The roles of speech errors, monitoring, and anticipation in the production of normal and stuttered disfluencies

Paul H. Brocklehurst

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

2011

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.

Paul H. Brocklehurst

Abstract .

In their Covert Repair Hypothesis (CRH), Postma and Kolk (1993) proposed that stuttering-like disfluencies arise, in both normal and stuttered speech, as a consequence of speakers' attempts to repair phonological-encoding errors before they start to speak. They posited that stutterers are particularly disfluent because they make larger numbers of such errors compared to normally-fluent speakers. To date, however, experimental research has provided little reliable evidence to support or counter this hypothesis. This thesis constitutes a systematic attempt to provide such evidence. Using a tongue-twister paradigm in conjunction with manipulations of auditory masking, it first documents (a) the vigilance with which normally-fluent speakers monitor for such errors; (b) the relative accuracy with which they detect them; and (c) the frequency with which they occur – in both inner and overt speech. A second set of experiments then extends the same investigation to a group of stutterers and matched controls and explores the relationship between the occurrence of participants' errors in the experimental paradigm and the frequency of their stuttering-like disfluencies in everyday speaking situations. Together, these experiments reveal that, compared to controls, participants who stutter monitor their speech with similar levels of vigilance; identify phonemic errors with similar degrees of accuracy; and, as predicted by the CRH, produce significantly more errors – in both their inner and overt speech. However, contrary to the predictions of the CRH, no relationship was found between the frequency of such errors in inner speech and the severity of participants' disfluencies. In a final set of experiments, a speech-recognition paradigm is employed to explore an alternative hypothesis: that stuttering-like disfluencies can be precipitated, in a speaker, by the mere anticipation that his words will result in communication failure. Results revealed that, for stutterers, stuttering decreased on words that were consistently followed by feedback implying correct recognition, but not on words followed by feedback implying incorrect recognition. For normally-fluent speakers, equivalent correlations were not found. The thesis concludes that slow or impaired phonological encoding may play a role in the development of the disorder. But, once established, the anticipation of communication failure may be a more important factor in determining where and when stuttering-like disfluencies actually occur. It then discusses implications of the experimental findings for hypotheses that posit a connection between phonological encoding and stuttering.

Acknowledgements

First I should like to thank the many people assisted me with various aspects of the PhD. In particular I am grateful to Martin Corley and Robin Lickley, my two supervisors who persuaded me to apply to do the PhD in the first place and supported me throughout. If they hadn't suggested it, I would never have contemplated such an undertaking. I am also especially grateful to my parents for assisting with funding throughout and making up short-falls wherever necessary. Again, without their help it would not have been possible.

I would like also to thank Suzy Moat for providing all sorts of help with things to do with computer programs and programming, Sam Miller who assisted with running participants in one of the tongue-twister experiments, Tony Buhr who provided valuable feedback on a number of theoretical issues, Louise Cotton for checking my coding of errors, and Gary Oppenheim whose studies on inner speech provided the inspiration for the tongue-twister experiments and whose feedback was especially valuable. I also wish to thank the British Stammering Association, and the regional self-help groups who put me in touch with many of the participants, and of course, the participants themselves, without whom the experiments could not have been carried out. Also thanks to my flat mates Garry Cai and Dina van der Hulst who, amongst other things, made sure I remembered to eat and drink, and to Elli Drake for ensuring I had regular walks.

Finally, I should like to thank the various funding bodies that supported me through the PhD, in particular the ESRC, who provided me with a postgraduate research studentship (no. ES/G01230X/1) without which the PhD would not have been possible. In addition, I received a Grindley grant which enabled me to attend the 2009 International Fluency Association conference in Rio de Janeiro, and a number of Research Support Grants, from The University of Edinburgh, School of Philosophy, Psychology and Language Sciences, that enabled me to attend the 2009 AMLAP conference in Barcelona and the 2010 European Symposium on Fluency Disorders in Antwerp.

Table of Contents

Figures

Tables

1. Introduction and Thesis Overview

1.1. Introduction

Persistent developmental stuttering is a disorder of fluency that affects approximately one percent of adults worldwide. The condition can significantly impair the ability of affected individuals to communicate effectively and to participate in society. Although a considerable amount is now known about the factors that influence stuttering, the mechanism through which episodes of stuttering occur is still not understood.

In 1993, a hypothesis was proposed by Postma and Kolk that linked stuttering to difficulties arising from an underlying impairment of the language production system. Specifically, it equated stuttering-like disfluencies in both normal and stuttered speech with speakers' attempts to repair phonological-encoding errors, covertly, before they start to speak; and it posited that stutterers are particularly disfluent because they make larger numbers of such errors compared to normally-fluent speakers. This so-called 'Covert Repair Hypothesis' differed from previous hypotheses insofar as it was the first ever to provide a detailed account of a possible mechanism through which the repetitions, prolongations and blocks that characterise stuttering may arise. To date, however, experimental research has provided little reliable evidence to support or counter this hypothesis. This thesis constitutes a systematic attempt to provide such evidence.

The Covert Repair Hypothesis is predicated on the following underlying presumptions about the nature of language production: (a) that speakers are able to monitor their speech plans for errors of phonological encoding prior to the onset of overt articulation; and (b) that, in stutterers, monitoring functions normally.

Because these underlying presumptions have been questioned, in this thesis we also investigate their validity. In order to do so, we make use of an experimental paradigm, originally developed by Oppenheim and Dell (2008), whereby speakers recite tonguetwisters in inner (and overt) speech and self-report their errors. However, the reliability of this paradigm has never been adequately established, and it is unclear to what extent speakers are able to perceive phonological encoding errors through pre-articulatory monitoring. Therefore, in our first three experiments we test the reliability of the paradigm across different groups of (normally-fluent) participants, by documenting the types and frequencies of inner and overt errors that they self report, and by comparing their (overt) self-reports to those of an independent rater. In so doing, these experiments also address some questions regarding the architecture of the language-formulation system.

The remainder of the thesis constitutes an attempt to provide answers to three key questions that relate directly to the Covert Repair Hypothesis...

1. Compared to normally-fluent speakers, do adults who stutter monitor their speech with an equal degree of vigilance?
2. Do adults who stutter make more errors of phonological encoding than normally fluent speakers?
3. Is the likelihood of stuttering on a word increased by stutterers’ perceptions that they need to speak the word more accurately?

The first two of these questions address two of the central tenets of the Covert Repair Hypothesis. The third is highly relevant to the Covert Repair Hypothesis because findings from studies that have investigated the moments when stuttering occurs suggest that the extent to which a speaker is likely to engage in covert error repair (and/or other related behaviours) may be driven more by his perception that he is not being understood, than by the presence of actual phonemic (or other) errors in his speech.

To investigate Questions 1 and 2 above, in our fourth experiment we adapt the Oppenheim and Dell tonguetwister paradigm so that it can be used to compare the frequency of phonological-encoding errors in the speech plans of adults who stutter and matched normally-fluent controls. As will be seen, this fourth experiment revealed that, compared to the normally-fluent controls, the stuttering participants monitored their speech with a similar level of vigilance, but self-reported significantly more errors, both in their inner as well as their overt speech. However, it also revealed that language-formulation errors are unlikely to account for more than a small proportion of the stuttering-like disfluencies that commonly occur in their everyday speech.

The latter of these findings led us to formulate Question 3 (above) and to consider the possibility that, rather than resulting from the speaker's formulation errors, stuttering may be more closely associated with the speaker's perceptions of the abilities of the listener (or speech-recognition system), and that perhaps stuttering-like disfluencies are precipitated by the speaker's anticipation that the words he is about to utter are likely to be misrecognised (and thus will fail to fulfil their intended function). In the final set of experiments, we developed a new experimental paradigm to test this possibility and, more specifically, to answer Question 3. In two experiments, the first involving stutterers and the second involving normally-fluent speakers, participants spoke single words into a computer that they believed contained speech-recognition software. The computer then provided (visual) feedback consistently indicating its misrecognition of certain words. The results of this final set of experiments revealed that, in adults who stutter, feedback suggesting misrecognition of previous iterations of a word increased the likelihood of future stuttering on that word. An equivalent experiment conducted with normally-fluent participants failed to produce the same results.

In the final chapters of the thesis, the implications of our experimental results for theoretical models of stuttering are evaluated within the context of other existing research. We propose that the symptoms of stuttering and the moments in which they occur can best be accounted for by the modification of currently existing psycholinguistic models such that, in addition to language-production difficulty, they also include the anticipation of communication failure as a key factor that disrupts the language-production system, and result in the production of stuttering-like disfluencies.

1.2. Chapter by chapter Overview

Chapter 2 starts with a discussion of the problems that arise when trying to define what stuttering is. It then provides an overview of the symptoms of stuttering and what they might tell us about its nature. In so doing, it also highlights key reasons for suspecting that both language-production impairment as well as stutterers' perceptions of the overall adequacy of their speech may be implicated in stuttering.

Chapter 3 provides a overview of three types (or categories) of hypothesis that have attempted to account for the occurrence and quality of stuttering-like disfluencies: (1) Hypotheses (including the Covert Repair Hypothesis)that postulate that stuttering stems from some form of underlying speech or language-production impairment; (2) Hypotheses that associate stuttering purely with abnormal perception or cognition. The chapter then goes on to describe a third category of 'Anticipatory Struggle' hypotheses, which view stuttering as a learned response that can be triggered by cues associated with past experiences of disfluency or communication failure. All three of these categories of hypothesis motivated experiments in this thesis. The chapter then goes on to describe the Covert Repair Hypothesis in detail and contains a review of published studies that provide evidence for or against the various tenets of the CRH. The main focus is on studies with experimental designs, although important correlational and corpus-based evidence is also outlined. Two alternative psycholinguistic hypotheses that have challenged the CRH are also described, together with supporting evidence: (a) The 'Vicious Circle Hypothesis' (Vasić & Wijnen, 2005), which attributes the production of stuttering-like disfluencies to excessive covert error repair activity resulting from hyper-vigilant error monitoring; and (b) the EXPLAN hypothesis

(Howell & Au-Yeung, 2002) which retains the CRH’s tenet that stuttering-like disfluencies arise as a consequence of slow language formulation, but challenges its assumption that they stem from covert repairs of language-formulation errors. The experiments contained in this thesis are relevant to both of these hypotheses.

Chapter 4 discusses how insight into the extent to which speakers employ covert editing and repair as a means of reducing errors of phonological encoding can be gleaned from the study of the frequency biases that such errors tend to exhibit. It begin with a description of two key frequency biases – the phonemic-similarity bias and lexical bias, and two different accounts of how these biases may arise. It then discusses inner speech, and how paradigms that investigate inner speech have the potential to reveal much about the extent to which speakers engage in covert editing and error repair. They also have the potential to reveal differences between stutterers and normally-fluent speakers with respect to the production, detection and covert editing of errors of phonological encoding.

Chapter 5 focuses on methodological considerations. It highlights the problems associated with the objective measurement of subjective experiences, including inner speech and the subjective components of stuttering and discusses the rationale behind the methods of measurement adopted in this thesis. It also provides a rationale for the use of mixed-effects modelling for the key statistical analyses conducted in each of the experiments.

Chapters 6 to 8 comprise descriptions of the experiments carried out during the course of the PhD. Specifically, Chapter 6 comprises a description of the first three tongue-twister experiments that were designed to investigate the reliability of the Oppenheim and Dell paradigm. Specifically, these three experiments measure inner and overt speech error rates and monitoring vigilance in groups of normally-fluent speakers. In addition to confirming the reliability of the paradigm as a means of measuring inner-speech errors, these experiments also provide some new insights into the architecture of the speech-production mechanisms of normally-fluent speakers. Chapter 7 describes our adaptation of the tonguetwister paradigm in which we compare speech errors in adult participants who stutter and matched controls. Chapter 8 describes the two experiments that investigate the role of anticipation of communication failure in the production of stuttering-like disfluencies both in adults who stutter and in normally-fluent speakers.

Chapter 9 comprises a general discussion of the findings of the three groups of experiments and what they tell us about the nature of stuttering and how current psycholinguistic models of stuttering may be adapted to accommodate the findings.

2. The nature of stuttering

In this chapter we discuss definitional issues surrounding stuttering and present a broad outline of its symptoms and the circumstances in which it occurs. Although the experimental studies in this thesis investigate stuttering in adults, to portray the adult condition in a meaningful perspective we have also described how the symptoms of stuttering first arise in childhood and how they change with age.

2.1. What is stuttering?

It would seem reasonable to start a thesis about stuttering with a definition of what it is. However, as yet, there is no such universally agreed definition. This reflects the fact that researchers have divided opinions with respect to a number of key issues regarding the nature of the disorder. In particular there are fundamental disagreements on: (a) what (if anything) distinguishes the ‘stuttered disfluencies’ produced by people who stutter from the ‘stuttering-like’ disfluencies that occur in normal speech; and (b) the extent to which stuttering really constitutes a disorder that is categorically distinct from the higher end of the spectrum of normal disfluency in the child and adult population.

To a large extent, definitional disagreements reflect differing views among researchers regarding the extent to which stutterers’ own subjective experiences should be taken into account. Thus for example, Perkins, Kent, and Curlee (1991) proposed that the core symptom of stuttering is a transient yet recurring experience of loss of control over articulation, which prevents the speaker from being able to fluently execute the words he has planned. This subjective component of stuttering is also reflected in an early, but nevertheless still widely used definition of stuttering, adopted by the World Health Organization, which describes the condition as:

“Disorders in the rhythm of speech in which the individual knows precisely what he wishes to say but at the time is unable to say because of an involuntary repetition, prolongation, or cessation of a sound”. (ICD9; World Health Organization, 1977, p. 202)

More recently, the above definition has fallen out of favour, in part because the experience of loss of control is entirely subjective, and is not amenable to independent verification, and also because there is there no way of telling whether young children (who have not yet acquired enough language to express such experiences) also experience a similar loss of control. Indeed, as will be discussed in Section 2.3, there are strong reasons to suspect that young children who stutter may not experience the same loss of control. On the contrary, there is evidence that their disfluencies may stem from language-formulation difficulties whereby they are slow to access appropriate words and/or to construct grammatical structures (see Bernstein Ratner, 1997 for a review).

For purely practical reasons the need for a definition of stuttering based on objectively verifiable criteria has led many researchers and clinicians to ignore these uncertainties and to base their working definitions of stuttering simply on the severity and frequency of ‘stuttering-like disfluencies’ in a person’s speech. Reflecting this change, the World Health Organization’s ICD10 definition focuses entirely on objectively measurable symptoms:

“Speech that is characterized by frequent repetition or prolongation of sounds or syllables or words, or by frequent hesitations or pauses that disrupt the rhythmic flow of speech. It should be classified as a disorder only if its severity is such as to markedly disturb the fluency of speech” (ICD10; World Health Organization, 1993).

The reliance of the ICD10 entirely on objectively measurable symptoms leads, however, to a dilemma: Whereas, from the perspective of the ICD9 definition of stuttering, stutterers and non-stutterers fall into two distinct categories, depending on whether or not they experience a loss of control over their ability to articulate words that they wish to say, from the ICD 10 perspective, there is no such clear dichotomy. Rather, whether or not someone is defined as a stutterer is likely to depend on the frequency or severity of their observable 'stuttering-like disfluencies' – a term first coined by Yairi and Ambrose (1992a) and now often used to denote the types of disfluencies that are particularly common in the speech of people who stutter, including part and whole word repetitions, prolongations and 'blocks'[1] (cf. Johnson, 1959; Wingate, 1988, 2001; Yairi & Ambrose, 2005)[2]. The point beyond which a speaker is classified as a stutterer is then, by necessity, arbitrary. Moreover, because all speakers produce some stuttering-like disfluencies, yet clearly not all speakers would be diagnosed as stutterers or would consider themselves to 'have a stutter', depending on the definition of stuttering that one adopts, stuttering-like disfluencies may or may not be symptoms of stuttering.

The validity of the status of stuttering-like disfluencies as symptoms of stuttering is especially relevant in relation to research into stuttering, not least because some individuals who consider themselves to be stutterers in fact produce fewer stuttering-like disfluencies than individuals who consider themselves to be normally-fluent (i.e. non-stutterers). However, 'stutterers' who do not produce abnormal numbers of stuttering-like disfluencies may go to great lengths to avoid people or situations where stuttering-like disfluencies are

---

[1] Yairi and Ambrose's (1992) use of the term *"stuttering-like disfluency"* (SLD) included (a) part-word repetition, (b) single-syllable word repetition, (c) disrhythmic phonation (sound prolongation and blocks), and (d) tense pause (audible tense vocalization between words).

[2] These authors disagree on exactly which types of disfluency should be considered 'stuttering-like', and in particular whether or not whole-word repetitions and silent pauses should be included in this category (see Howell, 2011b, Chapter 3, for a detailed discussion of this issue).

likely to occur, and the impact of the condition on their lives may be just as strong as it is on the lives of stutterers whose disfluencies are severe (Sheehan, 1953; Wingate, 2001).

First and foremost, the Covert Repair Hypothesis is a hypothesis that attempts to account for the occurrence of stuttering-like disfluencies; both in the speech of stutterers as well as normally-fluent speakers. The hypothesis considers stuttering to be simply the severe end of a continuum of otherwise normal disfluency and, in this respect, is predicated on the ICD10 definition cited above.

The question of the definition (and hence also diagnosis) of stuttering is discussed in more depth in Section 3.2.1 in the context of 'Diagnosogenic Theory' (Johnson, 1942, 1959) which posits that the self-perception of oneself as a stutterer can itself lead to the production of stuttering-like disfluencies. The working definition of stuttering adopted in this thesis is described in Section 5.1.2.

2.2. The onset of stuttering

Most commonly, stuttering is first diagnosed between 2 and 6 years of age in response to a sudden marked increase in the frequency and severity of stuttering-like disfluencies (Yairi & Ambrose, 1992b; Yairi & Lewis, 1984). Before this age, affected children generally appear to have been able to speak with an age-appropriate level of fluency. Onset frequently coincides with the 'grammar burst': the point at which time children start to use syntax and grammatical morphemes to encode meaning (Bernstein Ratner, 1997), and this observation has led to the proposal that stuttering in young children stems from difficulty with syntactic formulation (Bernstein Ratner, 1997; Bloodstein, 2002; see also Anderson & Conture, 2004, for experimental evidence in support of this hypothesis). There are, however, also other developmental milestones that occur between 2 and 6 years of age could be implicated in the onset of stuttering, including the adoption an incremental form of phonological encoding

(Byrd, Conture, & Ohde, 2007), and the development of a more sophisticated use of auditory feedback to self-monitor speech (Yeni-Komshian, Chase, & Mobley, 1968).

The symptoms of early childhood stuttering differ somewhat from those of persistent stuttering. They more frequently include whole-word repetitions, prolongation of continuants and the excessive use of fillers such as ‘um’ and ‘er’. Blocking is rare, although it does occur. The use of force is also relatively rare, as are signs of tension, and most young children who stutter appear relatively relaxed while stuttering. However, there are exceptions (Schwartz, Zebrowski, & Conture, 1990; Yairi, 1983). Between six years of age and adolescence, approximately eighty percent of young children who stutter spontaneously recover (Andrews & Harris, 1964). In the twenty percent in whom stuttering persists, there is generally an increase in signs of conscious awareness of difficulty, tension, concomitant movements associated with forcing words out, and a variety of behavioural adaptations, including the avoidance of specific sounds, words, topics, speaking situations and people (Bloodstein & Bernstein Ratner, 2008). Incipient stuttering is more commonly associated with whole-word repetitions, whereas in persistent stuttering blocking and part-word repetitions on content words are generally more evident (Howell, Au-Yeung, & Sackin, 1999). However, whether these differences represent a change in symptoms over time within individuals, or whether persistent stutterers always had an underlying tendency to block and produce part-word repetitions remains unclear.

In a minority of children stuttering first starts after six years of age, and teenage onset is not unknown (Chang, Synnestvedt, Ostuni, & Ludlow, 2010). A scenario occasionally reported by such children is that they first experienced difficulty articulating a word while reading out loud in front of a class and have continued to stutter from that point onwards. In these children the condition appears to involve blocking and the use of force right from the start (cf. Van Riper, 's 1982 description of 'track III' aetiology).

Both incipient stuttering and persistent developmental stuttering are currently idiopathic conditions. Nevertheless, the findings from twin studies of concordance rates for stuttering ranging from 20 and 90 percent in identical twins, compared to 5 to 19 percent in non-identical twins (Yairi, Ambrose, & Cox, 1996) are indicative of a strong genetic component underlying it. Genetic linkage studies (e.g. Kang et al., 2010; Suresh et al., 2006) also support this conclusion. In a small, but nevertheless significant minority of cases, the onset of stuttering can be traced to some form of trauma or injury. Most commonly, such 'neurogenic' stuttering is associated with strokes and closed head injuries, and often co-occurs with aphasia and/or apraxia of speech. Although neurogenic stuttering is usually considered to be an acquired disorder that affects adults, analysis of data from children with no family history of the disorder led Alm and Risberg (2007) to propose that onset in young children as a result of injury or trauma may not be uncommon.

Despite being idiopathic, a large amount is known about factors that influence the severity of stuttering and the likelihood of stuttering-like disfluencies occurring. These factors are summarized in Section 2.4.

2.3. Primary and secondary symptoms of stuttering

It is customary to divide the symptoms of stuttering into those that are 'primary' insofar as they are central to the disorder and often present at its onset, and more peripheral 'secondary' symptoms that tend to develop later on, perhaps as adaptation strategies in response to the primary symptoms. Thus, for example, prolongations, repetitions and blocks have tended to be regarded as primary, whereas the use of force to 'get words out' and any of the 'strategies' commonly used to avoid the primary symptoms are considered to be secondary. This distinction is, however, problematic, insofar as psycholinguistic studies of the occurrence of repetitions and prolongations in normally-fluent speakers suggest that these may in fact be learned behaviours that help maintain listeners' attention and the speaker’s conversation turn at times when the speaker, for one reason or another, is not yet able to say the next word (Blackmer & Mitton, 1991; Howell & Sackin, 2001). If this latter account is accurate, then only blocks should be considered primary symptoms, and repetitions and prolongations are secondary to blocks. In young children who stutter, however, visible signs of blocking are relatively infrequent, (Johnson, 1959; Yairi & Ambrose, 2005), whereas repetitions and prolongations are usually apparent. However, it is possible that young children who stutter use repetition and prolongation as strategies to maintain their conversation turn and the listener’s attention while they are trying to formulate what they want to say. Thus in young children, prolongations and repetitions may be ‘secondary' to language formulation difficulty, whereas in older children and adults they more often appear to be secondary to difficulty initiating overt execution of a phoneme or word that has already been formulated (Bernstein Ratner, 1997; Bloodstein, 2001, 2006).

For many years, the development of early-childhood stuttering into a persistent disorder was thought to be associated with the development of secondary symptoms (Johnson, 1942). This belief was reflected in the tendency of early researchers to describe early childhood stuttering as ‘primary stuttering’ and persistent stuttering as ‘secondary’ stuttering (Bluemel, 1932; Froeschels, 1943; Glasner & Vermilyea, 1953). More recently, however, a longitudinal study by Yairi, Ambrose, Paden, and Throneburg (1996), which documented the appearance of symptoms from around the time of onset, found that, contrary to what had previously been thought, the appearance of secondary symptoms did not predict persistence.

2.3.1. *The role of volition*

Brutten (1970) proposed that the secondary symptoms of stuttering result from operant conditioning and are (at least potentially) under the speaker’s volitional control. This seems to accord with clinical experience that stutterers can suppress secondary symptoms if asked to do so and fits well with the conceptualization of secondary symptoms as strategies (albeit generally maladaptive ones) perhaps originally adopted as a way of coping with the primary symptoms or as a way of maintaining listener attention or ‘holding the floor’(Maclay & Osgood, 1959). Brutten posited that the primary symptoms, on the other hand, are not under volitional control, and that these are classically conditioned. The picture is, however, blurred by the likelihood that, that as time goes on, even symptoms that were originally operantly conditioned (and effectively began as volitional strategies) are likely to occur increasingly automatically as they become firmly established, and in real-life speaking situations in which cognitive demands are relatively high, speakers may no longer experience them to be under their conscious and/or volitional control. The question of the extent to which stutterers are able to control their stuttering is thus a complicated one. In clinical settings, stutterers often give the appearance of being able to speak fluently when making a conscious effort to. However, this ability seems to be situation-specific and transient and does not carry over into real life.

2.4. When and where does stuttering occur?

People who stutter do not do so every time they speak. Rather, the likelihood of stuttering-like disfluencies occurring varies widely and has been shown to be influenced by a seemingly diverse array of factors, both internal and environmental. Our knowledge of the factors that influence stuttering comes from two distinct sources: (1) stutterers’ own self-reports, and (2) observations made by researchers and clinicians. As highlighted in Section 2.1, self-reports tend to document the situations in which stutterers experience of loss of control and difficulty speaking fluently, whereas observer reports document the occurrence of stuttering-like disfluencies. These two sources of information do not always overlap.

2.4.1. *Evidence from experimental and corpus-based studies*

Studies that have experimentally varied the propositional content of specific words (Eisenson & Horowitz, 1945; Schlesinger, Forte, Fried, & Melkman, 1965) support stutterer's own self-reports that stuttering is most likely to occur when trying to convey information. An experimental study by Kassin and Bjerkan (1982), similarly found stuttering was greater on 'critical words' - i.e. words that communicated non-trivial, non-redundant information than on words with comparable complexity and frequency that were not critical in the context in which they were used. However, the finding that stutterers also experience significant difficulty speaking (meaningless) nonwords to a listener (Packman, Onslow, Coombes, & Goodwin, 2001), further suggests that the difficulty is not confined to conveying propositional information, but also arises when trying to convey phonological or phonetic information.

Studies that have investigated the loci of stuttering-like disfluencies have shown that they are most likely to occur (or likely to be most severe) at the beginning of words or phrases, (although some cases of stuttering characterized by final-phoneme disfluencies have been reported; e.g. McAllister & Kingston, 2005) and on utterances that are linguistically complex (Bernstein Ratner, 1995, 1997; Brown, 1937, 1945; Soderberg, 1967). Furthermore, in young children, they most commonly occur on the function words preceding content words (Howell et al., 1999), whereas in older children and adults they are more likely to occur on content words, especially low-frequency (Palen & Peterson, 1982) or low-predictability ones (Schlesinger et al., 1965), or on the function words immediately preceding them (Dworzynski, Howell, Au-Yeung, & Rommel, 2004; Howell & Au-Yeung, 1995).

Although the evidence cited above indicates that there is definitely a tendency for the loci of stuttering-like disfluencies to coincide with the loci of points of language formulation difficulty, other findings suggest that language formulation difficulty is neither a necessary nor a sufficient causal factor in precipitating instances of stuttering. Particularly relevant in this regard is the robust finding that the frequency of stuttering-like disfluencies is greatly reduced when white noise prevents people who stutter from hearing the sound of their own voice, or when their auditory feedback is subject to delay or frequency shift (for reviews of studies investingating these effects see Kalinowski, Armson, Stuart, & Gracco, 1993; Soderberg, 1969; Wingate, 1970). Importantly, the increased fluency under such altered auditory feedback conditions can occur without any accompanying reduction in speech rate or increase in speech-error rate.

2.4.2. *Evidence from stutterers' self-reports*

There has been relatively little recent systematic investigation of stutterers' self-reports of the moments when stuttering occurs. Therefore, much of the available data relating to self-reports is indirect, coming from clinicians' and researchers' accumulated experiences of interacting with stutterers (e.g. Sheehan, 1970; Van Riper, 1982). Much of it is also somewhat dated and of dubious validity. Nevertheless, a comprehensive investigation of stutterers' self-reports of factors that influence their stuttering was conducted by Bloodstein (1950a, 1950b), who analysed data from 204 stutterers, each of whom had been systematically interviewed and had completed a questionnaire of their experiences of stuttering in 115 different speaking situations[3]. The data from the questionnaires and interviews was then further interpreted by Bloodstein (1950a) as being reducible to six underlying factors which lead to a reduction in the likelihood of stuttering, which he described as follows:

---

[3] The questionnaire was based on Bloodstein's (1949) literature review of conditions under which stuttering is either reduced or absent.

- Reduced communicative responsibility;
- The absence of unfavourable listener reactions;
- No urgency to create a favourable impression;
- Changes in accustomed speech pattern;
- Speech accompanied by bodily activity;
- The presence of intense and unusual stimulation.

Bloodstein further suggested that these six conditions themselves can be seen as falling into two underlying categories: (a) Situations in which it matters relatively little whether the subject stutters or not; and (b) Situations that distract the attention of the speaker away from his speech. He further concluded that these two factors themselves both reflect conditions where stutterers are least likely to be consciously attempting to avoid nonfluency.

With respect to the first of Bloodstein's 6 conditions, it appears that if nobody is listening, or if the person who is listening and/or the message being conveyed is not considered to be important, stutterers can generally articulate whatever they wish to say with ease. Similarly, they experience little difficulty saying words that were redundant or highly predictable in the context even when such utterances are syntactically and motorically complex. On the other hand, if a message contains non-redundant information and the speaker attaches some importance to it, even if it just consists of simple monosyllabic words it can pose major obstacles. Thus, for example, stutterers are unlikely to have difficulty saying yes or no if they are simultaneously able to signal these responses by nodding or shaking their heads. However, in situations where this is not possible, such as when speaking over the telephone, such words can pose major difficulties (James, Brumfitt, & Cudd, 1999). Saying names, even high-frequency monosyllabic ones, can be particularly difficult probably also because they are generally unpredictable and cannot be substituted. Words are most likely to be stuttered on the first time they are used in an utterance or conversation, and once a word has been introduced (by either the stutterer or his conversation partner) the likelihood of stuttering-like disfluency occurring on it decreases. Stuttering is more likely to occur when the speaker is trying to convey a message to listeners who are likely to have difficulty understanding what is said, for example to listeners who are hard of hearing, who are not familiar with the speaker’s language, or who are not paying attention (See Bloodstein & Bernstein Ratner, 2008, Chapter 10, for a detailed review).

Adults who stutter are generally able to predict in advance with a high degree of accuracy which words they will stutter on (Knott, Johnson, & Webster, 1937), and many (but not all) stutterers have a list of phonemes and words that cause them particular difficulty. However, these phonemes and words differ from person to person (Griggs & Still, 1979; Hendel & Bloodstein, 1973).

Stuttering generally occurs at times when the speaker is most conscious of the importance of not stuttering, and rarely occurs during utterances made during outbursts of extreme anger or when the stutterers’ attention is strongly focussed on something other than his speech or when making casual spontaneous remarks (Bloodstein, 1950a). Stutterers do not self-report stuttering in their inner speech (Netsell, Bakker, & Ashley, 2010).

2.5. Summary

In this chapter we considered two contrasting definitions of stuttering: one based on subjective experiences; and the other on objective measurement. We discussed how these can lead to conflicting interpretations of what stuttering entails and whether or not an individual stutters. We then outlined how the symptoms of incipient stuttering in young children differ from those of people with persistent developmental stuttering and raised the possibility that incipient and persistent stuttering may be two distinct disorders with distinct aetiologies: incipient stuttering being more often associated with language-formulation difficulties and persistent stuttering with execution difficulties.

We then focussed on the moments of stuttering – on when and where it tends to occur, and described how accounts of the types of factor that influence the moment of stuttering differ quite considerably, depending on whether their source is objective (from experimental and clinical studies) or subjective (from stutterers’ own accounts). When considered together, these two sources suggest that stuttering is most likely to occur (a) when the message to be conveyed is complex; (b) at loci where a high degree of phonetic accuracy is essential for successful communication; (c) when the speaker attaches importance to conveying a specific verbal message; and (d) when he doubts that he will succeed.

3. Hypotheses that account for the occurrence and characteristics of stuttering-like disfluencies.

Hypotheses that have been proposed to account for the occurrence of stuttering-like disfluencies can usefully be categorised into two major types: (1) Production-impairment hypotheses, that attribute their occurrence to the (direct or indirect) consequence of impaired language and/or speech production mechanism; and (2) perception-based hypotheses, which posit that stuttering-like disfluencies result from inappropriate compensatory behaviours initiated in response to the perception that the quality of speech is too poor to allow it to adequately fulfil its function, or the anticipation that its quality will be too poor to allow it to do so. Key examples of these two categories of hypothesis are briefly outlined in Sections 3.1 and 3.2.

Section 3.3 then focuses in more detail the subset of (perception-based) hypotheses that posit that the symptoms of stuttering arise from speakers' maladaptive responses to the anticipation of some form of upcoming difficulty speaking. Section 3.4 goes on to highlight a number of important theoretical themes and issues that arise from the various types of hypothesis introduced in this chapter. Then finally, in Sections 3.5 and 3.6, the Covert Repair Hypothesis is described in detail, together with two more recent hypotheses that challenge it: the Vicious Circle Hypothesis (Vasić & Wijnen, 2005) and EXPLAN (Howell & Au-Yeung, 2002).

3.1. Production impairment hypotheses

The idea that there is some form of underlying structural or functional weakness that predisposes some people to stutter has been around in one form or another for many years. In the nineteenth century a common view was that stuttering was due to anatomical faults in the vocal apparatus, and surgery was sometimes carried to remedy this (Büchel & Sommer, 2004). More recently stuttering has been associated with impaired speech motor control and impaired language production. Experimental evidence relating to these hypotheses is briefly summarized below.

3.1.1. *Impairment of speech motor control*

The first empirical evidence suggesting that stuttering may stem from an underlying impairment of speech motor control stems from an observation, by West (1929), that stuttering-like disfluencies sometimes appeared in syphilis patients together with a general sluggishness of muscle movements, and resolved when the underlying infection was successfully treated. This observation led West to conduct a series of experiments comparing maximal diadochokinetic rates in stutterers and matched controls. He found that stutterers, as a group, were significantly slower than controls with respect to jaw and brow movements (West, 1929).

More recently a number of studies have confirmed that speech-motor movements are more variable and/or less well coordinated in both Adults (Caruso, Abbs, & Gracco, 1988; Max, Caruso, & Gracco, 2003; Zimmermann, 1980) and Children (Alpermann & Zückner, 2008; Riley & Riley, 1986) who stutter. Moreover, there is now robust evidence from studies that have investigated finger-tapping and flexion (Max et al., 2003; Zelaznik, Smith, Franz, & Ho, 1997), and clapping (Olander, Smith, & Zelaznik, 2010), that slowness or poor co-ordination is not restricted to the muscle-movements involved in speech production, thus countering the possible argument that stutterers' slower articulatory rates are a learned response to stuttering.

3.1.2. *Language-production impairment hypotheses*

Since the 1980s, interest in the close association of stuttering-like disfluencies with a number of linguistic factors (see Section 2.4.1) has stimulated the development of a number of language-production impairment hypotheses (Karniol, 1995; Perkins et al., 1991; Postma

& Kolk, 1993; Wingate, 1988) and related research examining the possibility that language-formulation impairment could play an important and perhaps key role, over and above any limitations due to impairment of speech motor control, in limiting the capacity of stutterers to produce fluent speech. Currently, most of the evidence suggesting a link between language-production impairment and stuttering is correlational (see Sections 2.2 and 2.4.1) and does not provide any insights into causality. However, there is also some experimental evidence, from priming and rhyme judgement studies, of slow language formulation in people who stutter. This evidence is outlined in the literature review of the CRH (see Section 3.5.2).

Currently, the most influential psycholinguistic impairment-based hypothesis of stuttering is Postma and Kolk's Covert Repair Hypothesis (Kolk & Postma, 1997; Postma & Kolk, 1993). Its main strength over earlier psycholinguistic hypotheses is that it offers a highly-specified account of a potential mechanism that could potentially explain the full range of stuttering-like disfluencies (part-word and word repetitions, prolongations of continuants, and 'blocks'). The main tenets of the CRH and evidence for and against the hypothesis are presented in detail in Section 3.5. One of the main competitors to the CRH, The EXPLAN hypothesis (Howell & Au-Yeung, 2002), can also be classified as a language-production impairment hypothesis, insofar as, like the CRH, it is also predicated on the presumption that language formulation is abnormally slow in people who stutter. The EXPLAN hypothesis is described in Section 3.6.2.

3.1.3. *The Capacities and Demands Framework*

When considering whether stuttering can be accounted for in terms of an underlying language or speech-production impairment, it is important to remember that people who stutter do not stutter all of the time. Indeed, the majority of words and phrases spoken by the majority of stutterers appear, perceptually at least, to be as fluent as those produced by normally-fluent speakers. This intermittent nature of the symptoms of stuttering suggests that, if stuttering is caused by an underlying language or speech-production impairment, the impairment itself may be relatively mild, such that in many circumstances it does not cause a problem. Reflecting this conclusion, Starkweather (1987) proposed that it is most useful to consider stuttering in relation to a ‘Capacities and Demands’ framework, whereby stuttering-like disfluencies arise only when the overall rate or quality of speech production falls short of that which is required by the speaking situation. In situations where the demands are low, nearly everyone can produce fluent speech, whereas when demands are high, such as when discussing unfamiliar concepts in an unfamiliar language, nearly everyone will be disfluent.

The Capacities and Demands framework highlights the possibility that, in different individuals, different underlying impairments may be responsible for the breakdown of fluency, and that the exact nature of the impairment may not be as important as its net effect on the speed or accuracy with which speech can be produced. Thus, with respect to speech-production ability, people who stutter may fall towards the lower end of the spectrum for any of a number of underlying reasons, and may thus relatively frequently find themselves confronted by situations where the situational demands exceed their capacities.

3.2. Perception-based hypotheses

A number of major theories of stuttering can be classified as perception-based, insofar as they are predicated on the notion that the speaker perceives that his speech in some way falls short (or will fall short) of the standards required and, as a result, makes unnecessary adjustments or changes that result in stuttering. Of these, perhaps the best known and most influential has been the Diagnosogenic Theory (Johnson, 1942, 1959), outlined below.

3.2.1. *Diagnosogenic theory*

Analyses by Johnson (1942, 1959) of the frequency and severity of stuttering-like disfluencies in speech samples from young children diagnosed as stutterers consistently revealed substantial overlap with equivalent samples from age-matched normally-fluent controls. The existence of this overlap prompted Johnson to claim that there is no reliable difference between the disfluencies of young children who are diagnosed as stutterers and those who are not, and subsequently led to his formulation of the ‘Diagnosogenic Theory’. According to this theory, stuttering (the condition) starts as a purely cognitive disorder characterised by the (false) belief that one’s disfluencies are unacceptable. Johnson’s often quoted statement that “Stuttering begins not in the child’s mouth, but in the parent’s ear” (cited in Bloodstein & Bernstein Ratner, 2008, p. 58) refers to this false belief. Diagnosogenic theory identifies two stages in the development of the condition: First, the parents (or other primary care-givers) perceive the child’s (normal) disfluencies as in some way pathological and react to them accordingly; second, the child’s repeated experience of these negative reactions leads him to attempt to prevent himself producing disfluencies by exerting inappropriate effort to control his speech. His efforts to control his speech are counterproductive insofar as they result in an increase (rather than a decrease) in disfluencies. Diagnosogenic theory in its original form has been largely rejected on the basis of evidence that children who stutter produce significantly more disfluencies than normally-fluent children and that their disfluencies are different in character right from the start (Ambrose & Yairi, 1999; Yairi & Lewis, 1984) and the failure of the majority of studies than have investigated the relationship between parenting styles and stuttering to find any clear evidence that these factors are in any way linked (see Nippold & Rudzinski, 1995 for a review). However, the underlying idea, that trying to avoid, remove, or ‘repair’ disfluencies increases them, has been retained in a number of more recent hypotheses, the most relevant to the current thesis being Vasić and Wijnen’s (2005) ‘Vicious Circle Hypothesis’; a revised version of the Covert Repair Hypothesis, which draws from Diagnosogenic theory, insofar as it posits that stutterers’ earlier experiences lead them to monitor hyper-vigilantly and perceive and attempt to repair errors which are, in reality, not present. This hypothesis is covered in detail in Section 3.5.

3.2.2. ***Deficits in feedback and feedforward control***

A number of recent studies using minimal movement paradigms have found that compared to normally-fluent speakers, stutterers tend to rely more heavily on visual feedback for fine motor control. These findings suggest that poorly developed fine motor control skills in stutterers may sometimes result from an underlying sensory (kinaesthetic and/or proprioceptive) impairment (Loucks & De Nil, 2006a, 2006b; Loucks, De Nil, & Sasisekaran, 2007). It has, however, long been recognised that adequate control over the coordination of fast and complex muscle movements, including those required for speech, cannot be achieved through afferent (sensory) feedback alone (von Holst & Mittelstaedt, 1950/1973). This is because such feedback arrives too late to enable effective online corrections and adjustments.

A more plausible account is that motor coordination is achieved largely through ‘feedforward’ control whereby efference copy from motor commands is used to predict the sensory consequences of planned movements before the actual sensory consequences of such movements become available (von Holst & Mittelstaedt, 1950/1973). These predicted sensory consequences can then be compared with those that would be expected to arise from the intended movements, and if discrepancies are found, suitable online adjustments can be made. This type of feedforward regulation of muscle movements depends upon the availability of accurate internal models of the relationship between centrally generated motor commands and the sensory consequences of the resulting movements.

Max, Guenther, Gracco, Ghosh, and Wallace (2004), have proposed that stutterers have difficulty activating and/or maintaining such internal models. As a result they come to rely excessively on (slow and inefficient) ‘closed loop’ control via afferent (sensory) feedback, and as a result, they lack sufficient control over the coordination of the fast muscle movements required the production of speech. This (Max et al., 2004) hypothesis fits well with the experimental findings of excessive reliance on visual feedback in the minimal-movement paradigm experiments cited above. It is also supported by evidence from functional imaging studies that have, amongst other things, consistently found evidence of over-activation of the vermal region of cerebellum in people who stutter (Brown, Ingham, Ingham, Laird, & Fox, 2005), which is the region of the brain believed to be associated with the maintenance of internal models (Wolpert, Miall, & Kawato, 1998). Furthermore, computer modelling of speech production using the ‘DIVA’ model (Guenther, Ghosh, & Tourville, 2006) has demonstrated that an increased reliance on afferent feedback, rather than feedforward control, results in syllable and phoneme repetitions, and possibly also other disfluencies, similar to the stuttering-like disfluencies that characterise stuttered speech (Civier, Tasko, & Guenther, 2010).

Although the findings from motor-control studies are robust, it should be noted that in all such studies, some of the stuttering participants fall within the normal range for speed and accuracy, whereas some non-stutterers fall short of it. The presence of such group overlap implies that sluggish or inaccurate muscle movements may predispose to stuttering but cannot be considered a necessary cause (McClean, Tasko, & Runyan, 2004).

3.3. Anticipatory struggle hypotheses

The term “Anticipatory Struggle” was first used by Bloodstein (1958) to describe a range of hypotheses which share the common idea that stuttering occurs because “stutterers interfere in some manner with the way they are talking because of their belief that speaking is difficult” (Bloodstein & Ratner 2005, p43). The hypotheses vary in their accounts of how such a belief comes about and the details of the mechanisms by which stutterers’ responses to this belief manifest in the symptoms of stuttering. For example, it has been proposed that a child may come to believe that speaking is difficult, either because he finds it difficult to fulfil the unreasonably high expectations of his parents (Johnson, 1942), or because of a combination of factors, including high parental expectations, underlying language or speech impairments and a difficult or competitive speaking environment (Bloodstein, 1958, 1975; Johnson, 1959).

An important strength common to all of the various formulations of anticipatory struggle hypotheses is that they provide a parsimonious account for the onset of symptoms of stuttering even at times when actual demands do not exceed capacity, insofar as they imply that it is the *perception* of demands exceeding capacity that is important (rather than the reality).

Arguably, a weakness common to all the various formulations of anticipatory struggle hypothesis, is that the details of the mechanism by which the anticipation of struggle accounts for primary symptoms of stuttering, such as ‘tension and fragmentation (Bloodstein, 1975), the preparatory set (Van Riper, 1971), and conflicting drives (Sheehan, 1953) are rather vague, and current formulations of the anticipatory struggle hypothesis all fail to provide adequate explanations for precisely why the various symptoms of stuttering (repetitions, prolongations and blocks) occur in the way that they do. In contrast, the psycholinguistic hypotheses outlined in Section 3.1.1, that attribute stuttering to covert error repair or strategies to maintain the conversation turn, are far more successful at providing plausible, detailed accounts of how these symptoms come about.

3.4. Further key theoretical issues

3.4.1. *The Continuity Question*

A clinically relevant question that is often posed, and upon which researchers and clinicians still disagree, is the extent to which the disfluencies of stuttering arise in the same way (and for the same reasons) as do the normal disfluencies in normally-fluent speakers. Johnson’s (1942) Diagnosogenic theory is predicated on the belief that, in young children, stuttering-like disfluencies form a continuum of severity and there is no natural line of demarcation between what is and what is not stuttering. According to the theory, stuttering becomes a discrete condition when the speaker develops a fear of stuttering and then begins to make conscious efforts to avoid it. Diagnosogenic theory thus defines stuttering as a discrete condition characterised by the presence of the fear of stuttering.

An alternative view was proposed by Bloodstein (1975). Like Johnson, Bloodstein also proposed that the frequency of stuttering-like disfluencies in young children is best described as a continuum. However, Bloodstein did not equate stuttering with the acquisition of the fear of stuttering. Instead he suggested stuttering is simply a label used to describe “a more extreme degree of certain specific types of disfluency” (Bloodstein 1975, P49) in a similar way to how ‘fat’ is a label used to describe individuals on the heavy end of a continuum of body weight. In support of this ‘Continuity Hypothesis’, Bloodstein (1975) noted that, qualitatively, the stuttering-like disfluencies of ‘normally fluent’ children are not only similar to those of stuttering children, insofar as they sound alike and are likely to arise in the same places in an utterance, but also, that they too may be accompanied by signs of tension, fragmentation and struggle. According to Bloodstein’s Continuity Hypothesis, everyone stutters to some degree, and the identification of some people as stutterers and some as fluent speakers involves the imposition of an artificial and arbitrary point of division on what is in reality a pure continuum.

Since Johnson and Associates’ (1959) publication of their studies of symptoms associated with the onset of stuttering, a significant amount of research has been devoted to the identification of qualitative differences between the stuttering-like disfluencies of young children whose stuttering persists and those whose stuttering resolves. However, so far no reliable language or speech-related predictors of persistence have been found (Watkins & Yairi, 1997; Yairi & Ambrose, 2005; Yairi, Ambrose, Paden et al., 1996) and so the question of whether stuttering in young children constitutes a discrete disorder remains unresolved.

With respect to persistent stuttering in older children and adults, there is a more general agreement that it is a discrete disorder. This has probably been brought about largely as a result of the much clearer dichotomy between people who consider themselves as stutterers and those who do not. However, it remains unclear exactly how this dichotomy first arises. It may be related to the experience of blocking and the related loss of control, as outlined in Section 2.1. Certainly on the basis of stutterers' descriptions of the experience of stuttering, most researchers would agree that all persistent stutterers block occasionally, or at least would block were it not for the tendency to avoid words upon which they feel a block is likely to occur. Whether or not non-stutterers ever experience blocking in the same way is less clear. Unfortunately, systematic experimental research into the prevalence of blocking among people who do and do not stutter is lacking.

In recent years, the trend among speech therapists and researchers has been to reject the continuity hypothesis completely and to consider stuttering as being a discrete pathological condition that can exist in both young children and adults, and which may or may not be discernable from the percentage of stuttering-like disfluencies present in the speaker's overt speech. As a generally accepted rule of thumb, stuttering (the condition) is diagnosed in individuals where the percentage of stuttering-like disfluencies is three or above (e.g. Guitar, 2006; Ward, 2006). There are good pragmatic reasons for this, insofar as perceptually, it is approximately at this frequency that speech begins to sound disfluent to listeners (Yairi, 1981). However, for research which aims to explore the mechanisms behind stuttering-like disfluencies there is a strong argument for defining stuttering (the condition) in terms of the subjective experience of loss of control over the articulators, irrespective of the frequency of stuttering-like disfluencies. This is the operational definition of stuttering that we have adopted in the experiments conducted as a part of this thesis.

3.4.2. *Vicious circles*

This concept of stuttering being some sort of a vicious circle whereby a normal regulatory mechanism breaks down due to being pushed beyond its normal limits is a recurrent one. Johnson's Diagnosogenic theory (Johnson, 1942, 1959) in which the symptoms of stuttering were themselves the manifestation of a child's attempts to avoid stuttering implies a type of vicious circle, as does Bloodstein's (1958) conceptualization of the moment of stuttering as a 'tension and fragmentation' in which the speaker tries to avoid difficulty speaking by adopting a more difficult strategy to the one he would normally use. More recently, Vasić and Wijnen (2005) named their modified version of the CRH, 'The Vicious Circle Hypothesis', because if stutterers attempt to repair timing or prosodic irregularities, their repair attempts will themselves create further timing and prosodic irregularities, which will then prompt further repair attempts, and so on.

An important strength of these various vicious-circle hypotheses is their potential ability to account for why stuttering may persist beyond the time when any underlying impairment that originally caused it may have resolved. Such hypotheses also provide a potential resolution to the question of what distinguishes a stutterer from a non-stutterer (i.e. the continuity question) insofar as, although there may well be a continuum of frequency of stuttering-like disfluencies in individuals, there may well also be a critical frequency, beyond which a vicious circle is created, thus explaining how a dichotomy can arise out of continuum (cf. Dynamical Systems Theory; Van Lieshout, 2004).

3.4.3. *The differences between incipient and persistent stuttering*

A truly comprehensive theory of stuttering should be able to account for why it occurs in some people and not in others, why its symptoms manifest in the way they do and why about 80 percent of young children who stutter recover, whereas 20 percent do not. With respect to the first of these questions, it is clear, from the high concordance rates in identical twins compared to non-identical twins, that the majority of persistent stutterers have inherited a genetic predisposition to stuttering. This implies the presence, in at least some individuals with developmental stuttering, of structural differences that predispose to the condition. Such structural differences may be temporary, perhaps amounting to little more than developmental delays, and it is not yet possible to say to what extent the structural differences found in brain imaging studies of adults who stutter are the result of such underlying predispositions, and to what extent they represent changes resulting from a lifetime of stuttering. The fact that concordance rates in identical twins are less than 100 percent implies that environmental factors must also play a key role in determining whether or not stuttering manifests in an individual, and whether or not it persists. Such factors could include both neurological injuries and learning. The extent to which learning is implicated, and whether maladaptive learning may be responsible for the persistence of stuttering after any underlying impairments have resolved, remains unclear. Thus although, on the basis of current knowledge, it seems unlikely that purely perceptual hypotheses (such as the Diagnosogenic theory) can account for the onset of stuttering, it remains possible that beliefs, perceptions and associated patterns of behaviour learned in childhood may at least sometimes account for the persistence of stuttering after any underlying structural impairment has resolved (the 'Gone but not Forgotten Hypothesis'; Conture, Zackheim, Anderson, & Pellowski, 2004).

The dilemma facing the production-impairment hypotheses like the CRH and EXPLAN (described in detail in Sections 3.5 & 3.6, below) is that although there is substantial evidence of language impairment in young children who stutter, with the exception of a small number of individuals whose stuttering results from neurological injury, there is very little evidence of any significant degree of language impairment in adults who stutter. However, despite this lack of evidence, covert-repair and/or restart mechanisms, both of which are predicated on the notion of impaired language formulation, still provide the most highly specified and yet parsimonious explanations for the specific attributes of stuttering-like disfluencies in adults who stutter.

3.5. The Covert Repair Hypothesis

3.5.1. Overview

The Covert Repair Hypothesis (CRH; Postma & Kolk, 1993) is essentially a language-production impairment hypothesis, and as noted in Section 3.1.2, its main strength over earlier language-production impairment hypothesis is that it offers a highly-specified account of a potential mechanism that could potentially explain the full range of stuttering-like disfluencies (part-word and word repetitions, prolongations of continuants, and ‘blocks’). It is based on Levelt’s (1983, 1989) model of language production (Levelt’s model is described in detail in Chapter 4, Section 4.2). The CRH equates stuttering-like disfluencies in both normal and stuttered speech, with speakers’ attempts to repair phonological-encoding errors before the onset of overt articulation. Central to the CRH are the notions that speakers plan their utterances in advance, and that they are able to detect errors in those speech plans while they are temporarily stored in an articulatory buffer, awaiting overt execution. Detection of such errors is possible because the speaker can inspect these plans through an internal monitoring loop (roughly equivalent to inner speech).

Covert error repair involves a number of steps, including perceiving the error in the plan, cancelling the plan, retracing to a suitable point and formulating a new plan. Cancellation and reformulation may result in varying degrees of disruption to the forward flow of overt speech depending on where in the speech plan the error is located, and how long the speech plan is available in the buffer for monitoring before its overt execution commences. Thus, for example, sometimes the first segment or syllable of the plan may have already been overtly articulated before it is cancelled and reformulated. If this happens, a segment or syllable repetition or prolongation may occur. Several consecutive repetitions may occur in cases where a number of reformulations of the plan are needed before the correction is achieved. Repetition of continuants may occur without breaks in between, producing symptoms of prolongation rather than repetition. Alternatively, if the plan is cancelled before the onset of speech a block may ensue until the reformulated plan becomes available. In this way, the CRH accounts for the three main types of stuttering-like disfluency: repetitions, prolongations and blocks.

Because covert repairs of syntactic and lexical errors are more likely to be associated with larger retraces (Nooteboom, 1980), Postma and Kolk (1993) proposed that the part-word repetitions characteristic of persistent developmental stuttering most likely stem from repairs of errors of phonological encoding. However, they did not rule out the possibility that covert syntactic or lexical error repairs may also result in the production of stuttering-like disfluencies. A summary of how different types of disfluency may arise from different types of covert error repair is presented in Table 3-1.

Postma and Kolk (1993) proposed that the disfluencies that arise in the speech of both stutterers and normally-fluent speakers can be accounted for in terms of covert repair activity. The difference between stutterers and normally-fluent speakers is simply that stutterers make many more errors of phonological encoding and are thus more disfluent. Postma and Kolk further proposed that the reason stutterers make greater numbers of errors is because their phonological encoding proceeds relatively slowly, and as a result, at normal speech rates target phonemes are not sufficiently activated to ensure that they are successfully incorporated into the speech plan. This is explained in terms of Dell’s (1986) computer simulation of language production In Dell’s model, the activations of units (words and phonemes) gradually increase until they exceed those of any competing units. When overt speech is initiated, the units that happen to be most highly activated at that point of time are selected. Thus, the slower the speed of encoding, the greater are the chances that competing units will be selected in error. (This process is described in more detail in Chapter 4, Section 4.3).

The CRH thus rests on two basic tenets: (1) the speech plans of PWS contain abnormally high numbers of phonological-encoding errors; and (2) the covert repair of such errors accounts for the characteristically high numbers of Stuttering-like disfluencies (SLDs) in their utterances. However, since Postma and Kolk's (1993) formulation of the CRH, no direct evidence has been produced to confirm these tenets.

Table 3-1 Disfluencies presumed to result from differing types of covert error repair Adapted from Postma and Kolk (1993)

Internal Error	Covert Repair Restart strategy	Resulting disfluency
Semantic/syntactic error	Restart the phrase	Phrase repetition
Lexical error	Restart previous word	Word repetition
Phonemic error	Restart the syllable or phoneme	Blocks, prolongations and syllable or phoneme repetitions.
	Postponement strategy	
Semantic/syntactic/lexical error	Hold execution, reformulate	Silent pause (>200ms)
Phonemic error	Prolong current sound until proper continuation found	Prolongation of syllable non-initial sounds
Phonemic error	Hold execution of next sound until proper continuation found	Blocking in the midst of a syllable (broken words)

3.5.2. *Review of evidence for the CRH*

A weakness endemic to many of the experimental studies that purport to investigate language encoding in people who stutter is that the responses they measure are often dependent upon motoric as well as linguistic factors, and are therefore confounded. For example, a number of studies have compared the spontaneous utterances of children who stutter (CWS) and age-matched controls and have found evidence consistent with an interpretation of delayed/disordered phonological development or weaker encoding abilities in CWS (see Nippold, 1990; Nippold, 2001 for reviews). However spontaneous speech invariably reflects the combined contributions of a variety of (linguistic and motor) factors, so such studies are unable to pinpoint exactly what it is that gives rise to such performance differences. In a similar way, the finding that PWS are often slower than age-matched controls at picture naming tasks (e.g. Bernstein Ratner, Newman, & Strekas, 2009; Newman & Bernstein Ratner, 2007), may reflect either slower lexical access or slower motor responses.

There are similar difficulties with many of the studies that have focussed on the distribution of stuttering events. For example, stuttering is more likely to occur on more complex grammatical structures (Bernstein Ratner & Sih, 1987; Kadihanifi & Howell, 1992; Logan & Conture, 1997; Melnick & Conture, 2000; Yaruss, 1999); on lower-frequency words (Anderson, 2007; Hubbard & Prins, 1994; Newman & Bernstein Ratner, 2007; Palen & Peterson, 1982); in association with (non-systematic) errors of phonological encoding (Yaruss & Conture, 1996); and on wrongly articulated word-initial consonant clusters (Wolk, Blomgren, & Smith, 2000). But since articulatory complexity tends to co-vary with these linguistic factors, the causal picture remains unclear.

The extent of the difficulties associated with distinguishing poor performance due to linguistic impairment from poor performance due to impaired speech motor control, has meant that there are, in fact, very few studies that can be considered to contribute truly reliable evidence of the language encoding abilities of people who stutter. Evidence that arguably is relatively reliable comes from a series of priming and phoneme-monitoring studies, outlined below, all of which control for any possible influence of articulatory factors on response latencies through the use of repeated-measures paradigms in which the motor responses required of participants remain identical across all conditions.

Syntactic encoding

Two sentence-structure priming studies have produced evidence of slow syntactic encoding. In these studies, children (Anderson & Conture, 2004) and adults (Tsiamtsiouris &

Cairns, 2009) described action pictures immediately after hearing tape-recordings of utterances containing syntactic structures that were either similar or different to those required for their own utterances. In both studies, compared to controls, speech-onset latencies of participants who stutter were longer in the absence of priming but the difference between the stutterers and controls reduced significantly when syntactically similar primes were used, suggesting that syntactic encoding is slower and perhaps less robust in both adults and children who stutter.

Lexical access

Priming studies comparing the effect of lexical primes on lexical access in stutterers and controls have produced mixed results. In one study, Pellowski and Conture (2005) found that when young CWS and age-matched controls performed a picture-naming task in which semantically related or unrelated words were auditorily presented just prior to picture presentation, mean speech-onset latencies of controls were reduced by semantic primes, whereas those of CWS increased.

However, in a further study, comparing the effects of different types of lexical/semantic primes, Hartfield and Conture (2006) found that, in CWS, primes functionally related to target words interfered significantly less than those that were physically related. Hartfield and Conture (2006) concluded that these results, taken together, suggest that preschool CWS differ from controls in the speed and nature of lexical retrieval.

Using a similar picture-naming paradigm, Hennessey Nang and Beilby (2008) compared the effects on adults who stutter (AWS) and age-matched controls of a variety of semantic and phonological primes, but failed to find any differences between the two participant groups: In both, semantically related primes resulted in longer naming-onset latencies. (An interference effect is normal for semantically-related primes (Schriefers, Meyer, & Levelt, 1990)).

The findings of a number of dual-task studies suggest that group differences with respect to lexical-semantic encoding may exist, but may only become apparent under conditions of high cognitive load. In one such study that involved self-paced silent-reading (Bosshardt & Fransen, 1996), participants who stutter and controls read words and simultaneously compared them with a probe word that either rhymed, shared the same lexical category, or was identical to the target word. Bosshardt and Fransen (1996) found that compared to controls, participants who stutter were equally fast to read words in the 'identical probe' condition, and not significantly different in the rhyme judgement condition, but they took significantly longer, compared to controls in the condition that involved identifying lexical-category matches. Bosshardt and Fransen interpreted this as indicating that the stutterers were slower than controls to access semantic information (cf. Bosshardt, Ballmer, & De Nil, 2002).

Phonological encoding

Phoneme priming studies that have been carried out on children and adults who stutter have consistently found their responses to be slower compared to controls' responses. However, the actual priming effects have been mixed. Melnick, Conture and Ohde (2003) found that both CWS and age-matched controls similarly exhibited shorter picture naming latencies when phonologically-related (CV) primes were played prior to picture presentation, suggesting that the two groups' phonological encoding abilities were comparable. In adults who stutter, the results of the phonological manipulation of the Hennessey et al. (2008) study mirrored those of Melnick et al. (2003). In a priming study which compared the effect on speech-onset latencies of initial consonant (C), and consonant-vowel (CV) primes, Wijnen and Boers (1994) found that the two groups differed insofar as AWS only benefited from CV primes, whereas controls benefitted from both C and CV primes. However, Burger and Wijnen's (1999) larger-scale rerun of same paradigm failed to reproduce these findings.

Despite the mixed results from phonological priming studies outlined above, some evidence of phonological encoding differences between stutterers and controls has been found. A Priming study of normally-fluent children by Brooks and MacWhinney (2000) has provided evidence that, at an early age, their response latencies are sensitive only to ‘holistic’ phonological primes (consisting of larger units), whereas by six years of age their response latencies are sensitive to primes consisting of single phonemes. These findings have been interpreted as indicating a shift from holistic to segmental phonological processing between four and six years of age. Byrd, Conture, and Ohde (2007) used the same paradigm to investigate the sensitivity of young children who stutter to segmental and holistic primes. They found that, at six years of age, CWS continued to be responsive only to holistic primes, whereas six-year-old controls were responsive to segmental primes. They interpreted this finding as suggesting that CWS are slower to adopt the use of segmental phonology, and instead continue to code words holistically.

In adults who stutter, two well controlled phoneme-monitoring studies (Sasisekaran & De Nil, 2006; Sasisekaran, De Nil, Smyth, & Johnson, 2006) found that, compared to controls, stuttering participants were significantly slower to identify phonemes in words formulated in their inner speech. Importantly, there were no differences between the two groups’ performances with respect to auditory monitoring (of pure tones); picture naming; simple motor responses; or identifying phonemes when listening to tape recordings of the same words. These findings thus strongly suggest that the stutterers’ slower responses on the phoneme monitoring task stemmed from impaired phonological encoding and not from any general monitoring impairment or slow motor responses.

Finally, compared to normally fluent speakers, both CWS (Anderson, Wagovich, & Hall, 2006; Hakim & Bernstein Ratner, 2004) and AWS (Ludlow, Siren, & Zikira, 1997) have been found to be poorer at nonword repetition, although the extent to which poor performance on nonword repetition tasks may stem from impairment of phonological encoding remains somewhat unclear.

The frequency of phonological encoding errors in inner speech

To date, there has only been one study published (Postma & Kolk, 1992) that has investigated the frequency of errors in the inner speech of stutterers. In this study, participants were instructed to repeat a series of CV and VC strings aloud, both with and without auditory masking, and to press a button each time they noticed themselves making an error. It was presumed that under conditions of auditory masking, participants would have had little choice but to rely on internal monitoring of the speech plan for error detection (although the possibility remains that some errors may have been detected through monitoring of kinaesthetic or other forms of feedback). In both the normal speech and noise-masked conditions, although the PWS group reported a numerically higher proportion of phonemic errors than the controls, the difference was not statistically significant. However, the PWS group recited the strings more slowly, which may have enabled them to avoid errors which would have been manifest at a faster speech rate. Moreover, participants were not required to describe the errors that corresponded to button presses; effectively, the experimenter had to guess what errors had been perceived.

3.5.3. *Summary*

The experimental studies outlined above that have investigated the CRH have produced evidence to suggest that, compared to matched controls, syntactic encoding is slower in both adults and children who stutter. With respect to lexical access and phonological encoding, however, the evidence is mixed and somewhat inconclusive. There is no evidence, as yet, of a tendency for stutterers to make more errors in inner speech.

3.6. Challenges to the CRH

Despite its appeal, the Covert Repair Hypothesis does not account well for all of the symptoms of stuttering, such as why stutterers frequently have so much difficulty saying their own name and why stuttering is so sensitive to the status of the listener (See Section 2.4.2). Moreover, a particular difficulty with the CRH is that stutterers do not generally report being aware of large numbers of errors in their inner speech and do not equate stuttering with difficulty formulating words in their inner speech. Howell and Au Yeung (2002) criticised the Covert Repair Hypothesis on two counts: firstly, on the basis of parsimony, stuttering-like disfluencies can be adequately accounted for simply in terms of 'fluency failures' stemming from slow phonological processing without any need to additionally invoke a monitoring and repair process; and secondly, on the basis that it is impossible to test for the presence of covert error repairs, because covert repairs are, by their nature 'covert'.

The lack of evidence of any correlation between phonological encoding errors and stuttering-like disfluencies has led to the development of two alternative psycholinguistic hypotheses of stuttering: the "Vicious Circle Hypothesis" (Vasić & Wijnen, 2005) and EXPLAN (Howell & Au-Yeung, 2002), both of which are described below.

3.6.1. ***The Vicious Circle Hypothesis***

The Vicious Circle Hypothesis (Vasić & Wijnen, 2005) retains the CRH's basic tenet that SLDs arise as the by-products of covert error repair. However, it proposes that, rather than making different numbers of inner-speech errors, the difference between stutterers and normally fluent speakers is that stutterers monitor their (inner and overt) speech more vigilantly for errors, have a lower threshold for instigating repairs and tend to focus their attention excessively on speech timing (cf. Sherrard, 1975, for a similar proposal). Stutterers therefore perceive, and attempt to "repair", many minor timing irregularities that normally-fluent speakers would not be concerned about. A further consequence of such hyper-vigilant monitoring is that disfluencies resulting from the error-repair process are themselves likely to be identified as errors, triggering further (unnecessary) reformulations of the speech-plan and leading to a “vicious circle”.

The Vicious Circle Hypothesis also incorporates elements of Diagnosogenic theory (Johnson, 1942), insofar as Vasić and Wijnen (2005) speculated that some stutterers may have developed a tendency toward hyper-vigilant monitoring during childhood because someone drew their attention to their frequent disfluencies. Alternatively, they suggest that abnormal monitoring may have first arisen as a side-effect of imbalances between the maturation of their language development and speech production abilities in early childhood. A further possibility is that such hyper-vigilance may stem from the repeated experience of communication failure, the reasons for which may differ from individual to individual, as proposed in Section 3.1.3. Whatever the case, Vasić and Wijnen’s (2005) Vicious Circle Hypothesis has the potential to account for a much wider array of symptoms than the Covert Repair Hypothesis, particularly if hyper-vigilance is situation specific (although the Vicious Circle Hypothesis does not elaborate on this possibility).

Currently, experimental evidence in support of the Vicious Circle Hypothesis is limited. A number of dual-task experiments have investigated the effects of secondary tasks designed to distract attention away from monitoring their speech for errors, a procedure that should theoretically reduce cognitive resources available for speech-error monitoring and thus reduce the amount of covert repair activity speakers engage in. As a result, all other things remaining equal, if monitoring was previously hyper-vigilant, speakers’ disfluency rates should reduce and their overt error rates should remain stable, whereas if monitoring was not hyper-vigilant, disfluency rates should reduce but overt-error rates should increase. Findings, however, have been mixed. Arends, Povel and Kolk (1988) found that a visual-tracking task decreased disfluencies but only when the speaking task was highly demanding.

Two dual-task studies by Bosshardt (1999, 2002), both of which employed secondary linguistic task (mental calculation and a word-repetition tasks respectively) both found that the secondary tasks increased disfluencies. Vasić and Wijnen (2005) designed a dual-task study that systematically examined the different effects of linguistic and non-linguistic dual tasks. They found that stuttering was reduced compared to a baseline condition when stutterers' attention was distracted away from their speech by performing a concurrent non-linguistic task: playing 'Pong', a computer table-tennis game. It was also reduced when the focus of their attention was shifted onto a different aspect of speech by a secondary linguistic task: monitoring for occurrences of the (Dutch) pronoun 'die' in their speech. Furthermore, analysis of the relative proportions of various types of stuttering-like disfluencies for each of these conditions revealed that, compared to baseline, the proportion of 'blocks' was significantly reduced in both the Pong and 'die monitoring' conditions, with the magnitude of the reduction being significantly greater in the 'die' monitoring task. In contrast, whole-word and word-string repetitions increased in the *die*-monitoring condition compared to all other conditions. Vasić and Wijnen interpreted these findings as the net effects of two different processes: (1) overall, a reduction in monitoring vigilance, resulting in fewer blocks; and (2) in the *die*-monitoring condition, an increase in task difficulty, resulting in more repetitions. It is implied that the reduction in blocking resulted from a reduction in maladaptive covert error repairs, whereas the increase in repetitions may have resulted from the normal workings of an EXPLAN-like restart mechanism (See Section 3.6.2) in response to the increased burden on speech planning. This conclusion was further supported by the finding that the Control group also produced a similar increase in repetitions in the *die*-monitoring task. Unfortunately, although the dual-task studies outlined above documented disfluencies in detail, none of them have documented the effect of the secondary task on error rates.

In a magnitude estimation study, in which both stutterers and normally-fluent controls rated recordings of fluent speech that had been produced by stutterers and normally-fluent controls, Lickley, Hartsuiker, Corley, Russell, and Nelson (2005) found that, compared to controls, stutterers rated all recordings as more disfluent, and both stuttering and fluent raters rated recordings of stutterers' perceptually fluent speech as more disfluent than that of non-stutterers. These findings led them to propose that, (a) compared to normally-fluent speakers, stutterers monitored speech for disfluencies more vigilantly; and (b) that the perceptually-fluent speech of stutterers is in fact less fluent than that of normally-fluent speakers.

In a recent study, Arnstein, Lakey, Compton, and Kleinow (2011) used event-related potentials (ERPs) to compare error-monitoring in stutterers and normally-fluent controls, while performing a number of rhyme-judgement tasks of varying degrees of difficulty that were likely to result in differing numbers of errors. Arnstein et al. predicted that, for the more difficult tasks, compared to the control group the stuttering group would produce disproportionately larger spikes on the two ERP measures investigated (one believed to represent fast, automatic error monitoring and the other believed to represent slower, conscious error monitoring). Compared to controls, the stuttering group in fact produced larger spikes on both measures across all conditions, (there were no group by condition interactions). Although this could reflect the fact that monitoring by PWS was hyper-vigilant across all conditions and led to an increased perception of errors across all conditions, there was also some doubt as to whether the size of the effects really represented error-monitoring activity, or whether they simply reflected the amount of effort required for the tasks, in which case the results could also be interpreted in terms of the stuttering group simply making more effort to attend to the task. With regard to this, it is noteworthy that the mean age of the PWS group was 10 years older than that of the control group. So, the difference may have been a result of their age rather than their stuttering. Unfortunately, participants didn't give feedback on whether or not they had actually perceived the errors they made as errors.

3.6.2. *EXPLAN*

The EXPLAN hypothesis (Howell, 2011a; Howell & Au-Yeung, 2002) also attempts to account for researchers' failure to find firm evidence of abnormally high numbers of phonological encoding errors in the speech plans of PWS. It posits that stuttering-like disfluencies arise when speech planning proceeds at a slower rate than its execution and the speaker has effectively run out of speech plan to articulate. At such moments, speakers tend to repeatedly execute whatever speech plan is already available to them until more becomes available for execution. Because, according to EXPLAN, the words of a phrase are planned in parallel, shorter, more frequent, function words tend to become fully activated faster than the content words associated with them. Stuttering-like disfluencies arise as a result of either of two forms of response to the lack of an available complete plan: '*Stalling'* and 'A*dvancing'*. *Stalling* involves repetition of the function word prior to a problematic content word. It occurs when the activation level of the function word remains higher than that of the following (incompletely encoded) content word at the time when the content word should be articulated. As a result, the function word is repeated in place of the content word. Such function-word repetition may continue until the activation level of the content word rises above that of the function word. *Advancing* occurs when the speaker begins to articulate a word despite it not yet being completely encoded. Howell (2011) proposes that this occurs when the activation level of the initial phoneme(s) of the word are higher than those of any other word, despite (later) parts of it being incompletely encoded.

Howell and AuYeung (2002) posited that both young children who stutter as well as normally-fluent older children and adults generally exhibit Stalling behaviours whereby, if a required word is not ready for articulation, they repeat the entire word that precedes it, (usually a monosyllabic function word), whereas in speakers in whom stuttering becomes persistent, the tendency is to exhibit Advancing behaviours.

Advancing results in the production of rapid part-word repetitions and blocks, both of which are more disruptive to speech prosody. In normally fluent speakers, when Advancings occur, they trigger alert signals which are sent to a timekeeper (believed to be associated with the cerebellum) which responds by reducing the rate of utterance execution. However, in persistent stutterers, the speaker becomes desensitised to these alerts which consequently become ineffective. As a result, rather than slowing the rate of execution, the speaker continues to produce large numbers of Advancing disfluencies. During the course of a child's development, desensitization to these alerts may result in a gradual shift from a pattern of disfluent speech containing many (whole) function word repetitions, to one that contains large numbers of part-word repetitions. Desensitization to timing alerts is thus considered a key factor that differentiates persistent stutterers from normally fluent speakers (Howell et al., 1999).

An important difference between the EXPLAN hypothesis and hypotheses that invoke covert-repair mechanisms to explain the occurrence of stuttering-like disfluencies is that EXPLAN account does not involve speech-error monitoring and does not equate stuttering-like disfluencies with encoding errors. Hence, unlike the CRH, EXPLAN predicts that stutterers should not be aware of large numbers of encoding errors in their inner speech, and that a reduction in monitoring will not result in the production of large numbers of overt speech errors.

Experimental evidence that function-word repetitions constitute a stalling mechanism comes from a priming study by Savage and Howell (2008) involving picture naming. In the study, twelve children (all under nine years old) first heard an auditory prime and then saw an action picture that they had to describe. Required responses took the form of two function words followed by a content word, such as ‘he is swimming’ and ‘she is running’. The auditory primes were either the same two function words or the content word that the child was subsequently prompted (by the picture) to speak. It was hypothesized that because priming speeded the speech-onset latencies of the primed words, priming the function words would have the net effect of reducing the amount of time available for planning the content word, thus increasing the likelihood that the content word’s planning would not be complete when the time for when its execution arrived. Thus, because children under nine years of age produce mainly stalling disfluencies, when function words are primed, the children should produce *more* function-word disfluencies than when they are not primed. Priming a content word, on the other hand, should reduce the time needed for it to become completely activated, thus leading to a reduction in disfluencies on the content word. This was indeed the result they found. Importantly, the CRH would predict that priming of any word should *reduce* the likelihood of stuttering on it. Thus, the increase in function-word disfluencies in the function-word priming condition also constitutes evidence against the CRH.

A study by Howell and Sackin (2001) provided some evidence suggesting that stalling may serve to help speakers avoid silent pauses at times when language-formulation is slow. In the study, four normally-fluent speakers told a series of short stories. In the third trial, they were informed that they would receive feedback from a light bulb, which would light up whenever the quality of their performance was ‘poor’. In reality, the bulb was activated by silent pauses of more than 600ms. The presence of the light bulb did not increase the rate at which they told the story, but it did increase participants’ production of function-word repetitions and reduced the occurrence of silent pauses.

Exactly why some speakers begin to produce excessive numbers of advancing disfluencies is unclear. However, findings from a study by Howell and Sackin (2000) suggest that it is related to the experience of pressure to speak quickly. In Howell and Sackin’s Experiment 1, twelve normally-fluent participants were asked to produce a running commentary to a silent cartoon. When the sequence of events in the cartoon moved more rapidly the proportion of advancing disfluencies (part-word repetitions on content words) increased, suggesting that the increased time pressure caused speakers to attempt to execute content words before the plans for those words were fully formulated.

3.7. Summary

This chapter began by differentiating two broad classes of theories that have been put forward to explain stuttering: (1) Production-impairment hypotheses, such as the CRH, that attribute their occurrence to the (direct or indirect) consequence of impaired language and/or speech production mechanism; and (2) pure perception-based hypotheses, such as Diagnosogenic theory (Johnson, 1942), which posit that there is no underlying impairment. We highlighted the significance of the observation that, in people who stutter, stuttering-like disfluencies do not occur all of the time, and on the basis of this suggest that if there is some form of underlying impairment, it is most probably relatively mild. We then described how the pattern of their occurrence of stuttering-like disfluencies can, at least partially, be accounted for in terms of a ‘capacities and demands’ framework (Starkweather, 1987), whereby stuttering is most likely to occur when cognitive demands outstrip capacities.

We then discussed Diagnosogenic theory (Johnson, 1942) as an early example of a pure perception-based hypothesis, and how, although more recent evidence suggests that this type of hypothesis cannot account for the onset of stuttering, more recent variants of it, such as the Vicious Circle Hypothesis (Vasić & Wijnen, 2005) may nevertheless provide a parsimonious account of why it may persist into adulthood even when any signs of underlying production impairment have long since disappeared. In relation to this possibility we then outline a third broad class of hypotheses, described by Bloodstein (1958) as ‘anticipatory struggle hypotheses’ which posit that stuttering-like disfluencies occur at times when the speaker expects, on the basis of past experiences, that speaking (or communicating) will be difficult, and responds in a maladaptive way. The strength such hypotheses is that they provide a parsimonious account for the continued occurrence of symptoms of stuttering even when actual demands no longer exceed capacity.

We then discussed the unresolved issue ‘continuity’: whether or not stuttered disfluencies are categorically distinct from the stuttering-like disfluencies that occur in the speech of normally-fluent speakers. We suggest that the answer to this question may differ depending on whether the diagnosis of stuttering is based on objectively observable phenomena such as stuttering-like disfluencies (in which case there would appear to be a continuum) or on speakers’ own internal experiences of loss of control (in which case there would appear to be a dichotomy). We highlighted how the concept of a vicious circle crops up in a number of the hypotheses already discussed in the chapter, and how vicious-circle formation appears to be related to some form of inappropriate compensatory behaviour of one kind or another.

We briefly discussed the differences between incipient and persistent stuttering and the need for hypotheses of stuttering to address these differences. Then finally, the chapter provided more detailed descriptions of the CRH and its two main competitors: EXPLAN (Howell & Au-Yeung, 2002) and the Vicious Circle Hypothesis (Vasić & Wijnen, 2005), and described the key experimental studies that provide evidence for and against them.

4. Error biases and Language production

Chapter 3 highlighted the need for two key issues to be investigated in order to distinguish between the CRH and EXPLAN hypothesis: (a) the extent to which errors of phonological encoding occur in speakers' speech plans prior to overt articulation; and (b) the extent to which speakers engage in covert editing of those errors. Here in Chapter 4 we consider how these issues may be investigated. We discuss how insight into the extent to which speakers employ covert editing and repair as a means of reducing errors of phonological encoding can be gleaned from the study of the frequency biases that such errors tend to exhibit. We begin with a description of two key frequency biases – the phonemic-similarity bias and lexical bias, and two different accounts of how these biases may arise: one of which attributes them to the effects of activation-feedback within the production system, and the other to perceptually-mediated processes of monitoring, covert editing and repair. We then turn our attention to inner speech, and consider how paradigms that investigate inner speech, and that enable a comparison to be made between phonemic-error rates and biases in inner speech and those that can be observed in overt speech, have the potential to reveal much about the extent to which speakers engage in covert editing and error repair. Such paradigms also have the potential to reveal differences between stutterers and normally-fluent speakers with respect to the production, detection and covert editing of errors of phonological encoding.

4.1. Error biases in overt speech

Lexical bias is the tendency for phoneme substitution errors to result in the production of real words (rather than non-words) at a rate above chance. It has been consistently found in experimental studies designed to prime participants to produce real and non-word errors in equal numbers (e.g. Baars, Motley, & MacKay, 1975; Dell, 1986; Hartsuiker, Corley, & Martensen, 2005; Nooteboom, 2005a), and has also been found in some studies of naturally-occurring slips of the tongue (e.g. Dell & Reich, 1981; Nooteboom, 2005b). However, a small number of corpus studies have failed to show this effect (e.g. Del Viso, Igoa, & García-Albea, 1991; Garrett, 1976).

The phonemic-similarity bias is the tendency for errors involving similar-phoneme substitutions to occur more commonly than those involving dissimilar-phoneme substitutions (where 'similarity' is measured in terms of the number of features shared by the phonemes involved in the substitution). This bias has also been found consistently in experimental studies designed to show it (e.g. Nooteboom, 2005a; Nooteboom & Quené, 2008; Oppenheim & Dell, 2008; Wilshire, 1999) as well as in some corpus studies (e.g. Dell & Reich, 1981; Levitt & Healy, 1985; MacKay, 1970; Nooteboom, 1967; Shattuck-Hufnagel & Klatt, 1979)[4].

There are two different accounts of why these biases are found in overt speech. One is that they arise during the process of phonological encoding, in which case they should already be present in speech plans entering the articulatory buffer. A second is that they arise later, after phonological encoding is complete, as a result of pre-articulatory monitoring and editing of the speech plans while they are stored in the articulatory buffer. According to this second account, the biases arise because errors involving similar-phonemes and errors that constitute real-words are less salient and hence less likely to be perceived by the monitor and less likely to be edited out of the speech plan. As a consequence, they are more likely to be overtly articulated.

---

[4] Note that the phonemic similarity bias is not the same as the 'phonological similarity effect' that has been well documented in studies of working memory (e.g. Baddeley & Hitch, 1974; Conrad & Hull, 1964).

4.1.1. *The SLIP task*

The experimental study of error biases began in a systematic way in the early 1970s with a series of studies investigating semantic (Motley & Baars, 1976), syntactic (Motley, Baars, & Camden, 1981), and lexical biases (Baars et al., 1975) using an experimental paradigm known as the SLIP task. In this task, participants read silently a number of biasing word pairs, such as *ball dome,* each of which shares the same order of onset phonemes (in this case /b/, /d/). Then, occasionally a word pair is presented to be read aloud. The onset phonemes of such word pairs are sometimes manipulated so they occur in the opposite order to those of the biasing word pairs (such as *darn bore*). In such cases, priming from the biasing pairs, may cause the speaker to make onset substitution errors. Thus, in the above example, he may say *barn door* instead of *darn bore.* Importantly, in the paradigm, half of the errors that are primed-for in this way constitute real words and half constitute non-words. Therefore, if the proportion of primed-for errors actually made by participants is significantly more than half, this constitutes evidence of lexical bias.

The Baars, Motley and Mackay (1975) study (that focussed on lexical bias) contained two experiments. The key finding of the first experiment was that, in a SLIP task in which all the words to be read aloud and all the biasing pairs were real words, the proportion of onset errors that resulted in the production of real words was significantly higher than the proportion that constituted non-words. This confirmed, experimentally, the presence of lexical bias in a real-word context – analogous to that of everyday life. The second experiment investigated lexical bias in non-word and mixed contexts. In the non-word context, in which all the targets to be read aloud and all the biasing pairs were non-words, the proportion of onset errors that resulted in the production of non-words was not significantly different from the proportion that constituted words. However, when exactly the same non-word targets were read aloud, but the biasing pairs were real words (i.e. a mixed context), the lexical bias returned.

This finding from their second experiment, that error biases are modulated by the context in which they occur, was interpreted by Baars et al. (1975) as evidence for pre-articulatory monitoring and editing. More specifically, they proposed that before uttering a word aloud, speakers are able to check whether or not it conforms to certain simple criteria, such as the criterion ‘is this a word?’ and cancel execution of the plan if it does not. Importantly, however, they proposed that in contexts where the criterion is of no use, such as in the all non-word context, this form of monitoring can be turned off, in which case the lexical bias disappears.

Since Baars et al. (1975) performed these slip tasks, the monitoring account has been further elaborated by Levelt and colleagues and incorporated into their model of language production (Levelt, 1983, 1989; Levelt, Roelofs, & Meyer, 1999). As the CRH is to a large extent predicated upon Levelt’s model, we describe it in some detail below.

4.2. Levelt’s ‘Feed-forward’ model of language production

In this model (Levelt, 1983, 1989; Levelt et al., 1999), language production is conceptualised as a multi-level process in which information moves in a top-down (feed-forward) manner from higher to lower levels of processing. First, in a ‘Conceptualizer’, the concept behind a message is generated. Then, in a ‘Formulator’, a linguistic formulation of the message is generated (formulation is itself a multi-level process involving lexical selection, the building of a syntactic tree and the elaboration of the tree with phonological information). Having been formulated, messages may then be stored in an articulatory buffer until they are released to the ‘Articulator’ for overt articulation. The Formulator and Articulator are modular in nature and function independently of each other and in parallel. Thus, while one utterance is being articulated, the next utterance is already being planned and further utterances may already be undergoing conceptualization. Alongside this production system a comprehension system also operates which, in many ways, mirrors the production system, except that it works in the reverse direction (See Figure 4-1).

4.2.1. *The perception of errors in the speech plan*

Speech errors can arise at any of the three levels described above: During conceptualization, ‘appropriateness errors’ occur in which the wrong message is generated; during formulation, lexical, syntactic or phonological encoding errors can occur; and during motor execution, articulatory errors can occur. Phonological encoding errors generally stem from failures in the assignment of phonemes to the correct ‘slot’ in the syntactic frame. This may happen due to timing difficulties and background ‘noise’, both of which may result in interference from other words being encoded. Thus sometimes more than one phoneme is available to fill a slot in the syntactic frame, and the wrong phoneme is selected.

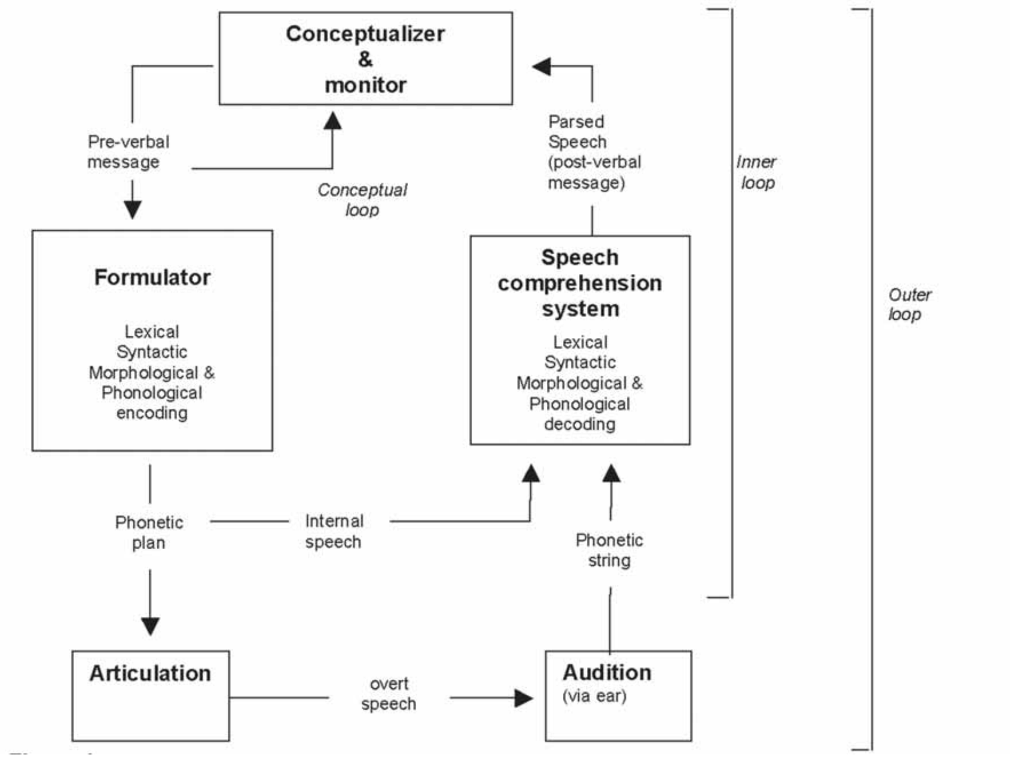


Figure 4-1 Levelt's model of speech production and monitoring (adapted from Levelt, 1989)

The model also contains three monitoring loops through which such errors can be detected: a ‘conceptual loop’, situated within the conceptualizer, which monitors the appropriacy of messages being generated; an ‘inner loop’ that allows monitoring of the speech plan stored in the articulatory buffer prior to their release for overt execution; and an ‘outer loop’ through which speakers can listen to their utterances in the same way as they can listen to those of other speakers. Both the inner and outer loops convey formulated messages to the comprehension system where they are parsed (i.e. converted back into phonological, syntactic and semantic representations). Each parsed message is then transferred back to the conceptualizer where it can be evaluated.

Two key features of this ‘perceptual monitoring’ are that it is (at least partially) accessible to conscious awareness, and it is flexible. Thus, once a parsed message has reached the conceptualizer, it can be evaluated against any of a number of general criteria, including “deviant sound form, deviant morphology and deviant syntax” (Levelt, 1989, p. 470), or alternatively, its meaning can be compared with the concept from which it was originally derived. However, perceptual monitoring is also attention-dependent and capacity-limited, and, because of this, some errors may be harder to perceive than others.

The existence of the inner monitoring loop in Levelt’s model allows for the possibility of covert error repair because incorrect messages can be detected, cancelled and reformulated before they are released from the articulatory buffer for overt execution. This process is achieved through a process described as being roughly equivalent to inner speech, although the details of how such inner speech is generated from the speech plan have never been specified.

Monitoring via the inner loop, together with the covert editing and reformulation of errors, is the mechanism posited by the CRH to result in the production of stuttering-like disfluencies. Importantly, because error-detection is conceptualised as a perceptually-driven process, dependent on available cognitive resources and mediated by attention, it potentially provides a parsimonious explanation for why stuttering-like disfluencies are so strongly influenced by the context in which an utterance is spoken. However, a weakness of the model is that, because error detection via the inner loop involves parsing via the comprehension system, and error repair requires reformulation of the speech plan, it questionable whether all this can be accomplished fast enough to account for disfluencies characterised by the very rapid syllable repetitions that frequently occur in stuttering, or for that matter in normal speech (Blackmer & Mitton, 1991).

This potential weakness of Levelt's model is due to its conceptualisation of the flow of activation within the formulator as entirely uni-dimensional. Thus, the only form of 'feedback' available through which encoding errors can be detected is that which occurs via the comprehension system. An alternative possibility that might allow for much faster 'repair' of errors is that activation flow within the formulator is bi-directional. This possibility has been elaborated most in Dell's (1986) 'interactive' model of language production, described below.

4.3. Dell's 'Interactive' model of language production

Dell's (1986, 1988) model of language production is the best-known exemplar of a number of 'interactive' (or spreading activation) models (see also MacKay, 1982; Stemberger, 1985), characterised by the bi-directional flow of activation between processing levels. Interactive models offer a different explanation for why lexical bias and phonemic similarity may occur; one that in some respects is more parsimonious than that offered by Levelt's model, insofar as it does not invoke separate monitoring, editing and self-repair mechanisms. Like Levelt's, Dells' model contains a number of processing levels (semantic, lexical, phonological and featural), and each level consists of a large number of nodes that are permanently available in long-term memory. However, unlike Levelt's model, activation can flow not only from higher to lower levels *but also* in the reverse direction, in which case it is often described as ‘activation feedback’. Thus nodes become primed through the spreading of activation throughout the entire network.

As with Levelt’s model, encoding errors may occur because the language-production system is noisy. This means that, during the early stages of formulation of an utterance, competing elements (e.g. words and phonemes) may have higher levels of activation than the correct, target elements, especially if they were used in recent utterances. However, in Dell’s (1986) model, activation of target elements is able to reverberate between different processing levels. This means that the activation level of target elements is maintained over time, whereas activation of competitor elements decays (See Subsection 4.5.1: ‘Activation-feedback accounts’ for more details). A consequence of this tendency for activation of target elements to reverberate between levels is that, if overt execution is delayed long enough, the activation of the target word or phoneme will eventually exceed that of its competitors, and errors will effectively repair themselves automatically. This tendency for the activation levels of target phonemes to gradually become reliably higher than those of competing phonemes is a key feature of interactive models.

The EXPLAN hypothesis is predicated on a spreading-activation mechanism of the sort proposed by Dell (1986) in conjunction with an additional autonomous restart mechanism (as described in Section 3.6.2). Postma and Kolk’s (1993) Covert Repair Hypothesis also draws from Dell’s model, insofar as it posits that the likelihood of errors being found in the speech plan at the point of time when it is selected for execution is a function of (a) how efficiently activation accumulates, and (b) the amount of time that passes between the onset of planning and the selection of the plan for execution. According to the Covert Repair Hypothesis, activation accumulates less efficiently in stutterers compared to normally-fluent speakers, with the result that, at the time of selection of the plan for execution, competitor phonemes are more likely to have activation levels comparable to those of the target phonemes. There is thus, in stutterers, a much higher probability of competitor phonemes being selected in error (see Figure 4-2).

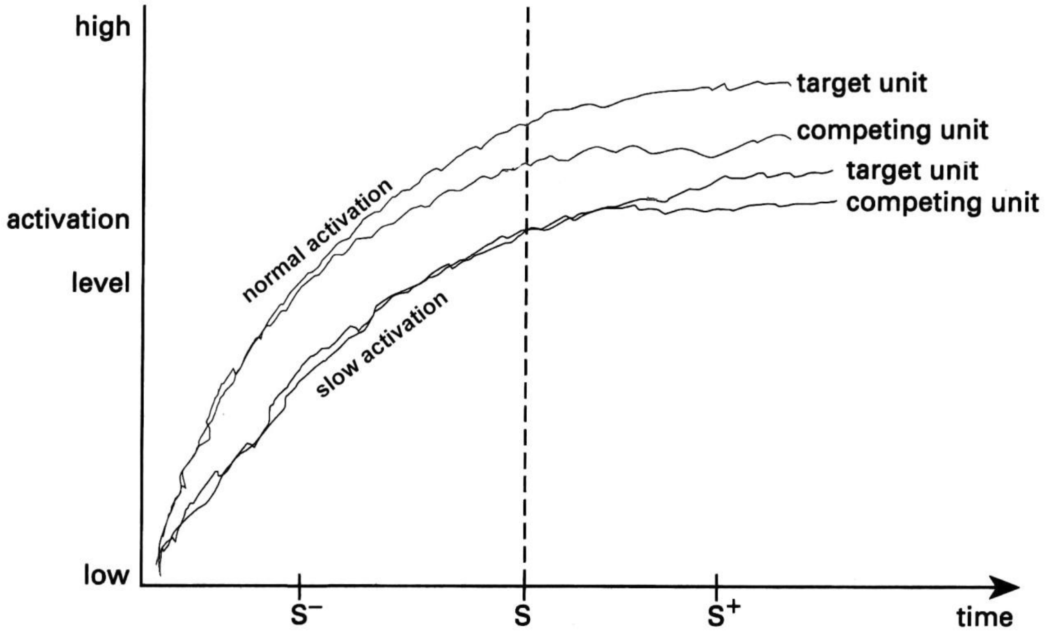


Figure 4-2 Phonological node activation and selection
Comparing a normal versus a slow build-up of activation and demonstrating the impact of phoneme selection at three time points: s– (an abnormally fast speech rate), s (a normal speech rate) and s+ (a slowed speech rate).
Adapted from Kolk & Postma, 1997.

The CRH’s tenet that slow phonological encoding leads to higher error rates thus presupposes an at least partially interactive speech production system, whereas its tenet that stuttering-like disfluencies result from a process of covert error repair is predicated on perceptual monitoring, which has been developed most fully in the context of a purely feed-forward language production system. The CRH is thus actually predicated on a hybrid model that incorporates both bi-directional spreading of activation as well as monitoring. Although such a hybrid model has not been computationally instantiated, it is nevertheless a plausible possibility (Goldrick, 2006; Hartsuiker et al., 2005; Rapp & Goldrick, 2000). In contrast, the EXPLAN hypothesis, which does not implicate monitoring in the production of disfluencies, is predicated entirely on Dell’s interactive model of language production.

It should be noted that a number of ‘Production monitoring’ models have also been developed (e.g. Laver, 1973, 1980) that can potentially also better account for fast rates of error detection and repair. Unlike Levelt’s perceptual monitoring account, these models posit that error monitoring is intrinsic to the language production system itself and that monitoring of the various stages of language formulation within the formulator can occur. As a result, the monitor may have the potential to detect errors at a much faster rate, in which case editing and repair may operate independently of consciousness. However, these accounts have been criticised (e.g. Levelt, 1989) as being unparsimonious insofar as they require the duplication of knowledge.

As a further alternative, Postma and Kolk (1993) considered the possibility that covert detection and editing of errors may occur within the speech-production mechanism as a consequence of bi-directional spreading of activation. They suggested that this could be achieved through a mechanism that compares the relative strengths of activation output from a node to that which is fed back to it from the level below [5]. This ‘monitoring’ would be sensitive to unexpected differences in these strengths and would respond to such differences by signalling the need to cancel and reformulate the utterance. This process has much in common with the production monitoring proposed by Laver (1980). For example, it would operate fast and efficiently, and potentially without the speaker being consciously aware of it. However, unlike Laver’s model, it does not require an external agent (homunculus) to

---

[5] A similar mechanism has also been proposed by Mackay (1987).

perform the monitoring and does not require reduplication of knowledge. It is thus more parsimonious, but may be less flexible and adaptive. As far as we know, there is, as yet, no experimental evidence to support the existence of this type of mechanism. (See Postma, 2000 for a detailed review of alternative monitoring mechanisms that have been postulated).

4.4. The nature and roles of inner speech

An underlying assumption implicit in Levelt’s and Dell’s models of language production is that speech plans that are formulated may not always be executed overtly. If they are not executed overtly, speakers may nevertheless be aware of them through their inner speech, and the plans themselves may nevertheless still fulfil a number of important functions, insofar as they may form the basis of the production of writing, personal thought, reasoning and memorization (e.g. Baddeley & Hitch, 1974; Ellis, 1988; Sokolov, 1972; Vygotsky, 1986).

Subjective accounts of individuals' experiences of inner-speech suggest that it frequently resembles overt speech, in that it appears to be acoustically-based and can vary in tempo, pitch and rhythm (MacKay, 1992). However, it has also been suggested (e.g. Vygotsky, 1986) that, compared to overt-speech, inner-speech is often attenuated, both syntactically, lexically and phonologically, and that the completeness with which inner speech is formulated may depend upon the purpose for which it is intended. Thus, for example, when it is being used purely for internal operations such as thought or analysis, there is no need to use the (first-person) personal pronouns ‘I’ and ‘we’ (cf. prodrop in 'diary style' writing; Haegeman & Ihsane, 2001). Furthermore, if there is no intention to speak out loud, then there would appear to be no real need to specify speech-plans below the phonological level. Indeed, leaving phonetic and/or articulatory details unspecified may bring a number of benefits, including a reduction in processing load and the possibility that internal speech may proceed at a significantly faster rate than the rate at which overt speech can be articulated.

Although there is evidence, from clinical and experimental studies, indicating that articulatory activity is not necessary for inner-speech (see MacKay, 1992; Sokolov, 1972, for detailed reviews), and there is also some experimental evidence suggestive of under-specification of speech-plans for inner-speech of normally-fluent speakers (Oppenheim & Dell, 2008, 2010; Wheeldon & Levelt, 1995), this area is relatively under-researched. Evidence of under-specification of speech plans is, however, particularly relevant to the current thesis because it has the potential to provide insights into what exactly speakers habitually monitor in order to detect errors in their inner speech. Thus, for example, findings from an inner-speech phoneme-monitoring study by Wheeldon and Levelt (1995) suggest that normally-fluent speakers at least sometimes monitor abstract (timeless) syllabified phonological representations, and it is possible that the use of this type of monitoring in such tasks may reflect the wider tendency (of normally-fluent speakers) to under-specify speech plans for inner speech. On the other hand, Vasić and Wijnen's (2005) proposal (in their Vicious Circle Hypothesis) that stutterers are excessively aware of subtle timing irregularities in speech plans, suggests that stutterers may have a stronger tendency to fully specify speech plans, irrespective of whether or not they are intended for overt speech[6]. Such group differences should be empirically testable, insofar as Dell's model of language production predicts that under-specification of speech plans at articulatory levels should result in a reduction in the strength of the phonemic-similarity bias in phonemic errors in inner-speech (See Section 4.5.3 below, for a full explanation of this). Thus, the finding of a

---

[6] For example, in line with the 'Levels of Processing argument (Craik & Lockhart, 1972), people with working memory impairments may routinely fully specify speech plans as a strategy to aid working memory.

stronger similarity bias in the inner-speech errors of stutterers compared to those of normally-fluent speakers would constitute evidence in favour of the Vicious Circle Hypothesis.

4.5. Error rates and biases: Predictions of language-production models

An important difference between Levelt's and Dell's models of language production is that they lead to differing predictions regarding the relative frequencies with which speakers are likely to make some types of errors of phonological encoding. This is in part because some errors are more salient than others and thus more likely to be missed by a perceptual monitor, and in part because some are 'encouraged' by the presence of activation-feedback (Hartsuiker, 2006). In the following subsections we focus on how these two language-production models predict the presence of lexical and phonemic-similarity biases in over speech.

4.5.1. *Phonemic-similarity and lexical biases in overt speech*

Monitoring and editing accounts

Levelt's model (Levelt, 1983, 1989; Levelt et al., 1999) accounts for lexical bias in overt speech as the result of an increased tendency (in real-word environments) to covertly edit-out and repair phonological-encoding errors that result in the production of non-words compared to errors that result in real words. As a result, proportionately more errors that result in real words are released for overt execution.

A similar process is likely to occur with respect to phonemically-similar errors: Because they are perceptually less salient than phonemically-dissimilar errors, they are less subject to covert editing and repair. In fact Levelt never explicitly stated whether the similarity bias in overt speech can also be accounted for in this way. However, as the monitor has access to word forms in the lexicon, it seems plausible that it can[7]. The net effect of this diminished salience is that (all other things being equal) more phonemic errors with similar onsets should slip through the covert editing process and be uttered in overt speech than phonemic errors with dissimilar onsets.

Activation-feedback accounts

Interactive models, in which activation is able to flow bi-directionally, such as Dell's (1986, 1988), also predict lexical bias and phonological similarity effects in overt speech. According to Dell's model, the two error biases result from the reverberation of activation between levels of language encoding: The lexical bias due to reverberation between lexical and phoneme levels; and the phonemic similarity bias due to reverberation between phoneme and feature levels.

---

[7] A potential mechanism through which this could occur is as follows: A word form accessed during speech formulation will still be relatively highly activated in the lexicon at the point of time when the comprehension system parses the corresponding input via the inner loop. Thus, if, for some reason, phonological encoding has gone wrong and an incorrect onset phoneme has been incorporated into the speech plan of that word, when it is monitored via the inner loop, the comprehension system will immediately detect a discrepancy between it and the most highly activated word-form in the lexicon (i.e. the intended word form). However, the above explanation presumes a single lexicon for comprehension and production, the existence of which remains an open question. See also Nooteboom (2005b) for an alternative monitoring account of lexical bias.

For example, if a speaker intends to say *dog*, he is more likely to accidentally say *log* than *pog.* The competing onset phoneme /l/ (which would result in erroneous production of the word *log*) is more likely to be erroneously selected than the competing onset phoneme/p/ (which would result in erroneous production of the nonword *pog*) because feedback from /l/ will activate (and reverberate with) the lexical entry for *log*, and will thus be sustained to some extent despite the fact that the /l/ phoneme is erroneous. Conversely, because there is no lexical entry for the nonword *pog*, there will be no reverberation of activation between the /p/ phoneme and any corresponding lexical entry. So activation for the /p/ phoneme will quickly die away.

A similar mechanism is also postulated for why, all other things being equal, phoneme-substitution errors are more likely to occur between similar sounding phonemes: Activation normally reverberates between a phoneme and the (sub-phonemic) features associated with it. The presence of such reverberation of activation between phoneme and feature levels means that competing phonemes that share some of the same features as the target phoneme will also receive some activation when the target phoneme becomes activated. The more features a competing phoneme shares with the target phoneme, the more activation it will receive and the more likely it is to be selected in error, especially if that competing phoneme already has some residual activation as a result of having recently been spoken (or heard). For example the target phoneme /p/ (bilabial, plosive, unvoiced) is more likely to be substituted by a phoneme with which it shares two features, such as /b/ (bilabial, plosive, voiced) than by a phoneme with which it shares only one feature, such as /d/ (alveolar, plosive, voiced).

Models incorporating both monitoring & feedback

Although Dell (1986) did not specify how his model interacts with the comprehension system, it is quite possible (indeed likely) that if the human speech production system is interactive in the way Dell described, perceptual monitoring via the comprehension system could nevertheless still be used for pre-articulatory detection and editing of errors. If this is the case, then it is possible that the error biases found in overt speech may be the net result of the combined influences of both monitoring and feedback.

Some experimental evidence in support of this combined monitoring and feedback account of lexical bias has been provided by Hartsuiker et al. (2005) who set out to replicate the Baars et al.'s (1975) SLIP task experiment, using a more tightly controlled design. Analysis of the raw numbers of errors obtained by Hartsuiker et al. (2005) in the various conditions best fitted an account whereby in the non-word context, the underlying influence of activation feedback, although still present, was masked by the (opposite) effect of pre-articulatory monitoring for errors that constituted real-words – which were successfully removed before articulation. Hartsuiker et al. (2005) concluded from their findings, that (a) the focus of monitoring can be adapted to the needs of the context; and (b) both activation feedback and monitoring can take place simultaneously and in some conditions can effectively cancel each other out.

4.5.2. *Error Biases in self-reports of overt speech*

All corpus studies and by far the majority of experimental studies that have documented the similarity and lexical biases that occur in speech have relied on transcriptions and codings made by independent raters. However, if speakers themselves are asked to provide online self-reports of the errors they make in their overt speech, their self-reports may differ from the off-line reports made by an independent rater. The extent of these differences will depend on how vigilant speakers are at perceiving and self-reporting their errors, and also on what exactly it is that they are self-reporting.

If a speaker's perception of his errors is less than perfect, then it could result in a 'reporting bias' whereby less salient errors are under-reported[8]. However, this tendency to under-report the less salient errors may occur together with the tendency to make more of them, because less salient errors are less likely to be noticed and therefore are less subject to pre-articulatory editing, Thus, insofar as the phonological-similarity and lexical biases stem from differing degrees of salience of the various kinds of error they involve, the overt error biases and the self-reporting biases will be in the opposite direction to each other and will tend to cancel each other out.

If, on the other hand, overt error biases stem purely from differing degrees of reverberation of activation feedback within the language production system itself, and monitoring is equally effective (or ineffective) for all types of error (i.e. a pure feedback account), then participants' own self-reports should be faithfully reflect the biases perceptible to an independent-rater in terms of their magnitude and direction, although the numbers may be slightly lower overall depending on how vigilant the self-reporter is.

It is noteworthy that the findings of three tongue-twister studies (Lackner & Tuller, 1979; Postma & Kolk, 1992; Postma & Noordanus, 1996) that have investigated the

---

[8] Strictly speaking the reporting bias may be influenced by two factors: (a) whether an error is perceived; and (b) whether the speaker bothers to report it. For the sake of simplicity of explanation, we will presume here that speakers are cooperative insofar as, if asked to do so, they report all the errors they perceive.

vigilance of online self-reporting (by asking participants to self-report their phonemic errors as they make them) all suggest that, when speakers are able to clearly hear what they have said (i.e. when they are able to monitor through the ‘external’ loop), although they do tend to somewhat under-report the numbers of errors they make, the extent of their under-reporting is not affected by the similarity of the error they are reporting to its original target. Similar, (although indirect) evidence that errors that do and do not result in the production of real words are equally likely to be perceived in overt speech has also been provided by a study by Nooteboom (2005b), that found that speakers were equally likely to self-repair their overt errors, irrespective of whether those errors constituted words or non-words. Together these findings suggest that, although self-monitoring of overt speech is not perfect, the extent to which errors are perceived when the outer loop is available is not influenced by the similarity of their onset phonemes to the original targets or by the lexical status of their outcomes.

The above findings may not, however, reveal the whole story. Implicit in Levelt’s (1989) account of pre-articulatory editing is the notion that, in speaking situations where a speaker attempts to perform pre-articulatory editing, the speaker may actually focus more on monitoring his inner speech (through the inner loop) than his overt speech, despite the fact that he is speaking aloud and feedback via the auditory loop is available. With this possibility in mind, it is noteworthy that the three studies mentioned above (Lackner & Tuller, 1979; Postma & Kolk, 1992; Postma & Noordanus, 1996) did not emphasise the need for accuracy and therefore did not encourage pre-articulatory editing. Moreover, they only failed to find any evidence of different degrees of salience of different types of phonemic error in overt speech when participants were able to clearly hear (and attend to) what they had said. In contrast, when their overt speech was masked by white noise (which prevented participants using the ‘external loop’ to monitor the sound of the strings they uttered), these studies found that speakers were less successful at self-reporting voicing errors compared to other types of errors.

These findings suggest that although voicing errors are comparably salient to other kinds of errors in overt speech, they may be less salient in inner speech. If this is the case, then speakers self-reports of errors are likely to be differentially biased depending on which monitoring channel the self-reporter has been using.

4.5.3. *Error Biases in self-reports of silent (non-articulated) speech errors*

Comparisons of detection rates of errors involving similar phoneme substitutions in the masked and unmasked overt speech conditions in the three studies outlined above (Lackner & Tuller, 1979; Postma & Kolk, 1992; Postma & Noordanus, 1996) provided some preliminary evidence that phonemically-similar errors may be less salient in inner speech than in overt speech. However, the inner speech that participants attended to in the masked overt conditions of these experiments may have differed from inner speech produced in the complete absence of any articulation. This is because, as explained in Section 4.4, speech plans produced in the absence of any intention to overtly articulate may lack sub-phonemic specification, and thus may not be influenced by activation-feedback from feature to phoneme levels of encoding. As a consequence, errors occurring in such speech plans may lack the similarity bias that may have otherwise been present, had the plan been intended for overt articulation. To date, the three studies that have investigated whether or not this is the case have provided conflicting findings: The comparison of self-reports of inner speech errors in the articulated an unarticulated conditions in Postma and Noordanus' (1996) study failed to find any differences, whereas comparable comparisons in two studies by

Oppenheim and Dell (2008, 2010) both found reduced similarity biases in self-reports of unarticulated inner speech[9].

The differences between Postma and Noordanus' and Oppenheim and Dell's findings with respect to the phonemic similarity effect in inner speech may have arisen due to differences in the details of how the experiments were conducted (the paradigms employed different tongue-twisters, different speech rates, different instructions, and were carried out in different labs), and it is quite possible that these differences in experimental design may have resulted in significant differences in the degree to which speech plans were specified and in how vigilant participants were in self-reporting their errors. Whatever the case, such differences highlight the need for extensive testing of a paradigm before using it to compare the performances of different participant groups. In recognition of this need, before comparing the performance of stutterers and controls with respect to error rates and biases, we therefore first performed three experiments that investigate the reliability and validity of the paradigm employed.

---

[9] In contrast to the similarity bias, Oppenheim and Dell (2008, 2010) found no evidence of a reduced lexical bias in inner speech: Specifically, Oppenheim and Dell (2008) found that the self-reported lexical biases in unarticulated inner speech and overt speech were of comparable magnitude and in the same direction, and the subsequent study by Oppenheim and Dell (2010) which used the same paradigm to compare inner speech with silent articulated speech found similar magnitudes of lexical bias in these two conditions as well. (These studies are described in more detail in Section 6.1.1).

4.6. Summary of Chapter 4

In Chapter 4 we set out to clarify why studies comparing phonemic error rates and error biases in inner and overt speech have the potential to inform us about how the human speech-production system functions and why errors of phonological processing may occur. We first outlined evidence from corpus studies and experiments that led to the identification of the lexical and phonological-similarity biases. We then described two different speech-production models that account for these biases in overt speech in two different ways – as the result of activation feedback or as the result of monitoring and editing. We then described how Dell's (1986) model, which incorporates activation-feedback, predicts a reduced phonemic-similarity bias in inner speech (in the absence of any overt articulation). We then distinguished between the biases actually present in overt speech and those present in speakers' self-reports of their errors, and explained how the two language-production models make different predictions regarding the relative strengths and directions of the biases when speakers are required to self-report their errors.

We concluded by cited the findings of a number of tongue-twister experiments that have investigated participants' self-reports of their errors in a variety of inner and overt speech conditions, including three that allow comparisons of errors in inner speech with and without accompanying articulation conflict. Of these, two (Oppenheim & Dell, 2008, 2010) found that the strength of the similarity bias in onset errors occurring in the tonguetwisters was reduced in the absence of articulation, whereas the third (Postma & Noordanus, 1996) found that the presence or absence of articulation had no effect. This difference in findings highlights the need to establish the reliability of the Oppenheim and Dell paradigm across populations before using it to compare the performance of stutterers and normally-fluent speakers.

Before continuing on to describe our own experiments (in Chapter 6), we first outline the rationales behind the reporting and statistical methods we have adopted in those experiments.

5. Reporting and statistical methods used in this thesis

5.1. The use of self-reported data

This thesis is largely concerned with the experience of inner speech, the errors that occur in inner speech, and the extent to which speakers experience difficulty speaking fluently. Because inner speech and the experience of difficulty are not amenable to any direct or reliable forms of objective measurement, in our experiments we have relied on participants' own self-reports of these experiences. In experimental conditions that involved overt speech, digital recordings were taken and an experienced coder (Paul Brocklehurst: referred to in the thesis as the 'independent rater' or 'experimenter') coded these recordings offline. During offline coding, recordings were replayed as often as necessary arrive at as accurate a judgement as possible. In Experiment 4 (in which stutterers and matched controls recited tongue-twisters) a random sample of ten percent of the tongue-twister recordings were also re-coded (by the experimenter) at a later date in order to enable a check of intra-rater reliability (See Section 7.2.4). In the overt speech conditions, the independent-rater codings of speech errors have been considered to constitute a stable baseline against which self-ratings can be judged. Although clearly they cannot be considered to be 'correct' in any ultimate sense, for operational purposes we have treated them as if they are correct. The justification for considering the independent-rater ratings in this way is a relative one, insofar as they were made carefully, offline under optimal conditions without time pressure, whereas in comparison, participants' self-ratings were made 'on the fly'.

5.1.1. *Method of self-reporting of inner and overt speech errors*

In the four tongue-twister experiments, participants were asked to provide verbal online self-reports of errors that had occurred in their inner speech and overt speech. Although this is the method of self-reporting used by Oppenheim and Dell (2008, 2010) in two experiments designed to investigate errors in inner speech, experience gleaned from experimenting with spoken and written (typed) modalities of self-reporting, together with the need for a mode of self-reporting that was not more difficult for participants who stutter, led to the adoption in subsequent experiments of typing as the sole modality of self-reporting. This eliminated any potential for error when transcribing participants' self-reports. It may also have influenced the accuracy with which participants made their self-reports and/or the likelihood that they would bother to make self-reports at the appropriate times. Because of these potential confounds, self-reporting modality is controlled for statistically in the combined analyses that involve both self-reporting modalities.

5.1.2. ***Criteria for classifying participants as People who stutter***

In the experiments which involved both stuttering and normally-fluent participants, the criteria we adopted for participants to be allotted to the stuttering group, were as follows: (a) participants must consider themselves to be people who stutter (more precisely, people with persistent developmental stuttering); and (b) they must also have been diagnosed (at some point in the past) as such by a speech therapist or speech pathologist. The assignment of participants to groups in the above manner is in line with the ICD9 categorical definition of stuttering as "Disorders in the rhythm of speech in which the individual knows precisely what he wishes to say but at the time is unable to say because of an involuntary repetition, prolongation, or cessation of a sound" (ICD9; World Health Organization, 1977, p. 202).

Measures of the percentage of syllables containing stuttering-like disfluencies were recorded for all participants (both for spontaneous conversation as well as reading). Composite stuttering-severity scores (SSI4 ; Riley, 2009), which include a variety of measures of both primary and secondary symptoms of stuttering, were also collected for stuttering participants. Self-ratings of difficulty communicating and of difficulty speaking fluently were also collected for all participants. In Experiments five and six we also used participants' self-reports to determine whether or not they had stuttered on individual words. Participants provided these self-reports immediately after uttering each single word. These latter four measures were used for statistical analyses but not for diagnostic purposes.

5.2. *The use of mixed-effects modelling*

In all our experiments, we were interested in quantifying the size and direction of the relationship between a variety of (dichotomous and continuous) independent variables and each dependent variable. Regression analyses are ideally suited for this type of analysis. Unlike fixed input variables, such as gender, in which all possible levels are represented, the levels of a number of the variables, such as the experimental subjects and items (all experiments involved repeated-measures designs), constitute a small (and by assumption, random) selection of a much larger potential population. To reflect this, mixed effects regression analyses were used whereby the baseline error-probability of random effects like subjects and items were treated as normally distributed and thus the variance of each random effect could be determined by the model. Mixed-effects modelling thus enabled us to build models that better accounted for such variance. More specifically, linear mixed-effects modelling was used for tests with continuous dependent variables and logistic mixed effects modelling was used for tests with dichotomous outcomes.

Around the time of writing-up the final two experiments, there was a change in recognised best practice with respect to mixed-effects modelling of psycholinguistic data of the type used in the thesis. Extensive testing of a number of approaches by Scheepers and Barr (2011), including ANCOVA and a number of mixed-effects modelling approaches, on simulated data suggest that optimally-low Type I and Type II error rates can be achieved by using mixed effects analyses that include additional random effects to quantify (between-participant) inequalities of variance across levels of fixed predictors. Therefore, in the analyses of our final two experiments, random effects to capture variance in *both intercepts and slopes* of predictors have been incorporated into the regression models. Our earlier experiments were analysed using mixed-effects modelling with random intercepts only. Where the results of these earlier analyses have already appeared in journal articles, we have not changed them in the main body of the thesis. However, where appropriate[10], equivalent models with random slopes have been provided and can be found in Appendix B. Notes alerting the reader to the existence of these alternative analyses have been added to the relevant table captions in the main body of the text.

Importantly, where it was possible to add random slopes to the models, their addition did not result in any substantive changes to the magnitude or direction of the coefficients or to the conclusions that could be drawn from them. The only instances where it was not possible to add random slopes to the models were the two analyses of 'Accuracy of self-reporting' (See Table 6-12 &Table 7-4). This was due to failure to achieve convergence, due to insufficient data for the number of degrees of freedom.

---

[10] For the first three experiments error numbers were low and there was insufficient power for the extra degrees of freedom that would have been added by random slopes. Therefore we only carried out analyses with random slopes on the combined analyses of these experiments.

6. Experiments 1 to 3: Assessing the reliability and validity of a paradigm to measure phonemic errors in inner speech.

6.1. Introduction

A central tenet of the Covert Repair Hypothesis (Postma and Kolk, 1993) is that, in stutterers, impaired flow of activation in the neural pathways responsible for language encoding results in the production of abnormally high numbers of phonological encoding errors. Postma and Kolk never specified exactly in what ways activation flow may be impaired, which is not surprising, because there are still ongoing uncertainties among researchers regarding how activation flows in such neural pathways in normally-fluent speakers. In particular it remains unclear whether flow is uni or bi-directional (as exemplified by the language production models discussed in Sections 4.2 and 4.3), and to what extent reverberation of feedback between encoding levels may play a role in reducing noise and minimizing the likelihood of encoding errors appearing in the speech plan.

Despite these ongoing uncertainties regarding the precise details of the functioning of the language-production mechanism, it may nevertheless be possible to determine whether language production in stutterers differs from that of normally-fluent speakers. This is because differences in the build-up and flow of activation in the system are likely to be reflected in differences in the relative frequencies with which stutterers and normally-fluent speakers produce various types of phonological-encoding errors. Such differences are likely to manifest most clearly in the relative strengths of the lexical and phonemic-similarity biases in the errors they make. Experiment 4 (Described in Chapter 0) investigates these differences using the Oppenheim and Dell (2008, 2010) tonguetwister paradigm.

The Oppenheim and Dell paradigm provides a systematic method for measuring: (a) phonological-encoding error-rates in inner and overt speech; (b) the relative magnitudes of the phonemic-similarity biases in inner and overt speech; and (c) the relative magnitude of the lexical biases in inner and overt speech. However, because it relies entirely on self-reports to document errors in inner speech, we first need to confirm that the data it provides are valid and reliable. Therefore, in the three experiments described in the current chapter we first attempt to replicate Oppenheim and Dell's (2008) original findings, using the paradigm on three different but highly similar groups of normally-fluent undergraduate students in our own lab in Edinburgh, UK. Having established its reliability and validity in this way, we then go on (in Experiment 4) to use it to compare the performances of a group of stutterers and (age, gender and education) matched controls.

In their experiments, Oppenheim and Dell (2008, 2010) only analysed data from participants' own self-reports. However, in our experiments, because stuttering has been equated with hyper-vigilant monitoring (Kamhi & McOsker, 1982; Vasić & Wijnen, 2005), we additionally compare participants' own self-reports of their overt errors to offline ratings (from digital recordings of the same utterances) made by the experimenter. This extra set of comparisons enables us to compare the vigilance and accuracy with which the different participant groups identify the (overt) errors they make; the presumption being that the vigilance and accuracy with which participants monitor their inner speech will, at least to a certain extent, impact upon the vigilance and/or accuracy with which they self-report their speech errors when speaking overtly, especially under noise-masked conditions.

6.1.1. ***The Oppenheim and Dell 2008 tongue-twister paradigm***

As detailed in Section 4.4, there are a number of reasons to suspect that, in the absence of any intention to articulate overtly, speech-plans for inner speech may be underspecified at articulatory (featural) levels, and there is some experimental evidence from a study by Wheeldon and Levelt (1995) to support this hypothesis. Converging evidence for such under-specification comes from computer instantiations of Dell’s (1986) spreading-activation model in which the featural level of the model is not activated. Such instantiations show that without featural activation, the phonemic-similarity bias disappears from onset-phoneme substitution errors but the lexical bias remains unaffected (Oppenheim and Dell, 2008).

To test whether a similar pattern is observed in real-life inner speech, Oppenheim and Dell (2008) performed an experiment in which participants were required to repeat a series of four-word tongue-twisters and to immediately stop and self-report each error that they made. The tongue-twisters were manipulated such that they intrinsically primed the speaker to make onset-substitution errors (henceforth referred to as ‘primed-for errors’), half of which would result in real words and half nonwords. In addition to this outcome-lexicality manipulation, half of the tongue-twisters were marked for silent repetition (in inner speech, without articulation), whereas the other half were to be recited out loud (overt speech).

In both the inner and overt speech conditions, participants’ self-reports of their onset errors revealed a lexical bias effect. In contrast, a phonemic-similarity bias was only found in their self-reports of their *overt* speech. Oppenheim and Dell (2008) interpreted the lack of a similarity bias in the inner-speech condition as providing support for the ‘Surface Impoverished Hypothesis’: that, in the absence of any intention to articulate, inner speech is underspecified at the feature level. One possible consequence of featural under-specification is that there is no feedback of activation from feature to phoneme levels of representation, no bottom-up activation of competitor phonemes and therefore no phonemic similarity bias. Oppenheim and Dell (2010) subsequently extended their (2008) findings using the same paradigm, this time comparing silent inner speech with silent articulated speech (mimed speech). The findings from their 2010 study confirmed the lack of any similarity bias in the inner speech only condition, but revealed a similarity bias in mimed speech that was similar to that found in the overt speech condition of their 2008 experiment.

As previously mentioned, because Oppenheim and Dell's paradigm provides a tightly-controlled experimental method for measuring the absolute and relative frequencies of the types of onset errors under consideration, it is ideally suited to testing Postma and Kolk (1993) hypothesis that stutterers make more phonological encoding errors than normally-fluent speakers. By quantifying the size of error biases, the paradigm also has the potential to highlight whether impaired flow of activation-feedback may account for any differences that may exist in stutterers' and normally-fluent speakers' speech error rates. However, two issues raised by Oppenheim and Dell's (2008, 2010) findings need further investigation before attempting to run the paradigm on participants who stutter: The first issue is that, as Oppenheim and Dell (2008) acknowledged, a plausible alternative account of the differences between their inner and overt speech conditions is that participants *under-reported* single-feature substitution errors in inner speech and, as a result, their self-reports of inner-speech errors failed to reveal the true extent to which such errors were present in their speech plans. This alternative account is supported by the study by Postma and Noordanus (1996), which found that a greater proportion of voicing errors were self-reported in overt-unmasked speech than in overt-masked speech. Their finding suggests that single-feature errors may be particularly susceptible to underreporting because there is no auditory or motor feedback to confirm that an error has been made (see also Borden, 1979; Lackner & Tuller, 1979; Postma & Kolk, 1992). On this interpretation, inner speech could be routinely fully specified at a featural level, irrespective of whether or not there is an intention to articulate it overtly.

The second issue is that Oppenheim and Dell's (2008) conclusions do not take into account the possibility that, in the overt condition, participants may have monitored their speech plans prior to articulation and edited out some errors that would otherwise have been produced overtly. They may also have engaged in covert error repair.

In our first three experiments, we therefore assess the relative extents to which under-reporting and covert editing are likely to occur in the paradigm, by comparing (normally-fluent) participants' own online self-reports of errors with independent-rater reports based on careful listening to recordings of the participants' overt iterations.

Furthermore, in Experiment 1, we also manipulate the ability of speakers to hear what they say, by employing auditory masking on half of the tongue-twisters. The intention behind this manipulation is to determine whether disproportionately fewer similar onset errors are self-reported when participants are unable to hear the words they say. Experiment 2 replicated the results of Experiment 1 using the tongue-twisters originally used by Oppenheim and Dell (2010). Experiment 3 investigated whether Oppenheim and Dell's use of an acoustic metronome (compared to our use of a visible metronome) may have caused their participants to self-report more errors. Then finally Experiment 4 (described in Section 0) uses the paradigm to compare the error phonemic rates and biases of a group of stutterers to a group of age, gender and education matched controls.

6.2. Experiment 1

Experiment 1 was closely modelled on Oppenheim and Dell's (2008) study. Participants produced a series of four-word tongue-twisters, designed in such a way that the onsets of words 1 and 4 primed participants to make onset errors on words 2 and 3. In addition, the codas of Word 3 were varied systematically to ensure that the primed-for errors on that word consisted of equal numbers of real words and non-words. We anticipated that, in line with Oppenheim and Dell (2008), our participants would report more primed-for onset-substitution errors on Word 3 that resulted in real words (compared to non-words) and also more that involved substitution of similar phonemes (compared to dissimilar phonemes). In an inner-speech condition, we expected the lexical bias to remain but similarity bias to disappear. Furthermore, with respect to overt speech, if, as Oppenheim and Dell (2008) concluded, the lexical and phonemic-similarity biases stem from activation feedback, we would expect the magnitude of these biases in participants' self-reports to resemble their magnitude in independent-rater reports. If on the other hand the biases stem in part from differential rates of under self-reporting of less salient errors, they should be less strong in self-reports compared to independent-rater reports.

As an extra manipulation, participants produced half of the tongue-twisters under conditions of auditory masking. If, as Oppenheim and Dell (2008) claimed, the differences between overt and silent conditions were because inner speech remains underspecified when there is no intention to speak aloud, we would expect the magnitude of the phonemic-similarity bias in participants' self-reported errors to be equal in the overt masked and overt unmasked conditions. If, on the other hand, the lack of a similarity bias in inner speech could be attributed to the difficulty of detecting single-feature errors in the absence of auditory feedback, the similarity bias would be expected to be smaller in overt speech under conditions of auditory masking.

6.2.1. *Method*

Participants

Thirty-two native speakers of English were recruited from the University of Edinburgh (Edinburgh, Scotland) student population for course credit. Ages ranged from 18 to 32 years old (median=18). In this and in the following two experiments, participants reported no speech, language, hearing, or visual impairments.

Materials.

Forty-eight matched sets of four-word tonguetwister sequences were generated following Oppenheim and Dell (2008) (See Appendix A). Candidate sequences were generated automatically from lists of CVC(C) (consonant–vowel– consonant[– consonant]) words with CELEX frequencies greater than 1 per million (Baayen, Piepenbrock, & Gulikers, 1995). Pronunciations were checked using the British English Example

Pronunciation Dictionary (Robinson, 1997), and also by hand. Words with ambiguous pronunciations were not used.

Each set comprised four sequences. To induce onset-phoneme substitution errors, the onset consonants of each sequence followed an ABBA pattern; however, the onsets of Words 1 and 4 (the A words) varied in each set, whereas those of Words 2 and 3 did not change. In two of the four sequences, the onsets of the A words were phonologically similar to those of the B words, differing by a single feature. In the other two, they were dissimilar, differing by two or more features.

In addition to the phonological similarity manipulation, the tongue-twister sequences were also manipulated within each set to vary the lexicality of error outcomes. This was achieved by varying the coda of Word 3 (traditionally, the most susceptible to onset substitutions in ABBA tongue-twisters; e.g. Wilshire, 1999), such that, if the onset of Word 3 was substituted with an A onset, the outcome would either be a word or a nonword.

Table 6-1: A matched set of tongue-twister sequences from Experiment 1

Outcome	Similar onsets				Dissimilar onsets			
Word	pat	cap	catch	pad	bat	cap	catch	bad
Nonword	pat	cap	cab	pad	bat	cap	cab	bad

A sample set of four tongue-twister sequences is shown in Table 6-1. In this example, /k/ differs from /p/ by one feature but from /b/ by two; substituting the onset of Word 3 with that of Word 4 would yield *patch* or *batch* in the word outcome conditions but *pab* or *bab* for nonword outcomes. Words in Position 3 were frequency matched across the experiment: Mean log frequency for the word outcome conditions was 0.97 and, for nonword outcome conditions, 0.97, *t*(47)= 0.001*, p*= .99. We also matched the frequencies of the primed-for real word outcomes (e.g. *patch* and *batch*) across similar and dissimilar conditions: Mean log frequency for the similar conditions was 1.14 and, for dissimilar conditions, also 1.14, *t*(47) =0.00, *p* =1.

Four lists were drawn up from the 48 matched sets of word sequences. Each list contained a different sequence from each of the 48 sets, arranged such that there were equal numbers of sequences in each condition in each list. Within each list, half of the sequences were assigned to the auditory-masking condition. Auditory masking was blocked, and four versions of each original list were drawn up such that all sequences appeared in masked and unmasked and in masking-first and masking-last conditions. Finally, each sequence in each of the resultant 16 lists was marked for either overt or silent recitation, such that there were equal numbers of each across other experimental conditions. This pattern was reversed to create an additional 16 lists, resulting in 32 lists of experimental items in a fully counterbalanced design.

Auditory masking was achieved using computer-generated pink noise, delivered through a set of Panasonic RP-HT225 stereo headphones. Participants' responses were captured on a Zoom H2 digital recorder and analyzed using Praat software (Boersma & Weenink, 2009).

Procedure.

The procedure was closely modelled on that of Oppenheim and Dell (2008), with three differences: (a) We used visual timing cues (white dashes on the screen) instead of an acoustic metronome to pace participants' repetitions of the word sequences (necessary because a metronome would not have been audible in conditions with auditory masking), (b) the tonguetwister sequence was not visible on the screen during experimental recitations (to eliminate potential orthographic interference), and (c) the timing cues stopped automatically after each recitation to ensure there were no distractions when participants reported errors.

Prior to beginning the experiment, participants underwent a computer-led tutorial and practice session, which included full instructions concerning the inner speech and overt speech procedures. To prevent a possible effect of silent articulation (cf. Oppenheim & Dell, 2010), the tutorial strongly emphasised the importance of not attempting to mouth sequences silently in the inner speech condition. At the beginning of each masked block, participants were instructed to adjust the fit and volume of the headphones to ensure that the loudness of the pink noise prevented them from hearing the sound of their own voice.

Tongue-twister sequences were presented in a random order on a 17-in. computer monitor. For each sequence, participants underwent a familiarization phase followed by a performance phase (See

Figure 6-1). In the familiarization phase, the tongue-twister sequence appeared in the centre of the screen, above an icon prompting participants to speak overtly (a mouth). Three seconds later, a series of four dashes appeared (one every second), acting as a visual metronome for the repetition of the words in the sequence. In the masked condition, pink noise began as the first dash appeared and lasted until the last of the four dashes disappeared. The dashes and mouth icon were then replaced by a single dot, which remained onscreen for an additional second before the mouth icon reappeared and the dash sequence started again. The dash sequence was repeated so that participants repeated each sequence aloud four times before the performance phase began.

During familiarization, participants were not aware whether repetition of the sequence during the subsequent performance phase would be silent or out loud. Once familiarization had ended, the sequence was moved to the top of the screen, and required activity was indicated centre-screen by means of the mouth icon (as used in familiarization) or a face icon representing silent repetition. At the same time, the words press ENTER to continue appeared below the icon. Pressing ENTER caused all text to disappear from the screen, leaving only the mouth or face icon visible. After 200 ms, a four-dash sequence began at a rate of one dash every 500 ms, acting as a visual metronome for the (overt or silent) repetition of the tongue-twister sequence, and (in the blocks with auditory masking) pink noise started to play over the participant’s headphones. Five hundred ms after the appearance of the fourth dash, the dashes disappeared, the pink noise (if any) ended, and the tongue-twister sequence reappeared at the top of the screen, together with the instruction report any errors and then press ENTER to continue at the bottom. Participants were instructed to report each error aloud, as fully as possible, saying, for example, “I said *pat cap patch pad*” (where *pad cap catch pad* was required). Once errors, if any, had been reported, pressing ENTER started the next four-dash sequence. Each performance phase included four repetitions of the four dashes before familiarization for the next word sequence began.

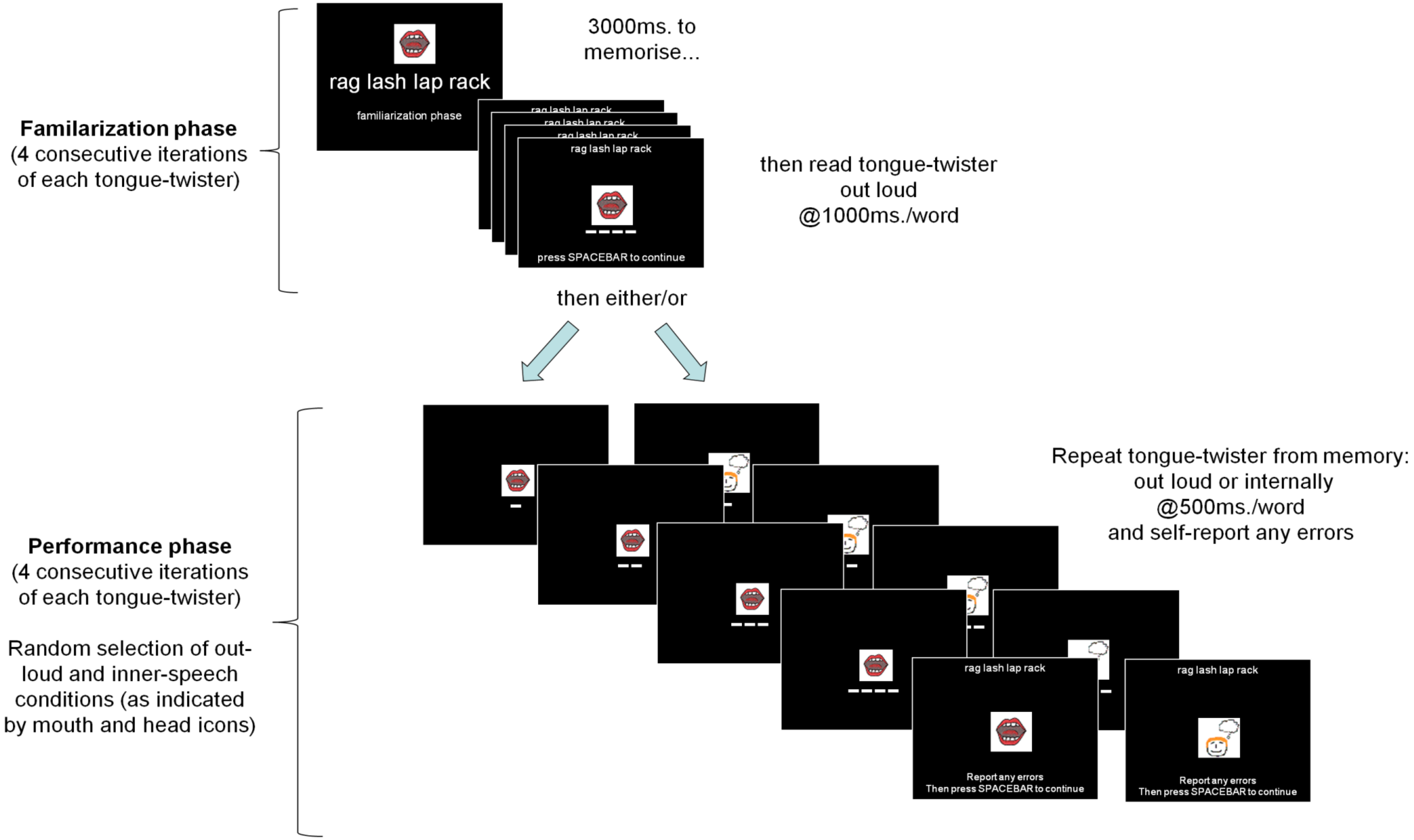


Figure 6-1. Experiment 1: Visual cues for familiarization and performance phases.

The figure depicts screen-shots of one familiarization and one performance iteration. Participants were required to speak a new word with the appearance of each new dash. In total, for each tongue-twister, participants produced four consecutive familiarization and four consecutive performance iterations before moving on to the next tongue-twister.

Transcriptions were made of participants' self-reports of their inner and overt speech errors. Additionally, the transcriber independently identified errors in the overt speech condition offline from digital recordings using PRAAT software (Boersma & Weenink, 2009). Errors were coded as primed-for onset substitutions, correct pronunciations, or other errors. To be considered as a primed-for onset substitution, an error had to consist of the substitution of the onset of a B word with that of an A word, with no concomitant change in the coda.

For example, in the tongue-twister "pat cap catch pad" → "pat cap *patch cad*" was considered to be a primed-for onset substitution, but"pat cap catch pad" → pat cap *pad catch"* was not (in practice, most of the latter type of error (in which both onset and coda changed) were indistinguishable from word-order errors. In cases where an error was followed by an overt self-repair, only the original error was coded for analysis.

Analyses.

Analyses were carried out using logistic mixed-effects regression modelling (Breslow & Clayton, 1993; DebRoy & Bates, 2004) using the lme4 package (Bates & Maechler, 2009) in R (R Development Core R Development Core Team, 2009). This approach allowed us to investigate the independent contributions of a variety of "predictor" variables (both naturally occurring and experimentally manipulated) to the (log) likelihood of making onset substitution errors relative to correct pronunciations (other errors were discarded from all analyses).

For each dependent variable of interest, we generated a base model that included an intercept, as well as random by-participant and by-item intercept variation. We then proceeded to add predictors stepwise to the model. Selection of models was based on two criteria. First, using χ^2 tests to compare model likelihood ratios, we assessed whether the fit of the model to the data was improved (as indicated by a significant increase in the absolute value of the model likelihood ratio) by the addition of each predictor. Predictors were retained if the model was improved but were removed from consideration if they did not improve the current best model. Second, where two or more predictors each significantly improved the current model, we selected the model that had the smallest log-likelihood. Once predictors representing the experimental manipulations and their relevant interactions had been exhaustively explored, the resulting model represented the best fit to the data, being a model that could not be improved by the addition of further predictors[11].

Each model includes coefficients representing the intercept and any effects of predictors. Where models were selected, the Wald statistic, calculated from each estimated coefficient and its standard error, was used to determine whether the coefficients differed significantly from zero (see Agresti, 2002).

Vigilance of self-reporting. Our first analysis evaluated the vigilance with which participants reported their own errors in overt speech. More specifically, it compared the likelihood of primed-for errors being reported in overt speech by the independent rater to the likelihood of them being reported in overt speech by the participants themselves. The independent rater's ratings were based on careful offline analysis of digital recordings of the tongue-twisters recited by participants and thus could be considered to constitute a stable baseline, whereas the participants' own ratings were made online. If participants self-

---

[11] To maintain consistency with data from these experiments already published (Corley, Brocklehurst, & Moat, 2011) random slopes have not been included in the models described in this chapter (See Section 5.2). However, for the combined analyses, equivalent models containing random slopes can be found in Appendix B (Section 12).

reported fewer errors than the independent rater, this could constitute evidence of hypo-vigilant monitoring, whereas if they self-reported more errors, then this could constitute evidence of hyper-vigilant monitoring. In addition to estimating overall vigilance, we also checked whether vigilance was affected by the addition of (a) masking; (b) similarity of onset phonemes: and (c) lexical status of the reported error (i.e. whether or not it constituted a real word).

Self-reports: Having quantified the vigilance of self-reporting of overt errors, we then carried out two further statistical analyses which focussed entirely on self-reports:

1. **Inner vs. overt speech -** focused on primed-for errors on Word 3 and compares the size of the (self-reported) phonemic similarity bias on primed-for errors in inner versus overt speech and also compares the size of the (self-reported) lexical bias in inner versus overt speech.
2. **Inner speech in isolation -** focused on primed for errors on Word 3 *in inner speech alone* and double-checks for the presence of phonemic similarity and lexical biases in inner speech.

6.2.2. *Results*

Out of a total of 6,144 experimental recitations (48 tongue-twisters, each repeated four times by 32 participants), 632 errors of any type were transcribed by the independent rater from recordings of participants' overt speech, whereas participants self-reported 510 errors in overt and 341 in inner speech.

Primed-for errors. Forty-one of the errors transcribed by the independent rater from recordings of participants' overt speech were unambiguous onset-phoneme substitution errors (cases in which the onset of Word 3 was substituted with that of an A word yet the coda remained unchanged). Participants self-reported 40 such errors in overt speech, and 37 in inner speech[12]13. Table 6-2 gives a breakdown of errors by experimental condition.

Table 6-2 Experiment 1: Onset Substitutions on Word 3

		Similar onsets			Dissimilar onsets		
Error outcome		Unmasked	Masked	(Total)	Unmasked	Masked	(Total)
Independent rater (overt speech)	Word	14	12	(26)	6	5	(11)
	Nonword	4	11	(15)	2	4	(6)
Self-reports (inner speech)	Word	9	7	(16)	7	5	(12
	Nonword	3	6	(9)	0	1	(1)
Self-reports (overt speech)	Word	8	10	(18)	4	4	(8)
	Nonword	4	6	(10)	1	3	(4)

Vigilance of self-reporting. We tested the following binomial (fixed) predictors and interactions in the order shown below. For each predictor a value of zero indicates '*no*', one indicates '*yes*':

- *similarity* (whether the substituted phonemes differed by just one feature)
- *lexicality* (whether an onset substitution would result in a word)
- *masking* (whether or not there was auditory masking),
- *rater-type* (whether or not the rating was made by the participant)
    - rater by similarity interaction
    - rater by lexicality interaction
    - rater by masking interaction

The best fitting model included the predictors: rater, lexicality and similarity: Improvement to the model fit due to *rater*: $\chi^2(1)=3.32$, $p=.069$[12]. Addition of the rater by similarity, or rater by lexicality interactions did not improve the model fit: both $ps \geq .880$, nor did addition of the lexicality by similarity interaction, $\chi^2(1)= 0.03$, $p=.852$.

Coefficients for predictors in the best-fit model (including rater) are given in Table 6-3. They reveal that after accounting for random variance, and irrespective of experimental condition, participants were only 0.65 times as likely to self-report (overt) errors compared to the independent rater. Irrespective of who was doing the rating, errors with lexical outcomes and those involving the substitutions of onset phonemes that differed by single features were reported in significantly greater numbers in overt speech. However, the lack of any rater by lexicality, and rater by similarity interactions indicates that the magnitude of the lexicality and similarity biases as calculated on the basis of the independent-rater's ratings, were not significantly different to their magnitude as calculated on the basis of the participants' self-ratings. Neither masking (as a main effect) nor the rater by masking interaction improved the model, suggesting that the ability to hear their own voices did not affect the number of primed-for errors made, nor did it affect the number self-reported. (for interactions, all $ps \geq .880$).

---

[12] The corresponding analyses presented in the Corley, Brocklehurst & Moat (2011) article included errors occurring on Words 2 (for which lexicality of outcome had not been determined). In contrast, the mixed-effects analysis presented in this thesis (Table 6-3, page 93) is based on Word 3 errors alone. As a result of this change, the addition of the 'rater' predictor only marginally improves the model, although the coefficient is nevertheless still reliable and so has been included here in the best-fit model. All other values remain substantially unchanged.

Table 6-3 Experiment 1: Vigilance of self-reporting. Model Coefficients and Probabilities for Best Fitting Models

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Overt speech (participants vs. independent rater), Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets, independent-rater	-5.57	0.36	<.001 ***
Lexicality	real-word outcome	0.64	0.23	.004 **
Similarity	similar onsets	0.93	0.24	<.001 ***
Rater	self rating	-0.43	0.22	<.050 *

Self-reports: Comparisons of inner and overt speech. The analyses of self-reported errors in inner and overt speech involved the following binomial predictors and interactions:

- *lexicality* (whether an onset substitution would result in a real word)
- *similarity* (whether the substituted phonemes differed by one feature or more)
    - lexicality by similarity interaction
- *overtness* (whether or not participants were speaking aloud),
    - overtness by lexicality interaction
    - overtness by similarity interaction

As in the previous analysis, the best fit model of self-reported primed-for errors on Word 3 included effects of lexicality and similarity. It was not improved by including their interaction: $\chi^2(1)= 2.12$, $p=.15$, overtness: $\chi^2(1)= 0.22$, $p=.64$, or the overtness by lexicality interaction $\chi^2(2)=0.78$, $p=.68$, and surprisingly, (in contrast to Oppenheim & Dell, 2008) there was no similarity by overtness interaction $\chi^2(2)=0.48$, $p=.79$.

Table 6-4 gives the coefficients of the best-fitting model and the probabilities that they differ from zero. These coefficients show that, once random variance was accounted for, participants were approximately 2.3 ($e^{0.85}$) times as likely to report errors when the outcome was a word and 2.2 times as likely when the substituted phoneme differed by a single feature, *irrespective of whether or not the error was made in inner or overt speech.*

Table 6-4 Experiment 1: Predictors of self-reported errors. Model Coefficients and Probabilities for Best Fitting Models

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Inner vs. overt speech, Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets	-5.59	0.31	<.001 ***
Lexicality	real-word outcome	0.85	0.26	<.001 ***
Similarity	similar onsets	0.79	0.26	.002 *
Inner speech only, Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets,	-7.10	1.10	<.001 ***
Lexicality	real-word outcome	2.53	1.12	.025 *
Similarity	similar onsets	2.24	1.14	.049 *
Lexicality X Similarity	real word & similar	-1.94	1.21	.110

Because of the unexpected lack of a similarity by overtness interaction, we ran an additional analysis, exploring inner speech in isolation (Word 3 only). In this analysis of inner speech in isolation, the best fitting model for errors on Word 3 included lexicality, similarity and their interaction (Improvement due to the addition of their interaction, $\chi^2(1)=4.13$, $p=.04$. The coefficients (see Table 6-4) reveal that when the error-outcomes were non-words, single feature onset-errors were 9.4 ($e^{2.24}$) times more likely to be reported than multiple-feature errors. In contrast, when the error-outcomes were real-words, although the trend was in the same direction, the magnitude of the similarity bias was not reliable.

6.2.3. *Discussion*

In Experiment 1, although, overall, participants tended to under-report the overt, primed-for errors they made, the lack of any rater-by-lexicality or rater-by-similarity interactions suggests that, at least in overt speech, such under-reporting occurred approximately to the same extent irrespective of how salient the errors were. This is in line with earlier findings by Nooteboom (2005b), that self-monitoring of overt speech is not sensitive to the lexical status of errors or to the phonetic distance between error and target. The finding that auditory masking did not affect the likelihood of primed-for errors being either self-reported or reported by the independent rater further suggests that the vigilance of self-reporting was not affected by the ability of participants to hear what they said. In this respect the current findings differ from those of Postma and Noordanus (2006).

In Experiment 1, both a phonemic similarity bias and a lexical bias were found in both overt and inner speech conditions, although in inner speech, the similarity bias was only reliable for errors that resulted in the production of non-words. The finding that the lexicality and similarity biases self-reported in inner-speech were in the same direction as those reported in overt speech (both by participants themselves as well as by the independent rater) fully supports an activation feedback account in which feedback of activation occurs not only from the phoneme to the lexical level of encoding *but also* from the feature to the phoneme level of encoding. This latter finding appears incompatible with Oppenheim and Dell's (2008) finding that there was no similarity bias in participants' inner-speech self-reports, although it is not incompatible with Oppenheim and Dell's (2010) finding that the similarity effect in inner speech is weaker than in overt speech, as there was a numeric trend in that direction.

These findings are potentially compatible with an activation-feedback account or with a mixed account, such as proposed by Hartsuiker et al. (2005) in which the biases are influenced by both feedback and editing effects. Whatever the case, they do not seem to be compatible with an account that attributes the biases entirely to poorer editing (and under-reporting) of less salient errors, because such an account would predict self-reporting biases in the opposite direction to the similarity and lexical biases that were found (see Section 4.5.2 & 4.5.3).

Overall, the numbers of errors self-reported in the experiment were substantially lower than the numbers reported by participants in the Oppenheim and Dell (2008) experiment[13] . It is not clear why they were lower, and it is possible that with higher error rates, different results may have emerged. Because of the difference, with respect to the phonemic-similarity bias, between the current findings and those of Oppenheim and Dell (2008), a primary motivation for Experiment 2 was to see if our findings (especially with respect to the phonological similarity effect in inner speech) could be replicated using Oppenheim and Dell's own set of materials.

6.3. Experiment 2

Experiment 2 was a replication of Experiment 1, using the same tongue-twisters that were used by Oppenheim and Dell (2010). Because there were fewer tongue-twisters than in our previous experiments, we used a larger number of participants (48) to maintain equivalent power. To enable a more direct comparison with Oppenheim and Dell (2008), we did not manipulate masking in this experiment. In all other respects, Experiment 2 was identical to Experiment 1.

6.3.1. Method

Participants

Forty-eight native speakers of English were recruited from the University of Edinburgh student population for course credit. Ages ranged from 18 to 23 years old (median=18).

---

[13] For comparison, Oppenheim & Dell's (2008) participants self-reported 1217 overt errors of any kind on 6144 recitations, including 125 primed-for errors on word 3.

Materials

We used the 32 matched sets of four-word tonguetwister sequences originally used by Oppenheim and Dell (2010); 31 of these were identical to those used by Oppenheim & Dell (2008). All sequences used real words, with onsets arranged in an ABBA pattern; as in Experiment 1, the onsets of the A words were manipulated such that they were either phonemically similar or dissimilar to the B onsets, and the coda of Word 3 was manipulated such that the substitution of the onset of Word 3 with an A onset would result in either a word or a nonword. Counterbalancing proceeded as in Experiment 1. Since there was no manipulating of auditory masking in the present experiment, the counterbalancing procedure resulted in eight experimental lists.

Procedure.

The procedure was identical to that of Experiment 1, with the exception that participants were not required to wear headphones since all tongue twisters were recited without auditory masking.

Analyses.

As in Experiment 1, we carried out mixed-effects analyses of reporting vigilance in overt speech (participants vs. independent rater), analyses of primed-for errors in inner versus overt speech (Word 3), and inner speech only (Word 3).

6.3.2. *Results*

Out of a total of 6,144 experimental recitations, Four hundred and thirteen errors were transcribed by the independent rater from recordings of participants’ overt speech, whereas participants self-reported 361 in overt and 225 in inner speech.

Primed-for errors. Fifty five of the errors transcribed by the independent rater from recordings of participants’ overt speech were unambiguous onset-phoneme substitution errors (cases in which the onset of Word 3 was substituted with that of an A word yet the coda remained unchanged). Participants self-reported 44 such errors in overt speech, and 31 in inner speech. Table 6-5 gives a breakdown of errors by experimental condition.

Table 6-5. Experiment 2: Onset substitutions on Word 3

Error outcome		Similar onsets	Dissimilar onsets	(Total)
Independent rater	Word	24	16	(40)
(overt speech)	Nonword	14	1	(15)
Self-reports	Word	16	7	(23)
(inner speech)	Nonword	6	2	(8)
Self-reports	Word	20	11	(31)
(overt speech)	Nonword	12	1	(13)

Vigilance of self-reporting. As in Experiment 1, we ran a set of analyses comparing participants' self-reported errors to those transcribed from the recordings by an independent rater. This time, once random variance was accounted for, the best fitting model included lexicality, similarity and also the lexicality by similarity interaction: Improvement due to adding the lexicality by similarity interaction: $\chi^2(1)=7.07$, $p<.008$.

Although participants self-reported lower numbers of primed-for errors than were recorded by the independent rater, adding *rater* as an additional predictor did not further improve the model fit: $\chi^2(1)=1.23$, $p=.267$. As in Experiment 1, neither the 'rater by lexicality' or' rater by similarity' interactions improved the model further: all $\chi^2(2)\leq 1.38$, *p*s≥.501, indicating that the magnitudes of the lexicality and similarity biases as calculated on the basis of the independent rater's ratings, were not significantly different to their magnitudes as calculated on the basis of the participants' ratings.

Coefficients for the predictors are given in Table 6-6. As in Experiment 1 they reveal significant phonemic similarity and lexical biases. In contrast to Experiment 1, they also additionally reveal an (overt) lexicality by similarity interaction. Specifically, after taking participant and item variance into account, irrespective of who is doing the rating, compared to ‘dissimilar’ onset-errors, ‘similar’ onset-errors were 10.3 ($e^{2.33}$) times as likely to be reported if they had non-word outcomes, or 1.6 ($e^{0.49}$) times as likely to be reported if they had real-word outcomes. Both ratios were reliably different from 1.

Table 6-6. Experiment 2: Vigilance of self-reporting. Model Coefficients and Probabilities for Best Fitting Models

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Overt speech (participants vs. independent rater), Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets,	-8.57	0.94	<.001 ***
Lexicality	real-word outcome	2.85	0.93	.002 **
Similarity	similar onsets	2.33	0.82	.005 **
Lexicality by similarity interaction		-1.84	0.87	.035 *

Self-reported errors: Comparisons of inner and overt speech. As in Experiment 1, we performed two separate analyses of participants’ self-reported errors (on Word 3 inner & overt speech combined, and on Word 3 inner speech in isolation). Together these analyses tested the effects of overtness, lexicality, and similarity and relevant interactions on the likelihood of participants self-reporting primed-for errors. Coefficients for each model and the probabilities that they differ from zero are given in Table 6-7.

In the analysis of primed-for errors in inner & overt speech combined, once random variance was accounted for, the best fit model included factors of lexicality and similarity but was not significantly improved by including their interaction, $\chi^2(1)= 2.60$, $p=.11$. Adding overtness marginally improved the model, $\chi^2(1)= 3.14$, $p=.08$, reflecting the fact that fewer errors were reported in inner speech. Importantly, overtness did not interact with either lexicality or phonemic similarity: both $ps \geq .78$. The coefficients for the best-fitting model reveal that, regardless of overtness and taking random effects into account, participants were approximately 2.7 (= $e^{0.99}$) times as likely to report errors when the outcome was a word and 2.7 times as likely when the substituted phoneme differed by a single feature.

When inner speech was analyzed in isolation, the best fit model of primed-for errors (on Word 3) included effects of lexicality: $\chi^2(1)= 5.78$, $p= .02$, and similarity: $\chi^2(1)= 5.53$, $p=.02$, but not their interaction: $\chi^2(1)= 0.07$, $p= .793$. The coefficients of the best-fitting model reveal that participants were 2.9 times as likely to report errors resulting in words and 2.5 times as likely to report the substitutions of similar phonemes.

Table 6-7. Experiment 2: Predictors of self-reported errors. Model Coefficients and Probabilities for Best Fitting Models

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Inner vs. overt speech, Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets	-6.43	0.39	<.001 ***
Lexicality	real word outcome	0.99	0.27	<.001 ***
Similarity	similar onsets	0.99	0.27	<.001 ***
Inner speech only, Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets	-6.24	0.54	<.001 ***
Lexicality	real word outcome	1.08	0.49	.027 *
Similarity	similar onsets	0.90	0.43	.038 *

6.3.3. *Discussion*

In Experiment 2, although overall, participants once again self-reported numerically fewer primed-for errors than the independent rater, the difference was not reliable. In common with Experiment 1, the lack of any rater-by-lexicality or rater-by-similarity interactions suggests that, even if under-reporting of primed-for errors did occur, it occurred to a similar extent irrespective of how salient the errors were. However, some caution needs to be exercised in arriving at such a conclusion because the overall number of errors involved was again low and there may not have been sufficient power to reveal such interactions.

Consistent with Experiment 1 and with Oppenheim and Dell (2008, 2010), we found a lexical bias in both overt and inner speech conditions. However once again, unlike Oppenheim and Dell (2008, 2010), there was not a significant interaction of phonemic similarity with overtness. In the current experiment, we found a *lexicality by similarity* interaction in overt speech (although in post-hoc analyses of independent-rater and self reports in isolation, the interaction was not reliable: independent-rater p=.073, self-ratings p=.133). This interaction was similar to in Experiment 1, in that the similarity effect was stronger on errors with non-lexical outcomes, but unlike Experiment 1 it was not found in inner speech.

As in Experiment 1, the overall numbers of errors reported were low, indeed, somewhat lower than Experiment 1 and significantly lower than Oppenheim and Dell's (2008, 2010) experiments (despite using their tongue-twisters). Importantly, overall, as in Experiment 1, there was a high degree of correspondence between participants' own self-ratings and the independent rater's ratings with regards to the strength of the (overt) similarity and lexical biases, suggesting that, at least in overt speech, participants were not disproportionately underreporting the less salient errors. However, these findings differ substantially from those of Oppenheim and Dell (2008) insofar as, in the inner-speech condition of their experiment, the phonemic similarity bias was entirely absent.

Having ruled out the possibility that the differences in the findings of our two labs resulted from our use of different tongue-twisters, Experiment 3 investigates the possibility that the differences may stem from the different types of metronomes used by the two labs: visual in our lab, and acoustic in the Oppenheim and Dell's lab. Experiment 3 thus replicated the procedure from our Experiment 1, using an acoustic metronome. A further motivation for performing this extra experiment was to provide extra data for a subsequent combined analysis, to investigate the reliability of the paradigm across participant groups and to determine whether our failure to find the same *similarity by overtness* interaction as Oppenheim and Dell (2008, 2010) is related to a lack of power due to the relatively low numbers of primed-for errors elicited in our experiments.

6.4. Experiment 3

Experiment 3[14] was a replication of Experiment 1, except that (a) we used an acoustic (instead of a visual) metronome to pace participants' iterations of the tongue-twisters; (b) the auditory-masking condition was not included; and (c) participants self-reported their errors by typing them directly into the computer (instead of stating them verbally).

6.5. Method

Participants

Thirty-two native speakers of English were recruited from the University of Edinburgh student population for course credit. Ages ranged from 17 to 45 years old (median=18).

Materials

We used the same 48 matched sets of four-word tonguetwister sequences as originally used in Experiment 1 (See Appendix A). Since there was no manipulation of auditory masking in the present experiment, the counterbalancing procedure resulted in eight experimental lists (as in Oppenheim and Dell, 2008).

---

[14] We wish to acknowledge the valuable help of Sam Miller with the running of participants for this experiment.

Procedure.

The procedure was identical to that of Experiment 1, except for the following: (a) participants did not wear headphones (these were unnecessary as all tongue twisters were recited without auditory masking); (b) instead of speaking in time to dashes as they appeared on the screen, participants spoke in time to a digital recording of an acoustic metronome beat delivered through the computer speaker; (c) participants self-reported their errors by typing instead of speaking. The change in the self-reporting mode brought it into line with the tongue-twister Experiment 4 which compared speech-errors in people who stutter and normally-fluent controls.

During both the familiarisation and testing phases, the acoustic metronome was activated in an identical way to the visible metronome in the previous experiments. The required speech rates were also identical to those of the previous experiments. During the testing phase, 500 ms after the onset of the final beat of each of the four tongue-twister recitations, the tonguetwister sequence reappeared at the top of the screen, together with an instruction to "type any errors and then press ENTER to continue" at the bottom. Participants were instructed to type, as fully as possible, what they had actually said, for example "rag lap rash rap" (when they should have said "rag lap *lash* rap"; the /r/ in rash being an anticipation of the /r/ in rap). They were instructed to type one or more question-marks in the relevant places if they could not remember what they had said for a word or 'X's if they had completely omitted all or part of a word. Once errors, if any, had been reported, pressing ENTER started the next four-dash sequence.

Analyses.

As in the previous experiments, we carried out mixed-effects analyses of primed-for errors in overt speech only (participants vs. independent rater); inner versus overt speech, and inner speech in isolation.

6.5.1. *Results*

Out of a total of 6,144 experimental recitations, nine hundred and seventy nine errors were transcribed by the independent rater from recordings of participants' overt speech, whereas participants self-reported 725 in overt and 483 in inner speech.

Primed-for errors. Eighty of the errors transcribed by the independent rater from recordings of participants' overt speech were unambiguous onset-phoneme substitution errors (cases in which the onset of Word 3 was substituted with that of an A word yet the coda remained unchanged). Participants self-reported 46 such errors in overt speech, and 45 in inner speech. Table 6-8 gives a breakdown of errors by experimental condition.

Table 6-8 Experiment 3: Onset substitutions on Word 3

Error outcome		Similar onsets	Dissimilar onsets	(Total)
Self-reports (inner speech)	Word	23	13	(36)
	Nonword	7	2	(9)
Self-reports (overt speech)	Word	19	14	(33)
	Nonword	10	3	(13)
Independent rater (overt speech)	Word	34	22	(56)
	Nonword	20	4	(24)

Vigilance of self-reporting. As in the previous experiments, we ran a set of analyses comparing participants' self-reported errors to those transcribed from the recordings by an independent rater. This time, once random variance was accounted for, the best fitting model included rater, and also lexicality, similarity, the lexicality by similarity interaction. Improvement due to adding rater: $\chi^2(1)= 6.01$, $p<.014$ (coefficients are given in Table 6-9). The coefficient for rater revealed that, irrespective of lexicality and similarity, and after accounting for random variance, participants were only 0.54 times as likely to report primed-for errors compared to the independent rater. As in the previous experiments, neither the 'rater by lexicality' nor' rater by similarity' interactions improved the model: both $\chi^2(1)\leq$

0.41, *ps*≥.520., indicating that the magnitudes of the lexicality and similarity biases as calculated on the basis of the independent rater's ratings, were not significantly different to their magnitudes as calculated on the basis of the participants' ratings.

Table 6-9. Experiment 3: Vigilance of self reporting. Model Coefficients and Probabilities for Best Fitting Model.

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Overt speech (participants vs. independent rater), Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets, independent-rater	-5.96	0.48	<.001 ***
Lexicality	real word outcome	1.79	0.44	<.001 ***
Similarity	similar onsets	1.74	0.44	<.001 ***
Lexicality by similarity interaction		-1.23	0.50	.015 *
Rater	self rating	-0.62	0.20	.002 **

The coefficients for lexicality, similarity and their interaction reveal that irrespective of who was doing the rating, compared to 'dissimilar' onset-errors, 'similar' onset-errors were either 5.7 ($e^{1.74}$) times as likely to be reported if they had non-word outcomes, or 1.7 ($e^{0.51}$) times as likely to be reported if they had real-word outcomes. Both odds ratios were reliably different from 1.

Self-reported errors: Comparisons of inner and overt speech. As in the previous experiments, we performed two separate analyses of participants' self-reported errors (on Word 3 inner & overt speech combined, and on Word 3 inner speech in isolation). Together these analyses tested the effects of overtness, lexicality, and similarity and relevant interactions on the likelihood of participants self-reporting primed-for errors. Coefficients for each model and the probabilities that they differ from zero are given in Table 6-10.

In the analysis of primed-for errors in inner & overt speech combined, once random variance was accounted for, the best fit model included factors of lexicality and similarity (improvement due to adding the similarity predictor: $\chi^2(1)= 10.305$ $p<.001$), but was not improved by the addition of the lexicality by similarity interaction: $\chi^2(1)= 2.26$, *p*=.133, or by addition of the overtness predictor: $\chi^2(1)= 0.34$, *p*=.562. As in the previous experiments, overtness did not interact with either lexicality or similarity: both $\chi^2(2) \leq 1.08$, *ps* ≥.583. The coefficients reveal that, regardless of overtness and taking random effects into account, participants were approximately 3.25 (= $e^{1.18}$) times as likely to report errors when the outcome was a word and 2.0 times as likely when the substituted phoneme differed by a single feature.

When inner speech was analyzed in isolation, the best fit model included effects of lexicality, and similarity (improvement due to adding the similarity predictor: $\chi^2(1)= 6.20$ *p*=.013). Adding the lexicality by similarity interaction did not further improve the model: $\chi^2(1)= 0.74$, . *p*=.389. The coefficients reveal that participants were 4.1 times as likely to report errors resulting in words and 2.2 times as likely to report the substitutions of similar phonemes.

Table 6-10. Experiment 3: Self reported errors. Model Coefficients and Probabilities for Best Fitting Models.

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Inner vs. overt speech, Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets	-5.69	0.33	<.001 ***
Lexicality	real word outcome	1.18	0.26	<.001 ***
Similarity	similar onsets	0.71	0.23	.002 **
Inner speech only, Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets	-6.00	0.47	<.001 ***
Lexicality	real word outcome	1.42	0.39	<.001 ***
Similarity	similar onsets	0.78	0.33	.020 *

6.5.2. *Discussion*

In Experiment 3, the overall numbers of primed-for rater-reported as well as self-reported error rates were marginally higher than in our previous two experiments (see the marginal effect of *metronome* in the combined analysis in 6.6.1, for confirmation of this) although there were still fewer error reports than in Oppenheim and Dell's (2008) experiment[13]. The increased numbers of errors may have been related to the adoption of an acoustic metronome similar to that used in Oppenheim and Dell's experiment[15]. Once again, participants under-reported their overt primed-for errors, and this time the difference reached significance. As in our previous experiments, the lack of any rater-by-lexicality or rater-by-similarity interactions suggests that under-reporting of overt primed-for errors occurred approximately to the same extent irrespective of how salient the errors were.

Consistent with our previous experiments and with Oppenheim and Dell (2008, 2010), we found a lexical bias in both overt and inner speech conditions. However once again, in contrast to Oppenheim and Dell (2008, 2010) but in common with our previous two experiments, there was no interaction of phonemic similarity with overtness. Indeed this time, numerically, the similarity bias was slightly *larger* in inner speech than in overt speech. As with our Experiment 2, a significant lexicality by similarity interaction was found in overt but not in inner speech.

Before considering the implications of our findings further, we present three sets of combined analyses of the combined data from the Experiments described above.

---

[15] The fact that there were also a higher number of independent-rater reports in this experiment appears to rule out the possibility that the typed self-reporting mode accounted for the higher numbers of errors. This conclusion is further supported by the finding of relatively low error rates in the 'controls' experiment of the next set of experiments described in Chapter 0, which used a visual metronome together with typed self-reports.

6.6. Analyses of data from Experiments 1, 2 and 3 combined

Because the number of primed-for errors that occurred in each of our individual experiments was relatively low, it is possible a lack of power may have contributed to our failure to find the overtness by similarity interaction found by Oppenheim and Dell (2008; 2010). Lack of power may also have meant that differences in the relative size of error biases across reporters and conditions may have been missed. Therefore, to check whether this was the case, and to determine whether the use of different groups of participants and/or any of the between-experiment manipulations made a significant difference to reporting vigilance or to the relative magnitude of the error biases, we report the findings of a series of combined analyses in which the results of the three experiments reported above are considered together.

6.6.1. *Vigilance of self-reporting.*

As with the individual experiments, this combined analysis of vigilance of self-reporting is based on Word 3 data from self and independent-rater reports of primed-for errors. We tested the same predictors as were used in the equivalent analyses of the individual experiments: i.e. *similarity*, *lexicality* and *rater*. In addition to these, we also tested two further factorial predictors: *experiment* (with 3 levels, one for each experiment) to test for gross differences between the experiments; and *metronome* (with 2 levels: visual and acoustic) to determine what (if any) effect the type of metronome had on the vigilance with which participants self-reported primed-for errors.

Results

The best fitting model for the combined analysis of vigilance of self-reporting included rater, lexicality, similarity, and the lexicality by similarity interaction. Improvement to the model as a result of adding rater: $\chi^2(1)= 6.97$, $p<.001$.

The coefficients of the predictors in the best-fit model (see Table 6-11) reveal that (irrespective of experiment) participants were only 0.64 times as likely to self-report errors as was the independent rater. The failure of the *rater by lexicality* and *rater by similarity* interactions to improve the model further (both $ps \geq .983$), confirmed that the vigilance of participants' self reporting of (overt) errors was not influenced to a different extent by how phonemically similar their errors were to the intended targets and did not depend on whether or not those errors constituted real words.

The coefficients also reveal that the similarity and lexical biases in participants' own self-reports were not significantly different to those found in the independent rater's reports. Specifically, after taking subject and item variance into account and irrespective of who was doing the rating, compared to 'dissimilar' onset-errors, 'similar' onset-errors were 4.7 ($e^{1.55}$) times as likely to be reported if they had non-word outcomes, or 1.8 ($e^{0.61}$) times as likely to be reported if they had real-word outcomes. Both ratios were reliable. (See Figure 6-3 b and c for graphical comparison of equivalent separate models for overt self and independent-rater ratings)

It is noteworthy that the addition of 'metronome' as a predictor marginally further improved the model, $\chi^2(1)= 3.07$, $p=.079$, its coefficient ($e^{0.54}$ $p=.076$) suggesting that the use of an auditory metronome in Experiment 3 may have resulted in higher speech error rates in that experiment (compared to the other two experiments combined). Adding the metronome by rater interaction did not improve the model $\chi^2(2)= 4.40$, $p=.111$, indicating that any increase in error rates that may have occurred due to the use of an acoustic metronome affected self-reports and independent-rater reports to a roughly similar extent and thus did not result in a loss of monitoring vigilance.

Table 6-11 Combined analysis Experiments 1, 3 & 4. Monitoring Vigilance. Model Coefficients and Probabilities for Best Fitting Model (see Appendix B Section 12.1, for equivalent analysis with random slopes)

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Overt speech (participants vs. independent rater), Word 3 onset errors				
(Intercept)	nonword outcomes, dissimilar onsets, independent rater	-6.24	0.31	<.001 ***
Lexicality	real word outcomes	1.51	0.27	<.001 ***
Similarity	similar onsets	1.54	0.27	<.001 ***
Lexicality by similarity interaction		-0.94	0.31	.002 **
Rater	self-rating	-0.44	0.13	<.001 ***

Discussion

This combined analysis of self- and independent-rater reports of primed-for errors in overt speech clearly shows that participants in all three experiments tended to under-report the primed-for errors they made. However, it also shows that the extent of under-reporting did not differ significantly across experiments or across conditions. Thus, in line with the analyses of the individual experiments in isolation, it did not reveal any evidence to suggest that the extent of participants' under-reporting was influenced by the lexical status of the error outcome (i.e. whether or not it constituted a real word) or its phonemic similarity to the original target word. The analysis also found that monitoring vigilance was not affected by the tonguetwisters used; the presence of auditory masking; or the type of metronome used (see Figure 6-3(b) & (c) on page 118 for graphical representations of the fitted models).

6.6.2. *Accuracy of self-reporting*

Although the combined analysis of *vigilance* of self-reporting (described above) revealed that participants self-reported fewer overt primed-for errors than the independent rater, it does not provide any insight into how *accurate* their self-ratings were. It is thus possible that occasions may have arisen where both the self rater and also the participant identified a word as containing an error, yet the description of that error provided by the participant may have differed from that provided by the independent rater. It is also possible that the degree of self-reporting accuracy may have been different in the different experimental conditions, despite the overall level of self-reporting vigilance being the same. To explore this possibility, we present below, an additional (combined) analysis which focuses specifically on the accuracy with which primed-for errors were self-reported.

This accuracy combined analysis is predicated on the presumption that the independent rater's reports can be considered to form a stable baseline against which self-reports can be judged for (relative) accuracy. This approach, which was also used in a comparable analysis of accuracy by Postma and Noordanus (1996), can be justified insofar as the independent-rater's reports were based on careful offline analyses of digital recordings whereby tokens were replayed as often as necessary in order to arrive at as reliable a judgment as possible. Unlike the combined analysis of vigilance, this analysis included only words containing primed-for onset errors (as identified by the independent-rater). Self-ratings of these same words were then classified as 'accurate' or 'inaccurate' according to whether or not they exactly matched the independent rater's rating. The dependent variable was thus the ratio of accurate to inaccurate self-ratings.

In the combined analysis of accuracy, there were fewer eligible self-reported primed-for errors on Word 3 compared to the combined analysis of vigilance (118 compared to 244)[16], so to compensate for this, we decided to additionally include primed-for errors on Words 2.

---

[16] The meta-analysis of vigilance contained 126 self-reported primed-for errors that were not coded as primed-for errors by the independent rater, and were thus not eligible for the accuracy analysis. The majority of these were errors in which the independent rater had identified both the onset and rime as different to the target word, but had been self-reported by the participant as only differing on the onset.

This provided an extra 22 self-reported and 41 rater-reported errors and so helped increase its power. However, it meant that the influence of the lexical status of error-outcomes on error likelihood could not be tested (because lexicality of error outcomes was not controlled for on Word 2). In addition to the predictors tested in the corresponding combined analysis of vigilance, an additional predictor for ‘*total errors*’ (the total number of independent-rater reported errors across the four iterations of the tongue-twister sequence in which the primed-for error occurred) was also tested. This extra predictor, the value of which could range from zero to sixteen, was added to test whether participants found it more difficult to accurately recall their errors when the overall density of errors on a tongue-twister was higher.

Results

Once random variance was accounted for, the best-fit model predicting the accuracy of self-reporting included the following predictors: *total rater-reported errors*; *similarity; metronome*; and the *similarity by metronome* interaction (improvement due to adding the similarity by metronome interaction: $\chi^2(1)= 7.16$ $p=.007$. Masking did not significantly improve the model: $\chi^2(1)= 3.31$ $p=.69$. Although addition of the *masking by similarity* interaction did further improve the model $\chi^2(2)= 6.09$ $p=.048$, the coefficient for the interaction was not reliable: $p=.132$, so masking and its interaction with similarity have been left out of the model cited below.

The coefficients for the best-fit model are shown in Table 6-12. The coefficient for *total errors* reveals that primed-for errors were less accurately self-reported by participants when they occurred in tongue-twister sequences that contained more (rater-reported) errors overall. On average, there were three rater-reported errors (of any kind in any position) across the four iterations of each tongue-twister sequence included in the analysis. After taking random effects into account, for each extra (rater-reported) error that occurred, participants were only 0.8 ($e^{-0.22}$) times as likely to accurately self-report the error.

Independently of the effect of the total number of errors, the coefficients for *similarity, metronome,* and their interaction further revealed that the accuracy of self-reporting was also influenced by the similarity of the onset-phonemes in a tonguetwister, and that the size of this effect differed depending on the type of metronome used. With an acoustic metronome, errors in tongue-twisters with similar onsets were 0.3 ($e^{-1.20}$) times as likely to be correctly recognised compared to those with dissimilar onsets, whereas with a visual metronome the likelihood of similar onset errors being correctly recognised was not reliably different to the likelihood of dissimilar onset errors being correctly recognised (p=.134).

Table 6-12. Combined analysis Experiments 1-3. Accuracy of Self Reporting. Model Coefficients and Probabilities for Best Fitting Model

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Overt speech: accuracy of self-reporting primed-for errors				
Intercept	3 rater-reported errors*, dissimilar onsets, acoustic metronome,	0.83	0.46	.069
Total errors reported by Independent rater	+1	-0.22	0.09	.013 *
similarity	similar onsets	-1.20	0.51	.019 *
Metronome (visual)	visual metronome	-0.68	0.61	.264
Similarity by metronome (visual)		1.86	0.68	.006 **

***For ease of interpretation, the predictor for *total errors* is centred on its mean (of 3).**

Discussion

The above combined analysis of participants' self-reporting accuracy reveals, unsurprisingly, that self-reporting accuracy declines as overall error-density increases, suggesting that participants had more difficulty recalling the details of the primed-for errors they made on the tongue-twisters in which they made more errors overall. Of particular interest is the finding that when an acoustic metronome was used, self-reporting accuracy of primed-for errors involving similar onset phonemes was significantly lower than self-reporting accuracy of errors involving dissimilar onset phonemes, whereas when a visual metronome was used, the difference between similar and dissimilar conditions was not significant. Figure 6-2, below, shows a graphical representation of the fitted model for accuracy in the four conditions (in terms of percentages of errors correctly recognised) for tongue-twisters containing an average number of rater-reported errors.

There were also signs of a similar (although not significant) trend suggesting a reduction in self-reporting accuracy for tongue-twisters with similar onsets under conditions of auditory masking. These findings provide some preliminary evidence to suggest that similar-phoneme substitution errors may be somewhat harder for speakers to accurately identify than dissimilar-phoneme substitution errors when auditory feedback is prevented or when the participant's attention is distracted by a secondary acoustic signal (such as an acoustic metronome).

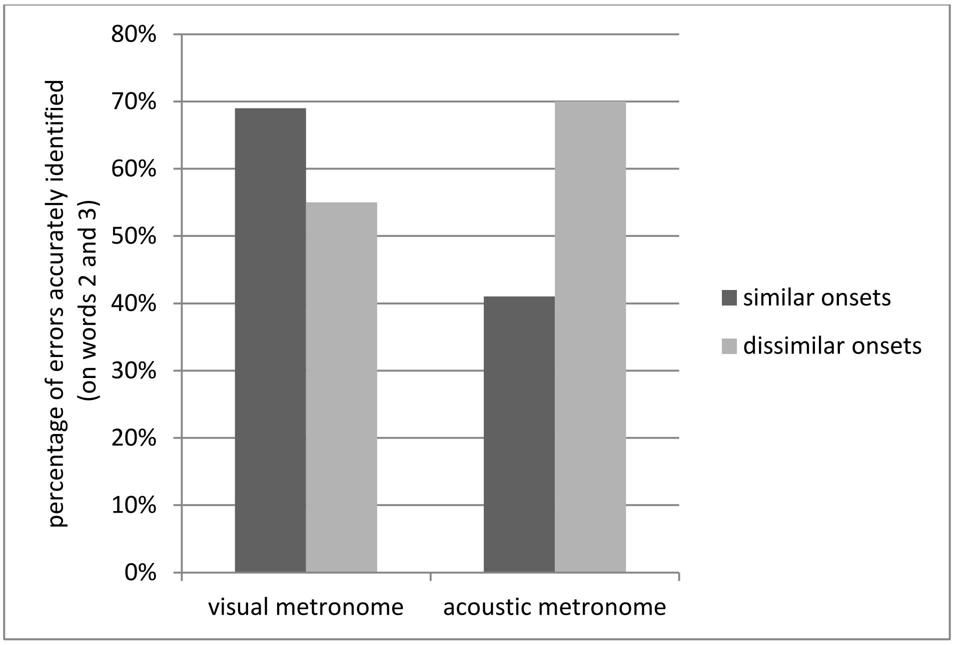


Figure 6-2. Experiments 1 to 3: The accuracy of self-reporting of primed-for errors on Words 2 and 3.
Influence of metronome type and the similarity of onset-phonemes. Fitted values are from the model presented in Table 6-12. The columns show percentages for an average tonguetwister sequence containing 3 rater-reported errors (of any kind) across the 16 words. Note. the effect of similarity for an acoustic metronome was significant: *p*=.019.

Taken together with the findings from the previous combined analysis (of self-reporting vigilance) these findings show that even when participants did identify errors they had made, they frequently failed to accurately report them. It is noteworthy that the reduction in accuracy of self-reporting was not accompanied by a corresponding reduction in the size of the self-reported phonemic-similarity bias (in either overt or inner speech). This suggests that participants were aware of the ABBA alternating pattern of the tongue-twister word onsets, and may have tended to use a situation-specific monitoring criterion: ‘did I get the onsets mixed up?’. It could be that they only monitored for exact matches of words if and when auditory monitoring was unimpeded (i.e. no masking and no distracting metronome beat). Such an approach to monitoring is broadly in line with the (Hartsuiker et al., 2005) ‘smart monitor’ account of self-monitoring (see also Nooteboom, 2005b).

6.6.3. *Self-reported errors: Lexical bias and Phonemic similarity.*

As with the individual experiments, we performed two separate combined analyses of participants’ self-reported errors: on Word 3 inner & overt speech combined; on Word 3 inner speech in isolation. Together these meta- analyses tested the effects of overtness, lexicality, and similarity and relevant interactions on the likelihood of participants self-reporting primed-for errors. We also tested a further factorial predictor: *metronome* (with 2 levels: visual and acoustic) to determine what (if any) effect the type of metronome had on the likelihood of participants self-reporting primed-for errors.

Results

In the combined analysis of primed-for errors on Word 3 (inner & overt speech combined), once random variance was accounted for, the best-fit model included predictors of lexicality and similarity and the lexicality by similarity interaction (improvement due to adding the lexicality by similarity interaction: $\chi^2(1)= 7.18$ p=.007, The model was not significantly improved by the addition of *experiment* or by its interactions with lexicality or similarity (all $\chi^2(4) \leq 2.46$, *ps*≥.651); nor was it improved by the addition of overtness: $\chi^2(1)= 2.62$ *p*=.106, or by the overtness by similarity interaction: $\chi^2(2)= 2.63$, *p*=.269. Nor was it significantly improved by addition of metronome or any other two or three-way interaction (all *ps*≥.128).

The model coefficients (see Table 6-13) show that, regardless of overtness and masking, and taking random effects into account, errors resulting in real-word outcomes were more likely to be self-reported, and errors involving substitutions of similar onset phonemes were also more likely to be self-reported. More specifically, for errors with non-word outcomes, those involving substitutions of similar onset phonemes were 4.7 (= $e^{1.55}$) times as likely to be self-reported than those involving substitutions of dissimilar onset phonemes, whereas for errors with real-word outcomes, those involving substitutions of similar onset phonemes were 1.9 times (= $e^{0.62}$) as likely to be self-reported than those involving substitutions of dissimilar onset phonemes.

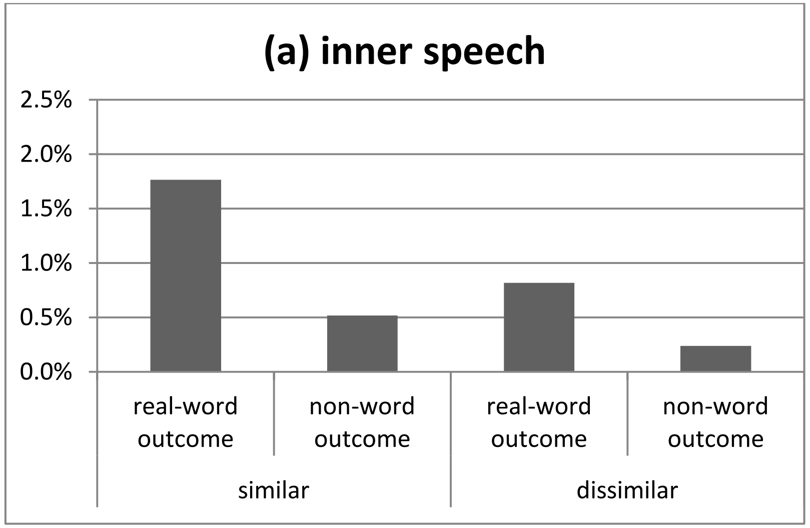
When inner speech was analyzed in isolation, the best fit model included effects of lexicality, and similarity (improvement due to adding the similarity predictor: $\chi^2(1)= 15.54$ *p*<.001). Addition of the lexicality by similarity interaction only marginally further improved the model, $\chi^2(1)= 3.31$ *p*=.069. Addition of experiment did not improve the model: $\chi^2(2)= 1.55$ *p*=.461. The coefficients of the best-fitting model are shown in Table 6-13. They show that participants were 4.1 times as likely to report errors resulting in words and 2.2 times as likely to report the substitutions of similar phonemes. See Figure 6-3(a) for a graphical representation self-reported errors in inner speech derived from the fitted model. (for comparison, Figure 6-3(b) & (c) on page 118 show the equivalent fitted models for overt self reported errors in isolation and for overt independent-rater reported errors in isolation)

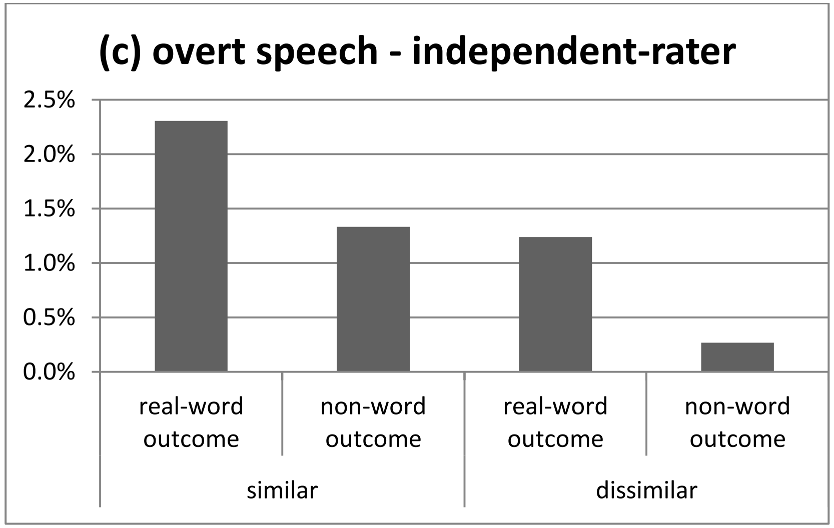
Table 6-13. Combined analyses of Experiments 1-3. Predictors of self-reported errors: Model Coefficients and Probabilities for Best Fitting Models. See Appendix B Section 12.1, for equivalent analysis with random slopes)

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Inner vs. overt speech, Word 3 onset errors				
(Intercept)	nonword outcome, dissimilar onsets	-6.33	0.25	<.001 ***
Lexicality	real word outcomes	1.59	0.23	<.001 ***
Similarity	similar onsets	1.55	0.23	<.001 ***
Lexicality by similarity interaction		-0.93	0.26	.013 *
Inner speech only, Word 3 onset errors				
(Intercept)	nonword, dissimilar,	-6.04	0.28	<.001 ***
Lexicality	real word outcomes	1.24	0.25	<.001 ***
Similarity	similar onsets	0.78	0.22	<.001 ***

Combined analyses: summary of key findings

Taken together, the results from the three combined analyses show the following: (a) Across experiments and conditions, participants' vigilance of self-reporting remained stable; (b) Compared to a visual metronome, use of an acoustic metronome resulted in a significant reduction in accuracy of self-reporting of primed-for errors, but only when such errors involved the substitution of similar onset phonemes. It also led to a marginal increase in speech-error rates; (c) Significant similarity and lexical biases were present in both inner and overt speech, the size of which remained stable across experiments.





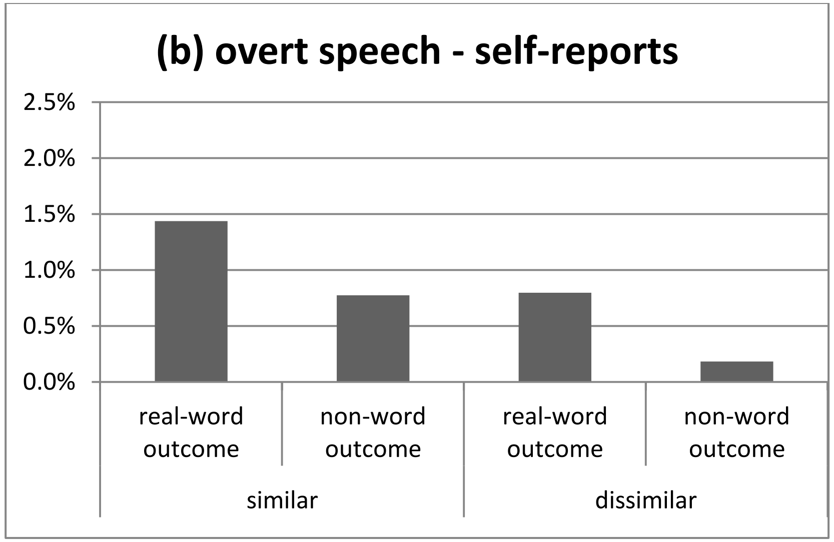


Figure 6-3. Primed for errors on Word 3: (a) inner speech (self-reports); (b) overt self-reports; and (c) overt independent rater reports.

Columns show the percentage of utterances (of Word 3) that were rated as containing primed-for errors, after adjusting for random participant and item variance. Fitted values are from the model presented in Table 6-11 and, to enable a direct comparison, from a model of inner speech that includes the lexicality by similarity interaction.

6.7. Experiments 1 to 3: General discussion

In these first three experiments we set out to determine; (a) whether the Oppenheim and Dell paradigm provides a reliable means of investigating the inner speech of people who stutter; (b) whether (or to what extent) participants under-report their errors when reciting tongue-twisters in the paradigm; and (c) whether under-reporting of errors occurs to similar extents across different conditions.

6.7.1. Monitoring vigilance and the under-reporting of errors

We performed three comparable experiments, with three separate (but similar) groups of first-year psychology students. In all three experiments, participants self-reported fewer overt primed-for errors than were reported by the independent rater. This was reflected most clearly by the strong main effect of *rater* in the corresponding combined analysis. The absence of any interactions between predictors for *rater* and *experiment* in that combined analysis further confirmed that under-reporting occurred to a similar extent in all three experiments, whereas the lack of any other interactions involving *rater* in the individual analyses further suggested that the extent of under-reporting of overt errors was also remarkably stable across conditions within experiments.

Furthermore, the complete lack, in any of the analyses or combined analyses, of interactions of rater or masking with either similarity or lexicality, strongly suggests that under-reporting of overt primed-for errors did not impact upon the relative frequency with which different types of errors were reported in overt speech. Taken together, these results indicate that, if participants' under-reporting of overt primed-for errors in these experiments stemmed from the reduced salience of the errors they made, it must have involved dimensions of salience that were not manipulated in the current set of experiments.

The negative effects of an acoustic metronome on the accuracy of self-reporting suggest that participants are more likely to pay attention to the precise details of errors made when there are no other auditory stimuli to distract them, whereas when distracted by an auditory metronome, participants were particularly likely to misreport the codas of errors with phonemically-similar onsets, despite still correctly identifying their onsets.

6.7.2. ***Error biases in overt and inner speech***

In all three experiments lexical biases and phonemic-similarity biases were present in both inner and overt speech. Although in Experiment 1 the similarity bias in inner speech was not reliable for errors with real-word outcomes, in the combined analysis of inner speech data from Experiments 1, 2 and 3 it was reliable, suggesting that Experiment 1 may have lacked sufficient power due to the small numbers of errors elicited.

Although the findings of each of the three experiments described here are consistent with each other, they differ substantially from the findings of Oppenheim and Dell (2008) with respect to the similarity bias: Oppenheim and Dell (2008) only found a significant phonemic-similarity bias in *overt* speech. The findings of our three experiments are, however, somewhat more similar to those of the subsequent study by Oppenheim and Dell (2010), which found a phonemic-similarity bias in inner speech. However, unlike our studies, the Oppenheim and Dell (2010) study also found a similarity by overtness interaction (the inner-speech similarity bias was significantly weaker than the overt similarity bias).

Our use of Oppenheim and Dell’s own set of tongue-twisters (in Experiment 2), seems to rule out the possibility that the differences in the findings of our two labs were related to the tongue-twisters used. Our finding that an acoustic metronome resulted in a reduction in self-reporting accuracy and a marginal increase in errors suggests that the differences in our labs findings may have been related to the type of metronome used. However, in the absence of any fully satisfactory explanations for this difference, it seems that further empirical studies are needed before this issue can be resolved.

6.7.3. *Evidence of interactivity and under-specification in the speech plan*

Irrespective of the differences between our own experimental findings (outlined in this chapter) and those of Oppenheim and Dell (2008), the overall finding of error biases in inner speech that resemble those found in overt speech (rather than being in the opposite direction) arguably best fits a model of speech production that attributes the biases (in both the self reports and independent-rater reports) directly to the functioning of the speech-production mechanism rather than to the results of under-reporting and covert error repair. That is not to say that participants didn't under-report their onset-substitution errors. On the contrary, the relatively lower numbers of self-reports of those errors compared to rater reports clearly shows that they did. Rather, the stability of the error biases across conditions and rater-type indicates that the extent of this under-reporting was not significantly affected by the salience of those errors.

The phonemic similarity bias in participants' (self-reports of) inner-speech errors constitutes the clearest evidence of the presence in the speech plan of a preponderance of phonological-encoding errors involving similar-phoneme substitutions. The findings of the current experiments thus strongly suggest that inner speech is specified at a featural level, even when there is no intention to articulate it overtly.

Whether or not inner speech is *fully* specified at sub-phonemic levels when there is no intention to articulate it overtly, remains an open question. In this regard, it is quite possible that more than one sub-phonemic level of encoding may exist. For example, there may be an acoustic and an articulatory level of features and, conceivably, the similarity bias may result from activation of (and/or) feedback from only the acoustic level, whereas articulatory features may remained unspecified or under-specified (cf. Schweppe, Grice, & Rummer, 2011).

The presence of robust lexical biases in self-reports of both inner and overt speech is fully in line with the findings of Oppenheim and Dell (2008, 2010). Finding a lexical bias in inner-speech strongly suggests that inner speech contains a predominance of phonological encoding errors that result in the production of real words (compared to non-words). The lexical bias found in onset errors in inner speech can only be explained (parsimoniously) by a feedback account as outlined in Section 4.5.3.

6.7.4. *Covert editing and error repair*

If covert editing and/or error repair had occurred to a significant extent in Experiments 1 to 3, we would have expected relatively more similar-phoneme substitution errors and errors resulting in the production of real words to have been self-reported in overt speech compared to in inner speech (in other words both the lexical and similarity biases should have been stronger in overt than in inner speech). This is because, in the inner-speech condition, participants would have self-reported whatever inner speech errors they perceived, whereas in the overt speech condition they would have only self-reported the ones that were not successfully edited-out prior to articulation. The above pattern was, however, not what we found. The most probable reason for the absence of any signs of covert editing or error repair is that participants simply did not engage in it to a significant extent in this paradigm. Their failure to do so may have been due to the time-pressure exerted by the need to speak in time to the metronome beat.

6.8. Experiments 1 to 3: Conclusions

Across three experiments, phonemic similarity and lexical biases were found in errors reported in both inner and overt speech. With respect to overt speech, the magnitude of these biases was similar in both self-reports and independent-rater reports, although overall, the independent rater consistently reported more errors than did the participants themselves. Together, these findings best fit a model of speech production that incorporates activation feedback between feature and phoneme levels as well as between phoneme and lexical levels of encoding, with the presence of such feedback being independent of the intention to articulate. The findings also suggest that under-reporting of errors by participants was consistently evident in participants' overt speech. The extent of under-reporting was not, however, influenced by the mode of reporting (written or typed), by the type of metronome used (acoustic or visual), by the presence or absence of auditory masking, by the tongue-twisters used, by the similarity of their onset-phonemes or by the lexicality of the primed-for error outcomes.

Having established that our adaptation of the Oppenheim and Dell paradigm produces consistent results across comparable groups of participants, and in particular that the patterns of onset errors self-reported by participants in their inner speech remain stable (provided the experiments are conducted 'within lab'); we are able to conclude that the paradigm constitutes a reliable experimental procedure for investigating phonemic errors that occur in inner speech. In the next experiment we therefore employ the procedure on a group of participants who stutter and a group of age, gender, and education-matched controls.

7. Experiment 4: Comparing encoding errors of stutterers and normally-fluent controls.

7.1. Introduction

In the three experiments described in Chapter 6, we tested the reliability and validity of the Oppenheim and Dell (2008) tongue-twister paradigm as a means of investigating the frequency with which errors of phonological encoding arise in the speech plans of normally-fluent speakers, and the extent to which speakers are aware of these errors prior to the onset of overt articulation. Using our own adaptation of the paradigm we, were able to demonstrate that the performances of three closely matched groups of normally-fluent participants did not differ significantly from one another with respect to: (a) the likelihood of phonemic errors occurring in their overt speech; (b) the likelihood of them self-reporting phonemic errors (in inner and overt speech); and (c) the magnitude of the phonemic-similarity and lexical biases in those errors and self-reports.

Having established in our own lab that the paradigm produces similar results across similar participant groups, in the current experiment (Experiment 4) we made use of the same paradigm to compare the performance of a group of people who stutter, with that of a group of matched, normally-fluent controls. The primary aim of Experiment 4 was to test the first tenet of the Covert Repair Hypothesis (Postma & Kolk, 1993): that stutterers make abnormally large numbers of errors of phonological encoding. However, it also investigated whether they also make abnormally large numbers of word-order errors, and whether there is a relationship between the rate at which they make encoding errors and how disfluent they are in everyday speech (as would be predicted if stuttering-like disfluencies stem from covert repairs of encoding errors). In order to ensure that our comparisons of the two groups' inner-speech error rates were valid, as in Experiments 1 to 3, we also investigated the two groups' monitoring vigilance and accuracy.

As in Experiments 1 to 3, in Experiment 4 we first investigated the vigilance (and accuracy) with which participants detected and self-reported their overt errors, and then compared the biases in their inner and overt self-reports. However, in the present experiment, our particular intention was to compare the performances of the two participant groups on these measures. In Experiment 4 we also performed three additional analyses. In the first, to maximize the chances of detecting between-group differences in onset-error rates we included all onset errors (both primed-for and non-contextual) that occurred – not only on Word 3 of the tongue-twisters, but also on Word 2. In the second, we analysed non-contextual errors in isolation. This was to follow up an empirical observation that between-group differences in error rates appeared to be strongest of all with respect to non-contextual onset errors. Then, in the third additional analysis we investigated whether the likelihood of making word-order errors (including word anticipations, perseverations and exchanges) also differed between the two groups. These additional analyses allowed us respectively: (a) to maximize our chances of finding evidence that participants who stutter have impaired phonological encoding compared to normally-fluent speakers; (b) to investigate a proposal originally made by Bosshardt (2006), that poor or incomplete modularization of the language formulator renders phonological encoding in stutterers vulnerable to noise from processes beyond encoding of the current word; and (c) to investigate whether language encoding in stutterers may be impaired in ways other than just phonological encoding.

Finally, in addition to these main analyses, we also conducted a post-hoc analysis that examined whether a correlation could be found between (a) the frequency with which individual participants who stutter self-reported onset errors in their inner speech; and (b) the frequency of their stuttering-like disfluencies in a variety of every-day speaking tasks. This post-hoc analysis was motivated by the second tenet of the CRH: that stuttering-like disfluencies stem directly from covert repairs of errors of phonological encoding. If this tenet is correct, we would expect to find a positive correlation between inner-speech errors and the frequency of stuttering-like disfluencies.

In Experiment 4, as in the previous three experiments, participants recited tonguetwisters both overtly and in inner speech, and immediately reported any errors they perceived themselves making in either condition. In order to determine whether the two participant groups differed in the extent to which they relied on auditory feedback for error detection, half of all tonguetwisters were spoken under conditions of auditory masking (following the same procedure that was used in Experiment 1 for non-stutterers). The use of auditory masking also necessitated the use of a visual (rather than an acoustic) metronome, which was therefore adopted throughout. Because participants who stutter may experience difficulty with verbal self-reporting of errors, all self-reporting of errors was done by typing, rather than speaking.

To determine monitoring vigilance and the possibility that stutterers differ from normally-fluent speakers with respect to monitoring vigilance, participants' overt recitations were also transcribed and coded by the experimenter. We also took forward digitspan measurements for all participants, in order to control for differences in the two participant groups' abilities to remember the tonguetwisters and to remember (and thus accurately self-report) their errors.

7.2. Method

7.2.1. Participants

Thirty-two participants who stutter (eight male) and thirty-two controls (nine male) matched for age, and education took part in the experiment. Participants were recruited through the University of Edinburgh student employment service and experimental subject-pool. Participants who stutter were additionally recruited through stuttering self-help groups, and some controls were recruited through an internet employment website. All participants were native speakers of English. PWS had a mean age of 38 (range 18 to 71); for PNS, the mean was 39 (range 18 to 68). Mean education level (on a scale where 1 corresponds to General Certificate of Secondary Education (GCSE) or equivalent, and 5 indicates a postgraduate degree) was 3.00 for stutterers and 2.97 for controls. Twelve participants from each of the groups were university students; four of the stuttering and five of the control group were retired. There was a marginal difference between the two groups in forward digit span: mean digit span for stutterers was 6.6, and for controls, 7.2 (*t* 62 = 1.67; *p* =.061).

All participants completed Section 3a of the Overall Assessment of the Speaker's Experience of Stuttering (OASES; Yaruss & Quesal, 2006), in which respondents rate their current difficulty in communicating verbally on a five-point scale in each of 10 commonly occurring situations. These include, for example, talking with another person one to one, initiating conversations, speaking to strangers, and continuing to speak regardless of how your listener responds to you. Mean scores for the OASES section 3a (communication difficulty) were: stutterers = 27.6; controls = 19.7 (*t* (62) = 4.62 *p* <.001). In addition to these ten OASES 3a questions relating to general communication difficulty, participants additionally provided ratings of "fluency difficulty" in the same ten situations. Specifically, for each situation, they were asked to rate how difficult it is to speak fluently, without stuttering and without avoiding words. Mean scores for fluency difficulty ratings were: stutterers = 31.9; controls = 18.0 (*t* (62) = 8.65 *p* <.001)

For all participants who stutter, full SSI4 (Riley, 2009) stuttering severity measures were derived from video recordings of samples of their spontaneous conversation and reading out loud. Mean SSI4 score was 20.7; range 8 to 36. Control participants only completed the reading portion of the above assessment. Mean number of SLDs per hundred syllables on the reading task was: stutterers 6.08 range 1.3 to 24; controls 0.43 range 0 to 1.85; (*t* (62) = 4.15 *p* <.001).

Apart from stuttering, participants reported no other speech, language, hearing or visual impairments.

7.2.2. *Materials*

The experimental materials were identical to those used in Experiment 1 (See Appendix A). They consisted of four-word tonguetwister sequences. The onsets of each tonguetwister sequence followed an 'ABBA' pattern to induce onset-phoneme substitution errors, e.g. *pink bid bit pick*. As in Experiment 1, to enable additional analyses of phonemic and lexical influences on speech errors, we created four variants of each tonguetwister sequence, and divided these between four lists, each list containing 48 tonguetwisters. Each participant recited tonguetwisters from just one of these lists (i.e. each participant recited 48 tonguetwisters). Every tonguetwister sequence was recited an equal number of times under identical conditions by (an equal number of) participants from both groups.

As in Experiment 1, within each list, half of the tonguetwisters were assigned to the auditory masking condition. Because auditory masking was blocked, four versions of each list were drawn up such that, in the experiment as a whole, all tonguetwisters appeared equally in masked and unmasked, and masking-first and masking-last conditions. Finally, two versions of each of the resultant 16 lists were created. In both versions, half of the tonguetwisters were marked for overt recitation and the other half for silent (inner-speech) recitation. Those that were marked for overt recitation in one version were marked for silent (inner speech) recitation in the other. This resulted in 32 lists of experimental items in a fully counterbalanced design.

Auditory masking was achieved using computer-generated pink noise, delivered through a set of Panasonic RP-HT225 stereo headphones. Participants' overt recitations of the tonguetwisters were captured on a Zoom H2 digital recorder and analyzed using Praat software (Boersma & Weenink, 2009).

7.2.3. *Procedure.*

The procedure was the same as the procedure in Experiment 1, with the exception that participants typed, rather than verbally reported, the details of any errors they perceived themselves to have made. Self-reporting through typing followed the same procedure as in Experiment 3 (See Section 6.5). The primary reason for adopting typing as the method of self-reporting was to remove the possibility that stuttering (or fear of stuttering) while verbally self-reporting may have led stuttering participants to under-report their errors. Moreover, difficulty providing verbal self-reports may reduce the accuracy with which stuttering participants self-reported their errors.

Prior to beginning the experiment, participants underwent a computer-led tutorial and practice session, which included full instructions concerning the inner speech and overt speech procedures. In all conditions, participants were instructed to place highest priority on speaking in time to the (visual) metronome, not to worry about making mistakes, and simply to skip words they felt likely to stutter on. For the inner-speech recitations, to prevent them from attempting to mouth sequences silently, participants were instructed not to move their mouths or any muscles associated with speech and, if possible, to keep their mouths completely closed.

At the beginning of each masked block, participants were instructed to adjust the headphones to ensure that the loudness of the pink noise prevented them from hearing the sound of their own voice. It was emphasized that participants should speak the overt tonguetwisters as quietly as possible throughout the masked block. The experimenter observed participants throughout the experiment and reminded them, where necessary, to adhere to the above instructions.

Tonguetwister sequences were presented in the same way as in Experiment 1 (See Section 6.2.1). However, after the (visual) metronome prompts for each of the four testing-phase recitations of each tongue-twister ended (i.e. 500 ms after the appearance of the fourth dash), the prompt disappeared, the pink noise (if any) ended, and the tonguetwister sequence reappeared at the top of the screen, together with an instruction to “type any errors and then press ENTER to continue” at the bottom. If participants perceived themselves to have made an error during a particular recitation, they were instructed to type, as fully as possible, what they had actually said, for example “rag lap *rash* rap” (when they should have said “rag lap *lash* rap”). They were instructed to type one or more question-marks in the relevant places if they could not remember what they had said for a word or ‘X’s if they had completely omitted all or part of a word. Once errors, if any, had been reported, pressing ENTER started the next four-dash sequence. Each performance phase included four repetitions of the four dashes, before the familiarization phase for the next word sequence began. (This self-reporting procedure was identical to the procedure adopted in Experiment 3)

As in the previous experiments, in addition to participants' self-reports of their inner and overt speech errors, the experimenter independently identified and transcribed errors in the overt speech condition (this was done online and then double-checked from recordings). Recordings from a random sample of five PWS and five PNS were transcribed a second time by the experimenter prior to analysis to enable intra-rater reliability to be calculated.

7.2.4. ***Coding***

Errors were coded into four categories: (1) primed-for onset errors; (2) non-contextual onset errors; (3) word-order errors; and (4) other/ambiguous errors. Errors were only ascribed to one of the first three categories if they were not in any way ambiguous. Thus, for example, the primed-for onset-error category only included instances where a ‘B’ word onset (i.e. the onset of words 2 or 3) was substituted by an ‘A’ word onset but the coda remained unchanged. The non-contextual onset-error category only included instances where a ‘B’ word onset was substituted by a non-contextual onset phoneme but the coda remained unchanged. Instances where the onset error resulted in production of one of the ‘A’ words (e.g. ‘dock *dock* notch dodge’ instead of ‘dock knock notch dodge’) were excluded, although such instances were rare because, in all but a few tonguetwisters, the codas of each of the four words differed.

As in the previous experiments, since the majority of errors occurred on words 2 and 3 (the ‘B’ words), counts of onset errors only included these words. This ensured that onset-exchange errors were only counted once. Counts of word-order errors included all four word positions, where one of the four words in the tonguetwister was either uttered twice, or exchanged positions with one of the other four words. Instances where the position change could potentially be accounted for as a result of a word omission (e.g. ‘rag lap rack xxxx’ instead of ‘rag lap lash rack’) and instances where the order-error could potentially have been a coda exchange (e.g. ‘rag *lash lap* rack’ instead of ‘rag *lap lash* rack’) were excluded from the counts. In cases where an error was followed by an overt self-repair, only the original error was coded for analysis.

7.2.5. ***Analyses.***

In total, six analyses were carried out in which the performance of the stuttering group was compared to that of the control group on the following measures:

1. The vigilance of self-reporting (Word 3 only).
2. The accuracy of self-reporting (Words 2 and 3).
3. The size and direction of lexical and phonemic-similarity biases (Word 3 only).
4. Likelihood of making an onset error of any kind (Words 2 and 3).
5. Likelihood of making a non-contextual onset error (Words 2 and 3).
6. Likelihood of making a word-order error (Words 2 and 3).

For all analyses we generated base models which included an intercept, and random effects representing by-participant and by-item intercept variation. Because t-tests revealed a marginal difference between the two participant-groups' mean digitspan scores, digitspan was controlled for by including it as a fixed effect in all analyses. We then proceeded to add further predictors stepwise to each model under consideration. As in the previous experiments, selection of models was based on two criteria. First, using χ^2 tests to compare model likelihood ratios, we assessed whether the fit of the model to the data was improved (as indicated by a significant decrease in the model likelihood ratio) by the addition of each predictor. (With the exception of digitspan, predictors were retained only if the current (best) model was improved). Second, where two or more predictors each significantly improved the current model, we selected the model which had the smallest log-likelihood. Once relevant predictors and their interactions had been exhaustively explored, the resulting model represented the 'best fit' to the data, being a model which could not be improved by the addition of further predictors.

7.3. Results

7.3.1. *Descriptive statistics*

Out of a total of 12,288 four-word tonguetwisters (48 tonguetwisters, each repeated four times by 64 participants) 1773 errors of any type were transcribed by the independent rater from recordings of participants' overt speech (PWS=1176, controls=597), whereas participants self-reported 1230 (PWS=821, controls=409) errors in overt and 971 (PWS=620, controls=351) in inner speech.

Primed-for onset-phoneme substitution errors (Word 3).

For participants who stutter (PWS), one-hundred-and-thirty of the errors transcribed by the independent rater from recordings of participants' overt speech were unambiguous onset-phoneme substitution errors (cases in which the onset of Word 3 was substituted with that of an A word yet the coda remained unchanged). PWS self-reported 92 such errors in overt speech, and 54 in inner speech. Table 7-1 gives a breakdown of errors by experimental condition. Figure 7-1 depicts the same data, together with that from controls, collapsed (a) across masking and lexicality of outcome and (b) across masking and similarity of onset.

Table 7-1. Experiment 4, PWS: Onset Substitutions on Word 3

		Similar onsets			Dissimilar onsets		
Error outcome		Unmasked	Masked	(Total)	Unmasked	Masked	(Total)
Self-reports (inner speech)	Word	15	13	(28)	4	6	(10)
	Nonword	3	5	(8)	5	3	(8)
Self-reports (overt speech)	Word	30	18	(48)	9	5	(14)
	Nonword	5	16	(21)	6	3	(9)
Independent rater (overt speech)	Word	37	28	(65)	11	16	(27)
	Nonword	10	18	(28)	7	3	(10)

For controls, fifty-two of the errors transcribed by the independent rater from recordings of participants' overt speech were unambiguous onset-phoneme substitution errors. Controls self-reported 35 such errors in overt speech, and 33 in inner speech. Table 7-2 gives a breakdown of errors by experimental condition (See also Figure 7-1).

Table 7-2 Experiment 4: Controls, Onset Substitutions on Word 3

		Similar onsets			Dissimilar onsets		
Error outcome		Unmasked	Masked	(Total)	Unmasked	Masked	(Total)
Self-reports (inner speech)	Word	8	10	(18)	3	1	(4)
	Nonword	3	5	(8)	0	3	(3)
Self-reports (overt speech)	Word	11	5	(16)	0	4	(4)
	Nonword	7	3	(10)	1	4	(5)
Independent rater (overt speech)	Word	13	11	(24)	2	4	(6)
	Nonword	9	6	(15)	1	6	(6)

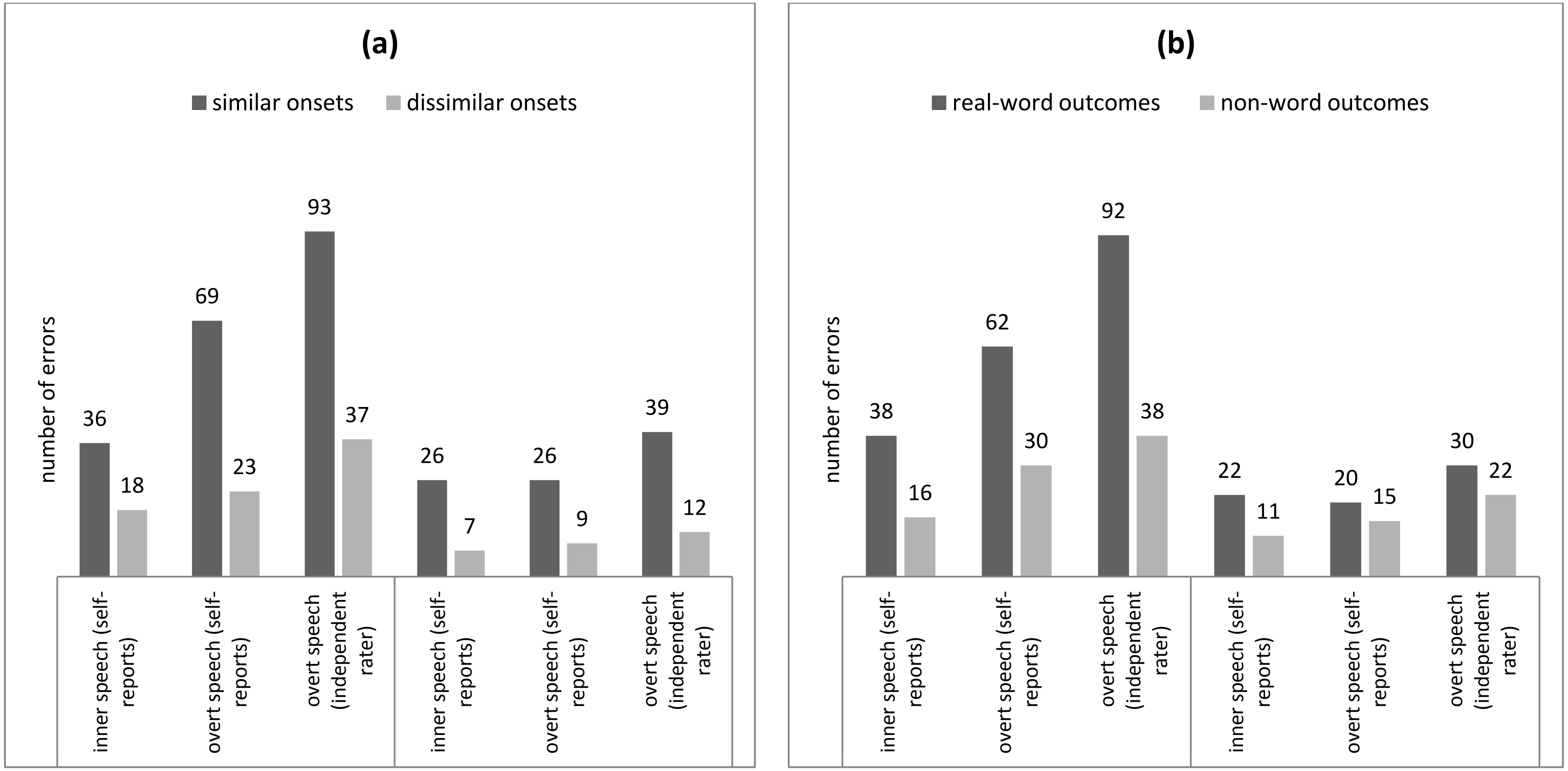


Figure 7-1. Experiment 4: (a) Phonemic similarity and (b) Lexical biases – on Word 3 onset-substitution error outcomes. Columns depict raw numbers of primed-for errors in (a) the similar and dissimilar onset conditions (collapsed across masking and lexicality of outcome); and (b) real-word and non-word outcome conditions (collapsed across masking and similarity of onset). In each chart separate columns are shown for stutterers and controls, for self-reports (inner and overt speech), and independent-rater reports (overt speech).

7.3.2. *Vigilance of self-reporting*

This analysis investigated whether the type of rater (self-rater or independent rater) influenced the likelihood of errors being reported, and whether there was a difference between the two participant groups with respect to how vigilant they were in self-reporting errors. The independent rater's ratings were considered to constitute a stable baseline. To establish whether the two participant-groups differed with respect to vigilance of self-reporting we used a (binomial) predictor: *group*, with values of 1 or 0 to represent participants who stutter and control participants respectively. In addition to estimating overall vigilance, we also checked whether vigilance was affected by the addition of (a) masking; (b) similarity of onset phonemes: and (c) lexical status of the reported error (i.e. whether or not it constituted a real word).

After controlling for digitspan, the following fixed predictors were tested (in a forward stepwise manner):

- Masking,
- Overtness,
- *Rater-type*, with 2 levels (experimenter-rating or self-rating),
- *Group*, with 2 levels (control or participant who stutters),
- two and three-way interactions.

The main interactions of interest were: group by rater-type (to clarify whether compared to controls, PWS self-report a greater proportion of their overt errors); and the three-way interactions containing group, rater-type and a third predictor (masking, lexicality or similarity)

The best fitting model comparing overt self- and independent-rater ratings of onset errors included the effects of *rater, digitspan, lexicality, similarity, group*, and the *lexicality by* *group interaction*: Improvement due to addition of *rater:* χ^2 (1) = 6.59, *p* =.010. Adding the rater by group interaction did not further improve the model: (χ^2 (1) = 0.00, *p* =.980), neither did the addition of any other two or three-way interactions: all *ps* ≥.253.

Details of the model coefficients are given in Table 7-3. Regarding the vigilance of self-reporting, the coefficient for *rater* indicates that, after controlling for random (participant and item) variance and irrespective of all other factors, participants (in both groups) only self-reported 0.64 ($e^{-0.45}$) times as many primed-for errors as did the independent rater. Amongst other things, the coefficients also reveal significant similarity biases (of similar magnitudes) in both participant groups' recitations. Interestingly, however, lexical bias was only present to a significant degree in the recitations of participants who stutter: more specifically, participants who stutter were 2.6 ($e^{0.95}$) times more likely to make (or self-report) errors with real-word outcomes than errors with non-word outcomes, whereas controls were not (reliably) more likely more likely to do so.

Table 7-3. Experiment 4: Vigilance of self-reporting of error biases.
Model Coefficients and Probabilities for Best Fitting Models. (see Appendix B Section 12.2, for equivalent analysis with random slopes)

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Overt speech (participants vs. independent rater), Word 3 onset errors				
(Intercept)	digitspan = 7, independent rater, controls, nonword outcome, dissimilar onsets	-5.69	0.32	<.001 ***
Digitspan	+1	-0.21	0.11	.043 *
Rater	Self rating	-0.45	0.12	<.001 ***
Group	PWS	0.85	0.34	.013 *
Lexicality	real word	0.38	0.23	.104
Similarity	similar	1.22	0.14	<.001 ***
Group by Lexicality interaction	real word & similar	0.57	0.28	.043 *

Note. In the equivalent model containing random slopes, the Group by Lexicality interaction does not reach significance. (see Appendix B Section 12.2)

The Lack of a significant 3-way *group by rater by lexicality* interaction indicates that this between-group difference with regard to lexicality was not due to control participants under-reporting their overt errors.

7.3.3. *Accuracy of self reporting*

Although the above analysis of self-reporting vigilance show that both PWS and PNS were around 0.64 times as likely to report primed-for onset errors as was the experimenter, they do not provide information on accuracy: There were, for example, some instances when a participant reported having made a different error from the one that was coded by the experimenter.

This second set of analyses therefore focussed on the *accuracy* with which individual overt errors were reported by participants, i.e. the extent to which participants' self-reports exactly matched the experimenter's transcriptions from recordings.

For this analysis we recoded all instances of onset errors that had been identified by the experimenter and marked each one as *a 'match'* if it was identical to the participant's self-report, or a *'mismatch'* if it differed or was not reported by the participant. We also counted false alarms (i.e. instances where the participant self reported an error which the experimenter coded as correct). However, since there were only eight of these in total they have been disregarded. To allow for a more powerful statistical analysis, we counted all primed-for as well as non-contextual onset errors that occurred on Words 2 and 3. This analysis is directly comparable to the equivalent combined analysis of accuracy in Experiments 1 to 3 (See Section 6.6.2). Numbers of matches and mismatches are shown in Table 7-4.

Table 7-4. Experiment 4: Accuracy of participants' coding of onset errors (words 2 and 3 combined).

		Matches (%)	Mismatches (%)	Missed (%)	Total independent-rater reported errors	False alarms
Onset errors						
masked	PWS	46 (52)	21 (24)	21 (24)	88	3
	PNS	18 (49)	7 (19)	12 (32)	37	5
unmasked	PWS	46 (60)	7 (9)	24 (31)	77	0
	PNS	19 (59)	5 (16)	8 (25)	32	0

Note: Percentages represent the percentage of experimenter reports that were also self-reported (in an identical manner) by participants.

In addition to the predictors tested in the corresponding combined analysis of vigilance, an additional predictor for '*total errors*' (the total number of independent-rater reported errors across the four iterations of the tongue-twister sequence in which the primed-for error occurred) was also tested. This extra predictor, the value of which could range from zero to sixteen, was added to test whether participants found it more difficult to accurately recall their errors when the overall density of errors on a tongue-twister was higher.

After controlling for digitspan, the best fitting model of included only *total errors* (the total number of experimenter-coded errors, of any type, in the tonguetwister) as a predictor of the likelihood of participants' ratings exactly matching the independent rater's ratings. The model was not significantly improved by the addition of masking ($\chi^2 = 2.67$, $p = .102$), and, most importantly, the model was not improved by adding group membership as a predictor ($\chi^2 = 0.04$, $p = .837$) or any of the interaction terms that include group membership (all $ps \geq .140$). This implies that, across experimental conditions, the two groups did not differ with respect to the likelihood that they would accurately report errors. Model coefficients are shown in Table 7-5. They show that, irrespective of which group participants belonged to, the likelihood of their reports of primed-for errors identically matching those of the experimenter diminished by a factor of 0.8 ($e^{-0.21}$) with each additional error (of any kind) that occurred during the four recitations of a tonguetwister.

Table 7-5. Experiment 4: Logistic mixed effects analyses of factors influencing the likelihood of participants accurately identifying onset errors.
(See Appendix B Section 12.2, for equivalent analysis with random slopes)

Predictor	Value	Coefficient	Std. Error	p (coefficient = 0)
likelihood of accurately identifying an error (onset errors)				
(Intercept)	Digitspan = 7 Total errors (across all four iterations tongue-twister) = 3	0.22	0.17	.204 *
digitspan	+1	-0.10	0.11	.383
Total errors	+1	-0.20	0.08	.010 **

7.3.4. *Self-reports*

Comparisons of inner and overt speech.

This third set of analyses investigated participants’ self-reports in inner and overt speech. These analyses are equivalent to the analyses carried out in Experiments 1 to 3, and enabled us to compare the magnitude of the lexical and phonemic-similarity biases in inner vs. overt speech. They additionally tested whether the two participant groups differed with respect to the size and direction of the error-biases.

The analyses of self-reported errors in inner and overt speech tested the following binomial predictors and interactions:

- *Group* (participant who stutters or control),
- *overtness* (whether or not participants were speaking aloud),
- *lexicality* (whether an onset substitution would result in a real word),
- *similarity* (whether the substituted phonemes differed by one feature or more),
    - lexicality by similarity interaction,

- lexicality by group interaction,
- similarity by group interaction,
- lexicality by overtness by group interaction,
- similarity by overtness by group interaction.

The best fitting model of self-reported primed-for errors on Word 3 included effects of lexicality, similarity, their interaction, overtness and participant group. Improvement due to the addition of participant group: $\chi^2(1)= 12.80$, $p<.001$. Adding other two or the three-way interactions did not further improve the model: all $ps \geq .635$. Table 7-6 gives the coefficients of the best-fitting model and the probabilities that they differ from zero.

Table 7-6. Experiment 4: Phonemic-similarity and Lexical biases. Predictors of self-reported errors. Model Coefficients and Probabilities for Best Fitting Models. (See Appendix B Section 12.2, for equivalent analysis with random slopes)

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Inner vs. overt speech, Word 3 onset errors				
(Intercept)	Digitspan=7 nonword outcome, dissimilar onsets inner speech, controls	-6.10	0.32	<.001 ***
Digitspan	+1	-0.18	0.09	.065
Overtness	overt	0.53	0.15	<.001 ***
Group	PWS	0.82	0.26	.002 **
Lexicality	word	0.28	0.28	.316
Similarity	similar	0.72	0.26	.006 **
Lexicality by similarity interaction*		0.68	0.33	.042 *
Inner speech only, Word 3 onset errors				
(Intercept)	Digitspan=7 nonword outcome, dissimilar onsets	-6.14	0.35	<.001 ***
Digitspan	+1	-.26	0.13	.054
Lexicality	word	0.92	0.25	<.001 ***
Similarity	similar	1.05	0.26	<.001 ***

***Note. In the equivalent model containing random slopes, the Group by Lexicality interaction does not reach significance. (see Appendix B Section 12.2)**

These coefficients show that, once random variance and digitspan is accounted for, participants who stutter were approximately 2.3 ($e^{0.86}$) times more likely to report errors than control participants. The lack of *group by similarity* and *group by overtness* interactions suggests that this tendency for participants who stutter to self-report more errors than control participants was not significantly different for errors that involved the substitution of similar vs. dissimilar onset phonemes; or for errors that were made in inner vs overt speech

.Inner speech in isolation

In the analysis of inner speech in isolation, the best fitting model included only lexicality and similarity (Improvement due to the addition of similarity, $\chi^2(1)=19.40$, $p<.001$, (see Table 7-6). The addition of the lexicality by similarity interaction only marginally further improved the model $\chi^2(1)= 3.68$, $p=.055$. The addition of group did not further improve the model $\chi^2(1)=2.69$, $p=.101$; nor did any of the two or three-way interactions involving group. Thus, with respect to Word 3 alone, participants who stutter did not make reliably more primed-for errors in inner speech than control participants. The numbers of these inner-speech errors were, however, low (PWS: 54, controls: 33) and it is possible that the analysis lacked sufficient power. Therefore, in the next analysis, to increase the statistical power, all unambiguous onset errors (on Words 2 and 3) were included.

All onset-phoneme substitution errors

This fourth set of analyses constituted a further investigation of whether the speech plans of participants who stutter contain more errors than those of controls. It focussed, once again, on onset errors, but this time all onset errors *on Words 2 and 3* were included, including both primed-for as well as non-contextual errors. Adding these extra errors meant that we could not investigate the effects of lexicality or similarity (and their interactions), but it brought the advantage of a substantial increase in statistical power.

For this analysis, in addition to random by-participant and by-item variation and digitspan, we tested the following predictors and interactions:

- auditory masking,
- *overtness* (whether or not participants were speaking aloud),
- *group* (PWS or controls),
- stuttering-like disfluencies per 100 syllables when reading aloud,
- self-ratings of difficulty speaking fluently,
    - overtness by group interaction,
    - the two overtness by disfluency interactions.

Two hundred and thirty four of the errors transcribed by the independent rater from recordings of participants' overt speech were unambiguous onset-phoneme substitution errors (cases in which the onsets of Word 2 or 3 were substituted by a primed-for *or* non-contextual phoneme yet the coda remained unchanged). Participants self-reported 160 such errors in overt speech, and 121 in inner speech. Table 7-7 gives a breakdown of onset errors by experimental condition.

Table 7-7. Experiment 4: Onset errors in inner and overt speech, stuttering and control participants

		Primed-for onset substitutions on words 2 & 3			Non-contextual onset substitutions on words 2 & 3			Total onset substitutions on words 2 & 3
		Un-masked	Masked	Total	Un-masked	Masked	Total	
Self-reports (inner speech)	PWS	33	34	(67)	4	10	(14)	(81)
	controls	15	22	(37)	2	1	(3)	(40)
Self-reports (overt speech)	PWS	51	48	(99)	6	12	(18)	(117)
	controls	22	21	(43)	1	0	(1)	(43)
Independent rater (overt speech)	PWS	73	77	(150)	4	11	(15)	(165)
	controls	31	35	(66)	1	2	(3)	(69)

After accounting for random variance and digitspan, the best-fit model for all self-reported onset errors included effects of overtness and group (improvement due to adding group: χ^2 (1) = 13.87, *p* <.001). Once again, the group by overtness interaction did not improve the model, nor did any other predictors or their interactions: all *ps*≥ .103. Table 7-8 gives the coefficients of the model, and the probabilities that they differ from zero.

Table 7-8. Experiment 4: Logistic mixed effects analyses of factors influencing the likelihood of occurrence of all onset errors. Model coefficients and probabilities are given for best-fitting models.
(See Appendix B Section 12.3, for equivalent analysis with random slopes)

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Inner vs. overt speech, All onset errors (Words 2 & 3)				
(Intercept)	digitspan=7, controls inner speech	-5.49	0.21	<.001 ***
Digitspan	+1	-0.16	0.08	.041 *
Group	PWS	0.86	0.22	<.001 ***
Overtness	overt speech	0.36	0.12	.004 **
Inner speech in isolation All onset errors (Words 2 & 3)				
(Intercept)	digitspan=7, controls	-5.36	0.23	<.001 ***
Digitspan	+1	-0.14	0.10	.186
Group	PWS	0.65	0.28	.020 *

After controlling for random variance and for digitspan and irrespective of overtness, stuttering participants were 2.37 (i.e. $e^{0.86}$) times more likely than controls to report errors. Independent of group participants were approximately 1.4 (i.e. $e^{0.36}$) times more likely to report errors in the overt condition. An equivalent analysis of inner-speech in isolation this time confirmed that, compared to controls, stuttering participants were indeed significantly more likely to report onset errors in inner speech (improvement due to adding group: χ^2 (1) = 5.32, *p*=.021 (see Table 7-9 for coefficients of the model).

Non-contextual onset errors

Inspection of the raw numbers of self-reported onset errors suggested that participants who stutter are especially prone to self-reporting non-contextual onset errors, disproportionately so under conditions of auditory masking (see Table 7-7). This observation was of interest because a number of studies by Bosshardt and his co-workers (see Bosshardt, 2006, for a review) has led them to propose that language formulation in stutterers may not be sufficiently modularized and may thus be particularly susceptible to phonemic influences beyond those involved in formulation of the current target word. This hypothesis motivated us to perform an additional analysis of factors affecting the likelihood of self-reporting non-contextual onset errors (on Words 2 and 3 combined). Predictors tested were: (digitspan), overtness, masking, group, and their interactions.

In addition to digitspan, the best-fit model for self-reported non-contextual onset errors included only the effect of group (improvement due to adding group: χ^2 (1) = 9.59, *p* =.002). The coefficient for group shows that, after controlling for random (participant and item) variance and digitspan, irrespective of overtness, participants who stutter were 5.9 times more likely to self-report non-contextual errors than controls (See Table 7-9). The addition of the masking by group interaction only marginally further improved the model: χ^2 (2) = 5.51, *p* =.063). Its coefficient was unreliable.

Table 7-9. Experiment 4: Logistic mixed effects analyses of factors influencing the likelihood of occurrence of non-contextual onset errors. Model coefficients and probabilities are given for best-fitting models.
(See Appendix B Section 12.3, for equivalent analysis with random slopes)

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Inner vs. overt speech, Non-contextual onset errors				
(Intercept)	digitspan=7 controls	-8.99	0.66	<.001 ***
Digitspan	+1	-0.43	0.19	.025 *
Group	PWS	1.78	0.70	.010 *

Word-order errors

This final set of analyses investigated word-order errors. As with the analyses of onset errors, we built two models. The first allowed a comparison of order-error likelihoods in inner vs. overt speech. The second analyses inner-speech in isolation, and was constructed to provide confirmation of the findings with respect to inner speech.

One hundred and fifty six of the errors transcribed by the independent rater from recordings of participants' overt speech were unambiguous word-order errors (where one of the four words in the tonguetwister was either uttered twice, or exchanged positions with one of the other four words. See Section 7.2.4 for a fuller description). Participants self-reported 126 such errors in overt speech, and 92 in inner speech. Table 7-10 gives a breakdown of word-order errors by experimental condition.

Table 7-10. Experiment 4: Word-order errors, stuttering and control participants

		Word-order errors		
		Unmasked	Masked	(Total)
Self-reports	PWS	27	46	(73)
(inner speech)	Controls	5	14	(19)
Self-reports	PWS	36	50	(86)
(overt speech)	Controls	20	20	(40)
Independent rater	PWS	46	60	(106)
(overt speech)	Controls	23	27	(50)

For this final analysis, in addition to random by-participant and by-item variation and digitspan, we tested the following predictors:

- *auditory masking*,
- *overtness* (whether or not participants were speaking aloud),
- *group* (PWS or controls),
- stuttering-like disfluencies per 100 syllables when reading aloud,
- self-ratings of difficulty speaking fluently,

- overtness by group interaction,
- the two overtness by disfluency interactions.

The analysis of self-reported word-order errors revealed significant effects of digitspan, masking, overtness, and group (improvement due to adding group: $\chi 2$ (1) = 12.20, p <.001). The group by overtness interaction did not further improve the model: $\chi 2$ (1) = 2.49, p =.114). No other predictors or interactions significantly improved the model: all ps ≥.097. The model coefficients are given in

Table 7-11.

Table 7-11. Experiment 4: Logistic mixed effects analyses of factors influencing the likelihood of occurrence of word-order errors. Model coefficients and probabilities are given for best-fitting models.
(See Appendix B Section 12.3, for equivalent analysis with random slopes)

Predictor	Value	Co-efficient	Std. Error	p (coefficient = 0)
Inner vs. overt speech, Word-order errors				
(Intercept)	digitspan=7, unmasked, controls, inner speech	-6.36	0.28	<.001 ***
Digitspan	+1	-0.26	0.10	.006 **
Masking	masked	0.44	0.14	.002 **
Group	PWS	1.11	0.31	<.001 ***
Overtness	overt speech	0.39	0.12	.006 **
Inner speech in isolation Word-order errors				
(Intercept)	digitspan=7, unmasked, controls	-6.80	0.28	<.001 ***
Digitspan	+1	-0.34	0.11	.003 **
Masking	masked	0.70	0.23	.003 **
Group	PWS	1.42	0.40	<.001 ***

They show a similar pattern to that observed for onset errors: Compared to controls, stuttering participants were 3.0 (i.e. $e^{1.11}$) times more likely to self-report errors. Independent of group, masking or digit span, participants were more likely to report errors in the overt condition. Across both participant groups, masking caused an increase in the numbers of self-reported word-order errors, marginally less so with respect to errors in overt speech.

As with total onset errors, an equivalent analysis of inner-speech in isolation provided additional confirmation that, compared to controls, stuttering participants were significantly more likely to self-report word-order errors in inner speech: χ^2 (1) = 13.49, *p* <.001.

7.4. Experiment 4: Discussion

The most important finding of Experiment 4 is that compared to controls, participants who stutter self-reported significantly more errors of phonological encoding in their inner speech. In light of the similar levels of self-reporting vigilance in the two groups, the above finding provides strong support for the first tenet of the CRH: that the speech plans of stutterers contain abnormally high numbers of errors of phonological encoding (See Section 7.4.1). Moreover, the additional finding of between-group differences in the frequency of word-order errors suggests that language encoding in stutterers may also be impaired at other levels in addition to phonology.

With respect to error biases, the finding that the strength and direction of the phonemic-similarity bias was similar in both participant groups suggests that both participant groups used a similar monitoring strategy for pre-articulatory detection of errors. The finding of a robust phonemic similarity effect in inner speech in both participant groups suggests that speech plans for inner speech were not impoverished even in the absence of any intention to articulate. The finding of a *lexicality by group* interaction (whereby significant lexical bias was only found in the stutterers' error reports) is likely to be an anomaly, bearing in mind the consistent finding of lexical bias in the self-reports of normally-fluent speakers in Experiments 1 to 3.

Each of the above conclusions is discussed in greater detail in the corresponding sub-sections below, as are the findings of a post-hoc analysis (to be presented in Section 7.4.3) which investigated correlations between the severity of participants' stuttering-like disfluencies (in everyday speech) and their inner-speech error rates.

7.4.1. ***Group differences in speech error rates***

As highlighted in the previous section, the most important finding from the current experiment is that the stuttering group self-reported significantly more onset and word-order errors than the control group, both in inner as well as in overt speech.

To establish whether the higher numbers of errors self-reported by the stuttering group reflected higher numbers of errors actually made by that group, we investigated whether or not the two groups differed with respect to the vigilance with which they reported such errors. Our analyses revealed that, compared to experimenter-ratings, both participant groups under-reported their errors to a similar extent. When we investigated the accuracy with which participants reported individual errors, there was again no difference to be found between stutterers and controls. Taken together, these analyses show that the proportions of experimenter-coded errors self-reported by each group do not differ, a finding in line with previous studies that have compared the abilities of stutterers and controls to detect phonemes and phonemic errors in recorded speech (Postma & Kolk, 1992; Sasisekaran & De Nil, 2006).

Support for the assertion that this relationship holds for inner speech can be gleaned from the performances of the two groups when reporting their *overt* errors under conditions of auditory masking, where self-reports *can* be compared to experimenter ratings. This is because, under conditions of auditory masking, speakers are deprived of auditory feedback and thus forced to rely largely on internal monitoring (i.e., monitoring through the "inner loop") to detect their overt errors in the same way that they monitor for errors in inner speech. The absence of any significant masking by group interactions in the analyses of overt speech suggests that the two groups are vigilant to a similar degree when monitoring their inner speech. Our finding therefore strongly suggest that the significantly greater numbers of onset-phoneme and word-order errors self-reported by the stuttering group in inner speech reflected the overall greater numbers of (phoneme and word-order) errors actually occurring in the speech plans of that participant group. They thus support the first tenet of the CRH.

Non-contextual errors

A comparison of (statistically-reliable) model coefficients reveals that, although participants who stutter produced greater numbers of errors of all types tested in the experiment, the greatest between-group difference was found in non-contextual (onset-phoneme) errors (See Table 7-12).

Table 7-12. Comparison of likelihood ratios for occurrence of different types of error across the two participant groups, ratios are from self-reports (of inner and overt speech combined).

Error type	likelihood of occurrence (stutterers/controls)
Primed-for onset errors (self-reports)	2.4/1
Non-contextual onset errors (self-reports)	5.9/1
Word-order errors (self-reports)	3.0/1

This suggests that the speech-plans of stutterers may be more vulnerable to interference from influences beyond the words/phonemes encoded in current utterance. This finding is in line with similar findings from a series of studies by Bosshardt and co-workers (see Bosshardt, 2006, for a review) that led them to propose that language formulation in stutterers may not be sufficiently modularized. Furthermore, the comparison of the raw numbers of non-contextual onset errors across groups and experimental conditions (see Table 7-7) suggests that this between-group difference in the likelihood of producing non-contextual errors may have been in some way sensitive to the imposition of auditory masking, although further research is needed before attempting to draw any conclusions from these preliminary observations.

7.4.2. *Group differences in error biases*

With respect to error biases, the only significant difference between the two participant groups was the stronger lexical bias in the stuttering group’s overt errors than in the control group’s overt errors. It is unlikely that this between-group difference can be ascribed to different levels of vigilance through the outer (auditory) loop, because the interaction was apparent in both overt self-reports as well as independent-rater’s reports. Nevertheless, the lack of a corresponding 2-way *group by lexicality* interaction in the analysis of inner-speech self reports means that we cannot rule out the possibility that it may have arisen because participants who stutter were somewhat more successful at detecting and editing-out errors with non-word outcomes in their inner speech. However, the overall numbers of errors were low, and it is also noteworthy that the magnitude of the stuttering participants’ lexical bias was not so different from that found in the combined analysis of student participants’ overt errors in Experiments 1 to 3. So, in light of these two factors, is not possible to draw any firm conclusions from these findings.

The comparison of the relative strength of the phonemic-similarity biases across the two participant groups had the potential to reveal group differences in the extent to which speakers habitually specify their speech plans for inner speech to include the sub-phonemic level of encoding. As noted in Section 4.4, Wheeldon and Levelt (1995) have provided evidence that normally-fluent speakers at least sometimes monitor abstract (timeless) syllabified phonological representations, and it is possible that this may reflect a tendency (of normally-fluent speakers) to under-specify speech plans for inner speech. In contrast, Vasić and Wijnen’s (2005) proposal (in their Vicious Circle Hypothesis) that stutterers are excessively aware of subtle timing irregularities in speech plans, suggests that stutterers may have a stronger tendency to fully specify speech plans, even when it is not necessary to do so. If this is the case, we might have expected to find a stronger inner-speech phonemic similarity bias in participants who stutter than in controls. However, the two groups' inner-speech similarity biases were not significantly different from one another. Indeed, the numeric trend was in the opposite direction (towards a stronger inner-speech similarity bias in controls than in stutterers).

7.4.3. *Covert error repair*

Experiment 4 was designed to minimize the potential for covert error repair activity through instructions to participants to give priority to speaking in time to the (visual) metronome rather than to maintaining a high level of accuracy. It was therefore anticipated that the error patterns of the two participant groups would not reflect error repair activity to any significant extent. The finding that, within both participant groups, significantly more overt than inner-speech errors were self-reported was fully in line with our expectations in this respect, insofar as it suggests that errors appearing in the original speech plan were not repaired prior to overt articulation. [17] Nevertheless, if covert error repair does play a significant role in causing disfluency in real-life speaking situations, we might have expected the participants with the highest scores on the measures of disfluency relating to real-life speaking situations to have made the most errors during the tonguetwister experiment. This was, however, not reflected in the results of our two analyses of all onset errors (See Section

---

[17] It further suggests that a proportion of participants' overt errors were likely to have originated downstream from the speech-plan (in processes related to the generation of motor commands).

7.3.4) which tested for effects of two different predictors that constituted measures of disfluency in everyday life. More specifically, whereas group membership (i.e. whether or not a participant was diagnosed as a stutterer) was retained as a predictor in the two best-fit models of onset-error likelihood, predictors for fluency-difficulty self-ratings, percentage of stuttering-like disfluencies in the reading task, and relevant interactions were not.

Because stutterers are often relatively fluent when reading, it is possible that the reading task did not provide an adequate test of the relationship between speech-errors and stuttering. It is also possible that predictors relating to disfluency were rejected from the regression models because they shared too much variance with other predictors already present in the models. To further explore these possibilities, we therefore performed three post-hoc analyses using the data from only the PWS group. Specifically, we plotted the stuttering group's inner and overt onset errors against their: (a) stuttering-like disfluencies per 100 syllables in the conversation task (Figure 7-2); (b) stuttering-like disfluencies per 100 syllables in the reading task (Figure 7-3); (c) self-ratings of difficulty speaking fluently (Figure 7-4).

Only two significant correlations were revealed in these plots: stuttering participants' stuttering-like disfluencies in conversational speech were correlated to their (self-reported) overt onset errors (*r=0.44, p*=.012), and to experimenter-reported overt onset errors (*r*=0.43, *p*=.015). Similar patterns were also evident in the plots of stuttering participants' overt onset errors and their self-ratings of difficulty speaking fluently (Figure 7-4); however two outliers prevented these correlations from reaching significance.

Figure 7-2. Scatterplots of total onset errors by percentage stuttering-like disfluencies in conversational speech, for participants who stutter.

Plots show, for each participant, total number of onset errors (x axis) plotted against stuttering-like disfluencies per 100 syllables as measured from the conversational speech task (y axis). From left to right, the three plots show raw numbers of onset errors in (1) inner speech; (2) overt speech (self-reports); and (3) overt speech (experimenter ratings).

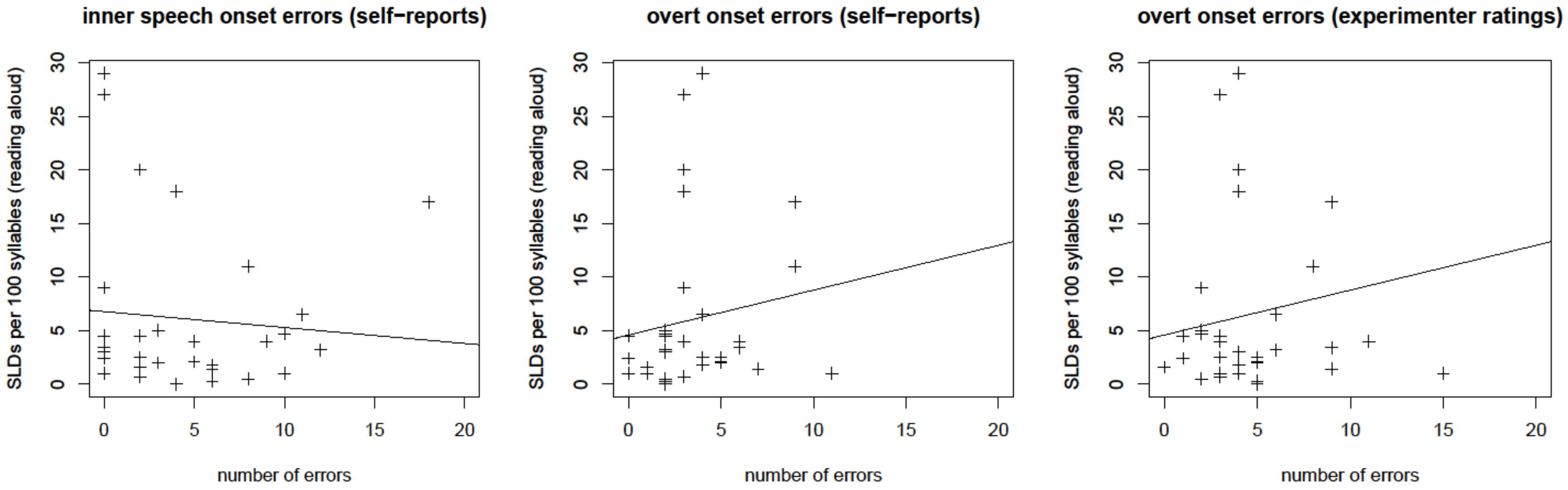


Figure 7-3. Scatterplots of total onset errors by percentage stuttering-like disfluencies in a reading task, for participants who stutter. Plots show, for each participant, total number of onset errors (x axis) plotted against stuttering-like disfluencies per 100 syllables, as measured from the reading task (y axis). From left to right, the three plots show raw numbers of onset errors in (1) inner speech; (2) overt speech (self-reports); and (3) overt speech (experimenter ratings).

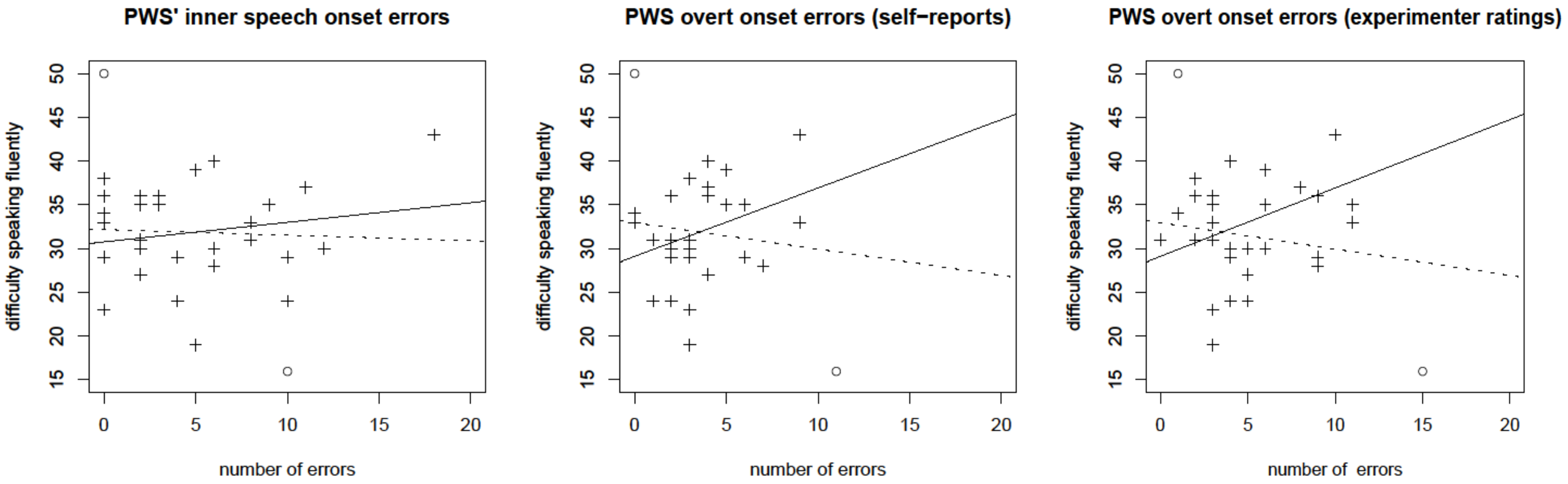


Figure 7-4. Scatterplots of total onset errors by self-ratings of difficulty speaking fluently, for participants who stutter.

Plots show, for each participant, total number of onset errors (x axis) plotted against self ratings of difficulty speaking fluently in ten commonly occurring speaking situation (y axis).From left to right, the three plots show raw numbers of onset errors in (1) inner speech; (2) overt speech (self-reports); and (3) overt speech (experimenter ratings). The unbroken regression line depicts the relationship between the two variables when two outliers (marked with ○) are excluded.

The lack of any correlation between participants’ inner-speech onset-error rates and their disfluency rates suggests that covert repairs of phonological encoding errors do not account for anything more than a minor proportion of the instances of stuttering-like disfluencies in their everyday speech. These findings, therefore, are not compatible with the second tenet of the CRH: that stuttering-like disfluencies in people who stutter are the result of covert repairs of large numbers of errors of phonological encoding. If stuttering does involve covert error-repair, the errors in question must either be so subtle that participants failed to report them, or they must originate downstream from the speech plan, and therefore only in inner speech when it accompanies overt articulation (cf. Max et al., 2004).

In Experiment 4, it was not feasible to attempt to measure whether or not the timing of the stuttering group’s inner-speech tonguetwister recitations was more variable than that of the control group. We therefore cannot rule out the possibility that stuttering may stem from inappropriate covert repair of small pauses and other prosodic markers in stutterers’ speech plans as posited in the Vicious Circle Hypothesis (Vasić & Wijnen, 2005; see also Lickley et al., 2005). Neither can we rule out the possibility that stutterers are hypervigilant towards such cues, although the experimental paradigm could potentially be adapted to explore these possibilities. However if, as Vasić and Wijnen (2005) propose, stuttering-like disfluencies in people who stutter are related to hyper-vigilance with respect to subtle prosodic cues, an explanation is needed for why our participants who stutter did not show a corresponding hypervigilance with respect to the phonological errors that they make in inner and overt speech, and why subtle prosodic cues or irregularities would lead to high levels of error-repair activity when gross phonological encoding errors apparently do not. One possibility is that hypervigilant monitoring in people who stutter is ‘situation specific’, insofar as it only occurs in situations where they perceive a need to speak with a high degree of accuracy. If this is the case, then the failure of the current experiment to reveal any difference in monitoring vigilance between stuttering and control participants may simply reflect the fact that the experimental paradigm did not provide the necessary stimulus to elicit such hypervigilance.

Our failure to find any correlation between the frequency of inner-speech errors and stuttering-like disfluencies is compatible with the EXPLAN hypothesis, and it is certainly possible that the higher numbers of onset and word-order errors made by stutterers were simply a side-effect of language encoding impairment but did not in themselves contribute to the production of stuttering-like disfluencies. In keeping with EXPLAN, it is also possible that a proportion of the perseveratory errors made by participants may have resulted from an established habit of repeating readily available segments or words in order to keep going when experiencing encoding difficulties (cf. Howell & Sackin, 2001). However, verification of these possibilities would require further research.

7.4.4. *The role of working memory*

Because the original intention behind the current experiment was to compare error rates of participants who stutter with those of matched controls, differences in participants' digitspan scores have been treated as a potential confounding variable. The main concerns were that participants with shorter digitspans may (a) find it more difficult to remember tonguetwisters and therefore make more errors, and (b) have more difficulty remembering their errors and therefore tend to under-report (or mis-report) them. Thus in all regression analyses, any variance attributable to differences in participants' digitspans was first co-varied out by entering digitspan as the first predictor. However, it has been proposed (e.g. Bajaj, 2007) that impairments of (various aspects of) working memory may be directly implicated in stuttering. In light of such proposals it is noteworthy that digitspan was itself a significant predictor of self-reported error-rates; and, in line with studies that have reported poor non-word repetition abilities in stutterers (Anderson et al., 2006; Hakim & Bernstein Ratner, 2004; Ludlow et al., 1997), our current findings suggest that working-memory limitations may play a role in the higher speech-error rates found in stutterers. This is perhaps not surprising bearing in mind the close association between working memory and phonological encoding, (Acheson & MacDonald, 2009a, 2009b).

7.5. *Caveats*

7.5.1. *Ecological validity of the study*

A number of studies have found that signs of language impairment only become apparent in stutterers under conditions of increased cognitive load (e.g. Bosshardt et al., 2002; Weber-Fox, Spencer, Spruill III, & Smith, 2004). In light of these findings, it would appear that the current experimental paradigm must have been sufficiently cognitively demanding to reveal differences between the two participant groups. Nevertheless, the tonguetwister paradigm did not require participants to compose their own utterances or to attend to prosodic encoding, and neither did it require them to monitor whether they were being understood by their conversation partners or to attend to conversation partners' responses. It may thus have been less cognitively demanding than many of the speaking situations commonly encountered by participants in real life. It thus remains possible that language encoding errors may occur substantially more frequently in stutterers in the cognitively demanding speaking situations encountered in their everyday life than they did in the tonguetwister experiment.

7.5.2. *Word omissions*

Together with the instruction to speak in time to the (visual) metronome, participants were also instructed to skip any words they felt likely to stutter on. This was to ensure that their overt recitations were uttered fluently and at the correct speech rate. We presumed that stuttering only occurs in overt speech, so we did not expect this instruction to cause participants to skip words in the silent (inner-speech) condition. However, we cannot rule out the possibility that, in the stuttering group, a proportion of words may have been omitted because they were associated with past experiences of stuttering. Had these words not been omitted, the inner-speech error rates of participants who stutter may have been even greater.

7.6. Experiment 4: Conclusions

When constrained to speaking at a fixed speech-rate, adults who stutter make more onset and word-order errors than do age, education and gender-matched, normally-fluent controls. Moreover, the majority of these errors have their origin in the speech plan, before the onset of articulation. These findings represent important evidence in line with the first tenet of the Covert Repair Hypothesis (Postma & Kolk, 1993).

However, the lack of any correlation between the frequency of participants' inner-speech errors and the frequency of their disfluencies in real-life speaking situations suggests that covert repair of errors of phonological encoding is unlikely to contribute significantly to the manifestation of stuttering-like disfluencies in everyday speech. These findings are thus incompatible with the second tenet of the Covert Repair Hypothesis.

It is possible that the higher numbers of onset and word-order errors made by stutterers are a side-effect of language encoding impairment but do not in themselves contribute to the production of stuttering-like disfluencies. It also remains possible that, in people who stutter, stuttering-like disfluencies may stem from covert repair of timing or prosodic errors, as proposed by Vasić and Wijnen (2005), or covert repairs of errors that occur downstream from phonological encoding, during the formulation of motor commands.

8. Experiments 5 and 6: The Role of Anticipation of Communication Failure in Stuttering

8.1. Introduction

In Experiment 4 (described in the previous chapter), which used a tongue-twister paradigm to investigate phonemic error rates in inner speech, we were able to show that participants who stutter made significantly more such errors than a closely matched group of normally-fluent controls. However, despite this group difference, we found no evidence of a relationship between the frequency with which participants who stutter made phonemic errors in their inner speech and the frequency of their disfluencies in everyday life. Moreover, some of the (normally-fluent) control participants made more inner-speech errors than some of the participants who stutter.

The extent of group overlap in inner-speech error rates, together with our failure to find any relationship between inner-speech errors and stuttering severity led us to conclude that covert error repair of errors of phonological encoding cannot adequately account for the majority of instances of stuttering-like disfluencies in the speech of people who stutter.

Therefore, in the final two experiments of the thesis we consider an alternative possibility, namely that a mechanism similar to covert error repair may be activated simply by the speaker's perception of a need to speak more accurately. We hypothesise that stuttering-like disfluencies may arise in response to any stimulus that results in the speaker arriving at this perception. More specifically, we focus on the extent to which factors external to the speaker which lead to the anticipation of immanent communication failure can influence the likelihood of stuttering.

To explore this issue, we employ a new experimental paradigm whereby participants repeatedly spoke single words into what they believed was a speech-recognition system.

The system was designed to give the impression that it had difficulty recognising certain words, thus instilling in participants the anticipation of communication failure on subsequent iterations of those words.

To help put Experiments 5 and 6 into a clearer perspective, we begin by considering some aspects of stuttering which production-deficit hypotheses like the Covert Repair and EXPLAN Hypotheses have difficulty accounting for, and we then consider whether these may be better explained by the alternative theoretical perspective; of stuttering as an anticipatory struggle response, as exemplified by the Anticipatory Struggle Hypothesis (Bloodstein, 1958, 1975).

8.1.1. *Stuttering as a direct consequence of a planning deficit*

As detailed in Section 3.5, psycholinguistic hypotheses have attempted to account for the occurrence of stuttering-like disfluencies, in both stutterers and normally-fluent speakers, as the unintended side-effects of ‘covert error repair’ (Postma & Kolk, 1993; Vasić & Wijnen, 2005), or ‘restart’ (Howell & Au-Yeung, 2002) mechanisms. Under normal conditions, such mechanisms may serve (with varying degrees of effectiveness) to regulate the flow of speech and ensure that it is relatively free of errors, thus helping the speaker to make himself understood and maintain his conversation turn during times of language-formulation difficulty.

These repair and restart mechanisms are not mutually exclusive, and both provide plausible explanations for the variety of stuttering-like disfluencies that occur in both stutterers as well as in normally-fluent speakers. They also provide compelling explanations for why the likelihood of stuttering tends to decrease on subsequent iterations of previously spoken words (the 'Adaptation Effect'; Brutten & Dancer, 1980; Johnson & Knott, 1937), why stutterers are particularly likely to stutter on word onsets, why the likelihood of stuttering occurring on a word is strongly influenced by its grammatical function (Bloodstein, 2006; Howell & Sackin, 2001), length, position in the sentence, frequency and predictability (Brown, 1937, 1945; Newman & Bernstein Ratner, 2007), and for why young children whose language and articulation skills lag behind those of their peers are more likely to stutter (See Section 3.5.2 for further details).

However, the CRH and EXPLAN hypotheses are less successful at accounting for a number of other observations in relation to the distributions of stuttering events, such as why older children and adults frequently stutter on isolated, commonly occurring single words, and have particular difficulty uttering their names, and why stuttering is influenced so strongly by the characteristics of the listener and the overall dynamics of the speaking situation (Bloodstein, 1949, 1950a; Bloodstein & Bernstein Ratner, 2008). Furthermore, the Covert Repair Hypothesis fails to account for the lack of any discernable correlation between the frequency with which stutterers produce inner-speech errors and their stuttering severity in everyday speaking situations, as described in Section 7.4.3 (See also Brocklehurst & Corley, 2011), or for why stuttering sometimes persists in older children and adults who no longer show any evidence of any underlying language or speech production impairment (Conture et al., 2004).

Thus it appears that compensatory responses to underlying language production difficulty of one kind or another cannot adequately account for many instances of stuttering in older children and adults (although they may adequately account for the stuttering-like disfluencies of young children). An alternative possibility, investigated in the current experiment, is that, in adults, stuttering-like disfluencies may occur as a side-effect of a compensatory response to the *anticipation* of difficulty, and such anticipation may stem from memories of having experienced difficulty speaking or communicating in similar situations in the past.

8.1.2. *Stuttering as an anticipatory struggle response*

As noted in Section 3.3, the term ‘anticipatory struggle’ was first used by Bloodstein in the 1950s to describe a broad category of hypotheses, all of which share the idea that stutterers believe that speaking is difficult and this belief in some way interferes with the smooth running of the processes that underpin fluent speech (see Bloodstein & Bernstein Ratner, 2008 for a review).

The attribution of stuttering to stimuli representative of past speech failures (rather than to any difficulty inherent in the current formulation and execution of specific individual words) is supported by evidence from a number of early studies by Johnson and his co-workers. In the first of these Johnson and Knot (1937) found that, on successive readings of the same passage, PWS tend to stutter on the same words (the ‘Consistency Effect’). Using the same materials, Johnson and Millsaps (1937) found that if, on subsequent re-readings, previously stuttered words are crossed out and participants told to omit them, stuttering then occurred on words that had previously been spoken fluently, and a consistency effect became established on those words too. In a further, related, reading experiment, Johnson, Larson, and Knott (1937; Experiment 3) marked the words a participant had stuttered on, and additionally marked five random, unstuttered words in the same way and told the participant that each mark represented a previously stuttered word. They then found that, on subsequent re-reading of the entire passage (including the marked words), stuttering on the five marked words was three times more common than on five randomly selected (unmarked) control words. This finding suggested that cues that were perceived by the participant to be associated with past stuttering had the power to induce stuttering on whatever words were associated with them.

Anticipatory struggle hypotheses have postulated a variety of mechanisms to account for how the anticipation of stuttering can lead to the production of stuttering-like disfluencies, including ‘approach-avoidance conflict’ (Sheehan, 1953); abnormal ‘preparatory sets’ (Van Riper, 1973), and ‘tension and fragmentation’ (Bloodstein, 1975).

Central to Bloodstein’s own (1975) ‘Anticipatory Struggle Hypothesis’ is the notion that the primary symptoms of stuttering (repetitions, prolongations and blocks) are essentially tensions and fragmentations in speech, which arise in response to stimuli representative of past speech failure, and which originally arose in response to the (normal) disfluencies of early childhood. Tension and fragmentation are regarded as the symptoms of “trying too hard”, and “taking the activity apart to do it piece by piece” (Bloodstein, 1975 p4) that characteristically occur when an individual wishes to execute a complex motor activity and yet doubts that he will be successful.

By conceptualizing stuttering in this way, Bloodstein’s (1975) Anticipatory Struggle Hypothesis provides a plausible explanation for why stuttering is more likely to occur in children with impaired or delayed development of linguistic skills and/or speech motor control. Importantly, unlike the Covert Repair and EXPLAN hypotheses, Bloodstein’s hypothesis does not attribute stuttering directly to the impairment or delay. Instead, it posits that stuttering only arises if the underlying impairment/delay serves to instil in the child the belief that, in particular situations, particular sounds or words will be difficult to speak. Thus it allows for the possibility that the exact nature of the impairment or delay that underlies stuttering may differ from child to child. It also provides a parsimonious explanation for how stuttering may persist even after any language or speech impairment/delay has resolved, by postulating that a vicious circle is established whereby the anticipation itself precipitates the struggle that was anticipated. Further, because it identifies stuttering as a disorder of communication in which the responses of the listener are every bit as important as the speech of the speaker, it provides a seemingly parsimonious explanation for a range of common observations in relation to stuttering, including why stutterers rarely have difficulty speaking to themselves or when they don’t care what the listener thinks of them or what they say; and conversely, why they may find it so much more difficult to speak fluently to certain people, about certain topics and in certain social situations (Bloodstein, 1949, 1950a).

Bloodstein conceptualized anticipatory struggle as being essentially a hypertonic avoidance response stemming from the desire to avoid negative listener responses (cf. Wischner, 1950). This conceptualization fits well with recent evidence of a strong link between social anxiety and stuttering (Craig, 1990; Craig, Hancock, Tran, & Craig, 2003; Iverach et al., 2009; Menzies, Onslow, & Packman, 1999; Messenger, Onslow, Packman, & Menzies, 2004) insofar as such evidence suggests that many older children and adults who stutter are indeed highly concerned about the negative reactions of listeners. It is also supported by findings of an early experimental study by Hansen (1955), outlined in Section 8.1.3, which found that negative listener responses can have a direct impact on the likelihood of producing stuttering-like disfluencies.

However, despite its appeal, Bloodstein's (1975) Anticipatory Struggle Hypothesis has two important weaknesses. Firstly, 'tension and fragmentation' is not well specified and fails to provide an adequate explanation for precisely why stuttering-like disfluencies manifest in the variety of ways that they do (as repetitions, prolongations or blocks). Secondly, by defining tension and fragmentation as a hypertonic avoidance response, it fails to address the possibility that approach behaviour may also play an important role[18] (cf. Sheehan, 1953).

---

[18] In other words, stuttering may stem not only from avoidance of punishing stimuli, but also from attraction to rewarding stimuli.

Regarding the second of these weaknesses, the common clinical observation that stutterers experience relatively little difficulty speaking fluently in situations where the semantic or propositional content is unimportant or already known to the listener (Bloodstein, 1949, 1950a) suggests that the (positive) desire to convey information is more likely to be the key precipitating factor, although there is, as yet, only very limited experimental evidence to support this view.

8.1.3. *Experimental investigation of the effects of anticipation of communication failure*

The final two experiments of this thesis constitute an experimental investigation of the effects of anticipation of communication failure on speakers' experiences of speaking. The first (Experiment 5) investigates whether the experience of stuttering can be precipitated by the anticipation of communication failure; and, the second (Experiment 6) investigates whether this effect extends to normally-fluent speakers (for whom there is no reason to suspect any underlying language or speech-production impairment). These two experiments are loosely based on a paradigm developed by Hansen (1955), originally designed to test the effects of different valences of audience response on stuttering severity. In Hansen's original experiment, participants who stutter performed a variety of reading and photograph-description tasks in front of an audience ranging from 12 to 25 people. The lighting was turned down so participants could not see the audience's faces. Positive or negative audience feedback was delivered to the speakers indirectly, by means of a series of green and red lights and corresponding counters, located on a table in front of the speaker. Each member of the audience was given a three-way switch to operate, with settings for positive, neutral and negative responses. The speaker and the audience were led to believe that the switches operated the lights and the counters summed the two types of response for the duration of speaking. In reality, the coloured-light activation and counter scores were regulated by the experimenter. Hansen found that, although overall there was a general decrease in stuttering over the duration of the experiment, the rate of decrease was greater where favourable stimuli were delivered than where unfavourable stimuli were delivered. These trends became noticeable after a short time lag, and were most noticeable during spontaneous speech when it was easier for the speaker to focus on the audience responses.

In our experiments (Experiments 5 & 6), instead of speaking to an audience, participants were instructed to speak into what they believed was speech-recognition software on a computer, and received automatic online feedback indicating whether or not the words they had spoken had been correctly recognized. As in the Hansen experiment, the feedback was, in reality, determined by the experimenter, and bore no relationship to the accuracy or fluency with which participants spoke. We chose to mimic a speech-recognition system (rather than using human confederates) because it enabled us to focus on the influence of the desire to communicate information (rather than the desire to avoid negative listener responses). Specifically, the use of such a system allowed us to create a communication situation in which participants were unlikely to feel any need or desire to elicit a favourable reaction from listeners. To further avoid any possibility that participants' performances may be affected by the fear of negative evaluation by potential listeners or over-hearers, participants were led to believe that they were not being recorded, that nobody was listening to them or able to hear them speak, and that the speech-recognition process was entirely automatic.

To maintain the illusion that participants' speech was being 'recognized', the software incorporated a voice-activated switch that was sensitive to participants' verbal responses. To ensure that the feedback from the software was plausible, we employed participants whose overt disfluencies were relatively mild and whose utterances were most likely to closely approximate the target utterances.

Because independent-rater judgments of stuttering are unreliable with respect to single-word utterances (many of the prosodic cues that alert listeners to stuttering in the multi-word utterances are absent), and because, when speaking into speech-recognition software, speakers frequently prolong or hyper-articulate words on purpose (Stent, Huffman, & Brennan, 2008), our primary analyses of stuttering were based on participants' self-reports. These primary analyses had the added benefit of enabling us to investigate directly the relationship between the anticipation of communication failure and the experience of 'loss of control' (see Section 5.1).

8.2. Experiment 5

In Experiment 5, all participants were adults who stutter. Participants used a simple cue, displayed on a computer screen, to identify a target word from a selection of five possible options (also displayed on the screen), and then spoke that target word four times consecutively. Immediately prior to the first iteration of a word, participants self-reported whether or not they expected to stutter on this word. Then, immediately following each iteration, participants (a) self-reported whether or not they had stuttered, and then (b) received feedback (on the computer screen) indicating whether the word they had spoken had been correctly recognized by the software. Across each of the four iterations, the feedback consistently indicated correct or incorrect recognition of the target word. Thus, it became increasingly clear to participants whether the remaining iterations were likely to be correctly recognized by the software.

We hypothesized that, due to the 'adaptation effect' (Brutten & Dancer, 1980; Johnson & Knott, 1937), there would be an underlying trend for self-reports of stuttering to decrease across iterations. If stuttered disfluencies result solely from a language or speech production deficit, this reduction would be unaffected by whether the software apparently failed, or succeeded, to recognize each word spoken, since lexical difficulty was held constant across conditions. Evidence that the feedback given to participants affected their performance differently in each of the conditions, however, would implicate an additional process. Regardless of whether or not participants who stuttered had production deficits, the anticipation of a struggle to communicate would make participants relatively more likely to produce stuttering-like disfluencies in the condition where the software apparently failed to recognise their productions of a particular word. We also hypothesised that these immediate effects would be evident over and above an underlying tendency to stutter on words that speakers had come to perceive as long term ‘problem words’, due to difficulties encountered in other earlier speaking situations (as revealed by participants’ responses to the question “will you stutter on this word?”).

8.2.1. *Method*

Participants

Fourteen participants were recruited through stuttering self-help groups and through the University of Edinburgh student employment website. Data from two participants was not analyzed because they failed to follow instructions and/or realized that they were not interacting with real speech recognition software. Mean age of the remaining 12 participants (9 male) was 32 (range 25 to 41); mean education level was 3.8 on a scale where 1 corresponds to General Certificate of Secondary Education (GCSE) or equivalent, and 5 indicates a postgraduate degree. Two participants were university students; all others were in paid employment.

All had previously been diagnosed with persistent developmental stuttering by a speech therapist and all considered themselves as still suffering from the condition. SSI-4 (Riley, 2009) stuttering severity scores were derived from video recordings of each participant speaking spontaneously and reading aloud. Mean SSI-4 score was 15.7, ranging from ‘very mild' (6) to 'severe' (34). Participants produced a mean of 5.8 stuttering-like disfluencies per 100 syllables when speaking (range 1 to 16), and 5.6 when reading aloud (range 0 to 24).

Participants also completed Section 3a of the Overall Assessment of the Speaker's Experience of Stuttering (OASES; Yaruss & Quesal, 2006), in which respondents rate their current difficulty in communicating verbally on a five-point scale in each of ten commonly occurring situations. These include, for example, talking with another person one to one, initiating conversations, speaking to strangers, and continuing to speak regardless of how your listener responds to you. Mean score was 28.1 out of 50 (range 21 to 38). Participants were also asked to rate how difficult it was to speak fluently, without stuttering or avoiding words, in the same ten situations, resulting in a mean score of 32.3 out of 50 (range 14 to 47)[19].

Apart from stuttering, participants reported no other speech, language, hearing or visual impairments.

Materials

The materials consisted of '*target words*' (that participants were required to identify and speak out loud), associated '*option words'*, '*cues*' and 'feedback words'. The materials were divided into three sets, two of which were used in the two experimental conditions and the third of which was used as fillers. Each set contained 16 *target words* to be spoken out loud (see Table 8-1 for examples). In the two experimental sets, each target word was associated

---

[19] For comparison, mean OASES 3a and 'fluency-difficulty' scores for normally-fluent participants were 19.7 and 18 respectively in Experiment 5; and 23.2 and 22.1 respectively in Experiment 6

with five ‘option words’, one of which was identical to the target word whereas the other four of which were distractors (that differed from the target word by just the onset phoneme). Each target word was also associated with a *cue* which participants used to distinguish it from its distractors: For example, for the target *prod* and the distractors *plod, pod, odd, mod,* the cue was *push with a finger or stick* (see Table 8-1). Each target word was also associated with a ‘feedback’ word. In one experimental condition the feedback word was always ‘Correct’ insofar as it was identical to the target word. In the other it was always ‘Incorrect’ insofar as it was identical to one of the four distractor words. In the filler set, each of the 16 target words was associated with five phonologically-different option words (four of which were distractors) and feedback words were always correct (i.e. identical to the target words).

Table 8-1. Examples of cues and their associated target words, distractors and feedback in the two experimental conditions and fillers.

Set	Cue	Option words		Feedback (‘Word recognized’)
		Target word	Distractors	
1. correct feedback	a vital organ	heart	art cart Bart art	HEART
2. incorrect feedback	Push with a finger or stick	prod	plod pod odd mod	PLOD
Fillers	Where someone is buried	grave	wick fan shrink mat	GRAVE

Note: All filler items were followed by correct feedback.

The feedback associated with each of the target words (i.e. the word that would be presented as having been ‘recognised’ once the participant had spoken) was predetermined, insofar as it was not influenced by the participant’s performance. For the filler set and for one of the experimental sets, whatever the participant said, the feedback would indicate that they had given the ‘correct’ response to the cue. For the other experimental set, for the first three iterations of each target word, feedback would always indicate that they had given an ‘incorrect’ response, and for the fourth iteration, it would indicate that they had given the ‘incorrect’ response 50% of the time. The inclusion of the occasional instance of ‘correct’ feedback in the incorrect condition was to prevent participants from concluding that in the *incorrect* condition later iterations would always be incorrectly recognised, and that it was therefore pointless trying to get the software to recognise their later iterations of words. The target words in all three sets were matched (overall) for frequency. The materials were counterbalanced insofar as a second version of the materials was drawn up in which the feedback associated with the two experimental sets was reversed. Half of the participants received one version and half received the other. Thus, across the entire experiment, each experimental target was equally likely to have been ‘correctly’ or ‘incorrectly’ named.

The experiment was controlled and administered using a laptop with a 15” screen. Participants made spoken responses via an integral headset and microphone; manual responses were made via a five-button response-box.

Procedure

Prior to the experiment, participants completed a demographic questionnaire. The experiment then began with a computer-led tutorial session. During this session, participants were informed that the investigation concerned their ability to answer questions using speech-recognition software, and that the software was pre-primed with appropriate lists of words, from which it would select the best match. They were informed that during the experiment proper, nobody would be able to hear them speak and that their speech would not be recorded. Finally, they were told that there was a financial reward if more than 71% of responses were correctly recognized (since the ‘correctness’ of each response was predetermined, each participant in fact scored 72%). These deceptions were necessary to minimize the likelihood that participants would be concerned about potential negative listener evaluations, and to maximize the likelihood that their sole motivation was to make the machine understand their responses. Fully-informed consent was obtained retrospectively, once the experiment was complete.

Following the tutorial, participants underwent a two-item practice session. The experimenter adjusted microphone sensitivity if necessary, and encouraged the participants to respond promptly where cued to speak, and to speak loudly enough for the software to register a response. Following the practice session the experimenter left the room, and the experiment proper commenced.

The procedure for each practice item, and for each of the 48 targets which followed, was identical, and consisted of four repetitions of a target-naming sequence. Each repetition began when five words, comprising the target and its four associated distractors, were displayed in an arbitrary order along the top of the computer screen. Simultaneously, the cue phrase which identified the target was displayed below the list. Immediately below the cue was the question "do you think you may stammer on this word?" (See Figure 8-1). Having used the cue to identify the target, the participant responded to this question using one of three response keys, labelled "no", "maybe", and "yes" on the response box. Pressing any of these keys caused a large hourglass to appear in the centre of the screen for 1000ms. Once the hourglass disappeared, the software began recording input from the microphone. After another 250ms, the screen turned green, and a large mouth icon appeared, prompting the participant to speak. At the same time, a voice-activated switch became potentiated.

The sequence continued in one of two possible ways, depending on whether or not the voice-activated switch was triggered. If the voice-activated switch was triggered within 1300ms, the green screen remained for 2500ms, after which it was replaced with a black screen and the hourglass icon. The hourglass disappeared after 2000ms, and was replaced 250ms later with the question "did you stammer on this word?" Participants answered using the response keys labelled "no", "maybe", and "yes". Pressing a response key triggered a feedback screen, described below.

The feedback screen showed the cue, followed by the list of target and distractors, followed by “you said...” and a preselected response (either the target, or one of the distractors). If the preselected response was the target, the screen background was green, and below the response the participant was informed that “the word you selected was correct!” For distractors, the screen was red, and participants were told that their selection was wrong. In each case, the screen additionally showed an online update of the “percentage correct so far”, followed by the words “you need 71% to win”. Below this, at the bottom, were the words “press any key to continue”. Pressing any of the response keys began the next sequence for the current target (or the first for the next target, if this was the fourth sequence). Targets were presented in a random order.

If the voice-activated switch was not triggered, the green screen was replaced 2000ms after it appeared with a red screen, and the message “sorry, I couldn’t identify what you said”. Immediately below this message was the question “did you stammer on this word?” 250ms after a response, the words “please try again” appeared for 1000ms. This was followed by a 250ms pause, after which the sequence started again at the green screen with the mouth icon. If the voice-activated switch failed to trigger a second time, the red screen appeared once more, but following the participant’s response to the stammering question, no feedback was given (the feedback screen did not appear). Instead, the iteration was abandoned and the next iteration was initiated. Thus participants were allowed up to two attempts at each of the four iterations of each target word before the sequence for the following target was initiated. Sound-files and response-box key presses were automatically recorded for all attempts, irrespective of whether or not the voice switch was successfully activated. (See Figure 8-1)

Once the experiment had finished, participants were fully debriefed, in part in order to ascertain whether or not they had realized that they had not been engaging with real speech-recognition software. Data from participants whose responses revealed that they had come to this conclusion were excluded from subsequent analyses.

The debriefing interviews were video recorded, for use as a spontaneous speech sample from which the percentage of syllables stuttered was estimated; after the debriefing, participants were also recorded reading a passage aloud. These recordings form the basis of the SSI-4 analyses of stuttering severity reported above.

Coding and Analysis

Irrespective of whether or not the voice-switch was activated, data from the first attempt at each iteration were included in the analyses (data from second attempts were not analyzed).

Recordings of participants' utterances were analyzed by the experimenter, using PRAAT (Boersma & Weenink, 2009). Two acoustic measures were taken: (1) Vowel onset latency[20], measured from the moment the recording was activated (250ms before the screen turned green and the 'mouth' icon appeared) to the onset of the first vowel sound, as determined by the onset of striations and associated formants on the spectrograph; and (2) word duration, measured from the beginning to the end of all evidence of speech-related activity on the spectrograph (duration measures thus included prolongations and repetitions but not silent blocks).

---

[20] As stuttering-like disfluencies tend to occur prior to the vowel onset, the vowel onset provided the most reliable measure of the onset of fluent speech.

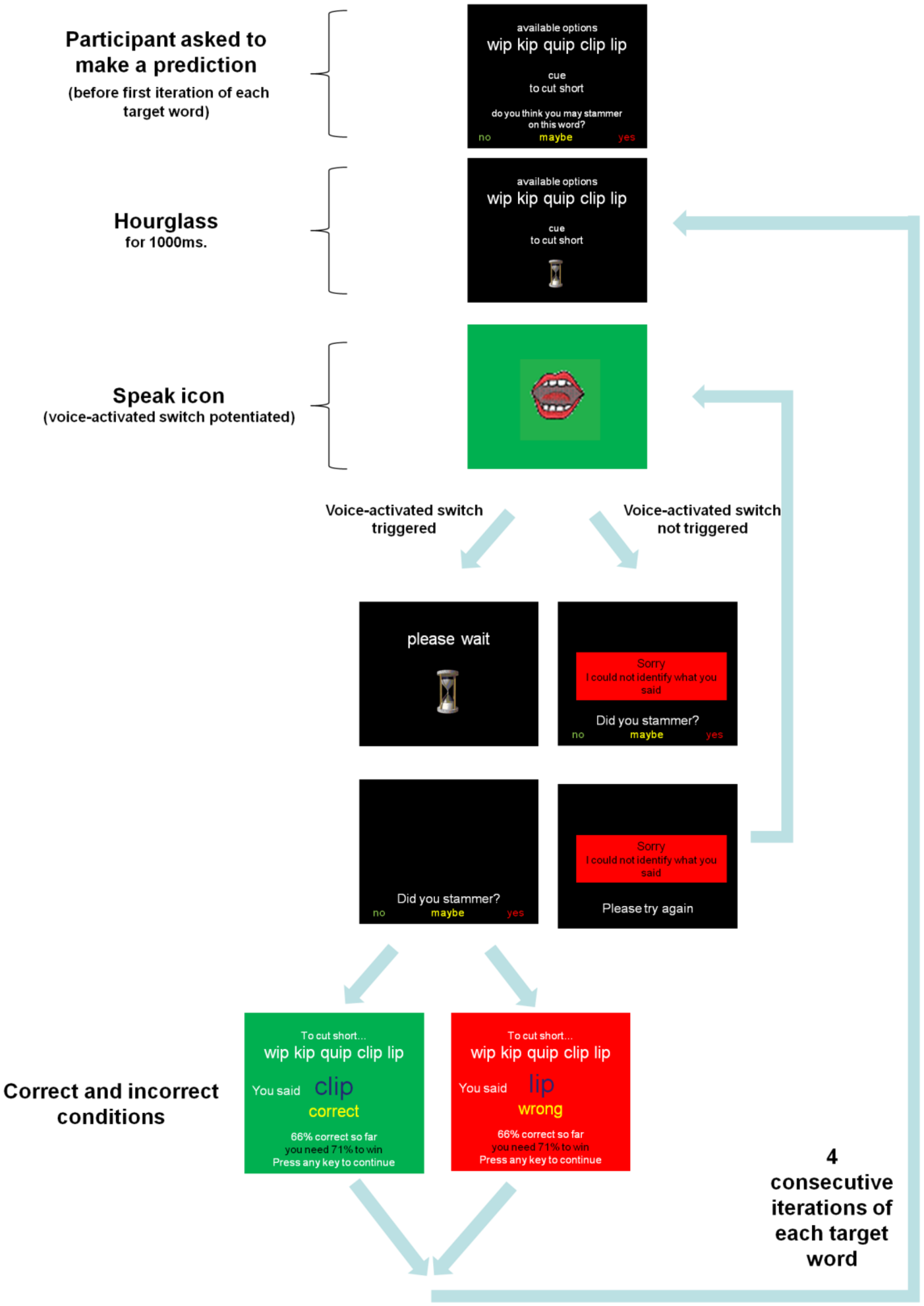


Figure 8-1. Experiments 5 & 6: Sequence of on-screen instructions to participants.

Statistical analyses of the acoustic data (vowel onset latencies and word durations) were carried out using linear mixed-effects regression modelling (Breslow & Clayton, 1993; DebRoy & Bates, 2004) using the lme4 package (Bates & Maechler, 2009) in R (R Development Core R Development Core Team, 2009). We generated base models which included an intercept with random by-participant and by-item variation, and then proceeded to add predictors stepwise to each model. For each predictor, we first added the random (by-participant) term, to test for inequality of variance across levels, then added the corresponding fixed term. The first predictor tested was ‘Condition’ (with two levels: correct recognition and incorrect recognition); then ‘Iteration’ (with four levels, one for each iteration), followed finally by a predictor for the Condition by Iteration interaction. Selection of models was based on χ^2 tests to assess whether the fit of the model to the data was improved (as indicated by a significant decrease in the model likelihood ratio) by the addition of each (random or fixed) predictor. Predictors were retained only if the current (best) model was improved. Where two or more predictors each significantly improved the current model, we selected the model which had the smallest log-likelihood. We iterated this process until we found a ‘best fit’ model which could not be improved by the addition of further predictors. Where models were selected, the *t* statistic, calculated from each estimated coefficient and its standard error, was used to determine whether the coefficients differed significantly from zero (see Agresti, 2002).

Analyses of participants’ self-ratings of having stuttered were carried out using logistic mixed-effects regression modelling, using the lme4 package (Bates & Maechler, 2009) in R (R Development Core R Development Core Team, 2009). This approach allowed us to investigate the independent contributions of predictor variables to the (log) likelihood of a word being self-rated as stuttered. Once again, we generated a base model which included an intercept, random by-participant and by-item intercept variation and then added random and fixed predictors stepwise to each model under consideration. The first (factorial) predictor to be tested in this way was participants' self-ratings of "Stuttering predicted" (with separate levels for 'no', 'maybe' and 'yes' responses to the question "do you think you will stutter on this word?"); then a predictor for 'Condition'; then 'Iteration', followed finally by a predictor for the Condition by Iteration interaction. Model selection proceeded as for the linear mixed-effects models of acoustic data described in the previous paragraph.

8.2.2. *Results*

Vowel onset latencies

In total, participants provided 1467 codable samples. Mean and standard deviations are provided in Figure 8-2. The best-fit model of onset latencies included 'Condition' as a fixed predictor and 'Iteration' as a random predictor (improvement due to adding 'Iteration: χ^2 (2) = 12.26, p <.002).

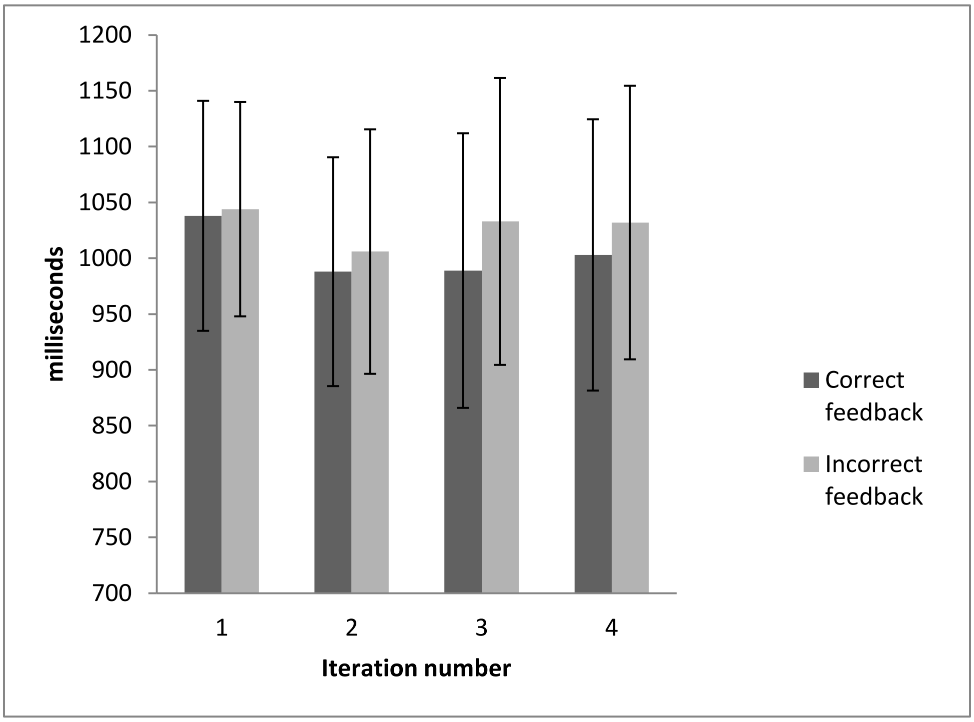


Figure 8-2. Experiment 5: Vowel-onset latencies in Participants who stutter, by condition and iteration.
Error bars show (global) standard deviations.

Adding further predictors did not improve the model (all *p*≥.441). Table 8-2 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

The model reveals that onset latencies became more variable across iterations, and that, irrespective of iteration, mean vowel onsets were 28ms longer in the ‘Incorrect’ condition.

Table 8-2. Vowel-onset latencies Exp6 - participants who stutter
Coefficients for best fitting model.

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = vowel onsets (milliseconds.)							
Intercept	correct feedback, iteration = 1	1041	29	<.001 ***	by word	1980	44
					by subject	9063	95
Condition	Incorrect feedback	28	10	.007 **			
Iteration	+1				by subject	305	17
					Residual	36826	192

Word durations

In total, participants provided 1423 codable samples. Mean durations and standard deviations are provided in Figure 8-3. The best-fit model of word durations included ‘Condition’ as a random predictor and ‘Iteration’ as a fixed predictor (improvement due to adding the ‘Iteration’ predictor: χ^2 (1) = 12.00, *p* <.001). Adding further predictors did not further improve the model (all *p*≥.129). Table 8-3 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

The model reveals that compared to the ‘Correct’ condition, durations of words uttered in the ‘Incorrect’ condition were more variable but not reliably longer. Irrespective of condition, the mean duration of words uttered increased across iterations by 11ms per iteration.

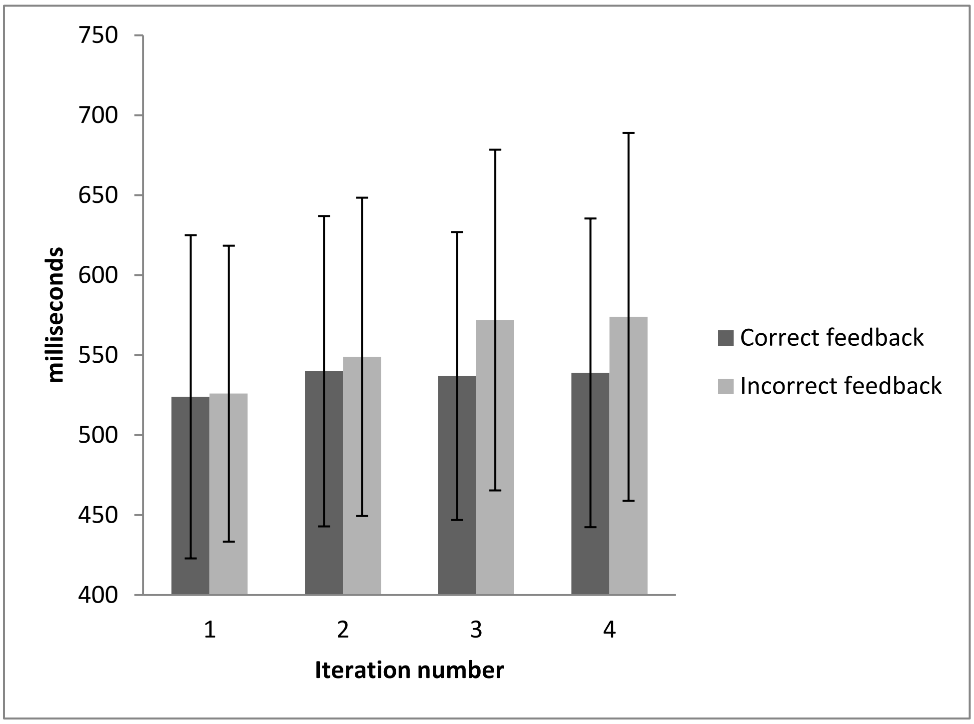


Figure 8-3. Experiment 5: Word durations in Participants who stutter, by condition and iteration.
Error bars show (global) standard deviations.

Table 8-3. Word durations: Exp6 - participants who stutter
Coefficients for best fitting model.

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = word durations (milliseconds.)							
Intercept	Iteration = 1 correct feedback	551	39	<.001 ***	by word	10717	104
Intercept					by subject	17639	133
Condition	Incorrect feedback				by subject	2840	53
Iteration	+1	11	3	<.001 ***			
					Residual	17077	131

Self reports of stuttering

In total, participants provided 1536 eligible self-reports (32 words, each repeated four times by 12 participants), out of which participants self-reported a total of 358 stutters (230 possible and 128 definite) (See Figure 8-4). Prior to the first iteration, there were 131 instances where participants predicted possible stuttering and 34 instances where they predicted definite stuttering.

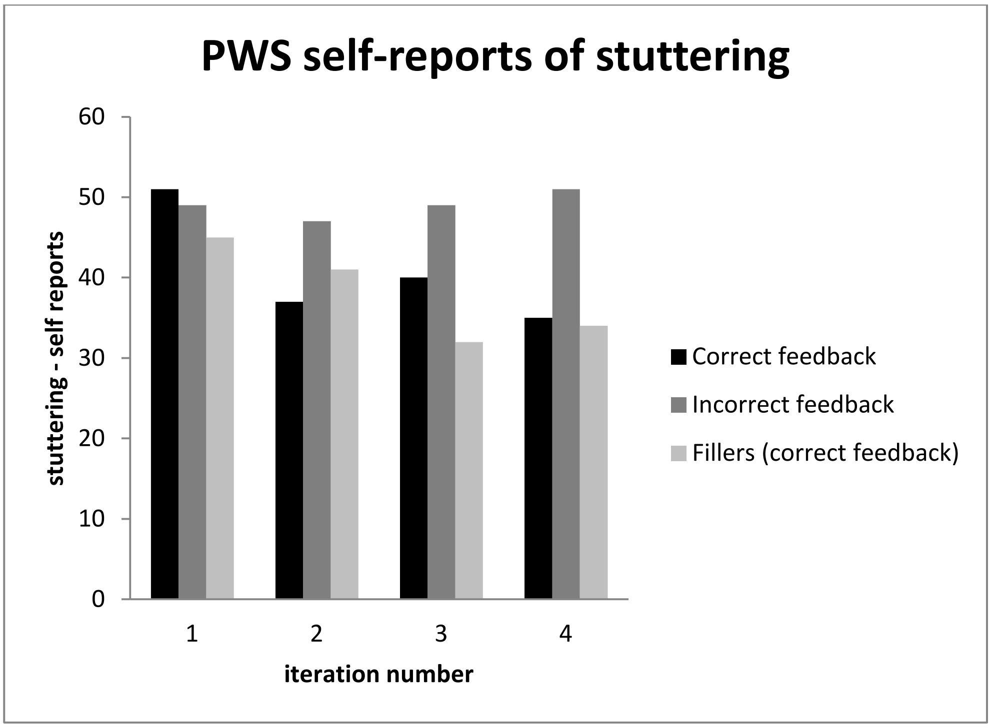


Figure 8-4. Experiment 5: Self-reports of stuttering in participants who stutter. Numbers of reports, across iterations, the two experimental conditions, and fillers.

The best-fit model of self-reported stuttering included the random predictors: ‘Stutter predicted’ and ‘Iteration’, and the fixed predictors: ‘Stutter predicted’, ‘Iteration’, ‘Condition’ and the ‘Iteration by Condition’ interaction (improvement due to adding the Iteration by Condition interaction: χ^2 (3) = 10.07, p =.018). Table 8-4 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

The model reveals that, independent of the feedback they received, once random variance was accounted for, participants were 9.9 times (i.e. $e^{2.29}$) as likely to self report stuttering on words upon which they had predicted (prior to the first iteration) that they would ‘maybe’ stutter, and 46.5 times as likely to self report stuttering on words upon which they had predicted (prior to the first iteration) that they would ‘definitely’ stutter. Independently of the above, the model also reveals that, overall, the likelihood of self-reporting words as ‘stuttered’ reduced across iterations. However, crucially, the significant ‘Condition by Iteration’ interaction confirms that participants reported less of a decrease in stuttering across iterations in the ‘Incorrect’ condition compared to the ‘correct’ condition: Compared to the ‘Correct’ condition, in the ‘Incorrect’ condition, once random variance was accounted for, the likelihood that participants would self-report stuttering increased by a factor of 1.47 (i.e. $e^{0.39}$) with each subsequent iteration.

Table 8-4. Likelihood of self-reporting stuttering: Exp6 - participants who stutter

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = likelihood of self-reporting stuttering							
Intercept	stutter not predicted Iteration = 1, correct feedback	-3.57	0.78	<.001 ***	by word	0.28	0.53
					by subject	5.57	2.36
Stuttering predicted	maybe	2.29	0.80	.004 **	by subject	5.64	2.38
Stuttering predicted	definitely	3.84	1.17	.001 **	by subject	11.01	3.32
Condition	incorrect	-0.03	0.39	.935	by subject	0.52	0.72
Iteration	+1	-0.33	0.12	.005 **			
Condition x iteration		0.39	0.16	.017 *			

Post-hoc analyses

Visual inspection of Figure 8-2 – Figure 8-4 suggests some degree of correspondence between participants’ mean vowel onset latencies, word durations and the pattern of their self reports of stuttering. However, because participants may have purposefully prolonged or stressed key sounds in order make it easier for the speech-recognition software to recognize the target words, it is unclear to what extent these vowel-onset and duration differences across conditions and iterations were the result of stuttering and to what extent they were strategic. To investigate this, we performed two additional linear (mixed-effects) regression analysis in which we tested whether participants' self-ratings of having stuttered predicted vowel onset latencies and/or word durations. The best-fit models from these two post-hoc analyses revealed that a positive response to the question "did you stutter" was associated with an increase in vowel-onset latency of 131ms (improvement due to adding vowel-onset latency as a fixed predictor χ^2 (1) = 27.85, *p* =.036), but was not associated with any increase in word duration (See Table 8-5 for the model coefficients, and the probabilities that they could have occurred by chance).

Table 8-5. Post-hoc analyses of vowel onset latencies and durations: Exp6 - participants who stutter

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = vowel onsets (milliseconds.)							
Post-hoc analysis (intercept)	no stuttering	1011	29	<.001 ***	by word	1280	36
					by subject	9325	97
Stuttering (self-reported)	Yes	131	33	.018 **	by subject	28564	169
					Residual	33157	182
DV = word durations (milliseconds.)							
Post-hoc analysis (intercept)	no stuttering	498	27	<.001 ***	by word	10029	100
					by subject	13449	116
Stuttering (self-reported)	Yes				by subject	20351	143
					Residual	16383	128

8.2.3. *Discussion*

The most important finding from Experiment 5 is that the likelihood of participants self-reporting stuttering on a particular iteration of a word was predicted not only by their anticipation of upcoming difficulty prior to the first iteration of that word, but was also independently predicted by the nature of the feedback they had received from their immediately preceding iterations of that word. Moreover, the feedback participants received influenced the likelihood of stuttering on subsequent iterations of that word despite the fact that, in this paradigm, the feedback was pre-determined independently of their actual performance.

These findings support the hypothesis that the anticipation of communication failure can precipitate stuttering in adults who stutter (AWS). The findings also suggest that, in AWS, the anticipation of communication failure is influenced both by feedback from the listener (or software) in response to their immediately preceding utterance(s) as well as by longer-term factors (as revealed by participants' responses to the question "will you stutter on this word?" which was posed before the first iteration).

Together, the results from these two post-hoc analyses suggest that the finding that participants' word durations were longer in the 'Incorrect' condition was largely the result of participants' strategic prolongations of those words, and not the result of stuttering. In contrast to this, stuttering self-reports *were* reliably associated with longer vowel onset latencies.

It is noteworthy that stuttering self-reports did not increase across the four iterations of the 'Incorrect' condition. Rather, the condition by iteration interaction was entirely due to the lack of any decrease in stuttering self-reports across iterations in that condition. Because of practical limitations, it was not possible to incorporate a 'no-feedback' condition into the experimental paradigm, so the paradigm does not inform us about how the likelihood of stuttering would have changed across iterations in the absence of any feedback whatsoever. However, in an earlier experimental study, in which participants read five consecutive iterations of each word and did not receive feedback, Brutten and Dancer (1980) found that stuttering decreased significantly across iterations. They attributed this 'Adaptation Effect' to motor learning/rehearsal.

In light of Brutten and Dancer's findings, it seems likely that the decrease in stuttering across iterations in the 'Correct' condition of the current experiment can be accounted for in terms of the adaptation effect (that does not require listener feedback). Thus the overall pattern of responses found in our current experiment most likely reflects the product of two simultaneous influences, such that, in the 'Incorrect' condition, the adaptation effect and that would otherwise have been apparent, is prevented or cancelled out by the experience of repeated communication failure. This conclusion is reinforced by the finding that the decrease in stuttering across iterations of filler words almost exactly matched the decrease in the 'Correct' condition (see Figure 8-4).

8.3. Experiment 6

8.3.1. *Method*

In the previous experiment (Experiment 5), the two analyses of word durations suggested that participants who stutter consciously prolonged or emphasized key phonemes to assist word recognition by the software. However, their self-reports of stuttering did not indicate that this strategy resulted in increased stuttering, as might have been expected in view of Bloodstein's (1959, 1975) attribution of stuttering to a process of 'tension and fragmentation' associated with exerting conscious control over complex fine motor tasks. It is possible, however, that the results were too noisy for the effect to be revealed.

Bloodstein's equation of the production of stuttering-like disfluencies with tension and fragmentation leads to the prediction that normally-fluent speakers should also experience a degree of difficulty speaking fluently in situations where the repeated experience of communication failure leads them to exert an increased level of conscious control over their speech. Experiment 6 was designed to test this prediction on a group of normally-fluent speakers. A key advantage of using normally-fluent speakers to test the relationship between communication failure and fluency difficulty is that the fluency of their responses is less likely to be confounded by any possible underlying language or speech production difficulty or by the anticipation of communication failure stemming from experiences prior to commencing the experiment.

Experiment 6 is essentially the same as Experiment 5, the only difference (apart from the participants) being that we made changes to the instructions to ensure that they were meaningful to participants who do not stutter. In Experiment 6 we thus predicted that the repeated experience of communication failure in the 'Incorrect' condition should lead to an increase, across iterations, in self-reports of difficulty speaking fluently relative to the correct condition, and that this increase may be associated with longer vowel onset latencies and durations.

Participants

Participants who do not stutter (PNS) were recruited through the Edinburgh University subject volunteer pool and student employment agency. Twelve participants (nine male) took part in the experiment. All participants were native speakers of English. Mean age of the participants was 36 (range 23 to 43); and mean education level was 3.50 on a scale where 1 corresponds to General Certificate of Secondary Education (GCSE) or equivalent, and 5 indicates a postgraduate degree. Three participants were university students; all others were in paid employment. t-tests revealed that this group was closely matched to the stuttering group in Experiment 5 (age: $t = 1.66$, $p = .11$; education: $t = 0.20$, $p = .85$).

Mean scores for the OASES section 3a (Communication Difficulty) was 23.2 (range 13 to 33). Mean score for Fluency Difficulty ratings was 22.1 (range 10 to 40). SSI4 (Riley, 2009) mean score was 6.7 (range 5 to 10). Mean number of SLDs per hundred syllables on the reading task was 0.3 (range 0 to 1.1); and on the speaking task was 1.5 (range 0 to 4.6).

Participants reported no speech, language, hearing or visual impairments that were likely to influence their results.

Materials

The experimental materials were identical to those used in Experiment 5.

Procedure

The procedure was identical to in Experiment 5 except that instead of being asked “do you think you will stammer” and “did you stammer”, participants were asked “will you be able to say this word fluently” and then “were you able to say this word in exactly the way you intended”. These questions were chosen because they summed up the subjective experience of stuttering without mentioning the label ‘stuttering’ (or ‘stammering’). The response-button labels were also adjusted accordingly.

Analyses

The same analyses were carried out as in Experiment 5.

8.3.2. *Results*

Vowel onset latencies

As in Experiment 5, only participants’ first attempts at each of the four iterations were coded. In total, participants provided 1526 codable samples. Mean and standard deviations are provided in Figure 8-5.

The best-fit model of onset latencies included the random predictor: ‘Condition’; and the fixed predictors: ‘Condition’, ‘Iteration’, and the Condition by Iteration interaction (improvement due to adding Condition by Iteration interaction: χ^2 (3) = 8.52, $p < .036$). Table 8-6 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

The model reveals that, irrespective of iteration, vowel onset latencies in the ‘Incorrect’ condition were longer than in the correct condition; vowel onset latencies became more variable across iterations; and furthermore, relative to the 'Correct' condition, vowel onset latencies in the 'Incorrect' condition increased by 14ms with each subsequent iteration.

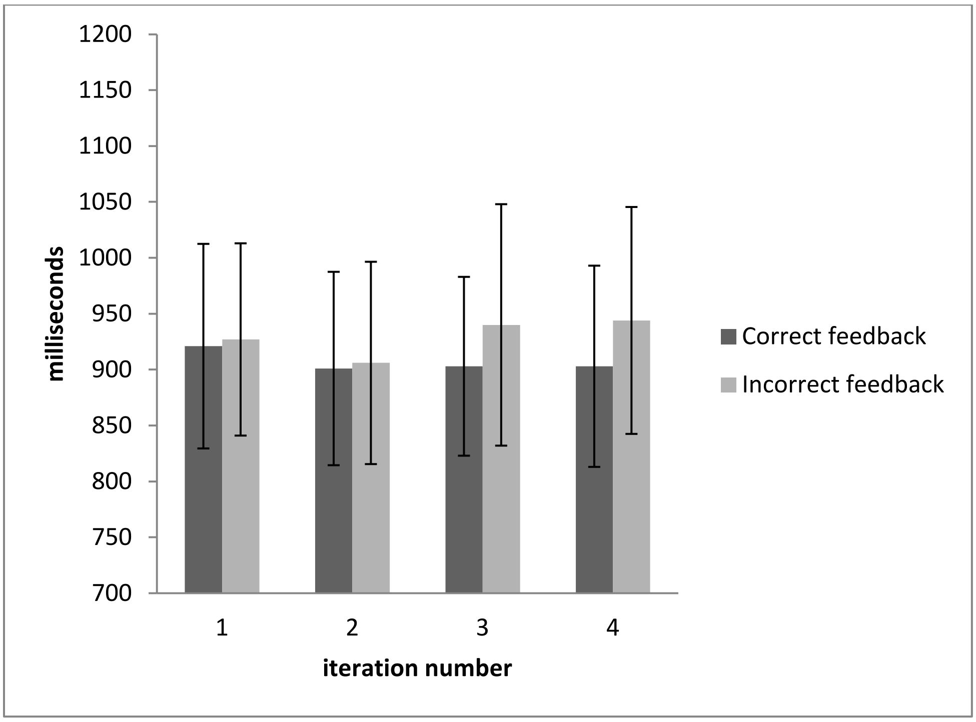


Figure 8-5. Experiment 6: Vowel-onset latencies in participants who do not stutter, by condition and iteration.
Error bars show (global) standard deviations.

Table 8-6. Vowel-onset latencies Exp7 - participants who do not stutter
Mixed effects analysis of factors influencing vowel onset latencies. Model coefficients and probabilities are given for best-fitting models, as determined by Chi squared model comparisons.

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = vowel onsets (milliseconds.)							
Intercept	Iteration = 1 correct feedback	916	36	<.001 ***	by word	2999	55
					by subject	13657	117
Condition	Incorrect feedback	1	15	.513	by subject	832	29
Iteration	+1	-5	5	.277			
Condition x iteration		14	6	.032 *			
					Residual	19412	139

Word durations

In total, participants provided 1526 codable samples. Mean durations and standard deviations are provided in Figure 8-6. The best-fit model of word durations included random predictors: 'Iteration' and 'Condition', and the fixed predictor: 'Iteration' (Improvement due to additionally adding 'Iteration' as a fixed predictor: χ^2 (1) = 8.47, *p* <.001). Adding further predictors did not further improve the model (all $p \geq .166$). Table 8-7 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

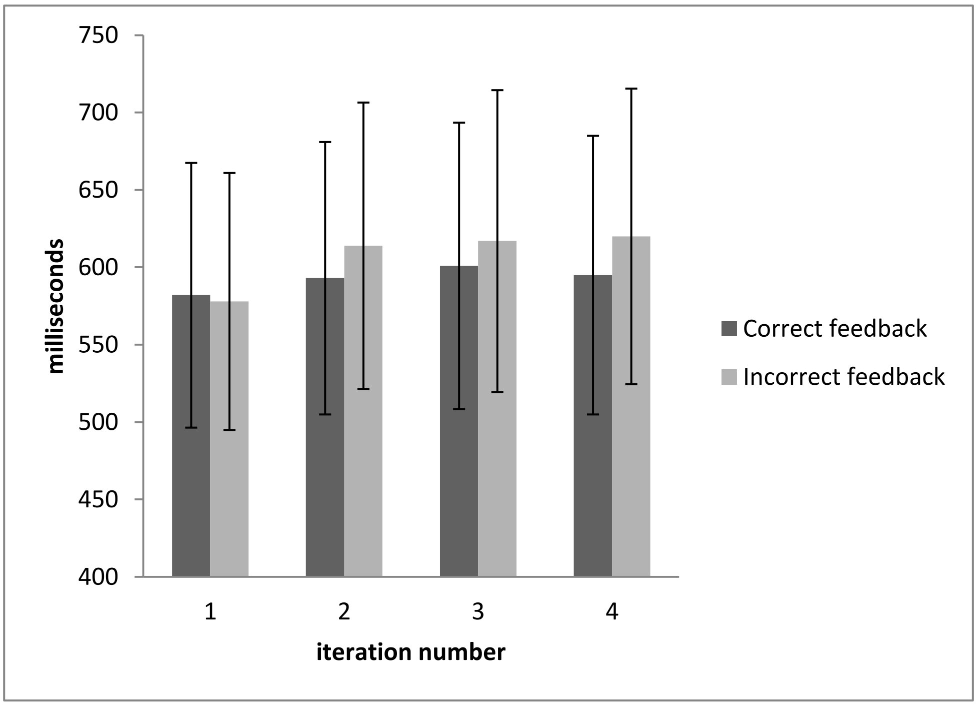


Figure 8-6. Experiment 6: Word durations in participants who do not stutter, by condition and iteration.
Error bars show (global) standard deviations.

The model reveals that word durations were more variable in the 'Incorrect' condition and on later iterations. Furthermore, irrespective of condition, word durations increased by 9ms with each subsequent iteration.

Table 8-7. Word durations: Exp7 - participants who do not stutter

Mixed effects analysis of factors influencing word durations. Model coefficients and probabilities are given for best-fitting models, as determined by Chi squared model comparisons.

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = word durations (milliseconds.)							
Intercept	Iteration = 1, correct feedback	587	34	<.001 ***	by word	13272	115
Intercept					by subject	8413	91
Condition	Incorrect feedback				by subject	1438	38
Iteration	+1	9	3	.001 **	by subject	18	4
					Residual	10818	104

Self-reports of difficulty speaking fluently

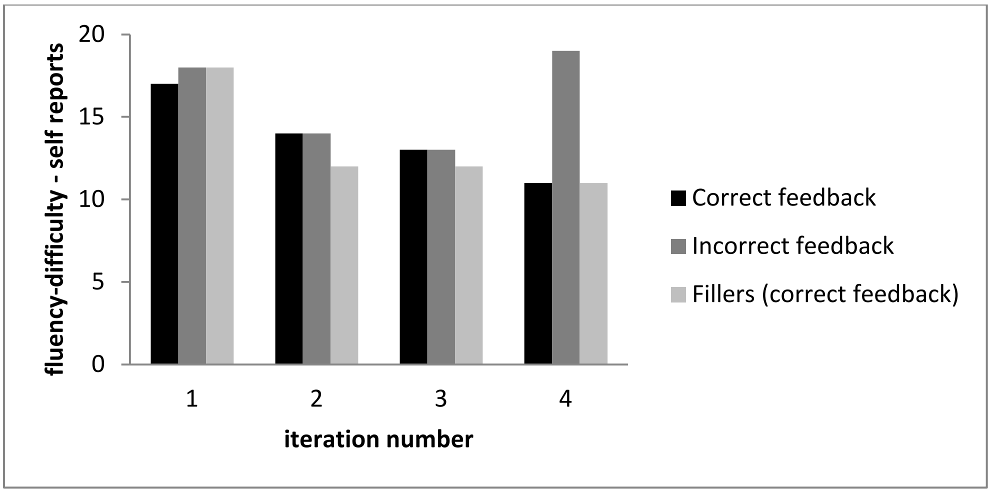


Figure 8-7. Experiment 6: Self-reports of difficulty speaking fluently: by participants who do not stutter.

Numbers of self-reports, across iterations, the two experimental conditions, and fillers.

In total, participants provided 1536 eligible self-ratings (32 words, each repeated four times by 12 participants) and on these, participants self-reported a total of 119 instances of difficulty speaking fluently (112 possible and 7 definite) (See Figure 8-7).

Prior to commencing the first iteration there were only two items for which participants predicted possible fluency-difficulty and no instances of predicting definite fluency-difficulty. The corresponding predictor was therefore not entered into the model.

The best-fit model of self-reported difficulty speaking fluently included only the random predictor: ‘Condition’ (Improvement due to additionally adding ‘Condition’ as a random predictor: χ^2 (2) = 7.06, *p* <.029). The addition of other random predictors or fixed predictors representing ‘Condition’, ‘Iteration’ or their interaction did not further improve the model (all *p* ≥.331).Table 8-8 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

The model reveals that participants’ were more variable in their self-ratings of difficulty speaking fluently’ in the ‘Incorrect’ than in the ‘Correct’ condition.

Table 8-8. Likelihood of self-reporting stuttering: Exp7 - participants who do not stutter.
Mixed effects analysis of factors influencing the likelihood of self-reporting stuttering. Model coefficients and probabilities are given for best-fitting models, as determined by Chi squared model comparisons.

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = likelihood of self-reporting stuttering							
Intercept	correct feedback	-6.12	0.12	<.001 ***	by word	1.25	1.12
					by subject	12.18	3.49
Condition	Incorrect feedback				by subject	1.31	1.14

Post-hoc analyses

As in Experiment 5, we also performed two linear mixed-effects regression analyses to test whether participants’ self-ratings of difficulty speaking fluently predicted vowel-onset latency or word duration. Across these two models, difficulty speaking fluently’ was only retained as a random predictor of word duration. (improvement over the base model containing only the intercept χ^2 (2) = 27.85, *p* <.001). See Table 8-9 for the model coefficients, and the probabilities that they could have occurred by chance.

Table 8-9. Post-hoc analyses: Experiment 6 - participants who do not stutter Mixed effects analysis of factors influencing vowel onset latencies and word durations. Model coefficients and probabilities are given for best-fitting models, as determined by Chi squared model comparisons.

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = vowel onsets (milliseconds.)							
Post-hoc analysis							
(intercept)		917	33	<.001 ***	by word	2330	48
					by subject	12106	110
					Residual	18692	137
DV = word durations (milliseconds.)							
Post-hoc analysis							
(intercept)	no difficulty speaking fluently	592	31	<.001 ***	by word	12046	110
					by subject	9333	97
Difficulty speaking fluently (self-reported)	Yes				by subject	2586	51
					Residual	10382	102

The best-fit models from these two post-hoc analyses revealed that a negative response to the question “were you able to say this word in exactly the way you intended?” was not associated with any difference in vowel onset latencies. It was, however, associated with more variable, but not reliably longer, word durations.

8.3.3. *Discussion*

As with the stuttering group in Experiment 5, the finding that difficulty speaking fluently was *not* associated with longer word durations suggests that, in the ‘Incorrect’ feedback condition, participants’ longer word durations reflected an intentional strategy whereby they prolonged or emphasized key phonemes to aid word recognition. In contrast to the stuttering group, the lack of any association between self-reports of difficulty speaking fluently and vowel onset latencies suggests that the relatively longer vowel onset latencies of participants on the later iterations of words in the incorrect condition were entirely strategic and not the result of difficulty.

Also in contrast to the stuttering group, the statistical analysis of the normally-fluent group's self-reports does *not* show that feedback consistently indicating speech-recognition failure increased the likelihood of experiencing difficulty speaking fluently. However, the number of self-reports of difficulty did rise on the forth iteration of the 'Incorrect' condition, suggesting that evidence for increased difficulty may have been found if the number of iterations providing feedback indicating failure had been greater. When asked about the impact of such feedback during the debriefing session, although participants often commented on it being a source of irritation, their comments did not suggest that it made speaking more difficult. Thus it seems that this participant group was resilient to the effects of incorrect feedback in a way that the stuttering group was not.

8.4. Coding of recordings by a naive independent rater (PWS and PNS groups combined)

Recordings of participants' utterances were coded by an independent rater who was naive to the differences between the two participant groups, to participant-group membership and to the nature and purpose of the experiment. She herself was a fluent speaker, but had had extensive experience of stuttering within her family (her father stuttered). Specifically, the independent rater was asked to identify all instances of (a) definite stuttering, and (b) possible stuttering, and code each of them as: 'Repetitions', 'Prolongations', 'Blocks', or 'Other'.

Because, in Experiments 5 and 6, the participant-groups' self-ratings were in response to slightly different questions (Did you stammer? vs. Were you able to say this word in exactly the way you intended?), it would have been misleading to analyze their responses together.

However, from a listener's perspective, participants' iterations of the target words are directly comparable, and so in the following independent-rater analysis of those iterations, we have combined the data from the two groups.

8.4.1. *Method*

Analysis

In this combined-group analysis of the naive independent rater's codings of recordings, logistic (mixed effects) regression analyses were carried out in the same way as in Experiments 5 and 6, but with the addition of an extra predictor representing participant group (PWS or PNS). Predictors representing 'voice activation, and 'stutter predicted' were not tested in this analysis because, unlike the participants themselves, the independent rater's reports could not have been influenced by these factors.

8.4.2. *Results*

In total, the independent rater was able to code 3052 recordings, from which she identified 118 repetitions (114 in PWS), 109 prolongations (28 in PWS), 29 blocks (12 in PWS) and 46 other stuttering-like disfluencies (16 in PWS) (See Figure 8-8). Seventy-one percent (i.e. 121 out of 170) of stuttering participants' recordings that the naïve rater rated as stuttering were similarly rated by the participants themselves. Six percent (8 out of 133) of normally-fluent participants' recordings that the naïve rater rated as stuttering were rated by those participants as having been difficult to speak fluently in the way intended.

The best-fit model of independent-rater reported stuttering included random predictors: 'Condition' and 'Iteration', and the fixed predictor 'Condition' and 'Iteration' (improvement due to adding Iteration as a fixed predictor: χ^2 (1) = 10.00, *p* =.001). Adding a predictor for the Iteration by Condition interaction did not further improve the model (χ^2 (1) =2.23, *p*

=.135). Table 8-10 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

Table 8-10. Factors influencing the naive rater's reports of stuttering: Experiments 6 & 7 combined.
Mixed effects analyses with random slopes. Model coefficients and probabilities are given for best-fitting models.

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
DV = likelihood of word being rated as 'stuttered' by the independent rater							
Intercept	Iteration = 1, correct feedback	-6.22	.88	<.001 ***	by word	0.20	0.45
					by subject	16.42	4.05
Condition	Incorrect feedback	1.73	0.33	<.001 ***	by subject	1.25	1.12
Iteration	+1	0.51	0.09	<.001 ***	by subject	0.06	0.24

The model reveals that, once random variance was accounted for, irrespective of group or iteration, words were 5.6 times as likely to be identified as stuttered in the 'Incorrect' condition than in the 'Correct' condition; and (irrespective of group or condition) each subsequent iteration of a word was associated with an increase by a factor of 1.7 in the likelihood of the word being identified as stuttered.

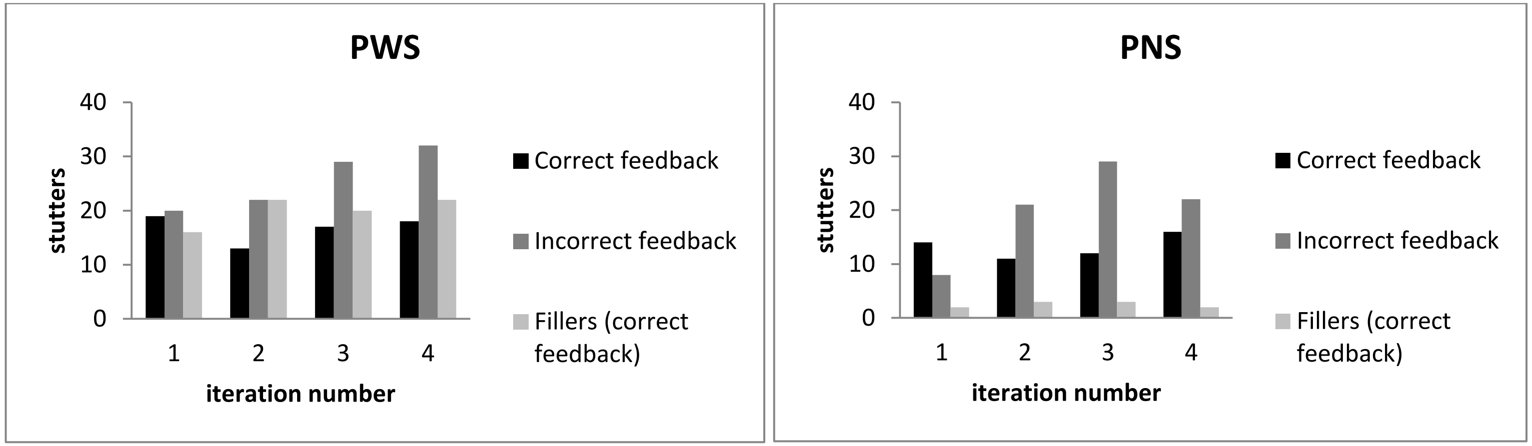


Figure 8-8. Naive-rater's ratings of stuttering (including prolongations) in participants who stutter (PWS) and do not stutter (PNS). Experiments 5 and 6 combined.

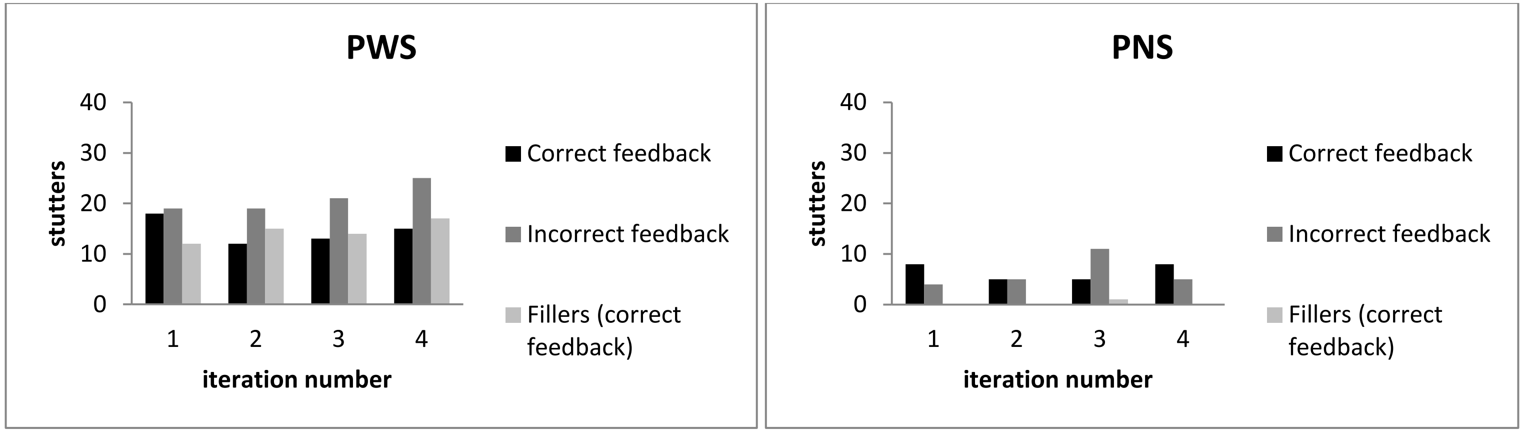


Figure 8-9. Naive-rater's ratings of stuttering (excluding prolongations) in participants who stutter (PWS) and who do not stutter (PNS). Experiments 5 and 6 combined.

8.4.3. *Discussion*

The naïve coder's ratings of stuttering reveal how difficult it can be for listeners to distinguish between stuttered disfluencies and intentional strategies intended to increase the listener's (or software's) chances of correct word recognition, especially with respect to prolongations, of which the naïve coder coded more instances in recordings from the PNS than the PWS group. Reports of 'blocks' were also more common in the PNS group, suggesting that the naïve coder probably missed many of the blocks actually made by the PWS group. As a result of these coding difficulties, the overall pattern of naïve ratings of stuttering in the two participant groups was remarkably similar, despite the large differences in the corresponding patterns of participants' self-ratings. The group differences only become apparent in the naïve-rater ratings when prolongations are excluded from the count (See Figure 8-9).

8.5. Experiments 5 and 6: General discussion

In Experiments 5 and 6 we set out to investigate the extent to which the experiences of stuttering and difficulty speaking fluently can result from the speaker's anticipation of his words being misrecognised. To do this we carried out two experiments in which people who stutter, and normally-fluent speakers, repeatedly spoke single words into what they believed was a speech-recognition system.

Experiment 5 revealed that the likelihood of stuttering occurring on a particular iteration of a word was independently predicted not only by participants' anticipation of difficulty prior to the first iteration of that word, but also by the nature of the feedback they have received about their immediately preceding iterations of that word; and moreover, the feedback received by participants influenced the likelihood of stuttering on subsequent iterations of that word *even when it bore no relation to their actual performance on that word*.

The latter of these two findings does not appear to be compatible with present formulations of the Covert Repair or EXPLAN hypotheses, both of which posit that stuttering arises as a side-effect of slow or impaired language encoding. It does not, however, rule out the possibility that language encoding deficits may sometimes play a role in the production of stuttering events.

Both findings are, however, compatible with an anticipatory struggle account, that posits that stutterers' recent memories of (apparent) word misrecognition (or miscomprehension) impact upon the amount of 'struggle' they engage in when attempting those words again. Furthermore, in light of the way the experiment was designed, it is reasonable to conclude that the 'anticipated struggle' associated with the increased likelihood of stuttering during the experiment was a 'struggle to get words correctly recognized' rather than a struggle to avoid negative listener reactions or to avoid stuttering. This is in line with accounts that portray stuttering as involving elements of 'approach' as well as 'avoidance' (e.g. Sheehan, 1953).

In this particular experimental paradigm, there was no semantic component to the recognition process, and word recognition would have appeared to participants to be entirely dependent on the accuracy or clarity with which they were able to articulate key sounds. The findings thus suggest that stuttering may be brought on by the experience of communication failure at a low level, and in this respect they could potentially account for why stutterers have as much difficulty speaking nonsense words as meaningful words (Packman et al., 2001).

Prolongation as a strategy to aid word recognition

The finding of significant random effects of 'Condition' in the word-duration data reveals that both participant groups tended to vary the length of iterations when they anticipated that those iterations may not be correctly recognized. It thus supports earlier findings by Stent et al. (2008), and extends them to people who stutter. Furthermore, the finding that word duration did not predict self-reports of either stuttering or difficulty speaking fluently indicates that prolongations were generally intentional and under volitional control and, as such, the study did not find any evidence suggesting that the PWS group was prone to interpret their own prolongations as stuttering. The extent of the mismatch between non-stuttering participants' own self-ratings and naïve-rater ratings supports Moore and Perkins' (1990) conclusion that subjective and objective reports of stuttering effectively measure different phenomena. They also reveal how difficult it can be for listeners to recognize the difference between a word that contains a stuttered disfluency and a word that contains an intentional prolongation.

It is noteworthy that although participants' self-reports of stuttering and difficulty speaking fluently were not significant predictors of word duration, in the PWS group they *were* a significant predictor of vowel onset latency, such that the experience of difficulty with word production was reflected in longer vowel onset latencies. This finding is in line with the frequently-cited observation that stuttering is closely related to difficulty with the initiation of words (Bloodstein & Bernstein Ratner, 2008) and/or with the association of onsets with rimes (Wingate, 1988).

Mechanisms that could result in disfluency as a result of anticipation of communication failure

If a speaker perceives that his words are likely to be misheard, irrespective of the actual cause of the anticipated communication failure, he is likely also to perceive that the pressure is on him to in some way adjust his speaking style to rectify the situation. Bloodstein (1959) proposed that speakers respond to this perception by resorting to a feedback-based approach that enables them to break the utterance down into easily manageable fragments (see also Mysak's 1960 account of deautomaticity). Although this hypothesis provides a plausible account of the mechanism by which speakers might purposefully prolong syllables and/or fragment them into their component sounds, the finding of the current study, that strategic prolongations were not associated with an increase in stuttering self-reports, does not support the hypothesis that tension and fragmentation associated with the deautomatization of speech results in the subjective experience of stuttering.

An account of how the anticipation of communication failure could cause stuttering-like disfluencies to arise, that better fits the findings of the current study, is provided by the ‘Vicious Circle Hypothesis’ (Vasić & Wijnen, 2005). This posits that past experiences of communication failure lead stutterers to develop a tendency to focus abnormally intensely on minor timing variations and infelicities in their inner (and/or overt) speech, and to set their thresholds for the initiation of covert error repairs at too low a level. In a similar vein, it is also possible that individual stimuli that remind the speaker of specific past experiences may lead to additional transient increases in monitoring vigilance, reductions in the error-repair threshold and consequently also to transient increases in error-repair activity.

Although, compared to the CRH, Vasić and Wijnen’s (2005) Vicious Circle Hypothesis is better able to account for the role of anticipation in precipitating stuttering, there are also two other accounts, both of which may be able to account for the variety of findings outlined in this paper in a more parsimonious way in that they do not require (either inner or overt) speech to be monitored.

The first of these alternative hypotheses is that the anticipation of communication failure leads directly to cancellation and reformulation of speech plans, irrespective of whether or not the plans themselves are scrutinized for errors. A direct anticipation-based mechanisms outlined above would result in many cancellations and reformulations being carried out unnecessarily, and could easily account for why stuttering-like disfluencies may continue to be produced despite the absence of any ongoing linguistic-formulation impairment.

The second of these alternative hypotheses, and one that potentially fits the experimental data and speaker's subjective experiences best of all, might be described as a 'Variable Release Threshold' hypothesis. In common with the Covert Repair and EXPLAN Hypotheses, this hypothesis is predicated on the idea, originally stemming from Dell's (1986) interactive model of language production, that the more time available for a speech planning prior to the initiation of overt articulation, the greater the probability that the speech-plan's target constituents (phonemes, words etc.) will reach levels of activation that are reliably higher than those of competitors, and the lower the likelihood that competing phonemes or words will be selected in error. A threshold mechanism that regulates the availability of words for overt execution has been proposed by Howell (2003, 2011b) in relation to the EXPLAN hypothesis. According to Howell's conceptualization of this mechanism, only plans for words that have attained an activation level above that of the buffer-release threshold can be executed. A logical extension to Howell's proposal is that the buffer-release threshold will rise whenever the speaker anticipates that a word will be misheard (or misunderstood). This rise in the buffer-release threshold would slow the rate at which the word(s) become available for execution and consequently maximize the likelihood that they will be executed clearly and accurately. In stutterers, anticipation of communication failure may lead to the buffer-release threshold mechanism generally being set at a relatively high level which would account for their longer onset latencies. It could also explain why stutterers may block if they are particularly concerned about the possibility of being misheard or misunderstood, insofar as the release-threshold may become set so high that some words fail to be released at all. Such a scenario might originally arise in response to some form of underlying language or speech production impairment and a tendency to make excessive phonological errors, but may then develop into a learned response that continues to be triggered by the anticipation of communication failure even after any underlying impairment has resolved. If this is the case, chances of successful communication would then be increased by the opposite response: i.e. by reducing the release threshold and allowing oneself to potentially make more errors. This is, however, probably not be the response which most stutterers attempt.

8.5.1. *Caveats*

Ecological validity of the study

Clearly, speaking single words into dummy speech-recognition software is a very different task to normal conversational speech. However, nowadays most speakers occasionally come across speech-recognition software in their daily lives, most commonly when providing gas or electricity readings over the telephone, or when accessing information over the telephone about cinema times etc. When participants were asked about their experiences of using such software the stuttering group consistently reported finding such experiences difficult and particularly likely to precipitate stuttering. It is, thus, perhaps surprising that the paradigm did not precipitate more stuttering than it did. A possible reason for the low incidence of stuttering during the paradigm was its relatively low ecological validity, and in particular the fact that the words they were required to utter were of no consequence to them in relation to their everyday lives. It seems that the £5 performance-related reward only acted as a limited incentive. Future studies would benefit from exploring ways of increasing the speakers' motivation, as the validity of the current findings is clearly compromised by the low power of the experiment.

How generalisable are the results?

A major difficulty encountered during the piloting of the software was to make it both engaging and difficult enough to precipitate stuttering and yet, at the same time, convincing enough, so that participants believed that their words were really being recognized (or misrecognized) by the software. In order for it to be convincing, it was important that participants did not accidentally utter the wrong word, and that they did not stutter so severely that a substantial number of occasions resulted where they activated the voice switch yet the sounds they made did not resemble the target word at all. Either of these scenarios would have likely resulted in the participant realizing that the software was not really sensitive to their utterances. Following extensive piloting, we found that only by using single single-syllable words and employing participants whose stuttering was relatively mild could we ensure that the paradigm was sufficiently convincing. However, as a result of these limitations we cannot be sure whether the main finding of the study (i.e. that stuttering is more likely to be experienced when speakers perceive that their words are not being recognized) applies to stutterers in general, or only to those whose overt disfluencies are relatively mild. Thus it is quite possible that ongoing language production difficulties and/or concern about negative listener reactions may play a greater role in precipitating the experience of stuttering in children and/or in people whose overt disfluencies are more severe.

The use of self-reported data

As noted in Section 5.1, generally speaking, stuttering researchers have tried to avoid reliance on self-reported data, as it is unverifiable. However, objective measures are unreliable, especially when single-word utterances are involved. The reliability of objective measures can be enhanced by the use of good quality video. But we rejected this solution during piloting when it became clear that the presence of a video camera (or any evidence made participants feel like they were being recorded) acted as a distraction that may have potentially confounded the results. An analogous situation is that, during telephone calls, stutterers often find the presence of an over-hearer strongly affects their ability to engage in the telephone conversation. As a compromise, participants' voices were recorded by a hidden microphone. However, this meant that only limited information was available with which the naïve coder could judge whether or not words were stuttered, and as a consequence, the naïve coder's judgments may have been less accurate than those of listeners in normal conversational situations. Nevertheless, it should be noted that, in everyday life, listeners do frequently make judgments based on partial or degraded information, and stutterers are likely to experience most difficulty speaking fluently when they are aware that the listener is only receiving an impoverished signal, such as when speaking over the telephone.

The effect of priming

In Experiments 5 and 6, in the correct-feedback condition, it is possible that the production of the target word was primed (and therefore facilitated) by the orthographic feedback from prior iterations, whereas in the incorrect condition it was not. Thus, in addition to the covert error repair and variable release threshold mechanisms outlined above, it is possible that priming effects could account for the ‘condition by iteration’ interaction found in Experiment 5 with respect to the likelihood of stuttering self-reports. Specifically, in the incorrect condition the orthographic feedback may have resulted in an interference effect because it consisted of one of the competitor option-words. Thus it is possible that in the correct condition, the words and phonemes may have been able to reach a higher level of activation more quickly than in the incorrect condition. Future studies could control for this potential confound by providing feedback that simply states that the participants utterance was ‘correct’ or ‘incorrect’ without providing orthographic feedback of the actual word that was recognized.

8.6. Experiments 5 and 6: Conclusions

Adults who stutter are more likely to stutter on a word when they anticipate that that word will be misrecognised. The anticipation that a word will be misrecognised may stem from a number of different sources, including not only past experiences of difficulty encoding and/or articulating the word but also past experiences of communication failure due to poor word recognition abilities of the listener (or speech-recognition software). This finding is consistent with the Vicious Circle Hypothesis (Vasić & Wijnen, 2005) insofar as such past experiences may lead to hypervigilant monitoring and unnecessary speech-error repair. However, it may potentially be more parsimoniously accounted for by the malfunctioning of a variable buffer-release mechanism that, under more normal circumstances, would serve to ensure the speaker achieves a high level of phonetic accuracy in situations where he believes he is likely to be misheard or misunderstood.

Both adults who stutter and normally-fluent adults may prolong or otherwise stress words when they anticipate communication difficulty, and these might be mistaken for stuttering by the listener.

9. General discussion and conclusions of the thesis

This thesis set out to investigate two key postulates of the Covert Repair Hypothesis (Postma & Kolk, 1993). The first is that the speech plans of people who stutter contain abnormally large numbers of errors of phonological encoding; the second is that the covert repair of such errors can account for the production of stuttering-like disfluencies. The results from our first four experiments led us to conclude that adults who stutter do indeed produce significantly more phonological encoding errors than normally-fluent speakers, but the covert repair of these errors cannot adequately account for the stuttering-like disfluencies they produce. In other words, the results support the first, but not the second tenet of the CRH.

The latter of the above two conclusions led us to consider an alternative possibility, namely that a mechanism similar to covert error repair may be activated simply by the speaker's perception of a need to speak more accurately. We hypothesised that stuttering-like disfluencies may arise not only in response to speech errors, but also in response to any stimulus that results in the speaker arriving at such a perception.

To test this hypothesis, our final two experiments employed a new experimental paradigm whereby participants repeatedly spoke single words into what they believed was a speech-recognition system. The system was designed to give the impression that it had difficulty recognising certain words, thus instilling in participants the anticipation of communication failure on subsequent iterations of those words. The results from these final two experiments led us to conclude that, in people who stutter, the anticipation of communication failure has the capacity to affect the likelihood of producing stuttering-like disfluencies and that such anticipation need not be directly associated with any currently existing impairment of language or speech production. The experiments we conducted, their findings, and the conclusions that we drew from them, are outlined in more detail in the discussion below.

To investigate the Covert Repair Hypothesis (Postma & Kolk, 1993), and in particular to determine whether the speech plans of adults who stutter contain abnormally large numbers of errors of phonological encoding, we adapted the Oppenheim and Dell (2008) tongue-twister paradigm in which participants self-report the errors that they perceive in their inner and overt speech while reciting the tongue-twisters.

Because the tonguetwister experiments relied on participants' self-reports of their own inner speech, our first three experiments were designed to determine to what extent such self-reports are reliable within the paradigm. They were carried out on three groups of normally-fluent participants with similar demographic profiles (all were psychology undergraduate students). By comparing participants' online self-reports of their overt phonemic errors with ratings made offline by an independent rater, these experiments provided data detailing the vigilance and accuracy of participants' self-reporting under a number of different speaking conditions. They also provided data detailing the extent of lexical and phonemic-similarity biases in participants' errors.

In these first three experiments, comparisons of participants' self-reported errors with those reported by an independent rater indicated that participants under-reported their overt errors. However, the extent of under-reporting did not vary significantly across conditions and it was not influenced by the imposition of auditory masking, which prevented participants hearing the sound of their own voices. Furthermore, the analyses of lexical and phonemic-similarity error biases revealed that the magnitude and direction of these biases in participants' self-ratings of inner speech errors were not significantly different to the (magnitude and direction of) these biases in their self-ratings of overt speech errors or in the corresponding independent-rater ratings. The consistency of under-reporting of errors across conditions, together with the stability of the error biases across raters and conditions strongly suggested that although participants' speech-error monitoring was not perfect (participants were generally hypo-vigilant), their self-reports were unlikely to have been distorted by monitoring biases and their hypo-vigilance did not compromise the reliability of the paradigm.

A further noteworthy observation arising from these preliminary experiments was that the patterns of error biases found in the inner-speech condition were strongly suggestive of a bi-directional flow of activation in the language-production system during phonological encoding, with feedback both from phoneme level to word level as well as from feature level to word level. This latter finding contrasted with the findings of Oppenheim and Dell (2008; 2010) and with their conclusion that inner speech is under-specified at the featural level.

Having established the reliability of the Oppenheim and Dell tongue-twister paradigm in Experiments 1 to 3, Experiment 4 then made use of the same paradigm to compare phonological encoding in stutterers and age-matched controls. The key finding from this experiment was that, as postulated by Postma and Kolk (1993) in the first tenet of their Covert Repair Hypothesis, stutterers, as a group, do indeed make significantly more phonological encoding errors than do normally-fluent speakers, both in inner as well as in overt speech. More specifically, they make more phoneme-substitution errors of all kinds, but in particular they make more non-contextual phoneme-substitution errors. The disproportionately larger numbers of non-contextual phoneme substitution errors suggest that their production systems are ‘noisy’, insofar they are more susceptible to interference from phonemes that do not belong in the target utterance.

In Experiment 4, despite the stuttering group making more errors than the control group, the size and direction of the phonemic similarity bias in the two groups’ reports of inner-speech errors were remarkably similar[21]. This suggests that that the two groups did not differ with respect to the extent to which they monitored their speech plans and to the extent to which those speech plans were specified at sub-phonemic (featural) levels. This is a non-trivial finding, insofar as activation-feedback from feature to phoneme level (which is dependent upon full featural specification) and monitoring may both potentially play key roles in minimizing the likelihood of phonemic errors occurring during speech production. The finding of comparable phonemic-similarity biases in both groups is also of interest because we might have expected group differences to be apparent in these biases if the two groups had employed different monitoring strategies. In particular, if stutterers normally focus more strongly on details of timing, as proposed by Vasić and Wijnen's (2005) Vicious Circle Hypothesis, we might have expected the speech plans they produced (in the inner speech condition) to show signs of greater sub-phonemic specification, in which case the similarity bias should have been stronger than for normally-fluent speakers.

A further finding of Experiment 4 was that the frequency with which participants who stutter perceived errors of phonological encoding in their inner speech was *not* related to the (self-reported) severity of their stuttering or the frequency with which they produced stuttering-like disfluencies in their everyday speech. This finding appears to be incompatible with the second tenet of Postma and Kolk's (1993) Covert Repair Hypothesis: that stuttering-like disfluencies are the (perceptible) side-effects of speakers' attempts to repair errors of

---

[21] Although a weaker lexical bias (on overt errors) was found in the control group than in the stuttering group, the strength of lexical bias in the stuttering group was similar that of the groups of (normally-fluent) student participants in Experiments 1 to 3. This suggests that the weaker lexical bias in the control group was anomalous.

phonological encoding before they become expressed in overt speech. Rather, it suggests that the link between stuttering like disfluencies and the tendency to make abnormally large numbers of phonological encoding errors is probably much more tenuous than that posited by the CRH. This conclusion is supported by the finding, in Experiment 4, that, although overall, the stuttering group made more phonemic errors than the control group, both participant groups exhibited a wide range of individual differences and some members of the stuttering group made fewer onset errors than some of the controls. This finding implies that impaired phonological encoding is neither a necessary nor a sufficient factor to account for stuttering in adults. It mirrors similar findings in studies that have compared the speech motor-control abilities of stutterers and controls (See Section 3.1).

Together, these findings from Experiment 4 led us to reject the Covert Repair Hypothesis in its original form and to consider the broader possibility that a mechanism similar to covert error repair may be activated simply by the speaker's perception of a need to speak more accurately.

Experiments 5 and 6 were designed to investigate this possibility. We hypothesised that stuttering-like disfluencies may arise not only in response to speech errors, but also in response to any stimulus that causes the speaker to perceive the need to speak more accurately. To test this hypothesis we employed a new experimental paradigm whereby participants repeatedly spoke single words into what they believed was a speech-recognition system. The system was designed to give the impression that it had difficulty recognising certain words, thus instilling in participants the anticipation of communication failure on subsequent iterations of those words.

Our decision to focus on anticipation of communication failure in Experiments 5 and 6 (as opposed to anticipation of stuttering or of negative listener responses) was because communication failure is likely to have been a common experience for individuals with a past history of impairments of language production and/or speech motor control, insofar as both of these impairments are likely to result in speech that is relatively difficult for listeners to understand. Specifically, the paradigm was designed to instil into participants the anticipation that words will be misrecognised by providing them with feedback (on a computer screen) that informed them that their immediately preceding iterations of specific words had been misrecognised. Importantly, this feedback was predetermined and bore no relationship to the actual quality of those previous iterations, or whether or not they had stuttered on them.

The results from Experiments 5 and 6 revealed that, for participants who stutter, the usual reduction in stuttering that occurs when words are repeated was negated when feedback from previous iterations of that word suggested that it had been misrecognised. Because the paradigm was counterbalanced, it clarified that this effect of anticipation was related purely to the nature of the feedback received (from the computer) and did not depend on the semantic, lexical, phonemic or articulatory properties (or complexity) of the individual words. However, regression analysis of factors contributing to the likelihood of stuttering also confirmed that, in participants who stutter, this feedback-related anticipatory effect exerted its influence alongside (and independently from) a pre-existing anticipatory effect that may indeed stem from more distant experiences of difficulty formulating or articulating specific words.

When considered together with the more general findings of studies that have investigated the moments when stuttering events occur (See the review in Section 2.4), the findings of Experiments 5 and 6 suggest that stuttering-like disfluencies may be precipitated by the anticipation of communication failure, and that this may itself potentially stem from one or more of a number of different factors. Such factors may include a tendency to make more phonological (and/or word-order) encoding errors (as revealed in Experiment 4), and/or a tendency for articulation, and possibly fine motor control in general, to be somewhat less precise (as outlined in Section 3.1). However, it is noteworthy that the anticipation of communication failure need not necessarily always stem from any functional impairment of the speaker's abilities to produce speech. Evidence suggesting a higher incidence of stuttering in bilingual families (see Van Borsel, Maes, & Foulon, 2001, for a review) suggests that it may sometimes stem from the limited capacity of listeners to understand the language spoken, and it may also result from an infelicitous speaking environment or negative responses of the listener, for any of a variety of reasons (cf. Johnson, 1942; 1959).

9.1.1. *Mechanisms that could account for the production of stuttering-like disfluencies*

How might the experience of communication failure lead to the production of stuttering-like disfluencies? In the discussion of the findings of Experiments 5 and 6 (See Section 8.4.3), a number of possible mechanisms were outlined, all of which could be activated by the perception of the need to speak more accurately. One possibility, that is essentially similar to Vasić and Wijnen's (2005) Vicious Circle Hypothesis (VCH), is that the anticipation that a word will be misunderstood leads to hypervigilant monitoring of the speech plan containing that word. As a result, minor, trivial phonetic irregularities in the plan may constitute sufficient stimuli for it to be cancelled and reformulated; the audible consequences of such cancellations and reformulations being repetitions, prolongations and blocks. It seems to us most plausible that such hyper-vigilance is momentary, insofar as it is may only occur at times or in situations where the speaker is concerned that his words will be misunderstood (Vasić & Wijnen's formulation of the VCH appears to suggest that stutterers are more generally hyper-vigilant with respect to monitoring of speech plans). Such transient hyper-vigilance might better explain why the occurrence of stuttering-like disfluencies is so dependent on speakers' perceptions of the demands posed by the speaking situation (See Section 2.4.2). Thus, for example, our failure to find any evidence of hyper-vigilant monitoring in participants who stutter in the tonguetwister paradigm we used for Experiment 4 may have been because participants did not perceive any need for their words to be understood.

A second possible way in which the experience of communication failure could lead to the production of stuttering-like disfluencies, also outlined in the General Discussion of Experiments 5 and 6 (See Section 8.4.3), is that the mere anticipation of a word being misrecognised or misunderstood is in itself sufficient to lead to cancellation and reformulation of the speech plan. This explanation could be considered as a hybrid, incorporating aspects of the CRH, VCH and EXPLAN hypotheses: In common with EXPLAN it does not require the monitoring of speech plans, and so avoids the problem of how to account for very rapid repetitions; in common with the CRH and VCH, it does involve the cancellation and reformulation of plans.

Although both of the mechanisms described above constitute plausible accounts of how stuttering-like disfluencies may arise, we would argue that neither of them is fully satisfactory. Stutterers generally do not report being aware of excessive numbers of errors in their inner speech, nor do they report perceiving their inner speech as disfluent. On the contrary, it appears that at least the majority of them consider their inner speech to be normally fluent (Netsell et al., 2010), although systematic larger scale studies are needed to confirm this. The apparent lack of stuttering-like disfluencies in the inner speech of stutterers suggests, as highlighted in the ICD9 definition of stuttering (See Section 2.1), that rather than being unable to formulate adequate speech plans, their main difficulty is that they are unable to execute the plans they have formulated at the time when they are needed.

In the General Discussion of Experiments 5 and 6 (Section 8.4.3), we proposed a ‘variable buffer-release threshold mechanism’ that could account for the apparent difficulty stutterers experience executing speech plans at an appropriate moment in time. This mechanism is congruent with the findings of our own studies and with the broader literature on stuttering, and is also in-line with this subjective experience of people who stutter. This mechanism is postulated to be responsible for determining the moment at which a speech plan that is stored in the articulatory buffer is released for overt execution. Specifically, we hypothesised that speech plans stored in the buffer have a particular level of activation, and they are only released when that level exceeds a particular threshold: the ‘release threshold’. The process of execution can be initiated by the lowering of that threshold to a level that is lower than the level of activation of the stored plan. It could be conceptualised as being like a dam that functions to hold water in a reservoir, the height of which can be varied according to needs. It is conceivable that the variable buffer-release threshold mechanism may develop in childhood and that it normally also plays an important role in allowing verbal thought and reasoning to proceed without overt articulation (cf. Vygotsky, 1986). As such, control over the buffer-release threshold would constitute an important cognitive and social skill attained during childhood.

As discussed in Chapter 4, Section 4.3, the more highly-activated a speech plan, the higher the probability that it will be error-free. Therefore, in situations where the speaker perceives that it is important to ensure a high quality of speech, the tendency may be for the buffer-release threshold to be set at a higher setting. Although such a strategy should increase the accuracy of speech, a side-effect may be that it is then more difficult for speech plans to be released for execution. Potentially, if the mechanism is set too high, some speech plans may not be released at all – in which case the speaker will find that, although he can access speech plans perfectly easily in inner speech, some of these he may nevertheless find himself unable to execute.

If such a threshold mechanism exists, the findings of Experiment 5 suggest that the level at which the threshold is set is likely to be raised by cues that are associated with past experiences of communication failure. The appropriateness of such a rise in the threshold is dependent upon the validity of the cues in the current speaking situation. Thus, for example, if the release threshold is raised in speaking situations where a high degree of accuracy is required, the likelihood of successful communication of the intended message may be increased. On the other hand, if a high degree of fluency is necessary in order to convey the intended message, communication may fail as a result of a rise in the setting of the release mechanism. Thus problems may arise if a speaker has difficulty determining which of the two scenarios is responsible for communication failure. As proposed by the VCH, such difficulty may arise because the speaker has come to interpret disfluencies as errors.

9.1.2. *The meaning of anticipation*

A common theme underlying the various formulations of anticipatory struggle hypothesis that have appeared over the years is that moments of stuttering occur when "stutterers interfere in some manner with the way they are talking because of their belief that speaking is difficult" (Bloodstein & Bernstein Ratner, 2008, p43). Bloodstein proposed that initially stutterers interfere with their speech in response to the direct experience of difficulty, but then as the condition progresses they start to interfere with their speech in response to cues that have become associated with previous experiences of difficulty. He equated the presence of such cues with the anticipation of upcoming difficulty, although he also noted that the cues may exert their influence with a minimal degree of conscious awareness (Bloodstein & Bernstein Ratner, 2008, pp247-248)

According to the Variable Release Threshold Hypothesis, the presence of cues associated with past experiences of communication failure leads specifically to a rise in the activation threshold that words and phonemes have to reach before they can be released for execution. Importantly such cues may already lead to a rise in the release threshold of a planned utterance before the speaker attempts to initiate overt articulation of that utterance. If this is the case, overt articulation of the utterance may effectively already have been rendered impossible even before the speaker attempts to initiate it. However, although stutterers can generally accurately predict when stuttering events will occur, it is noteworthy that a number of early studies have shown that they nevertheless sometimes find themselves blocking on words quite unexpectedly without any prior premonitions, and they also sometimes wrongly anticipate that they will stutter (Knott et al., 1937; Milisen, 1938; Silverman & Williams, 1972).

Together, such observations suggest that the conscious anticipation that one is about to block does not stem directly from an awareness of the level at which the variable buffer release mechanism is set. Rather, they suggest that conscious anticipation of stuttering most likely involves a form of forward modelling whereby the speaker draws on past experiences to calculate the probable impact of cues associated with past instances of stuttering on his current ability to utter a specific word in a specific speaking situation. It would make sense that the speaker’s level of skill at performing such forward modelling increases with age and experience, and it is likely that such forward modelling constitutes the underlying mechanism behind the development of avoidance behaviours that are particularly characteristic of persistent stuttering in older children and adults. The gradual increase in avoidance behaviours with age that is characteristic of developmental stuttering is fully in line with the finding that school-age children who stutter are able to accurately predict less than half of their instances of stuttering (Silverman & Williams, 1972), whereas adults who stutter are able to accurately predict about 95% of such instances (Knott et al., 1937).

In summary, it seems likely that anticipation influences stuttering behaviour in two ways: Firstly through an unconscious automatic process whereby cues associated with past experiences of communication failure lead to a rise in the release threshold for speech plans; and secondly, through a more conscious process of forward modelling whereby stutterers learn to anticipate the situations in which the release threshold will rise to the point where blocking will occur.

A further question that arises in relation to the nature of the cues that become associated with experiences of stuttering or communication failure is – What determines how much impact specific cues can have? Bearing in mind that some cues are likely to be more reliable than others and some are more readily available than others, it would perhaps not be surprising to find that some cues have a stronger impact on stuttering than others. As discussed in Section 2.4.1, it has long been known that stuttering is strongly influenced by the quality and availability of auditory feedback and that its severity is greatly reduced when auditory feedback is masked or distorted (for example, by delay of frequency shift) to the point where it no longer provides useful information (Kalinowski et al., 1993; Soderberg, 1969; Wingate, 1970). This would suggest that many of the cues that stutterers come to associate with communication failure are perceived as a result of monitoring of auditory feedback. However, this cannot be the whole story because, although auditory masking (and other forms of altered auditory feedback) can greatly reduce the severity of stuttering, they do not always have such an effect, and sometimes severe stuttering continues to occur in the complete absence of auditory feedback (e.g. Lincoln, Packman, Onslow, & Jones, 2010). In light of these observations it seems likely that speakers attend to whatever cues to the likelihood of communication failure are likely to be most reliable and most readily available. For the majority of stutterers most of the time, auditory feedback provides a source of cues that is readily available and reasonably reliable. Occasionally, however, visual or other forms of feedback may provide a more reliable indication of the likelihood of communication failure, in which case they may take priority.

9.1.3. *The anticipation of communication failure*

Previous researchers have proposed that moments of stuttering occur in response to the anticipation of: (a) negative listener reactions; (b) stuttering; (c) difficulty speaking; and (d) difficulty verbally conveying propositional content (Bloodstein, 1958, 1975, 1997; Johnson, 1942, 1959). Each of these proposals is supported by evidence both from stutterers own self-reports (e.g. Bloodstein, 1950a; Bloodstein, 1950b) as well as from experimental studies (see Bloodstein & Bernstein Ratner, 2008 Chapter 10, for a review). A key question that needs to be answered is – What do these various forms of anticipation have in common with each other?

In this thesis, we have argued that the common thread that connects the above forms of anticipation is that they are all manifestations of an underlying anticipation of communication failure. However, we have avoided explicitly stating what exactly we mean communication failure. A possible operational definition of it is: ‘failure (of the recipient) to apprehend the main illocutionary force of an utterance’. Thus communication failure may occur on a variety of levels, depending on the nature of the message. Although the illocutionary force may sometimes be adequately conveyed simply through correct recognition of the words spoken, in many speaking situations correct interpretation of the intended message depends on the subtle nuances of prosody and timing with which the words are spoken. In such situations the correct selection of phonemes may be relatively unimportant. In light of the specific role that variable release threshold and covert error repair mechanisms are likely to play in regulating the phonetic accuracy of executed utterances, if people who stutter have developed a tendency to adjust the settings of these mechanisms whenever they anticipate communication failure of any kind, many of the adjustments they make will be inappropriate, especially if a side-effect of such adjustments is a reduction in fluency and a reduced control over prosody and timing of utterances.

9.1.4. ***False beliefs***

Irrespective of whether stuttering stems from the maladaptive settings of a variable buffer release mechanism or an excessive tendency towards cancellation and reformulation of speech plans, either way the findings from our Experiments 5 and 6 suggest that the anticipation of communication failure and the associated desire to speak more accurately both play key roles in the disorder. As previously noted, anticipation of communication failure may arise because of underlying difficulty with the speaker's language or speech-production abilities, or because of an infelicitous speaking environment, or it may arise due to some sort of difficulty on the part of the listener. Whatever the cause, some of the time, the likelihood of successful communication can be increased by making more effort to speak accurately. However, in some situations, speaking more accurately will not help, and the belief that it will help constitutes a false belief. It is in such situations, where the speaker fails to recognise that speaking more accurately will not increase successful communication, that excessive error repairs or a maladaptive raising of the buffer-release threshold would be most likely to occur. The failure to recognise such situations may then lead the system to become hypersensitive generally, such that speech may become disfluent whenever the speaker perceives even a slight possibility of communication failure. Moreover, because disfluency can itself lead to communication failure, such hypersensitivity is likely to become self-sustaining; in which case, as Conture et al. (2004) suggested in their 'Gone but not Forgotten Hypothesis', stuttering may persist, even though the initial reasons for it may long since have disappeared.

9.1.5. *The stuttering 'block'*

In Chapter 2 we discussed how difficult it is to differentiate between what constitutes the underlying disorder of persistent stuttering, what symptoms arise directly from it, and what symptoms arise indirectly as a result of the speaker's attempts to adapt to it. Although most researchers would agree that we do not yet know what the underlying disorder is, few would question the traditional belief that repetitions, prolongations and blocks – the so-called 'primary' symptoms of stuttering – are symptoms that arise directly from it, whereas concomitant movements and avoidance of words, people, and situations – the so-called 'secondary' symptoms – arise indirectly, from the speaker's attempts to adapt to the disorder.

An alternative view, discussed in Section 2.3, is that the stuttering ‘block’ is the only primary symptom of persistent stuttering, whereas prolongations and repetitions are secondary symptoms of adaptation to the block. Although this latter view has rarely been considered, it nevertheless is the view that fits best with the Variable Release Mechanism Hypothesis insofar as the stuttering block is the only direct symptom of the speaker’s inability to release a stored plan for execution. According to this alternative view, repetitions and prolongations may represent learned responses that help the speaker maintain the attention of the listener and maintain his conversation turn (cf. Blackmer & Mitton, 1991; Howell & Au-Yeung, 2002). Thus, according to the Variable Release Mechanism hypothesis, the speaker only finds himself unable to initiate execution of the specific units (phonemes, syllables or words) that fail to achieve a level of activation above the threshold required for their release, whereas the units preceding (and also following) them can be executed without difficulty. It is likely that the release threshold is set differently for each unit, depending on the speaker’s perceptions (based on prior experiences) of how important it is to ensure error-free selection of that unit in order to avoid communication failure. Units that are perceived as essential for successful communication are likely to be assigned the highest release thresholds and may thus be the most difficult to initiate articulation of. In contrast, non-essential units, including function words and unstressed units would be assigned lower release thresholds and can therefore be initiated without difficulty; hence their availability for use as fillers.

The above conceptualization of repetitions and prolongations as secondary to blocks does not, however, fit well with the pattern of symptoms normally associated with incipient stuttering of early childhood, which is characterised by little, if any, evidence of blocking (Bloodstein, 2001, 2006). The lack of blocking in incipient stuttering, together with the stronger evidence of an association between incipient stuttering and weak or delayed development of syntactic formulation skills (Bernstein Ratner, 1997) suggest that the

Variable Buffer-Release Mechanism Hypothesis may only account well for the symptoms of persistent stuttering (i.e. stuttering in older children and adults). In contrast, the symptoms of incipient stuttering of early childhood may be more in line with the standard EXPLAN account (Howell & Au-Yeung, 2002), whereby the stuttering-like disfluencies of young children who stutter arise on occasions when the speech plan is simply not sufficiently complete at the point of time where the child attempts to initiate its execution. An incomplete speech plan may account for why incipient stuttering is rarely accompanied by blocking and signs of struggle. Thus it could be that ‘stalling behaviours’ are characteristic of incipient stuttering because the child has not yet accessed the full form of the word he wants to say and effectively needs to stall for time. ‘Advancing behaviours’, on the other hand, may be characteristic of persistent stuttering because the stutterer knows exactly the word he wants to say, but is unable to execute it. His instinctive response, therefore, is to try to use force to push the word out, rather than to stall.

9.2. Conclusions and future directions

Through a series of tongue-twister studies investigating inner-speech errors in adults who do and do not stutter, we were able to show that adults who stutter make more errors of phonological encoding and also more word-order errors than normally-fluent speakers. However, the rate with which individual participants made such errors was not correlated with the severity of their stuttering, and some participants who stutter make fewer phonological encoding errors than some normally-fluent speakers. We therefore concluded that impaired phonological encoding constitutes, at most, only one out of a number of underlying factors that may predispose speakers to persistent stuttering.

Our experiments that involved inducing the anticipation that words will misrecognised provided preliminary evidence suggesting that the experience of stuttering may be associated with stutterers’ inappropriate attempts to reduce their speech errors and/or to speak more accurately. In this thesis we have proposed a number of mechanisms that could account for how the perception of the need to speak more accurately may result in the production of stuttering-like disfluencies. These include two variants of the Vicious Circle Hypothesis (both of which involve cancellation and reformulation of speech plans), and a third model in which the perception of a need for increased accuracy results in a slower release of words from the articulatory buffer. We have posited that any of these mechanisms could potentially cause stuttering if they are inappropriately engaged by speakers, such as in situations where successful communication is not dependent on how accurately or clearly the speaker speaks.

9.2.1. ***Future research***

Although our word-misrecognition experiments provided evidence in support of the role of anticipation of communication failure in the production of stuttering-like disfluencies, it was not possible within the time frame of the thesis to design and run experiments to provide detailed evidence for either of the posited mechanisms through which this might occur. Future research could investigate the above-mentioned mechanisms further through an in-depth analysis of how stutterers perceive their speech plans in inner speech, and in particular whether or not they perceive them to accurately represent exactly what they wish to say, whether they perceive any differences in speech plans that are followed by stuttering compared to speech plans that are followed by fluent speech, and whether the experience of stuttering is associated with difficulty formulating adequate plans or difficulty releasing formulated plans for execution.

Insofar as any of these posited mechanisms account for the production of stuttering-like disfluencies, they lead to two important questions, both of which have a direct relevance to therapy: (1) To what extent do stutterers try to speak more accurately than they need to? And (2) to what extent do they have the capacity to vary how accurately they try to speak?

Future research could investigate these issues by examining (and comparing) how stutterers and normally-fluent speakers perceive the relative importance of accuracy and fluency in situations where they anticipate communication failure, and by examining whether they have more difficulty than normally-fluent speakers in discriminating between situations where efforts to adjust one's speech are likely to improve the chances of successful communication and situations where they are not.

In light of the findings of the current thesis, it seems likely that, for people who stutter, an improved level of communication effectiveness may be achievable through a more adaptive awareness of the relative importance of accuracy and fluency in specific speaking situations. Therefore, from a clinical perspective it would be beneficial for researchers to explore the extent to which therapy can help clients who stutter develop such an awareness.

10. References

Acheson, D. J., & MacDonald, M. C. (2009a). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language, 60*(3), 329-350.

Acheson, D. J., & MacDonald, M. C. (2009b). Verbal Working Memory and Language Production: Common Approaches to the Serial Ordering of Verbal Information. *Psychological Bulletin, 135*(1), 50-68.

Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ: Wiley-Interscience.

Alm, P. A., & Risberg, J. (2007). Stuttering in adults: The acoustic startle response, temperamental traits, and biological factors. *Journal of Communication Disorders, 40*(1), 1-41.

Alpermann, A., & Zückner, H. (2008). Speech Motor Skills of Stuttering Children= Sprechmotorische Fähigkeiten stotternder Kinder. *Sprache, Stimme, Gehör, 32*(1), 36-40.

Ambrose, N. G., & Yairi, E. (1999). Normative Disfluency Data for Early Childhood Stuttering. *Journal of Speech, Language and Hearing Research, 42*(4), 895-909.

Anderson, J. D. (2007). Phonological neighborhood and word frequency effects in the stuttered disfluencies of children who stutter. *Journal of Speech, Language and Hearing Research, 50*(1), 229.

Anderson, J. D., & Conture, E. G. (2004). Sentence-structure priming in young children who do and do not stutter. *Journal of Speech, Language, and Hearing Research, 47*(3), 552-571.

Anderson, J. D., Wagovich, S. A., & Hall, N. E. (2006). Nonword repetition skills in young children who do and do not stutter. *Journal of Fluency Disorders, 31*(3), 177-199.

Andrews, G., & Harris, M. (1964). *The syndrome of stuttering*. Oxford, England: Spastics Society Medical Education.

Arends, N., Povel, D. J., & Kolk, H. (1988). Stuttering as an attentional phenomenon. *Journal of Fluency Disorders, 13*(2), 141-151.

Arnstein, D., Lakey, B., Compton, R. J., & Kleinow, J. (2011). Preverbal error-monitoring in stutterers and fluent speakers. *Brain and Language*, *116*(3), 105-115.

Baars, B. J., Motley, M. T., & MacKay, D. (1975). Output editing for lexical status in artificially elicited slips of the tongue*. *Journal of Verbal Learning and Verbal Behavior, 14*(4), 382-391.

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2)[cd-rom]. *Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor]*.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The psychology of learning and motivation, 8*, 47-89.

Bajaj, A. (2007). Working memory involvement in stuttering: Exploring the evidence and research implications. *Journal of Fluency Disorders, 32*(3), 218-238.

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–32.

Bernstein Ratner, N. (1995). Language complexity and stuttering in children. *Topics in Language Disorders, 15*(3), 32.

Bernstein Ratner, N. (1997). Stuttering: A psycholinguistic perspective. *Nature and treatment of stuttering: New directions*, 99–127.

Bernstein Ratner, N., Newman, R. S., & Strekas, A. (2009). Effects of word frequency and phonological neighborhood characteristics on confrontation naming in children who stutter and normally fluent peers. *Journal of Fluency Disorders*, *34*(4), 225-241.

Bernstein Ratner, N., & Sih, C. (1987). Effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders, 52*(3), 278.

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition, 39*(3), 173-194.

Bloodstein, O. (1949). Conditions under which stuttering is reduced or absent: A review of literature. *Journal of Speech and Hearing Disorders, 14*(4), 295-302.

Bloodstein, O. (1950a). Hypothetical conditions under which stuttering is reduced or absent. *Journal of Speech and Hearing Disorders, 15*(2), 142.

Bloodstein, O. (1950b). A rating scale study of conditions under which stuttering is reduced or absent. *Journal of Speech and Hearing Disorders, 15*(1), 29-36.

Bloodstein, O. (1958). Stuttering as an anticipatory struggle reaction. In J. Eisenson (Ed.), *Stuttering: A symposium* (pp. 3-69). New York: Harper & Row.

Bloodstein, O. (1975). Stuttering as tension and fragmentation. In J. Eisenson (Ed.), *Stuttering: A second symposium* (pp. 1-96). New York: Harper & row.

Bloodstein, O. (1997). Stuttering as an anticipatory struggle reaction. In R. F. Curlee & G. M. Siegel (Eds.), *Nature and treatment of Stuttering: New Directions (2nd edition)* (pp. 167-181). Needham Heights, MA: Allyn Y Bacon.

Bloodstein, O. (2001). Incipient and developed stuttering as two distinct disorders: Resolving a dilemma. *Journal of Fluency Disorders, 26*(1), 67-73.

Bloodstein, O. (2002). Early stuttering as a type of language difficulty. *Journal of Fluency Disorders, 27*(2), 163-167.

Bloodstein, O. (2006). Some empirical observations about early stuttering: A possible link to language development. *Journal of Communication Disorders, 39*(3), 185-191.

Bloodstein, O., & Bernstein Ratner, N. (2008). *A handbook on stuttering* (6th ed.). NY: Delmar.

Bluemel, C. (1932). Primary and secondary stuttering. *Quarterly Journal of Speech, 18*, 178–200.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 07). [Computer software manual]. Retrieved 10th September 2011, 2009, from http://www.praat.org/

Borden, G. J. (1979). An interpretation of research on feedback interruption in speech. *Brain and Language, 7*(3), 307-319.

Bosshardt, H. (1999). Effects of concurrent mental calculation on stuttering, inhalation and speech timing. *Journal of Fluency Disorders, 24*(1), 43-72.

Bosshardt, H. (2002). Effects of concurrent cognitive processing on the fluency of word repetition: comparison between persons who do and do not stutter. *Journal of Fluency Disorders, 27*(2), 93-113; quiz 113-114.

Bosshardt, H. (2006). Cognitive processing load as a determinant of stuttering: Summary of a research programme. *Clinical Linguistics & Phonetics, 20*(5), 371-385.

Bosshardt, H., Ballmer, W., & De Nil, L. (2002). Effects of category and rhyme decisions on sentence production. *Journal of Speech, Language, and Hearing Research, 45*(5), 844.

Bosshardt, H., & Fransen, H. (1996). Online sentence processing in adults who stutter and adults who do not stutter. *Journal of Speech and Hearing Research, 39*(4), 785.

Breslow, N., & Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9-25.

Brooks, P. J., & MacWhinney, B. (2000). Phonological priming in children's picture naming. *Journal of Child Language, 27*(2), 335-366.

Brown, S. (1937). The influence of grammatical function on the incidence of stuttering. *Journal of Speech and Hearing Disorders, 2*(4), 207.

Brown, S. (1945). The loci of stutterings in the speech sequence. *Journal of Speech and Hearing Disorders, 10*(3), 181.

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta analysis of functional neuroimaging studies. *Human Brain Mapping, 25*(1), 105-117.

Brutten, G. (1970). Two factor theory and therapy. In S. F. o. America (Ed.), *Conditioning in stuttering therapy* (pp. 37–56). Memphis, TN: Speech Foundation of America.

Brutten, G., & Dancer, J. E. (1980). Stuttering adaptation under distributed and massed conditions. *Journal of Fluency Disorders, 5*(1), 1-10.

Büchel, C., & Sommer, M. (2004). What causes stuttering? *PLoS Biology, 2*(2), e46.

Burger, R., & Wijnen, F. (1999). Phonological encoding and word stress in stuttering and nonstuttering subjects. *Journal of Fluency Disorders, 24*(2), 91-106.

Byrd, C. T., Conture, E. G., & Ohde, R. N. (2007). Phonological priming in young children who stutter: holistic versus incremental processing. *American Journal of Speech-Language Pathology, 16*(1), 43-53.

Caruso, A., Abbs, J., & Gracco, V. (1988). Kinematic analysis of multiple movement coordination during speech in stutterers. *Brain, 111*(2), 439.

Chang, S. E., Synnestvedt, A., Ostuni, J., & Ludlow, C. L. (2010). Similarities in speech and white matter characteristics in idiopathic developmental stuttering and adult-onset stuttering. *Journal of Neurolinguistics, 23*(5), 455-469.

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production. [doi: 10.1016/j.jfludis.2010.05.002]. *Journal of Fluency Disorders, 35*(3), 246-279.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*(4), 429-432.

Conture, E. G., Zackheim, C. T., Anderson, J. D., & Pellowski, M. W. (2004). Linguistic processes and childhood stuttering: many's a slip between intention and lip. In B. Maasen, R. Kent, H. F. M. Peters, P. v. Lieshout & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 253-281). Oxford: Oxford University Press.

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 162-175.

Craig, A. (1990). An investigation into the relationship between anxiety and stuttering. *Journal of Speech and Hearing Disorders, 55*(2), 290.

Craig, A., Hancock, K., Tran, Y., & Craig, M. (2003). Anxiety levels in people who stutter: A randomized population study. *Journal of Speech, Language, and Hearing Research, 46*(5), 1197.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

DebRoy, S., & Bates, D. M. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis, 91*(1), 1-17.

Del Viso, S., Igoa, J. M., & García-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20*(3), 161-185.

Dell, G. (1986). A spreading activation theory of retrieval in language production. *Psychological Review, 93*, 283–321.

Dell, G. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model* 1. *Journal of Memory and Language, 27*(2), 124-142.

Dell, G., & Reich, P. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 611-629.

Dworzynski, K., Howell, P., Au-Yeung, J., & Rommel, D. (2004). Stuttering on function and content words across age groups of German speakers who stutter. *Journal of Multilingual Communication Disorders, 2*(2), 81-101.

Eisenson, J., & Horowitz, E. (1945). The influence of propositionality on stuttering. *Journal of Speech and Hearing Disorders, 10*(3), 193.

Ellis, A. W. (1988). Normal writing processes and peripheral acquired dysgraphias. *Language and Cognitive Processes, 3*(2), 99-127.

Froeschels, E. (1943). Pathology and therapy of stuttering. *Nervous Child, 2*, 148-161.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (Vol. 30, pp. 231-256). Amsterdam, North-Holland.

Glasner, P. J., & Vermilyea, F. D. (1953). An Investigation Of The Definition And Use Of The Diagnosis,'Primary Stuttering'. *Journal of Speech and Hearing Disorders, 18*(2), 161.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes, 21*(7), 817-855.

Griggs, S., & Still, A. W. (1979). An Analysis of Individual Differences in Words Stuttered. *Journal of Speech and Hearing Research, 22*(3), 572-580.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. [doi: 10.1016/j.bandl.2005.06.001]. *Brain and Language, 96*(3), 280-301.

Guitar, B. (2006). Stuttering: An integrated approach to its nature and treatment: Lippincott Williams & Wilkins.

Haegeman, L., & Ihsane, T. (2001). Adult null subjects in the non-pro-drop languages: Two diary dialects. *Language Acquisition, 9*(4), 329-346.

Hakim, H., & Bernstein Ratner, N. (2004). Nonword repetition abilities of children who stutter: an exploratory study. *Journal of Fluency Disorders, 29*(3), 179-199.

Hansen, H. T. (1955). The effect of a measured audience reaction on stuttering behavior patterns. (Unpublished doctoral dissertation). University of Wisconsin, Madison.

Hartfield, K. N., & Conture, E. G. (2006). Effects of perceptual and conceptual similarity in lexical priming of young children who stutter: preliminary findings. *Journal of Fluency Disorders, 31*(4), 303-324.

Hartsuiker, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes, 21*(7), 856-891.

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al.(1975). *Journal of Memory and Language, 52*(1), 58-70.

Hendel, D., & Bloodstein, O. (1973). Consistency in relation to inter-subject congruity in the loci of stutterings. *Journal of Communication Disorders, 6*(1), 37-43.

Hennessey, N., Nang, C., & Beilby, J. (2008). Speeded verbal responding in adults who stutter: Are there deficits in linguistic encoding? *Journal of Fluency Disorders, 33*(3), 180-202.

Howell, P. (2003). Is a perceptual monitor needed to explain how speech errors are repaired? *Gothenburg Papers in Theoretical Linguistics, 90*, 31-34.

Howell, P. (2011a). Language–Speech Processing in Developmental Fluency Disorders. In J. Guendouzi, F. Loncke & M. J. Williams (Eds.), *A handbook on psycholinguistics and cognitive processes: Perspectives on Communication Disorders.* (pp. 435-462). Hove, Sussex: Taylor & Francis.

Howell, P. (2011b). *Recovery from Stuttering*: Psychology Press/Taylor & Francis Group.

Howell, P., & Au-Yeung, J. (1995). Syntactic determinants of stuttering in the spontaneous speech of normally fluent and stuttering children. *Journal of Fluency Disorders, 20*(4), 317-330.

Howell, P., & Au-Yeung, J. (2002). The EXPLAN theory of fluency control applied to the diagnosis of stuttering *Clinical linguistics: Theory and applications in speech pathology and therapy* (pp. 75–94).

Howell, P., Au-Yeung, J., & Sackin, S. (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language, and Hearing Research, 42*(2), 345-354.

Howell, P., & Sackin, S. (2000). Speech rate modification and its effects on fluency reversal in fluent speakers and people who stutter. *Journal of Developmental and Physical Disabilities, 12*(4), 291-315.

Howell, P., & Sackin, S. (2001). Function word repetitions emerge when speakers are operantly conditioned to reduce frequency of silent pauses. *Journal of Psycholinguistic Research, 30*(5), 457-474.

Hubbard, C., & Prins, D. (1994). Word familiarity, syllabic stress pattern, and stuttering. *Journal of Speech and Hearing Research, 37*(3), 564.

Iverach, L., O'Brian, S., Jones, M., Block, S., Lincoln, M., Harrison, E., et al. (2009). Prevalence of anxiety disorders among adults seeking speech therapy for stuttering. *Journal of Anxiety Disorders, 23*(7), 928-934.

James, S. E., Brumfitt, S. M., & Cudd, P. A. (1999). Communicating by telephone:: Views of a group of people with stuttering impairment. *Journal of Fluency Disorders, 24*(4), 299-317.

Johnson, W. (1942). A study of the onset and development of stuttering. *Journal of Speech and Hearing Disorders, 7*(3), 251.

Johnson, W. (1959). *The onset of stuttering*. Minneapolis MI: University of Minnesota press.

Johnson, W., & Knott, J. (1937). Studies in the Psychology of Stuttering: I: The Distribution of Moments of Stuttering in Successive Readings of The Same Material. *Journal of Speech Disorders, 2*(1), 17.

Johnson, W., Larson, R., & Knott, J. (1937). Studies in the Psychology of Stuttering: III: Certain Objective Cues Related to the Precipitation of the Moment of Stuttering. *Journal of Speech and Hearing Disorders, 2*(1), 23.

Johnson, W., & Millsapps, L. (1937). Studies in the psychology of stuttering: VI. The role of cues representative of past stuttering in the distribution of stuttering moments during oral reading. . *Journal of Speech Disorders, 2*(2), 101-104.

Kaasin, K., & Bjerkan, B. (1982). Critical words and the locus of stuttering in speech. *Journal of Fluency Disorders, 7*(4), 433-446.

Kadihanifi, K., & Howell, P. (1992). Syntactic Analysis of the Spontaneous Speech of Normally Fluent and Stuttering Children. *Journal of Fluency Disorders, 17*(3), 151-170.

Kalinowski, J., Armson, J., Stuart, A., & Gracco, V. L. (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. *Language and Speech, 36*(1), 1.

Kamhi, A. G., & McOsker, T. G. (1982). Attention and stuttering: do stutterers think too much about speech? *Journal of Fluency Disorders, 7*(3), 309-321.

Kang, C., Riazuddin, S., Mundorff, J., Krasnewich, D., Friedman, P., Mullikin, J., et al. (2010). Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. *New England Journal of Medicine, 362*(8), 677-685.

Karniol, R. (1995). Stuttering, language, and cognition: A review and a model of stuttering as suprasegmental sentence plan alignment (SPA). *Psychological Bulletin, 117*, 104-124.

Knott, J. R., Johnson, W., & Webster, M. J. (1937). Studies in the Psychology of Stuttering: II: A Quantitative Evaluation of Expectation of Stuttering in Relation to the Occurrence of Stuttering. *Journal of Speech and Hearing Disorders, 2*(1), 20.

Kolk, H., & Postma, A. (1997). Stuttering as a covert repair phenomenon. In R. F. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2 ed., pp. 182-203). Boston; MA: Allyn & Bacon.

Lackner, J., & Tuller, B. (1979). Role of efference monitoring in the detection of self-produced speech errors. *Sentence Processing*, 281–294.

Laver, J. (1973). The detection and correction of slips of the tongue. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 132-143). The Hague: Mouton De Gruyter.

Laver, J. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 287–305). New York, NY: Academic Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*(1), 41-104.

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(01), 1-38.

Levitt, A. G., & Healy, A. F. (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors* 1. *Journal of Memory and Language, 24*(6), 717-733.

Lickley, R., Hartsuiker, R., Corley, M., Russell, M., & Nelson, R. (2005). Judgment of disfluency in people who stutter and people who do not stutter: Results from magnitude estimation. *Language and Speech, 48*(3), 299.

Lincoln, M., Packman, A., Onslow, M., & Jones, M. (2010). An Experimental Investigation of the Effect of Altered Auditory Feedback on the Conversational Speech of Adults Who Stutter. *J Speech Lang Hear Res, 53*(5), 1122-1131.

Logan, K. J., & Conture, E. G. (1997). Selected temporal, grammatical, and phonological characteristics of conversational utterances produced by children who stutter. *Journal of Speech, Language and Hearing Research, 40*(1), 107-120.

Loucks, T. M. J., & De Nil, L. F. (2006a). Anomalous sensorimotor integration in adults who stutter: a tendon vibration study. *Neuroscience Letters, 402*(1-2), 195-200.

Loucks, T. M. J., & De Nil, L. F. (2006b). Oral kinesthetic deficit in adults who stutter: A target-accuracy study. *Journal of Motor Behavior, 38*(3), 238-247.

Loucks, T. M. J., De Nil, L. F., & Sasisekaran, J. (2007). Jaw-phonatory coordination in chronic developmental stuttering. *Journal of Communication Disorders, 40*(3), 257-272.

Ludlow, C., Siren, K., & Zikira, M. (1997). Speech production learning in adults with chronic developmental stuttering. In W. Hulstijn, H. M. F. Peters & p. h. h. m. v. Lieshout (Eds.), *Speech production: Motor control, brain research and fluency disorders* (pp. 221–230). Nijmegen, The Netherlands: Nijmegen University Press.

MacKay, D. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*(3), 323-350.

MacKay, D. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review, 89*(5), 483.

MacKay, D. (1987). The organization of perception and action : A theory for language and other cognitive skills. New York ; London: Springer-Verlag.

MacKay, D. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition, 1*(3), 199-225.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*(59), 19-44.

Max, L., Caruso, A., & Gracco, V. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech, Language, and Hearing Research, 46*(1), 215.

Max, L., Guenther, F., Gracco, V., Ghosh, S., & Wallace, M. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders, 31*, 105–122.

McAllister, J., & Kingston, M. (2005). Final part-word repetitions in school-age children: Two case studies. *Journal of Fluency Disorders, 30*(3), 255-267.

McClean, M. D., Tasko, S. M., & Runyan, C. M. (2004). Orofacial Movements Associated With Fluent Speech in Persons Who Stutter. *Journal of Speech, Language and Hearing Research, 47*(2), 294-303.

Melnick, K., & Conture, E. G. (2000). Relationship of length and grammatical complexity to the systematic and nonsystematic speech errors and stuttering of children who stutter. *Journal of Fluency Disorders, 25*(1), 21-45.

Melnick, K., Conture, E. G., & Ohde, R. (2003). Phonological priming in picture naming of young children who stutter. *Journal of Speech, Language and Hearing Research, 46*(6), 1428-1443.

Menzies, R., Onslow, M., & Packman, A. (1999). Anxiety and stuttering: Exploring a complex relationship. *American Journal of Speech-Language Pathology, 8*(1), 3.

Messenger, M., Onslow, M., Packman, A., & Menzies, R. (2004). Social anxiety in stuttering: measuring negative social expectancies. *Journal of Fluency Disorders, 29*(3), 201-212.

Milisen, R. (1938). Frequency of Stuttering with Anticipation of Stuttering Controlled. *J Speech Disord, 3*(4), 207-214.

Moore, S., & Perkins, W. (1990). Validity and reliability of judgments of authentic and simulated stuttering. *Journal of Speech and Hearing Disorders, 55*(3), 383.

Motley, M. T., & Baars, B. J. (1976). Semantic bias effects on the outcomes of verbal slips. *Cognition, 4*(2), 177-187.

Motley, M. T., Baars, B. J., & Camden, C. T. (1981). Syntactic criteria in prearticulatory editing: Evidence from laboratory-induced slips of the tongue. *Journal of Psycholinguistic Research, 10*(5), 503-522.

Mysak, E. (1960). Servo theory and stuttering. *Journal of Speech and Hearing Disorders, 25*(2), 188.

Netsell, R., Bakker, K., & Ashley, E. (2010). *The inner speech of persons who stutter*. Paper presented at the International Motor Speech Conference.

Newman, R. S., & Bernstein Ratner, N. (2007). The role of selected lexical factors on confrontation naming accuracy, speed, and fluency in adults who do and do not stutter. *Journal of Speech, Language and Hearing Research, 50*(1), 196-213.

Nippold, M. A. (1990). Concomitant Speech and Language Disorders in Stuttering Children - a Critique of the Literature - Reply. *Journal of Speech and Hearing Disorders, 55*(4), 803-804.

Nippold, M. A. (2001). Phonological disorders and stuttering in children: what is the frequency of co-occurrence? *Clinical Linguistics & Phonetics, 15*(3), 219-228.

Nippold, M. A., & Rudzinski, M. (1995). Parents' speech and children's stuttering: A critique of the literature. *Journal of Speech and Hearing Research, 38*(5), 978.

Nooteboom, S. (1967). Some regularities in phonemic speech errors *IPO Annual Progress Report, 2* (Vol. 2, pp. 65-71). Eindhoven: Institute for Perception Research.

Nooteboom, S. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand.* (pp. 87–95). New York: Academic Press.

Nooteboom, S. (2005a). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication, 47*(1-2), 43-58.

Nooteboom, S. (2005b). Listening to oneself: Monitoring speech production. In R. J. Hartsuiker, Y. Bastiaanse, A. Postma & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 167-186). Hove: Psychology Press.

Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language, 58*(3), 837-861.

Olander, L., Smith, A., & Zelaznik, H. N. (2010). Evidence That a Motor Timing Deficit Is a Factor in the Development of Stuttering. *Journal of Speech, Language, and Hearing Research, 53*(4), 876.

Oppenheim, G., & Dell, G. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*(1), 528-537.

Oppenheim, G., & Dell, G. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & cognition, 38*(8), 1147-1160.

Packman, A., Onslow, M., Coombes, T., & Goodwin, A. (2001). Stuttering and lexical retrieval. *Clinical Linguistics & Phonetics, 15*(6), 487-498.

Palen, C., & Peterson, J. M. (1982). Word frequency and children's stuttering: The relationship to sentence structure. *Journal of Fluency Disorders, 7*(1), 55-62.

Pellowski, M. W., & Conture, E. G. (2005). Lexical priming in picture naming of young children who do and do not stutter. *Journal of Speech, Language and Hearing Research, 48*(2), 278-294.

Perkins, W., Kent, R., & Curlee, R. (1991). A theory of neuropsycholinguistic function in stuttering. *Journal of Speech and Hearing Research, 34*(4), 734.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77*(2), 97-132.

Postma, A., & Kolk, H. (1992). Error monitoring in people who stutter: evidence against auditory feedback defect theories. *Journal of Speech and Hearing Research, 35*(5), 1024-1032.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research, 36*(3), 472-487.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech, 39*(4), 375-392.

R Development Core Team. (2009). R: A language and environment for statistical computing. *Foundation for Statistical Computing, Vienna, Austria*. Retrieved from http://www.R-project.org

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460.

Riley, G. (2009). Stuttering Severity Instrument for Children and Adults. (SSI-4) (4th ed). Austin TX: Pro-Ed

Riley, G., & Riley, J. (1986). Oral motor discoordination among children who stutter. *Journal of Fluency Disorders, 11*(4), 335-344.

Robinson, A. (1997). The British English Example Pronunciation Dictionary, v0. 1. Retrieved 4th July 2009, 2009, from ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/

Sasisekaran, J., & De Nil, L. F. (2006). Phoneme monitoring in silent naming and perception in adults who stutter. *Journal of Fluency Disorders, 31*(4), 284-302.

Sasisekaran, J., De Nil, L. F., Smyth, R., & Johnson, C. (2006). Phonological encoding in the silent speech of persons who stutter. *Journal of Fluency Disorders, 31*(1), 1-21; quiz 19.

Savage, C., & Howell, P. (2008). Lexical priming of function words and content words with children who do, and do not, stutter. *Journal of Communication Disorders, 41*(6), 459-484.

Scheepers, C., & Barr, D. (2011). Type I and II error probabilities of various approaches to categorical data analysis: Results from Monte Carlo simulations. Paper presented at the University of Edinburgh, Psycholinguistics Coffee.

Schlesinger, I., Forte, M., Fried, B., & Melkman, R. (1965). Stuttering, information load, and response strength. *Journal of Speech and Hearing Disorders, 30*(1), 32.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86-102.

Schwartz, H. D., Zebrowski, P. M., & Conture, E. G. (1990). Behaviors at the onset of stuttering. *Journal of Fluency Disorders, 15*(2), 77-86.

Schweppe, J., Grice, M., & Rummer, R. (2011). What models of verbal working memory can learn from phonological theory: Decomposing the phonological similarity effect. *Journal of Memory and Language*.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18*(1), 41-55.

Sheehan, J. (1953). Theory and treatment of stuttering as an approach-avoidance conflict. *The Journal of Psychology, 36*(1), 27-49.

Sheehan, J. (1970). *Stuttering: Research and therapy*. Oxford, England: Harper & Row.

Sherrard, C. (1975). Stuttering as “False Alarm” Responding. *International Journal of Language & Communication Disorders, 10*(2), 83-91.

Silverman, F. H., & Williams, D. E. (1972). Prediction of Stuttering by School-Age Stutterers. *J Speech Hear Res, 15*(1), 189-193.

Soderberg, G. A. (1967). Linguistic Factors in Stuttering. *Journal of Speech and Hearing Research, 10*(4), 801-810.

Soderberg, G. A. (1969). Delayed auditory feedback and the speech of stutterers: A review of studies. *Journal of Speech and Hearing Disorders, 34*(1), 20.

Sokolov, A. (1972). *Inner Speech and Thought*: Plenum Publishing Corporation.

Starkweather, C. (1987). *Fluency and stuttering*: Prentice-Hall, Inc.

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143-186). Hove, UK:: Lawrence Erlbaum Associates Ltd.

Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication, 50*(3), 163-178.

Suresh, R., Ambrose, N., Roe, C., Pluzhnikov, A., Wittke-Thompson, J. K., Ng, M. C. Y., et al. (2006). New complexities in the genetics of stuttering: significant sex-specific linkage signals. *The American Journal of Human Genetics, 78*(4), 554-563.

Tsiamtsiouris, J., & Cairns, H. S. (2009). Effects of Syntactic Complexity and Sentence-Structure Priming on Speech Initiation Time in Adults Who Stutter. *Journal of Speech, Language and Hearing Research, 52*(6), 1623-1639.

Van Borsel, J., Maes, E., & Foulon, S. (2001). Stuttering and bilingualism: A review. *Journal of Fluency Disorders, 26*(3), 179-205.

Van Lieshout, P. (2004). Dynamical systems theory and its application in speech. In B. Maasen, R. Kent, H. F. M. Peters, P. v. Lieshout & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 51–82). Oxford: Oxford University Press

Van Riper, C. (1971). *The nature of stuttering*. Englewood Cliffs N.J.: Prentice Hall.

Van Riper, C. (1973). *The treatment of stuttering*: Prentice-Hall Englewood Cliffs, NJ.

Van Riper, C. (1982). The nature of stuttering, 2nd edn. Englewood Cliff: NJ: Prentice-Hall, Inc.

Vasić, N., & Wijnen, F. (2005). Stuttering as a monitoring deficit. In R. J. Hartsuiker, Y. Bastiaanse, A. Postma & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 226–247). Hove, East Sussex Psychology Press.

von Holst, E., & Mittelstaedt, H. (1950/1973). The reafference principle. In R. Martin (Ed.), *The behavioural physiology of animals and man: The collected papers of Erich von Holst* (Vol. 1, pp. 139-173). Coral Gables, FL: University of Miami Press.

Vygotsky, L. S. (1986). Thought and language (rev. ed.). *Cambridge, MA.*

Ward, D. (2006). Stuttering and cluttering: Frameworks for understanding and treatment. Hove, Sussex: Psychology Press.

Watkins, R., & Yairi, E. (1997). Language production abilities of children whose stuttering persisted or recovered. *Journal of Speech, Language and Hearing Research, 40*(2), 385.

Weber-Fox, C., Spencer, R., Spruill III, J., & Smith, A. (2004). Phonologic processing in adults who stutter: electrophysiological and behavioral evidence. *Journal of Speech, Language, and Hearing Research, 47*(6), 1244.

West, R. (1929). A neurological test for stutterers. *British Medical Journal, 1*(38), 114.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time-course of phonological encoding. *Journal of Memory and Language, 34*(3), 311-334.

Wijnen, F., & Boers, I. (1994). Phonological priming effects in stutterers. *Journal of Fluency Disorders, 19*(1), 1-20.

Wilshire, C. (1999). The" tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech, 42*(1), 57.

Wingate, M. E. (1970). Effect on stuttering of changes in audition. *Journal of Speech and Hearing Research, 13*(4), 861-873.

Wingate, M. E. (1988). The structure of stuttering: A psycholinguistic analysis. Berlin: Springer Verlag.

Wingate, M. E. (2001). SLD is not stuttering. *Journal of Speech, Language and Hearing Research, 44*(2), 381-383.

Wischner, G. J. (1950). Stuttering behavior and learning: A preliminary theoretical formulation. *Journal of Speech and Hearing Disorders, 15*(4), 324.

Wolk, L., Blomgren, M., & Smith, A. B. (2000). The frequency of simultaneous disfluency and phonological errors in children: A preliminary investigation. *Journal of Fluency Disorders, 25*(4), 269-281.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences, 2*(9), 338-347.

World Health Organization. (1977). Manual of the international statistical classification of diseases, injuries, and causes of death (Vol. 1): WHO Geneva.

World Health Organization. (1993). *International statistical classification of diseases and related health problems*: World Health Organization, Geneva, Switzerland.

Yairi, E. (1981). Disfluencies of Normally Speaking Two-Year-Old Children. *J Speech Hear Res, 24*(4), 490-495.

Yairi, E. (1983). The onset of stuttering in two-and three-year-old children: A preliminary report. *Journal of Speech and Hearing Disorders, 48*(2), 171.

Yairi, E., & Ambrose, N. (1992a). A longitudinal study of stuttering in children: A preliminary report. *Journal of Speech and Hearing Research, 35*(4), 755.

Yairi, E., & Ambrose, N. (1992b). Onset of stuttering in preschool children: Selected factors. *Journal of Speech and Hearing Research, 35*(4), 782.

Yairi, E., & Ambrose, N. (2005). *Early Childhood Stuttering*. Austin, TX: PRO-ED, Inc.
.

Yairi, E., Ambrose, N., & Cox, N. (1996). Genetics of stuttering: a critical review. *Journal of Speech and Hearing Research, 39*(4), 771.

Yairi, E., Ambrose, N. G., Paden, E. P., & Throneburg, R. N. (1996). Predictive factors of persistence and recovery: Pathways of childhood stuttering* 1. *Journal of Communication Disorders, 29*(1), 51-77.

Yairi, E., & Lewis, B. (1984). Disfluencies at the onset of stuttering. *Journal of Speech and Hearing Research, 27*(1), 154.

Yaruss, J. (1999). Utterance length, syntactic complexity, and childhood stuttering. *Journal of Speech, Language, and Hearing Research, 42*(2), 329.

Yaruss, J., & Conture, E. G. (1996). Stuttering and phonological disorders in children: Examination of the covert repair hypothesis. *Journal of Speech and Hearing Research, 39*(2), 349-364.

Yaruss, J., & Quesal, R. (2006). Overall Assessment of the Speaker's Experience of Stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency Disorders, 31*(2), 90-115.

Yeni-Komshian, G., Chase, R. A., & Mobley, R. L. (1968). The Development of Auditory Feedback Monitoring: II. Delayed Auditory Feedback Studies on the Speech of

Children Between Two and Three Years of Age. *Journal of Speech and Hearing Research, 11*(2), 307-315.

Zelaznik, H. N., Smith, A., Franz, E. A., & Ho, M. (1997). Differences in bimanual coordination associated with stuttering. *Acta Psychologica*, *96*(3), 229-243.

Zimmermann, G. (1980). Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech & Hearing Research*.

11. Appendix A

Materials for Experiment 1

Outcome and similarity for word 3 onset substitution															
word, similar				nonword, similar				word, dissimilar				nonword, dissimilar			
fan	van	vat	fad	fan	van	valve	fad	man	van	vat	mad	man	van	valve	mad
pole	coast	cope	poke	pole	coast	comb	poke	soul	coast	cope	soak	soul	coast	comb	soak
till	kid	kin	tinge	till	kid	kiln	tinge	bill	kid	kin	binge	bill	kid	kiln	binge
seep	heath	heel	scene	seep	heath	heave	scene	keep	heath	heel	keen	keep	heath	heave	keen
rig	link	limb	rip	rig	link	limp	rip	dig	link	limb	dip	dig	link	limp	dip
pat	cap	catch	pad	pat	cap	cab	pad	bat	cap	catch	bad	bat	cap	cab	bad
busk	Puff	puck	bunk	busk	puff	pub	bunk	musk	puff	puck	monk	musk	puff	pub	monk
cob	Golf	gone	cot	cob	golf	goth	cot	yob	golf	gone	yacht	yob	golf	goth	yacht
finch	Ship	shin	fill	finch	ship	shift	fill	pinch	ship	shin	pill	pinch	ship	shift	pill
meal	bead	beak	mean	meal	bead	beach	mean	weal	bead	beak	wean	weal	bead	beach	wean
dove	Gulf	gull	dump	dove	gulf	gut	dump	love	gulf	gull	lump	love	gulf	gut	lump
wail	range	rake	waist	wail	range	race	waist	tale	range	rake	taste	tale	range	race	taste
pink	bid	bit	pick	pink	bid	bib	pick	kink	bid	bit	kick	kink	bid	bib	kick
come	tut	tub	cuff	come	tut	tuck	cuff	hum	tut	tub	huff	hum	tut	tuck	hufi
conk	toss	top	cog	conk	toss	tongs	cog	honk	toss	top	hog	honk	toss	tongs	hog
reap	leap	leach	reef	reap	leap	leash	reef	beep	leap	leach	beef	beep	leap	leash	beef
dock	tod	tot	dodge	dock	tod	tom	dodge	lock	tod	tot	lodge	lock	tod	tom	lodge
peck	ketch	keg	pet	peck	ketch	kelp	pet	beck	ketch	keg	bet	beck	ketch	kelp	bet
gust	cusp	cut	gum	gust	cusp	cup	gum	rust	cusp	cut	rum	rust	cusp	cup	rum
face	vein	vale	feign	face	vein	vague	feign	race	vein	vale	cane	race	vein	vague	cane
pang	tank	tack	patch	pang	tank	tap	patch	hang	tank	tack	hatch	hang	tank	tap	hatch
hunk	thump	thug	hump	hunk	thump	thud	hump	junk	thump	thug	jump	junk	thump	thud	jump
rot	watt	wad	rob	rot	watt	was	rob	not	watt	wad	knob	not	watt	was	knob
tape	pain	pale	take	tape	pain	paid	take	nape	pain	pale	knave	nape	pain	paid	knave
tut	done	duck	tug	tut	done	dove	tug	mutt	done	duck	mug	mutt	done	dove	mug
dock	knock	knot	dodge	dock	knock	notch	dodge	lock	knock	knot	lodge	lock	knock	notch	lodge
sill	tick	tip	sick	sill	tick	tint	sick	chill	tick	tip	chick	chill	tick	tint	chick
deck	wreck	wren	dead	deck	wreck	realm	dead	tech	wreck	wren	ted	tech	wreck	realm	ted
run	duck	dub	rum	run	duck	dud	rum	son	duck	dub	some	son	duck	dud	some
wench	wreck	red	well	wench	wreck	rev	well	bench	wreck	red	bell	bench	wreck	rev	bell
kale	gauge	gape	cake	kale	gauge	gait	cake	shale	gauge	gape	shake	shale	gauge	gait	shake
bag	dad	dash	back	bag	dad	damp	back	sag	dad	dash	sack	sag	dad	damp	sack
mull	buff	buck	much	mull	buff	bulge	much	dull	buff	buck	dutch	dull	buff	bulge	dutch
roam	lone	lope	role	roam	lone	loaf	role	dome	lone	lope	dole	dome	lone	loaf	dole
wade	range	reign	wait	wade	range	wraith	wait	maid	range	reign	mate	maid	range	wraith	mate
sit	zing	zip	sick	sit	zing	zinc	sick	knit	zing	zip	nick	knit	zing	zinc	nick
puff	buff	bunch	punk	puff	buff	bulge	punk	huff	buff	bunch	hunk	huff	buff	bulge	hunk
rip	width	witch	rim	rip	width	wish	rim	hip	width	witch	hymn	hip	width	wish	hymn
dock	toss	tot	dosh	dock	toss	top	dosh	wok	toss	tot	wash	wok	toss	top	wash
delve	wreck	ref	dead	delve	wreck	realm	dead	shelve	wreck	ref	shed	shelve	wreck	realm	shed
wreck	wet	west	wren	wreck	wet	wedge	wren	peck	wet	west	pen	peck	wet	wedge	pen
fame	safe	sail	fade	fame	safe	sage	fade	maim	safe	sail	maid	maim	safe	sage	maid
bell	peg	pet	beck	bell	peg	pep	beck	knell	peg	pet	neck	knell	peg	pep	neck
pad	tank	tack	patch	pad	tank	tab	patch	mad	tank	tack	match	mad	tank	tab	match
teem	seep	seek	teach	teem	seep	siege	teach	beam	seep	seek	beach	beam	seep	siege	beach
hub	thump	thug	hush	hub	thump	thud	hush	rub	thump	thug	rush	rub	thump	thud	rush
jug	chuck	chump	just	jug	chuck	chub	just	lug	chuck	chump	must	lug	chuck	chub	must
rot	loft	lock	rob	rot	loft	loll	rob	not	loft	lock	knob	not	loft	loll	knob

12. Appendix B. Combined analyses with random slopes

12.1. Experiments 1 to 3

Vigilance of self-reporting

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
Overt speech (participants vs. independent rater), Word 3 onset errors							
Intercept	nonword outcomes, dissimilar onsets, independent rater	-10.73	1.50	<.001 ***	by tongue-twister	1.32	1.15
					by participant	31.89	5.65
Lexicality	real word outcomes	5.32	1.51	<.001 ***	by participant	33.20	5.76
Similarity	similar onsets	5.80	1.51	<.001 ***	by participant	32.47	5.70
Lexicality by similarity interaction		-4.41	1.53	.004 **	by participant	34.70	5.89
Rater	Self-ratings	-0.45	0.12	<.001 ***			

Accuracy of self-reporting

Insufficient data for the number of degrees of freedom.

Inner vs. overt speech

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
Inner vs. overt speech, Word 3 onset errors							
Intercept	(Iteration 1, correct feedback)	-6.86	0.36	<.001 ***	by tongue-twister	0.52	0.72
					by participant	1.77	1.32
Lexicality	real word outcomes	1.94	0.34	<.001 ***	by participant	0.07	0.26
Similarity	similar onsets	1.99	0.35	<.001 ***	by participant	0.59	0.77
Lexicality by similarity interaction		-1.06	0.38	.005 **			

Inner speech only

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
Inner speech only, Word 3 onset errors							
(Intercept)	nonword, dissimilar,	-7.49	0.62	<.001 ***	by tongue-twister	0.18	0.42
					by participant	3.10	1.76
Lexicality	real word outcomes	2.42	0.62	<.001 ***	by participant	1.65	0.41
Similarity	similar onsets	2.21	0.65	<.001 ***	by participant	2.44	1.56
Lexicality by similarity interaction		-1.31	0.67	.052			

12.2. Experiment 4: Lexical bias and phonemic similarity

Vigilance of self-reporting

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
Overt speech (participants vs. independent rater), Word 3 onset errors							
Intercept	Digitspan=7 independent rater, controls nonword outcomes, dissimilar onsets,	-6.29	0.40	<.001 ***	by tongue-twister	1.05	1.03
					by participant	2.83	1.68
digitspan	+1	0.00	0.05	.963			
Rater	Self-ratings	-0.46	0.13	<.001 ***			
Group	PWS	1.06	0.41	.010 **			
Lexicality	real word outcomes	0.40	0.35	.259	by participant	1.47	1.21
Similarity	similar onsets	1.40	0.23	<.001 ***	by participant	1.58	1.26
Group by Lexicality interaction		0.72	0.44	.102	by participant		

Inner vs. overt speech

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
Inner vs. overt speech (participants vs. independent rater), Word 3 onset errors							
Intercept	Digitspan=7 Inner speech, controls nonword outcomes, dissimilar onsets,	-5.58	0.35	<.001 ***	by tongue-twister	0.72	0.85
					by participant	1.75	1.32
Digitspan	+1	-0.17	0.09	.063			
Overtness	overt speech	0.54	0.18	.003 **	by participant	0.48	0.69
Group	PWS	0.10	0.26	.010 **			
Lexicality	real word outcomes	0.56	0.30	.058	by participant	0.27	0.52
Similarity	similar onsets	0.82	0.27	.003 **	by participant	0.08	0.29
Group by Lexicality interaction		0.58	0.34	.091			

Inner speech only

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
Inner speech only, Word 3 onset errors							
Intercept	Digitspan=7, nonword outcomes, dissimilar onsets,	-6.16	0.35	<.001 ***	by tongue-twister	0.87	0.93
					by participant	0.83	0.91
Digitspan	+1	-0.26	0.13	.054			
Lexicality	real word outcomes	0.92	0.25	<.001 ***			
Similarity	similar onsets	1.04	0.26	<.001 ***			

12.3. Experiment 4 – all onset errors and word-order errors

Vigilance of self-reporting

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
Overt speech (participants vs. independent rater), Onset errors							
Intercept	digitspan=7, independent rater, controls	-3.70	0.66	<.001 ***	by tongue-twister	0.69	0.83
					by participant	0.59	0.77
digitspan	+1	-0.18	0.09	.018 *			
rater	Self rating	-0.47	0.11	<.001 ***	by participant	0.00	0.05
group	PWS	1.04	0.25	<.001 ***			
Overt speech (participants vs. independent rater), Word-order errors							
(Intercept)	digitspan=7, independent rater, controls	-5.71	0.30	<.001 ***	by tongue-twister	0.77	0.88
					by participant	1.55	1.25
digitspan	+1	-0.19	0.12	.115			
rater	Self rating	-0.27	0.12	.032 *			
group	PWS	1.03	0.37	.005 **			

Accuracy of self-reporting

Insufficient data for the number of degrees of freedom.

Inner vs. overt speech

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
All onset errors							
(Intercept)	digitspan=7, controls inner speech	-5.52	0.21	<.001 ***	by tongue-twister	0.46	0.68
					by participant	0.47	0.68
Digitspan	+1	-0.16	0.08	.038 *			
Group	PWS	0.87	0.22	<.001 ***			
Overtness	overt speech	0.37	0.14	.006 **	by participant	0.13	0.36
Non-contextual onset errors							
(Intercept)	digitspan=7, controls	-.899	0.66	<.001 ***	by tongue-twister	1.56	1.25
					by participant	0.76	0.87
digitspan	+1	-0.43	0.19	.025 *			
group	PWS	1.78	0.70	.010 *			
Word-order errors							
(Intercept)	digitspan=7, unmasked, controls, inner speech	-6.45	0.28	<.001 ***	by tongue-twister	0.35	0.59
					by participant	0.80	0.89
Digitspan	+1	-0.29	0.09	.002 **			
Masking	masked	0.52	0.17	.003 **	by participant	0.35	0.59
Group	PWS	1.19	0.30	<.001 ***			
Overtness	overt speech	0.24	0.17	.197	by participant	0.53	0.73

Inner speech only

Predictors	Value	Fixed effects			Random effects		
		Co-efficient	Std. Error	p (coef.= 0)	Random Analysis	Random variance	Std. Dev.
All onset errors							
(Intercept)	digitspan=7, controls	-5.36	0.23	<.001 ***	by tongue-twister	0.44	0.66
					by participant	0.44	0.66
digitspan	+1	-0.14	0.10	.186	by participant		
group	PWS	0.65	0.28	.20 *	by participant		
Word order errors							
(Intercept)	digitspan=7, controls	-7.05	0.41	<.001 ***	by tongue-twister	0.26	0.51
					by participant	1.54	1.24
digitspan	+1	-0.34	0.11	.003 **	by participant	0.93	0.96
masking	masked	0.10	0.30	<.001 ***			
group	PWS	1.38	0.39	<.001 ***			

13. Appendix C Publications