Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10–12 September 2005, Aix-en-Provence, France, pp. 93-98.

Simulations of the types of disfluency produced in spontaneous utterances by fluent speakers, and the change in disfluency type seen as speakers who stutter get older

Peter Howell and Olatunji Akande**

* Department of Psychology, University College London

Abstract

The EXPLAN model is implemented on a graphic simulator. It is shown that it is able to produce speech in serial order and several types of fluency failure produced by fluent speakers and speakers who stutter. A way that EXPLAN accounts for longitudinal changes in the pattern of fluency failures shown by speakers who stutter is demonstrated.

1. Introduction

There are several ways in which loss of speech control can be accounted for. Three accounts have been discussed [15], one of which was proposed by the speech team in the Department of Psychology, University College London (UCL). This account, the EXPLAN model of fluency control [7, 9, 11], represents planning and execution as independent processes (the acronym for the model takes its name from these two processes). According to the model, fluency fails when a) planning is slow or b) execution rate is rapid. EXPLAN is a model of fluent speech control and how fluency can fail. The types of fluency failure observed in children who stutter (CWS) are similar to those seen in fluent children (though fluency failures occur more frequently in the CWS, which makes them useful for studying these events). Stuttering in adults who stutter involves events that are comparatively rare in fluent children and CWS. If the theory is right about its proposal as to how fluency failures arise in all children, it also needs to account for how and why fluency failures change in type when CWS become adults who stutter.

EXPLAN is assessed in this paper by simulating the way planning and execution processes interact according to the model. The performance of the model is examined to see if it can produce a) a fluent sequence of speech, and b) the different types of disfluency that arise when planning and execution are perturbed (slowed and speeded respectively). The paper reviews the history of EXPLAN in section two. Section three outlines the basic model that generates fluent speech according to the principles underlying the account (separate planning and execution components). In section four, the parameters representing planning and execution are perturbed to see whether incidence and type of disfluency are affected in the predicted manner. Finally, section five shows how the features seen in stuttering as it persists into adulthood, could be an adaptive change to the patterns of fluency failure seen in normal fluency development when the processes that led to these failures persist into adulthood.

2. EXPLAN theory

2.1. Planning

Typically, the elements in an utterance (the words, for example) vary in how complex they are to plan. Complexity can be reflected at different processing stages in utterance formulation or planning, including syntactic, lexical (word class and word frequency), phonological, prosodic and phonetic levels. Views that maintain that fluency failure is a result of planning complexity abound in the stuttering literature, and they usually emphasize the role of one linguistic level. For instance, Bernstein Ratner [3] emphasizes the syntactic level, a group at East Carolina consider word frequency is paramount, and Wingate [26] has promoted the view that the phonological level (and prosody in particular) is the primary source of the problem.

The planning component in EXPLAN allows complexity to be affected by any of these levels (though the evidence suggests that a pure syntactic deficit is not likely to be a determinant of stuttering [13, 14, 22, 23 for reviews]. To allow all the remaining levels to influence fluency failure, we have examined lexical forms (function and content words) separately. The reason for this is that lexical class correlates with the other factors that could specify complexity, so using lexical type integrates contributions from these several sources. For instance, content words are the only type of words stressed in English, so investigating lexical class is effectively also an examination of stress (and lexical class is easier to determine objectively than is stress).

Word frequency is conceived in a different way to stress in EXPLAN. Content words tend to be low frequency, so frequency correlates with lexical class. However, a case has been made that, in work on isolated words, word frequency can be dissociated from lexical class (there are low frequency function words) and that low frequency words are more likely to be stuttered than high frequency ones, irrespective of their lexical status. The implication to drop lexical status that is carried by this view is problematic because types of fluency failure depend on the type of word they occur on. Function words are repeated in their entirety or have pauses preceding them, content words, involve disfluency on their first part which can either be prolonged, repeated or have a break inserted between the initial (onset) and subsequent (rhyme) parts. It is hard to conceive how word frequency that operates independent of lexical class could account for why different types of disfluency are linked with different word types. Word frequency differences are idiosyncratic and also have an ephemeral property in an individual's speech. Though these properties make infrequent words unamenable to systematic study, there are some observations that can be made about the acquisition of words of different frequency: Function word usage will be reasonably stable once a mature syntactic system has been established. Content word usage will increase in frequency throughout life. Content word vocabulary in early life consists of relatively high frequency such words. Expansion of vocabulary at the start and throughout adulthood tends to involve adding low frequency content words. The latter aspect, in particular, that is specifically a feature of content words, has a critical role (as indicated in section four) in the current simulation for explaining the ontogenesis of the disorder into adulthood.

Content words tend to be phonetically and phonologically more complex than function words. The UCL team has used phonological and phonetic measures as a way of quantifying different levels of difficulty within function and content word classes. Incidence of phonetic properties varies between these word classes and, more surprisingly, the incidence of phonetic properties varies across the age range 6 years to adulthood. This is shown in Figure 1 for words with difficult manners, words which are long and words which have contiguous consonants (shown separately for content and function words). Figure 1a (content words) shows significant increases over ages in use of each factor whilst there is no systematic increase for function words (Figure 1b).

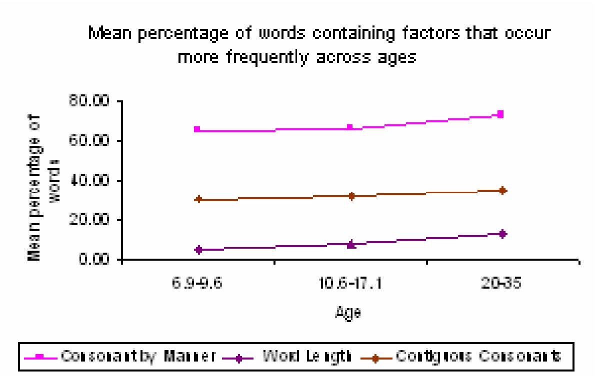

Figure 1a. Mean percentage of content words containing difficult manners, long words or contiguous consonants that occur more frequently in the speech of older speakers.

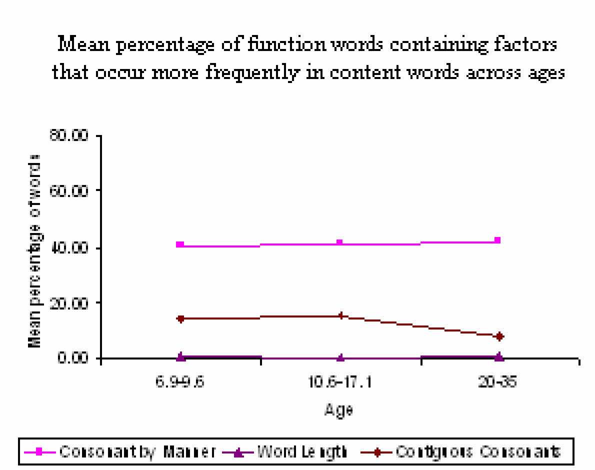

Figure 1b. Mean percentage of function words containing difficult manners, long words or contiguous consonants.

Comparison of percentage of words containing each of these factors also shows a big imbalance between usage of these phonetic properties between word classes (all factors occur less frequently in function words). The change over word type and ages for content words is probably a reflection that vocabulary increases with age and content words that are acquired later are more complex than those acquired early. Note that these variables also relate to word frequency as words acquired late are likely to be used infrequently. The phonetic and phonological properties correlate with stuttering rate for content words, but not for function words. This is shown in Figure 2 where phonetic difficulty is represented as the sum over four factors [18] (the three indicated earlier plus whether the word contained a dorsal consonant) for words marked as difficult (e.g. if a word contains a contiguous string of consonants, it scores a point).

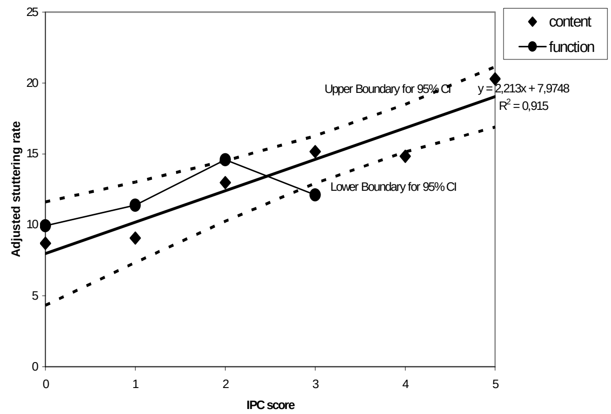

Figure 2. Adjusted stuttering rate (ordinate) versus number of times the four factors marked as difficult occurred (abscissa) for speakers aged over 18 years. The straight line is fitted to the content words and the upper and lower bounds around this line are indicated by the dashed line. The function word points are connected by a solid line.

There is a significant correlation for content, but not function words. The graph also shows that function words have a more limited range of phonetic difficulty. The lack of correlation with the difficulty measure for the function words underlines the importance of examining word types separately.

In summary, planning of function words appears to differ from the planning of content words: Content words show different form of fluency failure (disfluency on parts), this class includes attributes that relate to stuttering (stress, phonetic and phonological indices), across the members of the class, word frequency is low and this is also reflected at the phonetic level (for example, there is variation in usage of phonetic features between age groups for content words). Function words are more stable insofar as disfluencies on them involve complete words, this class does not carry stress in English nor do function words usually have complex phonetic properties and the incidence of the phonetic properties of words in this class does not vary across age groups. We assume that more difficult words take longer to plan and, if difficulty is measured phonetically, there are going to be influences specific to content words that vary with age (even into adulthood, as seen in Figure 1) that impact on fluency (Figure 2). Reference to function/content words can also be regarded as comment on stress and phonetic properties. Thus, if instead of lexical type, words were represented as bundles of phonetic features and divided into classes with high frequency of phonetic features, we would have come full circle insofar as these properties would define content and function types.

2.2. Execution

If timing constraints lead to stuttering and the speech execution system operates independent of planning, increasing execution rate should exacerbate fluency problems and decreasing execution rate should decrease the chance of fluency problems [16, 17]. As planning and execution are represented independently in EXPLAN, a speaker can be planning a different segment to that currently being executed. This necessitates examination of how planning and execution interact which is the third component in the model (discussed next).

2.3. Interaction between planning and execution and disfluency types

Representing planning and execution processes separately allows speech to be planned in advance of the extract of the utterance currently being emitted. Though it is desirable to allow planning and execution to proceed on different timescales, some specification needs to be made about the point at which they interact.

If the plan for a later word in an utterance is not ready, fluency is likely to fail (though this is not inevitable, depending on the execution model). To flesh out this view, the contextual units within which utterances are planned and delivered need to be specified. We have used units that have been developed for other purposes in phonology. The contextual unit is the phonological, or prosodic, word, PW [24]. PW, as defined by Selkirk, have an obligatory content word that can be preceded and followed by various numbers of function words [See 10 for an alternative definition of PW]. Examples of PW are 'in the spring' (two function words precede the content word), 'I hit him' (one function word preceding and one function word following the content word). Speakers could start by planning 'in', and when complete can start its execution, while they plan 'the' and so on. As long as there is enough time to plan the next word, speech will proceed fluently.

Disfluencies occur when speakers are in the limiting situation and there is not sufficient time to plan the next word during the time allowed for the current word to be executed. The disfluencies that occur within PW can be conveniently divided into those which involve whole, and those which involve parts of, words. Whole word repetition has either been taken as signifying that speaker restarts an utterance because an error was made [19, 20] or stalls because a future word is not ready in time [5, 7, 9, 11, 21]. This class tends to involve function words which are more likely than not to start an utterance (more than 50% of utterances start with a function word in [2]) and contain the majority (more than 90%) of all types of disfluency including those on function words. Stalling explains why this happens on initial, not final, function words [25]. Stalling is also consistent both with the view that the speaker restarts at the boundary of a constituent that contains an error and that stalling occurs before a content word.

Disfluencies in content words are rarer. They often involve part of such words, in particular repetition or prolongation of the initial part (they are called advancings). Ambrose and Yairi [1] show that these make up less than 1% of disfluency for three age groups - fluent. Speakers who stutter have a higher rate (up to eight times). This type of disfluency occurs away from the constituent boundary [2] and occurs on words where phonetic complexity is high [6] and is seen mainly in older speakers who stutter [12].

3. Implementation representation of planning and execution and how they interact and lead to disfluency

The initial version of EXPLAN accounted for fluent production of a series of words and stalling and advancing fluency failure. Planning in this early version was represented as a line indicating the time taken (so the line for a content word was greater than that for a function word). Execution was represented similarly, though it has to occur with an offset relative to planning. The offset between the two indicated the interaction and the way this could lead to fluency failure.

Figure 3 shows the particular situation where planning of one segment is completed during the time the preceding segment is being executed, thus leading to fluent speech.

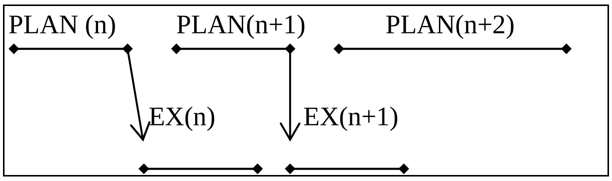

Figure 3. Diagrammatic representation of the temporal relationship between planning and execution for speech produced fluently. Time is along the abscissa. The epoch during which planning (PLAN) and execution (EX) occur are shown as bars in the top and middle rows respectively. Planning of adjacent words is shown in series for simplicity. Execution of word n commencing after its plan is complete and that there is sufficient time to plan the following word (n+1) while word (n) is being executed.

The case shown is at the limit where the next plan is just ready in time (when the speaker has pre-planned speech well in advance, that speech will also be fluent). Speakers need to gain extra time for planning when the subsequent segment is not ready. They can do this by pausing or repeating one or more prior segments (referred to as 'stallings'). If the segments that are planned rapidly are function words and the segments that take longer to plan are content words, the hesitation or repetition occurs around the function words. The function words are produced fluently during stalling and, for this reason, stalling is the least risky way of dealing with fluency failures and predominates in children's speech (whether they are fluent or stutter). This situation is depicted in Figure 4.

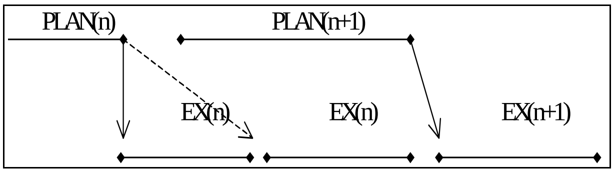

Figure 4. Representation of stalling as using the same conventions as in Figure 3. Execution of a prior word is complete before the plan for the following word is completed. In this case, after word n has been spoken the first time, it is repeated to allow more time to complete the plan of word (n+1).

Alternatively, the speaker can commence execution of the partly-prepared word [20]. If speakers do this, the plan may run out part way through the word (resulting in part-word disfluencies as shown in Figure 5).

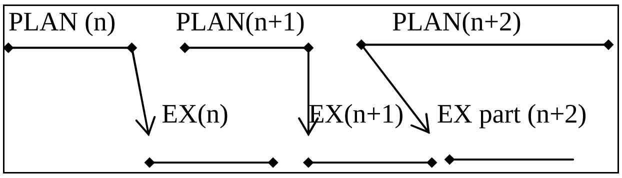

Figure 5. Representation of advancing as in Figure 3. After the first two words (n and n+1) have been completed, even though the plan of the next word is not ready, the speaker commences word (n+2) and the plan runs out resulting in part-word disfluency.

The part-word disfluencies that result would occur on segments that take a long time to plan, which would be content words in this case. The units within which this interaction occurs are PW (in particular those with initial function words).

In the next version of the model [8], planning was represented as an activation profile (such as that which would arise in a spreading activation account). Content words were assumed to have slower activation rates than function words. Selection of a word for execution was based on maximum activation as long as activation was above a minimum threshold. The approach taken was to show diagrammatically that there were occasions on which words were produced in correct sequence (fluent serial order), where preceding function word had higher activation than the content word that was next in sequence (which led to function word repetition) and when the content word had highest activation but was not at maximum resulting, if it was executed, in a part content word disfluency. In this model, planning differences were limited to gross differences between function and content words. The current version of EXPLAN includes more subtle planning rate differences and their effect on output when they are perturbed. In the first activation profile model, execution processes were not explicitly identified (they are complex and fill the interval represented by the difference between word initiation of two words which is affected by the point where a word was initiated, planning rate and threshold at which execution could be initiated).

4. Perturbation of parameters representing planning and execution

The range of disfluencies examined in the previous activation profile model was limited and the perturbations to the planning and execution processes were not examined systematically EXPLAN has now been implemented using a dynamic graphics package and speech output interface. A sequence of words is input (shown at left of each section in Figures 6-8 for ‘in the morning’ – a PW of the form FFC). Each red line represents buildup (rising line) and decay of a single word. The last point of the three red lines represents the activation level of each word at selected points in time. The threshold a word has to reach to be eligible for production is indicated. Figures 6, 7 and 8 show snapshots after ‘in’, ‘the’ and ‘morning’ have been produced (the way activations builds up dynamically will be demonstrated in the presentation).

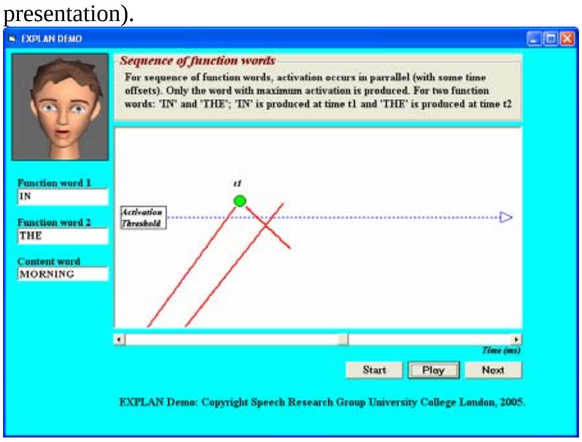

Figure 6. Activation rate (ordinate) over time (abscissa) for two function words followed by a content word. This still picture shows the situation after “in” has been executed and “the” is about to be produced.

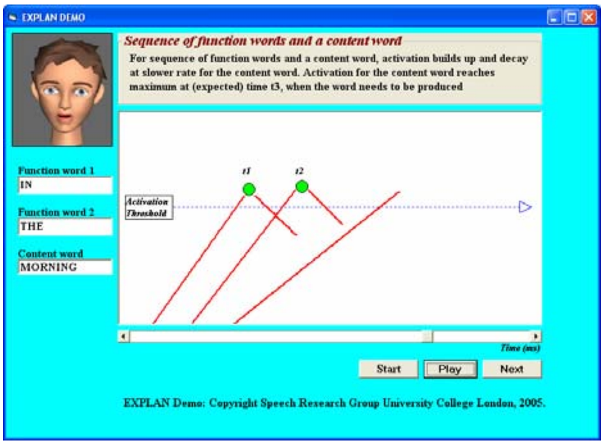

Figure 7. As Figure 6 showing the situation after “the” has been completed.

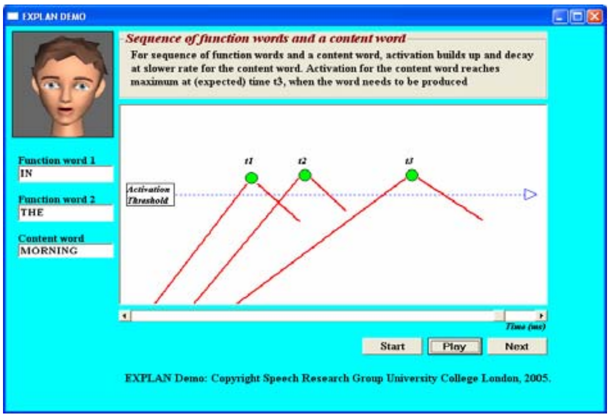

Figure 8. As Figure 6 showing the situation after “morning” has been completed.

Next the execution and planning inputs were perturbed in ways that should lead to the different types of disfluency. Execution rate changes were simulated by shortening the gap between words. As shown in Figure 9, this leads to the activation of the second function word being still above threshold after it has been produced once, while the content word is below activation threshold. Consequently, ‘the’ is repeated. After repetition of ‘the’, the content word is above activation threshold and can be produced.

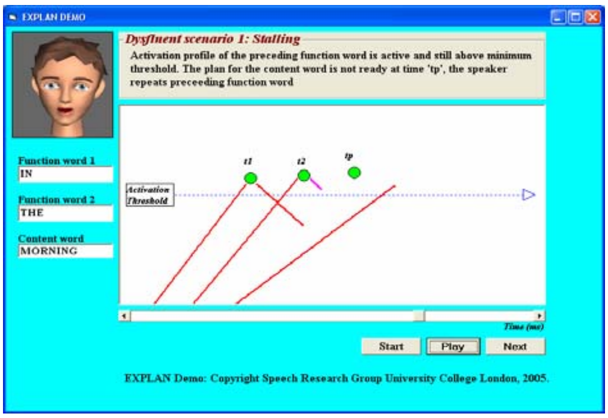

Figure 9. Diagram of the situation leading to function word repetition (see text for description).

Multiple repetitions of the same function word arise due to reactivation of the plan as proposed by Blackmer and Mitton [4]. The function word is reactivated however many times as are needed for the content word to reach its threshold for activation.

The detailed structure of phrase repetitions has not been described previously, so analysis was made of such events. The two main observations are: 1) Phrase repetitions occur most often on PW that start with two function words (as in the 'in the morning' example). This is predicted by EXPLAN (both words can be used for stalling). 2) PW of this form allow repetition of the first ('in in the morning') or second ('in the, the morning') word as well as phrase repetition ('in the, in the morning'). The PW of six CWS were examined for PW that started with two or more function words and which had repeated words. Repetition was at the phrase level for 9.7% of the sample (as in "in the, in the morning". Word repetition (90.3%) only involved the first word (as in "in, in the morning") (never the second). Repetition of the first word is consistent with Levelt's [20] main interruption rule, MIR (this rule states that speakers interrupt as soon as possible before difficulty is experienced). MIR is included in the simulator to produce phrase repetitions.

Planning rate changes were simulated by decreasing the slope of content word activations relative to function words. Decreasing the slope eventually results in the content word having a higher (but not full) activation than the function word which leads the speaker to produce an advancing type disfluency as shown in Figure 10. In this situation, only the plan of the first part of the content word is available to be produced.

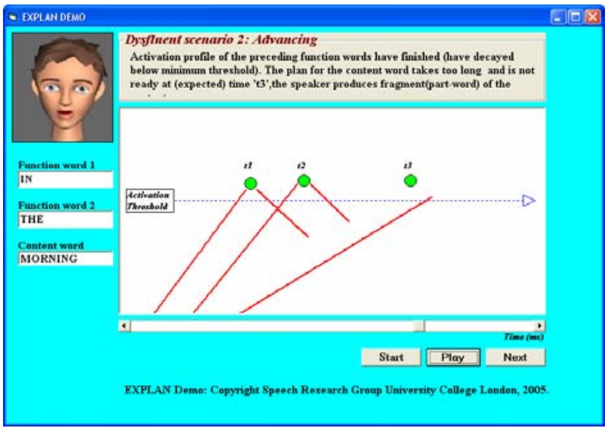

Figure 10. Diagram of the situation leading to part-content word disfluency (see text for description).

Variation in activation rates of word in the content word category occurs because of phonetic difficulty (Figure 2) which correlates with frequency with which these properties occur in content words over ages (Figure 1a). This is used in the simulations in section five to predict change in disfluency type for speakers who stutter over ages.

Recycling utterances already available, when applied to part of a plan, would lead to prolongation and part-word repetition – prolongation would arise when the onset consonant alone is available (which happens mainly on fricatives, laterals and nasals) and part-word repetitions when the plan is complete up to the onset-nucleus boundary or to onset plus nucleus, but typically not beyond that point in the syllable (i.e. not to the coda). The question that arises if this account is correct, is how speakers break out of the prolongation or repetition loop. This question did not arise in connection with whole word repetitions as planning of the subsequent content word continues and its execution takes over when it is completed. Here, however, the element being prolonged or repeated is also the element being planned. There are several ways that part of a word can be, for example, prolonged while at the same time its planning can continue. One way of achieving this is to allow the elementary constituents of a word that are available (e.g. up to the end of the onset consonant or the start of the coda) to be the elements that are reinitiated rather than the word forms. Planning of the rest of the word can then proceed if it is not complete.

5. Simulation of change from stalling to advancing with age

As noted, speakers who stutter change from producing stallings to advancing as they get older [12]. One explanation for why this occurs is that it arises because of the impact that changes in the frequency of usage of content words (function word usage remains constant once a child has learned the syntax of a language). Speakers who are inclined to produce a high incidence of stallings (CWS) use more content words as they get older. This dilutes the frequency of occurrence of all content words, making all words rarer and new words infrequent when they are acquired late, compared with when they are acquired early, in life. If the activation rate is determined in part by difficulty, it would tend to decrease for all words (this would apply to the words learned recently too). In turn this would tend to reduce the chances of content words reaching full activation and have the effect of increasing incidence of advancing-type disfluencies).

6. Summary and Conclusions

EXPLAN has been outlined and simulations of the main types of disfluency shown by fluent speakers and speakers who stutter have been presented. It has been shown how advancing type disfluencies could be a natural response to vocabulary changes in speakers prone to produce a high rate of disfluencies. As the emergence of this form of disfluency is based on processes that occur in fluent speakers as well as speakers who stutter, there would not appear to be structural central nervous system abnormalities that lead to the disorder.

7. Acknowledgement

This research was supported by Wellcome Trust grant 072639.

References

[1] Ambrose, N.G. & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language and Hearing Research, 42,* 895-909.

[2] Au-Yeung, J. Howell, P. & Pilgrim, L. (1998). Phonological words and stuttering on function words. *Journal of Speech, Language and Hearing Research, 41,* 1019-1030.

[3] Bernstein Ratner, N. (1997). Stuttering: A psycholinguistic perspective. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd edition). Needham, MA: Allyn & Bacon.

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

[5] Clark, H. & Clark, E. (1977). *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace.

[6] Dworzynski, K. & Howell, P. (2004). Predicting stuttering from phonetic complexity in German. *Journal of Fluency Disorders, 29*, 149-173.

[7] Howell, P. (2002). The EXPLAN theory of fluency control applied to the Treatment of Stuttering by Altered Feedback and Operant Procedures. In E. Fava (Ed.), *Current Issues in Linguistic Theory series: Pathology and therapy of speech disorders* (pp.95-118). Amsterdam: John Benjamins.

[8] Howell, P. (2003). Is a perceptual monitor needed to explain how speech errors are repaired? *Gothenburg papers in Theoretical Linguistics, 90*, 29-32.

[9] Howell, P. (2004a). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communicative Sciences and Disorders, 39*, 122-139.

[10] Howell, P. (2004b). Comparison of two ways of defining phonological words for assessing stuttering pattern changes with age in Spanish speakers who stutter. *Journal of Multilingual Communication Disorders, 2*, 161-186.

[11] Howell, P. & Au-Yeung, J. (2002). The EXPLAN theory of fluency control and the diagnosis of stuttering. In E. Fava (Ed.), *Current Issues in Linguistic Theory series: Pathology and therapy of speech disorders* (pp.75-94). Amsterdam: John Benjamins.

[12] 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*, 345-354.

[13]Howell, P. Davis, S. & Au-Yeung, J. (2003). Syntactic development in fluent children, children who stutter, and children who have English as an additional language. *Child Language Teaching and Therapy, 19*, 311-337.

[14] Howell, P. & Dworzynski, K. (in press). Planning and execution processes in speech control by fluent speakers and speakers who stutter. *Journal of Fluency Disorders.*

[15] Howell, P. Hayes, J. & Akande, O. (in press). Factors that determine the form and position of disfluencies in spontaneous utterances. Disfluencies in Spontaneous Speech.

[16] 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*, 457-474.

[17]Howell, P. Au-Yeung, J. & Pilgrim, L. (1999). Utterance rate and linguistic properties as determinants of speech dysfluency in children who stutter: *Journal of the Acoustical Society of America, 105*, 481-490.

[18] Jakielski, K. J. (1998). *Motor organization in the acquisition of consonant clusters*. PhD thesis, University of Texas at Austin. Ann Arbor Michigan.

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

[20] Levelt, W. J. M. 1989. *Speaking. From Intention to Articulation*. Cambridge, Massachusetts: MIT Press.

[21] MacWhinney, B., & Osser, H. (1977). Verbal planning functions in children's speech. *Child Development, 48,* 978-985.

[22] Nippold, M. A. (1990). Concomitant speech and language disorders in stuttering children: A critique of the literature. *Journal of Speech and Hearing Disorders, 55,* 51-60.

[23] Nippold, M. A. (2001). Phonological disorders and stuttering in children: What is the frequency of co-occurrence? *Clinical Linguistics and Phonetics, 15*, 219-228.

[24] Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

[25] Strenstrom, A-B. & Svartvik, J. (1994). Imparsable speech: Repeats and other nonfluencies in spoken English. In N. Oostdijk & P. de Haan (Eds.), *Corpus-based research into language*. Atlanta, GA: Rodopi.

[26] Wingate, M. (1984). Stutter events and linguistic stress. *Journal of Fluency Disorders, 9*,295-300.