theory london: Blackwell. In S. Lappin (Ed.), (p. 271-298). London: Blackwell.

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience*, *12*, 664-678.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, *2*, 437-442.

Sanford, A. J., Clegg, M., & Majid, A. (1998). The influence of types of character on processing background information in narrative discourse. *Memory and Cognition*, *26*, 1323-1329.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Scences*, *6*, 382-386.

Sanford, A. J. S., & Molle, J. (2006, August). Disfluencies and selective attention in speech processing. In *Archetectures and mechanisms in language processing* (p. 183). Turku, Finland.

Sanford, A. J. S., Sanford, A. J., Filik, R., & Molle, J. (2005). Depth of lexical-semantic processing and sentential load. *Journal of Memory and Language*, *53*, 378-396.

Sanford, A. J. S., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*, 109-130.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, *60*, 362-367.

Schachter, S., Rauscher, F., Christenfeld, N. J. W., & Crone, K. T. (1994). The vocabularies of academia. *Psychological Science*, *5*, 37-41.

Scharf, B., & Buus, S. (1986). Handbook of perception and human performance, volume i. In K. Boff, L. Kaufman, & J. Thomas (Eds.), (p. 14.1-14.71). New York: Wiley.

Scherg, M., Vajsar, J., & Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, *1*, 336-355.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the twenty-eighth meeting of the cognitive science society.* Vancouver, Canada.

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-prime user's guide.* Pittsburgh, PA: Psychology Software Tools, Inc.

Schröger, R. (1997). On the detection of auditory deviants: A pre-attentive activation model. *Psychophysiology*, *34*, 245-257.

Shriberg, E. S. (1995, August). Acoustic properties of disfluent repetitions. In *Proceedings of the international congress of phonetic sciences* (Vol. 4, p. 384-387). Stockholm, Sweden.

Shriberg, E. S. (1996). Disfluencies in SWITCHBOARD. In *Proceedings international conference on spoken language processing* (Vol. Addendum, p. 11-14). Philadelphia, PA.

Shriberg, E. S. (2001). To "errrr" is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*, 153-169.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, *5*, 644-649.

Simons, R. F., Graham, F. K., Miles, M. A., & Chen, X. (2001). On the relationship of P3a and the Novelty-P3. *Biological Psychology*, *56*, 207-218.

Smith, V., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25-38.

Spence, C., & Driver, J. (1997). On measuring selective attention to an expected sensory modality. *Perception and Psychophysics*, *59*, 389-403.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli. *Electroencephalography*, *38*, 387-401.

Stracuzzi, D., & Kinsey, J. (2003). *Eye Track (Version 0.7.6).* http://www.umass.edu/psychology/div2/eyelab/.

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*, 882-888.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*, 1187-1188.

Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listenting. *American Journal of Psychology*, *77*, 533-546.

Treisman, A. M., & Riley, J. G. (1969). Is selective attention selective perception or selective response? A further test. *Journal of Experimental Psychology*, *79*, 27-34.

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*, 86-94.

Underwood, G., & Moray, N. (1971). Shadowing and monitoring for selective attention. *The Quarterly Journal of Experimental Psychology*, *23*, 284-295.

Urbach, T. P., & Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology*, *39*, 791-808.

Van Petten, C. M., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*, 380-393.

Van Petten, C. M., & Kutas, M. (1991). Influences of semantic and syntactic context on open and closed class words. *Memory and Cognition*, *19*, 95-112.

Van Petten, C. M., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, *3*, 131-150.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition.*, *31*, 443-467.

Van Berkum, J. J. A., Zwisterlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*, 701-718.

Vaughan, B., & Yantis, S. (1999). Information preservation during occlusion. *Investigative Ophthalmology & Visual Science*, *40*, 415.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal* *of Experimental Psychology: Human Perception and Performance*, *23*, 1656-1674.

Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Memory & Cognition*, *35*, 73-86.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48-64.

Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Change Blindness and Visual Memory*, *7*, 297-322.

Woldorff, M. G., Hackley, S. A., & Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, *28*, 30-42.