The influence of anticipation of word misrecognition on the likelihood of stuttering

Citation for published version:
Brocklehurst, PH, Lickley, RJ & Corley, M 2012, 'The influence of anticipation of word misrecognition on the likelihood of stuttering', *Journal of Communication Disorders*, vol. 45, no. 3, pp. 147-160.
https://doi.org/10.1016/j.jcomdis.2012.03.003

Digital Object Identifier (DOI):
10.1016/j.jcomdis.2012.03.003

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Communication Disorders

The influence of anticipation of word misrecognition on the likelihood of stuttering.

Abstract:

This study investigates whether the experience of stuttering can result from the speaker's anticipation of his words being misrecognised. Twelve adults who stutter (AWS) repeated single words into what appeared to be an automatic speech-recognition system. Following each iteration of each word, participants provided a self-rating of whether they stuttered on it and the computer then provided feedback implying its correct or incorrect recognition of it. Each word was repeated four times. Unbeknown to participants, 'Correct' and 'Incorrect' recognition of words by the system was pre-determined and bore no relation to the actual quality of participants' iterations of those words. For words uttered in the 'Correct recognition' condition, the likelihood of AWS self-reporting stuttering on a word diminished across iterations, whereas for words in the 'Incorrect recognition' condition it remained static. On the basis of the findings it is argued that: (a) In AWS, the anticipation that a word will be misrecognised increases the relative likelihood of stuttering on that word in the future; and (b) This effect is independent of the degree of difficulty inherent in the formulation and motor execution of the word itself, although it may interact with it. Mechanisms that can account for these findings and yet are also congruent with the wider range of evidence from psycholinguistic and speech motor control domains are discussed. It is concluded that stuttered disfluencies may best be explained as resulting from the inappropriate functioning of covert repair and/or variable release threshold mechanisms in response to the anticipation of communication failure.

Published 2012 in the Journal of Communication Disorders
Brocklehurst, P. H., Lickley, R. J., & Corley, M. (2012). The influence of anticipation of word misrecognition on the likelihood of stuttering. Journal of communication disorders, 45(3), 147–160. http://dx.doi.org/10.1016/j.jcomdis.2012.03.003

1 Introduction

People who stutter (PWS) do not stutter all the time. Rather, stuttering moments are more likely to occur on specific words, with specific conversation partners and in specific speaking situations, such as talking over the telephone, before groups, etc. (Bloodstein & Bernstein Ratner, 2008, Chapter 10). The exact pattern of their occurrence may, however, vary considerably from one PWS to another, and a different pattern is found in young children who stutter compared to older children and adults (Bloodstein, 2001; Bloodstein & Grossman, 1981; Dworzynski, Howell, Au-Yeung, & Rommel, 2004; Howell, Au-Yeung, & Sackin, 1999).

In young children who stutter (CWS), stuttering is most likely to occur on utterances that are linguistically or motorically complex (Bernstein Ratner, 1997; Bloodstein & Grossman, 1981; Logan & Conture, 1997; Yairi & Ambrose, 2005), in line with the view that the language or speech production systems of young CWS are not yet sufficiently developed to enable them to fluently produce utterances with an age-appropriate level of complexity (e.g., Bernstein Ratner, 1997; Conture, Zackheim, Anderson, & Pellowski, 2004). In older children and adults who stutter (AWS) evidence of impaired language production or speech motor control is more equivocal. Although experimental studies have found that, compared to controls, AWS tend to have slower speech-onset latencies (e.g., Burger & Wijnen, 1999; Lieshout, Hulstijn, & Peters, 1996; Sasisekaran & De Nil, 2006; Tsiamtsiouris & Cairns, 2009), these could simply reflect speakers' attempts to adapt to the disorder. AWS have been found to make more phonological-encoding and word-order errors, in both inner and overt speech (Brocklehurst & Corley, 2011) and show more variability in fine motor coordination (e.g., Kleinow & Smith, 2000; Loucks, De Nil, & Sasisekaran, 2007; Max, Caruso, & Gracco, 2003). However, in all such studies there is a large degree of overlap between the stuttering and control participant groups. Thus, it seems likely that, in AWS, stuttering events may sometimes occur even in the absence of any significant ongoing underlying impairment in language formulation or speech motor control (Conture et al., 2004).

In the current study we explore the extent to which stuttering-like disfluencies can be precipitated on specific words independently of any formulation or articulation difficulty that production of those words might entail. We describe an experiment designed to test whether the likelihood of stuttering increases when participants produce specific words which they have been led to believe will be difficult (for a speech-recognition system) to recognize.

To put the study into perspective, we begin with an overview of two very different theoretical perspectives on the causes of moments of stuttering: Stuttering as a symptom of adaptation to underlying formulation or production impairment, as exemplified by the Covert Repair and EXPLAN Hypotheses (Howell & Au-Yeung, 2002; Kolk & Postma, 1997; Postma & Kolk, 1993), and stuttering as an anticipatory struggle response, as exemplified by the Anticipatory Struggle Hypothesis (Bloodstein, 1958, 1975).

1.1 Stuttering as a symptom of adaptation to underlying impairment

Findings from brain imaging research suggest that, as a group, PWS have both structural and functional weaknesses in areas of the brain associated with syllable planning and production (see Watkins, Smith, Davis, & Howell, 2008, for a review). The accumulation of such evidence has stimulated the development of a number of hypotheses that posit that PWS have underlying language or speech production deficits and that stuttered disfluencies arise as the unintended side-effects of their attempts to adapt to those deficits (Civier, Tasko, & Guenther, 2010; Howell & Au-Yeung, 2002; Max, Guenther, Gracco, Ghosh, & Wallace, 2004; Postma & Kolk, 1993; Vasić & Wijnen, 2005). Most commonly the adaptations that lead to stuttering are believed to involve overburdened 'covert error repair' or 'restart' mechanisms which, under more normal conditions serve 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.

Perhaps the best known of these hypotheses is the Covert Repair Hypothesis (CRH: Kolk & Postma, 1997; Postma & Kolk, 1993) which is predicated on the view that speakers audit their inner speech to check their planned utterances for encoding errors (Levelt, 1983, 1989). Because speech planning takes place somewhat in advance of motor execution, if an error is detected in inner speech, the speaker may have time to stop and reformulate the plan, and thus repair the error before starting to speak. The CRH accounts for the different symptoms of stuttering (whole and part-word repetitions, prolongations and blocks) by postulating that these are the overt symptoms of covert repairs that have been only partially successful because there was insufficient time to repair the error. Thus if cancellation of the erroneous speech plan occurs just as the first phoneme is about to be uttered, a silent pause, or ‘block’, may result while the speaker reformulates it, whereas if cancellation occurs after the first phoneme, syllable or word has already been uttered, a (phoneme, syllable or word) repetition may result, and if this happens several times in a row, then multiple repetitions may occur. More recently a similar mechanism, involving error detection and ‘motor resets’, has been postulated to operate at the level of speech motor control (Civier et al., 2010; Max et al., 2004), and an alternative, threshold-based mechanism whereby stuttered disfluencies arise in response to speakers’ attempts to execute speech-plans are simply incomplete or insufficiently activated, rather than containing actual errors has been posited – in the EXPLAN hypothesis (Howell, 2003, 2011; Howell & Au-Yeung, 2002).

Such mechanisms provide plausible explanations for the variety of stuttering-like disfluencies that occur in both PWS 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)[1], why PWS 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); for why stuttering is more common on utterances that are longer and/or more complex (Logan & Conture, 1995, 1997; Newman & Bernstein Ratner, 2007); and for why young children whose language and articulation skills lag behind those of their peers may be more likely to stutter (Bernstein Ratner, 1997; however, cf. Nippold, 1990; Nippold, 2001).

However, adaptation hypotheses, such as the CRH, EXPLAN and their motor-control equivalents, are less successful at accounting for other observations in relation to the distributions of stuttering events in older children and adults. In particular, the Covert Repair Hypothesis fails to account for the lack of any discernible correlation between the frequency with which AWS produce inner-speech errors and their stuttering severity in everyday speaking situations (Brocklehurst & Corley, 2011), and more generally, adaptation hypotheses fail to account for why older children and adults frequently stutter on isolated, commonly occurring single words; why they have particular difficulty uttering their names; why they are influenced so strongly by the characteristics of the listener and the overall dynamics of the speaking situation, and in particular, why they appear to be able to speak complex utterances perfectly fluently when there is no listener present; and why some speakers with severe language or speech production disorders do not stutter. (see Bloodstein & Bernstein Ratner, 2008, Chapter 10 for an extensive review of such observations).

Thus it appears that, although compensatory responses to underlying difficulties in language production or speech motor control may plausibly account for the stuttering-like disfluencies of young children, they cannot fully account for the

---

[1] Nb. Brutten and Dancer's (1980) use of the term 'Adaptation Effect' is unrelated to the notion of stuttering as a 'symptom of adaptation' to underlying impairment.

persistence of stuttering in older children and adults. An alternative possibility investigated in the current study is that, in adults, stuttering-like disfluencies may occur as a side-effect of compensatory responses (of one type or another) 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 (cf. Conture et al., 2004).

1.2 Stuttering as an anticipatory struggle response

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 PWS 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, Chapter 2, for a review).

Anticipatory struggle hypotheses have proposed 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 experience of difficulty with speech, language, and/or communication in 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 adaptation hypotheses, Bloodstein's hypothesis does not attribute stuttering directly to the speaker's attempts to overcome current instances of production difficulty. Instead, it posits that stuttering arises in response to 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 that belief may differ from child to child, and in some individuals, the impairment or delay that originally caused the belief to become established may no longer be present. Thus, Bloodstein's hypothesis 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 PWS rarely have difficulty speaking to themselves or when they do not 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, 1950b).

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). In comparison, the more recent psycholinguistic hypotheses, outlined above, are much more successful. And secondly, although the notion of anticipatory struggle provides a parsimonious explanation for the observational data and self-reports regarding the moments when stuttering occurs, it has proved particularly difficult to test experimentally.

1.3 The current study

The current study constitutes an experimental investigation of the influence of anticipation on the likelihood of stuttering. Specifically, it investigates whether the experience of stuttering can be precipitated on specific words by instilling, in the speaker, the anticipation that those words will not be recognized by the recipient.

Our experiment was 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. The speaker was led to believe that feedback was controlled by the audience, whereas in reality, it was manipulated 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 positive feedback was delivered than where negative feedback was 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 feedback.

In our experiment, instead of speaking to an audience, participants who stutter spoke single words into what they believed was speech-recognition software on a computer, and received automatic online feedback indicating whether or not those words had been correctly recognized. We designed the paradigm in this way because we specifically wanted to investigate the effect of anticipation of word misrecognition (rather than anticipation of a negative listener response). To avoid any possibility that participants' performances may be affected by the fear of negative evaluation by potential listeners or over-hearers, participants provided their own self-reports of stuttering (or difficulty speaking fluently) and 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.

As in Hansen's (1955) experiment, feedback was, in reality, predetermined, and bore no relationship to the accuracy or fluency with which participants spoke. Participants were prompted to utter each target word four times, receiving feedback after each attempt. Across the four iterations, the feedback consistently indicated either correct or incorrect recognition of the target word. Thus, participants could predict with increasing confidence whether or not the remaining iterations of the target word were likely to be correctly or incorrectly 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 an underlying language or speech production impairment, 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 word-recognition feedback appearing on the computer screen affected their performance on subsequent iterations of the same word would, however, implicate an additional process. If that process is related to the anticipation of a struggle to articulate words sufficiently well for them to be recognized, then regardless of whether or not participants who stuttered had underlying production deficits, participants should be relatively more likely to produce stuttering-like disfluencies in the condition where the software apparently failed to recognize their productions of a particular word.

2 Method

2.1 Participants

Fourteen participants were recruited through stuttering self-help groups and through the University of Edinburgh student employment website. Data from two participants were 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). 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 undergone some form of speech therapy following diagnosis. All considered themselves as still suffering from the condition. Mean SSI-4 (Riley, 2009) stuttering severity score for the group was 15.7, with participants 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). SSI-4 stuttering severity scores were derived from video recordings of each participant answering questions and reading a passage aloud during the debriefing session immediately following their participation in the experiment. Combining the SSI4 tests with the debriefing interview helped reduce the overall time required of participants.

Apart from stuttering, participants reported no speech, language, hearing or visual impairments that were likely to influence the results.

2.2 Materials

The materials consisted of '*target words*' (that participants were required to identify and speak out loud), associated '*distractor 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 1 for examples). In the two experimental sets, each target word was associated with four distractor words 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 1). The purpose of the cues and distractor words was twofold: (a) to increase the ecological validity of the task by introducing an element of choice; and (b) to alert participants to the need for accurate articulation in order to sufficiently distinguish the target word from its competitors. Each target word was also associated with a 'feedback' word informing the participant which out of the five option words the software had recognized. In one experimental condition the feedback word was always correct insofar as it was identical to the target word. In the other it was almost 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 words (one of which was a target word, the other four of which were distractors) and the feedback words were always correct (i.e., identical to the target words). Adding the fillers in this way meant that, for the paradigm overall, the majority of the target words were correctly recognized by the software. This helped to create the illusion that the software was moderately successful at recognizing participants' speech.

Table 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 word (i.e., the word portrayed as having been ‘recognized’ by the software) 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 participants, each experimental target appeared in both the Correct and Incorrect conditions an equal number of times. The additional set of filler targets was added to increase the overall proportion of ‘correctly recognized’ targets to make it appear that the speech recognition system was better than chance: overall, target words were more than twice as likely to be ‘correctly recognized’ than ‘incorrectly recognized’.

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. Unbeknown to participants their responses were recorded by a hidden microphone.

To maintain the illusion that participants’ speech was being ‘recognized’, the software incorporated a voice-activated switch that was sensitive to participants’ verbal responses.

2.3 Procedure

Prior to the experiment, participants completed a demographic questionnaire. Participants were then informed that during the experiment proper they would provide their own self-reports of stuttering; nobody would be able to hear them speak; and their speech would not be recorded. 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. They were informed that for each trial the software was pre-primed to recognize five possible responses: the five 'option words' and that on each trial it would select, from these, the one that best matched their response. Finally, they were informed 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 recognize 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. 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. Before the first iteration of each target word participants rated whether or not they anticipated they would stutter on that word. Then, immediately following each iteration, they (a) self-reported whether or not they actually had stuttered on it, and then (b) received feedback (on the computer screen) indicating whether the word they had spoken had been correctly recognized by the software. Each repetition began when five option 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 1).

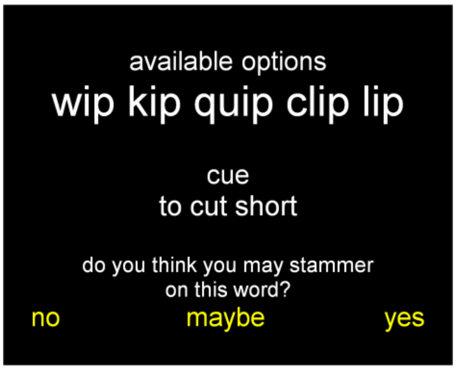

Figure 1. On-screen instructions visible to participants prior to the first iteration of each word.

Having used the cue to identify the target, the participant responded to this question using one of three response keys, labeled "no," "maybe," and "yes." 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 labeled “no,” “maybe,” and “yes.” Pressing a response key triggered a feedback screen. 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 they selected was ‘correct’. If it was one of the 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” (See Figure 2). 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.

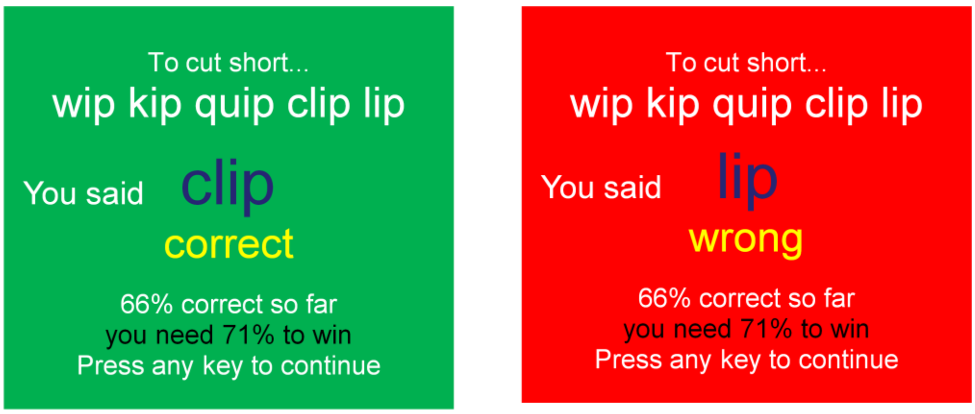

Figure 2: ‘Correct’ and ‘Incorrect’ feedback that appeared on the screen following each iteration of each word.

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 the participant’s 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.

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 two 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.

2.3.1 Coding and Analysis

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 analysis of stuttering was based on participants' self-reports. This approach 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 Moore & Perkins, 1990 for a detailed discussion of the validity of subjective ratings). In addition to participants' self-reports, we also obtained objective measures of their vowel onset latencies and word durations. Analyses of this objective data were carried out primarily in order to provide some insights into the acoustic correlates of the experience of stuttering on single word utterances.

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

Analyses of participants' self-ratings of having stuttered were carried out using logistic mixed-effects regression modeling, 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. 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. 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 increase in the model likelihood ratio) by the addition of each (random or fixed) predictor. Each random and fixed predictor was retained in the model only if it led to an improvement of the model fit. 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). 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 'Iteration'; then 'Condition'; followed finally by a predictor for the Condition by Iteration interaction.

Recordings of participants' utterances were analyzed by the experimenter, using PRAAT (Boersma & Weenink, 2009). Two acoustic measures were taken: (1) Vowel onset latency[2], 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 spectrogram; 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).

Analyses of acoustic data (vowel onset latencies and word durations) were carried out using linear mixed-effects regression modeling (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 'Iteration' (with four levels, one for each iteration), then 'Condition' (with two levels: correct recognition and incorrect recognition), followed finally by a predictor for the Condition by Iteration interaction. Model selection proceeded as for the logistic mixed-effects model of self-reported data described in the previous paragraph.

3 Results

3.1 Stuttering self-reports

Prior to the first iteration, participants provided a total of 384 predictions (each of the 12 participants provided predictions for 32 different words), including 131

[2] As virtually all evidence of struggle occurred prior to the vowel onset, the vowel onset provided the most reliable available landmark for comparing latencies for the onset of fluent speech.

instances where participants predicted possible stuttering and 34 instances where they predicted definite stuttering. They then went on to provide 1536 eligible self-reports of their actual performance (32 different words, each repeated four times by each of the 12 participants), out of which they self-reported a total of 358 stutters (230 possible and 128 definite).

For the statistical analysis, self-ratings of iterations as ‘maybe’ and ‘definitely’ stuttered were pooled due to their low numbers. (See Figure 3 for an overview of the resulting total number of stuttering self-reports on each of the four iterations, collapsed across participants).

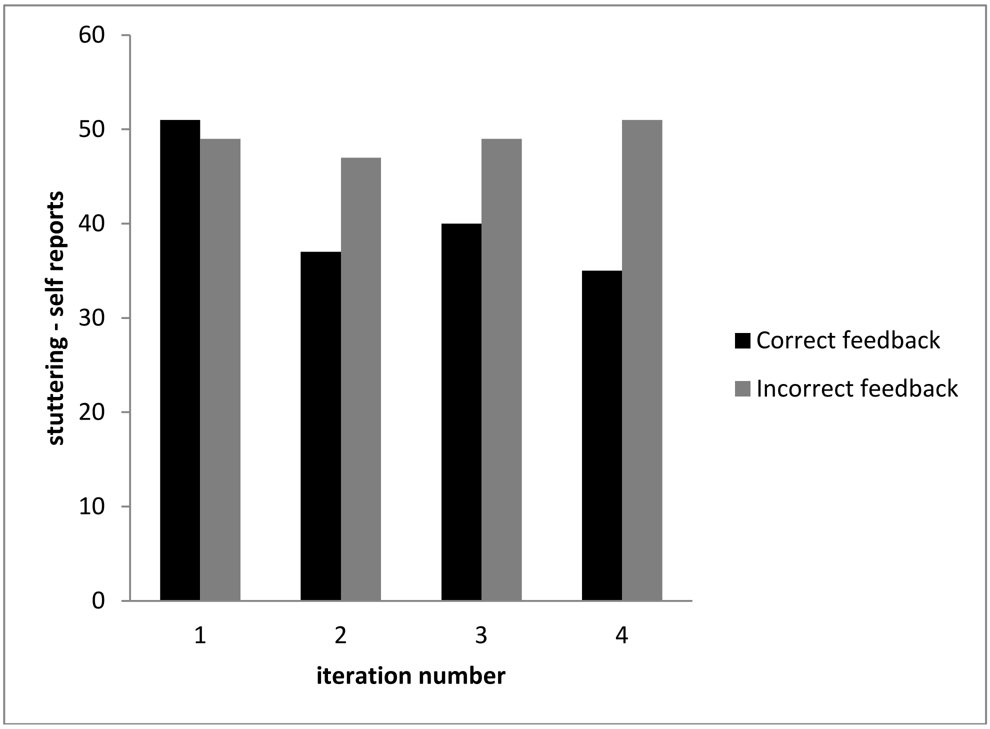

Figure 3: Total number of words self-reported by participants as stuttered, across the four iterations and two conditions.

After controlling for random effects, the best-fit model of self-reported stuttering included ‘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 2 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

Table 2: Mixed effects analyses of random and fixed factors influencing the likelihood of stuttering. Data shown are for the 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**
DV = likelihood of self-reporting stuttering						
Intercept	(stutter not predicted, Iteration 1, correct feedback)	-3.23	0.81	<.001 ***	by word	0.28
					by subject	5.57
Stuttering predicted	maybe	2.29	0.80	.004 **	by subject	5.64
Stuttering predicted	definitely	3.84	1.17	.001 **	by subject	11.01
Condition	incorrect	-0.42	0.50	.405	by subject	0.53
Iteration	+1	-0.33	0.12	.005 **	by subject	NS.
Condition x iteration		0.39	0.16	.017 *	by subject	NS.

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 a relative increase 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.

3.2 Vowel onset latencies

In total, participants provided 1467 codable samples. Mean and standard deviations are provided in Figure 4.

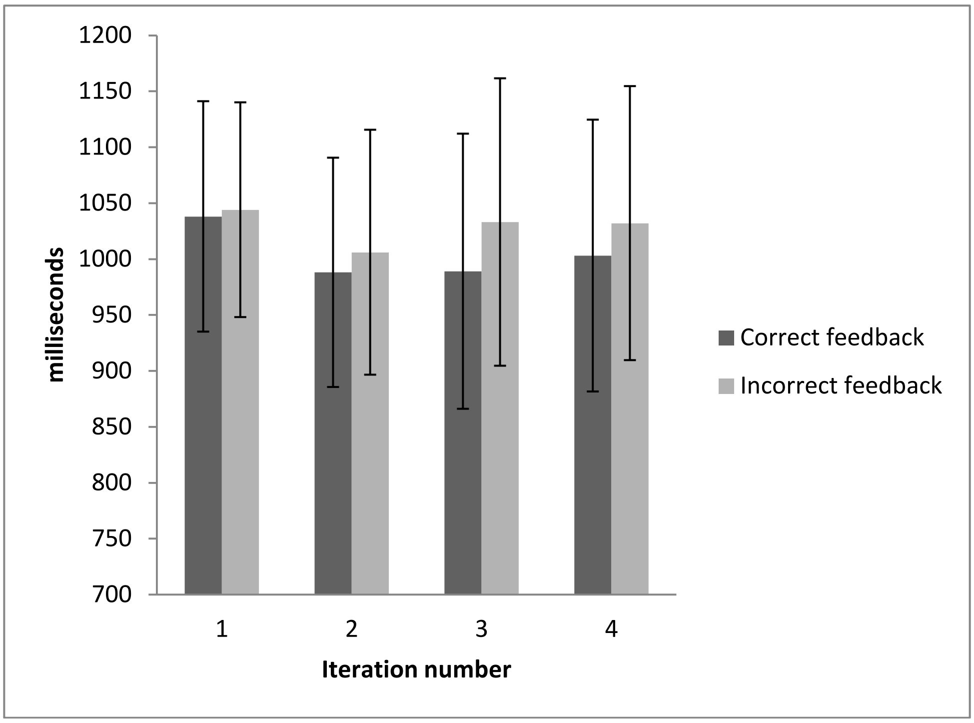

Figure 4: Mean vowel onset latencies with (grand) standard deviations, across the four iterations and the two experimental conditions.

After controlling for random effects, the best-fit model of onset latencies included only 'Condition' as a predictor (improvement due to adding 'Condition: χ^2 (1) = 7.26, p <.007). Adding further predictors did not improve the model (all $p \geq .441$). Table 3 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.

3.3 Word durations

In total, participants provided 1423 codable samples. Mean durations and standard deviations are provided in Figure 5.

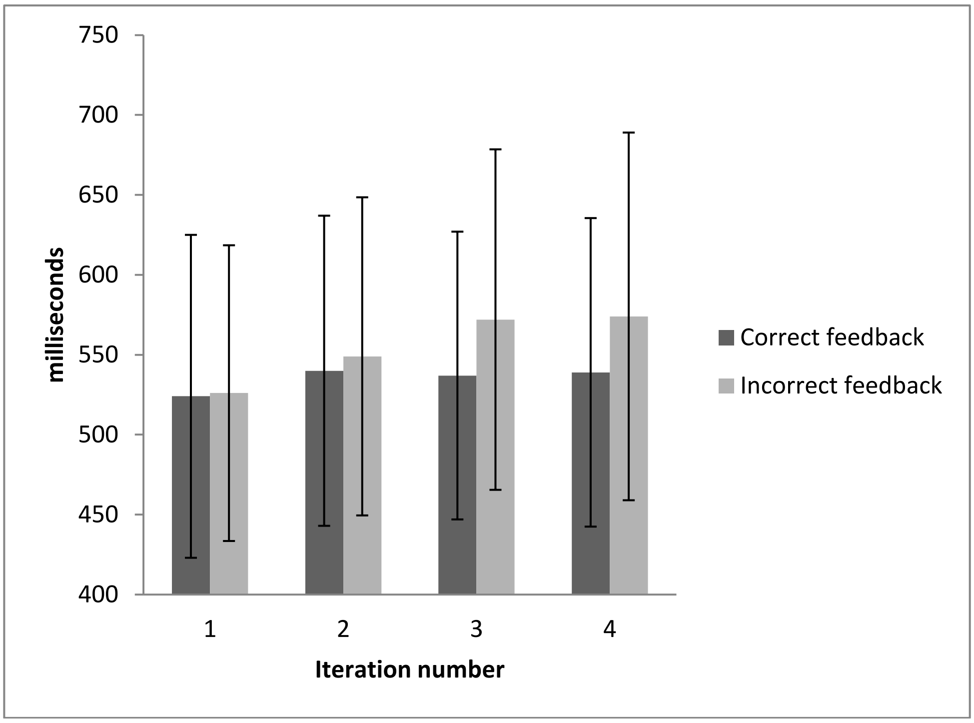

Figure 5: Mean word durations with (grand) standard deviations, across the four iterations and the two experimental conditions.

After controlling for random effects, the best-fit model of word durations included only 'Iteration' as a predictor (improvement due to adding 'Iteration': χ^2 (1) = 12.00, p <.001). Adding further predictors did not further improve the model (all $p \geq .129$). Table 3 gives the coefficients of the model, and the probabilities that they could have occurred by chance.

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

Table 3: Mixed effects analyses of random and fixed factors influencing vowel onset latencies and word durations. Data shown are for the 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**
DV = word onsets (milliseconds.)						
Intercept	(Iteration 1, correct feedback)	1041	29	<.001 ***	by word	1980
					by subject	9063
Condition	Incorrect feedback	28	10	.007 **	by subject	NS.
Iteration	+1	NS.	NS.	NS.	by subject	305
					Residual	36826
DV = word durations (milliseconds.)						
Intercept	(Iteration 1, correct feedback)	540	39	<.001 ***	by word	10717
Intercept		NS.	NS.	NS.	by subject	17639
Condition	Incorrect feedback	NS.	NS.	NS.	by subject	2840
Iteration	+1	11	3	<.001 ***	by subject	NS.
					Residual	17077

3.4 Post-hoc analyses

Visual inspection of Figures 3, 4, and 5 reveals some degree of correspondence between participants’ mean vowel onset latencies, word durations and the pattern of their 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, but was not associated with any increase in word duration (See Table 4 for the model coefficients, and the probabilities that they could have occurred by chance). These post-hoc analyses thus suggest that the longer vowel onset latencies in the incorrect condition may have resulted from difficulty in initiating words. In contrast, the increases in word duration that occurred across iterations (in both experimental conditions) were most likely strategic, and not a result of stuttering.

Insert Table 4 here

4 Discussion

In this study we set out to investigate the extent to which the experience of stuttering can result from the speaker's anticipation of his words being misrecognised. To do this we carried out an experiment in which people who stutter repeatedly spoke single words into what they believed was a speech-recognition system.

The most important finding of this experiment is that the likelihood of a participant self-reporting stuttering on a particular iteration of a word was predicted not only by whether or not s/he anticipated that s/he would stutter (prior to their first iteration of that word), *but also* by whether or not the speech recognition system had 'correctly recognized' his/her previous iterations of that word. Moreover, the feedback participants received from the speech recognition system influenced the likelihood of stuttering on subsequent iterations of that word despite the fact that, in this paradigm, the feedback was not influenced by their actual performance.

These findings support the hypothesis that the anticipation of communication failure can precipitate stuttering in AWS. The findings also suggest that, in AWS, the anticipation of communication failure is influenced both by feedback from 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.

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 canceled out by the experience of repeated communication failure.

The finding that feedback received by participants influenced the likelihood of stuttering on subsequent iterations of that word even though it bore no relation to their actual performance on that word does not appear to be compatible with present formulations of the Covert Repair or EXPLAN hypotheses which posit that instances of stuttering arise directly in response to incomplete, insufficiently activated or erroneous speech plans. It does not, however, rule out the possibility that language or speech planning deficits may play a role in the production of stuttering events in AWS.

The above finding is, however, compatible with an anticipatory struggle account. Furthermore, in light of the way the experiment was designed, it is reasonable to conclude that the '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.

In this particular experimental paradigm, there was no semantic component to the word recognition process, and it would have appeared to participants that successful word recognition was entirely dependent on the phonetic accuracy and clarity of their productions. 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, under certain circumstances, PWS can have as much difficulty speaking nonsense words as meaningful words (Packman, Onslow, Coombes, & Goodwin, 2001).

4.1 Mechanisms that could result in stuttering following anticipation of communication failure

If a speaker anticipates that his words are likely to be misrecognised or misunderstood, he is likely also to perceive that the pressure is on him to adjust his speaking style in some way to rectify the situation. The findings of the current study suggest that in PWS, at least with respect to single-word utterances, some of those adjustments result in stuttering. Bloodstein proposed that PWS are likely to stutter on potential problem words due to an anticipatory struggle response that involves excessive tension and fragmentation. However, as mentioned in the introduction, the 'tension and fragmentation' account fails to adequately explain the specific forms that stuttered disfluencies characteristically take – namely repetitions, prolongations and blocks. In light of this, the question arises as to whether the anticipation of communication failure could lead to an increase in stuttered disfluencies through mechanisms similar to those that have been posited by the CRH and EXPLAN hypotheses. We now consider this possibility.

With respect to covert error repair, Vasić and Wijnen (2005), have posited that past experiences of difficulty lead PWS to develop a tendency to focus abnormally intensely on minor timing variations and infelicities in their speech, and to set their thresholds for the initiation of covert error repairs at too low a level. This may result in a vicious circle whereby the disfluencies resulting from covert error repairs themselves spark off further covert error repair activity. Crucially, this 'Vicious Circle Hypothesis' can be extended to account for the influence of anticipation by positing that PWS are most likely to monitor hypervigilantly at times when they anticipate communication failure. Thus, if this is the case, past experiences of difficulty with specific sounds or words may trigger unnecessary covert error-repair activity on those specific sounds or words.

The EXPLAN hypothesis (Howell & Au-Yeung, 2002) rejects the idea that stuttering stems from excessive covert error repair activity. Instead it posits that, in PWS, the activation levels of speech plans build up abnormally slowly and so are not ready for execution at an appropriate time. Implicit in this hypothesis is the notion of a threshold mechanism whereby words can only be executed after their activation exceeds a certain level (e.g., Howell, 2003, 2011). This hypothesis can be modified to account for the influence of anticipation of communication failure simply by additionally proposing that the release threshold is *variable* and will rise whenever the speaker anticipates that a word will be misheard (or misunderstood). Because it takes time for activation levels of words to rise, a rise in the release threshold would normally slow the rate at which the word(s) are released for execution and consequently maximize the likelihood that they will be executed clearly and accurately. However, in PWS, anticipation that a word is likely to be misrecognised may cause the release threshold of that word to rise so high that the word fails to be released at all. This would explain why PWS are more likely to find themselves unable to initiate execution of a sound if they are particularly concerned about the possibility of it being misheard or misunderstood. Such a scenario might first arise in response to any underlying language or speech production impairment. However, it may then develop into a learned response that is triggered by anticipation.

An important prediction of the two psycholinguistic versions of the anticipatory struggle hypothesis that are described above is that, at times of anticipated difficulty, it may be possible to reduce stuttering and increase the chances of successful communication by making less effort to articulate potential problem words accurately. This is because a reduction in the level of accuracy aspired to should lead to a reduction in covert repair activity and also to a lowering of the 'release threshold' that plans need to exceed in order to be executed. It is noteworthy that such a strategy is counterintuitive and thus probably the opposite to that which most PWS normally attempt.

4.2 Caveats

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. During the debriefing session, when, participants were asked about their experiences of using such software in real life situations they consistently reported finding such experiences difficult and particularly likely to precipitate stuttering if the software failed to 'understand' what they said. 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.

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 PWS in general, or only to adults who stutter 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.

5 Conclusions

Adults who stutter are more likely to stutter on single words when the speaking circumstances lead them to anticipate communication failure on that word. This effect is independent of the degree of difficulty inherent in the production of the word itself, although it may interact with it. This finding is incompatible with hypotheses that posit that stuttering in adults occurs as a direct result of language or speech production difficulty alone.

The findings are most consistent with the hypothesis that persistent stuttering is characterized by inappropriate functioning of covert repair and/or variable release threshold mechanisms that, under more normal circumstances, may 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.

6 References

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

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

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

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

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.

Bloodstein, O., & Grossman, M. (1981). Early stutterings: Some aspects of their form and distribution. *Journal of Speech and Hearing Research, 24*(2), 298.

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/

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

Brocklehurst, P. H., & Corley, M. (2011). Investigating the inner speech of people who stutter: Evidence for (and against) the Covert Repair Hypothesis. *Journal of Communication Disorders, 44*, 246-260.
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.
Brutten, G., & Dancer, J. E. (1980). Stuttering adaptation under distributed and massed conditions. *Journal of Fluency Disorders, 5*(1), 1-10.
Burger, R., & Wijnen, F. (1999). Phonological encoding and word stress in stuttering and nonstuttering subjects. *Journal of Fluency Disorders, 24*(2), 91-106.
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.
Conture, E., Zackheim, C., Anderson, J., & Pellowski, M. (2004). Linguistic processes and childhood stuttering: many's a slip between intention and lip. In B. Maassen, R. Kent, H. Peters, P. v. Lieshout & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 253-281). Oxford: Oxford University Press.
DebRoy, S., & Bates, D. M. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis, 91*(1), 1-17.
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.
Hansen, H. T. (1955). *The effect of a measured audience reaction on stuttering behavior patterns. (Unpublished doctoral dissertation).* University of Wisconsin, Madison.
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. (2011). *Recovery from Stuttering*: Psychology Press/Taylor & Francis Group.
Howell, P., & Au-Yeung, J. (2002). The EXPLAN theory of fluency control applied to the diagnosis of stuttering. In E. Fava (Ed.), *Clinical linguistics: Theory and applications in speech pathology and therapy*
(pp. 75–94). Amsterdam; Philadelphia: J. Bechamin's.
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. (2001). Function word repetitions emerge when speakers are operantly conditioned to reduce frequency of silent pauses. *Journal of Psycholinguistic Research, 30*(5), 457-474.
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.
Kleinow, J., & Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research, 43*(2), 548.

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.

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.

Lieshout, P. H. H. M. v., Hulstijn, W., & Peters, H. F. M. (1996). Speech Production in People Who Stutter: Testing the Motor Plan Assembly Hypothesis. *Journal of Speech and Hearing Research, 39*(1), 76-92.

Logan, K. J., & Conture, E. G. (1995). Length, grammatical complexity, and rate differences in stuttered and fluent conversational utterances of children who stutter. [doi: 10.1016/0094-730X(94)00008-H]. *Journal of Fluency Disorders, 20*(1), 35-61.

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., & Sasisekaran, J. (2007). Jaw-phonatory coordination in chronic developmental stuttering. *Journal of Communication Disorders, 40*(3), 257-272.

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.

Moore, S., & Perkins, W. (1990). Validity and reliability of judgments of authentic and simulated stuttering. *Journal of Speech and Hearing Disorders, 55*(3), 383.

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.

Packman, A., Onslow, M., Coombes, T., & Goodwin, A. (2001). Stuttering and lexical retrieval. *Clinical Linguistics & Phonetics, 15*(6), 487-498.

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.

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

Riley, G. (2009). Stuttering Severity Instrument for Children and Adults. (SSI-4) (4th ed). Austin TX: Pro-Ed

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.

Sheehan, J. (1953). Theory and treatment of stuttering as an approach-avoidance conflict. *The Journal of Psychology, 36*(1), 27-49.

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.

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 Riper, C. (1973). *The treatment of stuttering*: Prentice-Hall Englewood Cliffs, NJ.

Vasic, N., & Wijnen, F. (2005). Stuttering as a monitoring deficit. *Phonological encoding and monitoring in normal and pathological speech*, 226–247.

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.

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain, 131*(1), 50-59.

Yairi, E., & Ambrose, N. (2005). *Early Childhood Stuttering*. Austin, TX: PRO-ED, Inc. .