Europe PMC Funders Group
Author Manuscript
Contemp Issues Commun Sci Disord. Author manuscript; available in PMC 2008 February 06.

Published in final edited form as:
Contemp Issues Commun Sci Disord. 2004 ; 31: 122–139.

Europe PMC Funders Author Manuscripts

Assessment of Some Contemporary Theories of Stuttering That Apply to Spontaneous Speech

Peter Howell
University College, London

Abstract

In this article, a selection of theoretical approaches about stuttering is examined. One way of characterizing theories is in terms of whether the problem of stuttering arises at the linguistic or motor levels or in the interaction between the two. A second contrast between theories is in terms of whether they link production together with perception (linked theories, e.g., the covert repair hypothesis) or they consider that the production system works independent of perception (autonomous theories, e.g., EXPLAN). It is argued that many features of stuttering can be explained in an autonomous production model in which the problem arises at the point where linguistic and motor processes interact.

Keywords

EXPLAN; covert repair hypothesis; spreading activation; altered auditory feedback

The issue of what goes wrong with the speech production skills of people who stutter (PWS) has been approached from a number of theoretical perspectives. One dimension along which to characterize these theories is in terms of whether production processes are considered to be *autonomous* from perceptual processes or whether there is an essential *link* between perception and production (see Howell, 1996; Howell & Harvey, 1983, for full descriptions of these alternatives). In autonomous models, no information from perception is used in production. Linking perception to production could, potentially, assist production. For instance, the perceptual system could be responsible for ascertaining whether the speaker has made an error. When an error is detected in production, speech could then be interrupted and a correction made. Production would not play a part in detecting any errors that are made. Though a scheme in which production is linked with perception benefits from this advantage, there are problems with such theories that do not arise in an autonomous model (detailed later in this article).

A second dimension on which theories can be characterized is in terms of the production processes that are functionally involved in the problems experienced by PWS. A broad distinction will be drawn between cognitive–linguistic processes (planning level) and motor levels involved in organizing plans for output (execution). Speech is not the only way that speech plans can be output; the plans can be written or represented by manual signs. This suggests that planning and execution have some measure of independence. The problem of stuttering could arise in the planning processes, in the motor processes, or in the way they operate together (often referred to as the language-speech interface).

Europe PMC Funders Author Manuscripts

Spring 2004 © NSSLHA

Contact author: Peter Howell, Department of Psychology, University College London, Gower St., London WC1E 6BT, England. E-mail: p.howell@ucl.ac.uk.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

It can now be seen that there are several approaches to the question of what goes wrong with production skills in PWS. Does the problem arise in an autonomous speech production system, or is the perceptual system involved in monitoring speech (a linked account of stuttering)? Irrespective of whether an autonomous or linked theory is favored, is the problem a reflection of problems in planning processes or execution processes or in the way the two interact? These issues are taken up in the remainder of this article. The review does not pretend to be comprehensive in the theories it covers, nor to be exhaustive about those theories that it does cover, as most attention will be given to the author's EXPLAN theory.

WHAT UNDERLYING PROBLEM OR PROBLEMS IN THE LANGUAGE GENERATION AND OUTPUT SYSTEMS COULD LEAD TO STUTTERING?

This examination of accounts of stuttering starts with the issue of where in the planning and execution processes the problem in speech control arises (planning, execution, or their interaction). Those who take the view that stuttering arises in the language processing system have tried to establish a deficit or deficits at one or more level/s within the overall planning system. Within the planning system, syntactic, lexical, and phonetic levels have been examined. A deficit at any of these levels can be established by seeing whether PWS perform less well on tests that examine the requisite level as compared with fluent controls.

The case has been made that there are some children with specific language impairment who have a syntactic, but no other, processing deficit (van der Lely & Stollwerck, 1997), and the same might apply to PWS. Syntactic performance in PWS could be examined by requiring participants to match a sentence with one of a number of pictures presented to them (as in Bishop's 1983 Test of the Reception of Grammar, TROG). A participant's competence on items that vary in syntactic complexity could then be used as an index to establish whether PWS have poorer syntactic abilities than controls. Children who stutter and matched controls have been compared on their performance in the Reception of Syntax Test (ROST). ROST has some similarities with the TROG test (Howell, Davis, & Au-Yeung, 2003), but it is simpler (requiring choice between two, rather than TROG's four, pictures). This allows it to be used with children younger than those tested with TROG. Howell et al. (2003) showed that performance on ROST in PWS is of a similar level of accuracy as that of controls, suggesting that PWS do not have a syntactic deficit. Nippold (1990, 2001) examined the wider literature on syntax and concluded that a syntactic deficit is unlikely to be a cause of the problem. A more moderate conclusion is suggested here: A syntactic deficit that affects this component in the language processing system alone is not likely to be the cause of stuttering. Phrasing the conclusion this way does not rule out syntax affecting performance when the linguistic and motor systems operate together (which would imply a problem in dealing with material that varies in syntactic complexity at the language–speech interface).

Studies have also been conducted to establish whether there is a lexical or phonetic deficit in stuttering. Some of these studies have examined whether there is any relationship between phonetic properties of material and the likelihood of that material being stuttered. Unlike syntactic tests that seek a deficit that arises when speech output is not required (a pure syntactic deficit), the influence of phonetic difficulty has been examined in these studies when speakers have to produce output. Therefore, these studies also involve execution and the language–speech interface. In the first modern study on whether stuttering is related to the phonetic complexity of material, Throneburg, Yairi, and Paden (1994) used a metric that characterized the difficulty of a word with respect to whether it contained consonant strings (CS), whether it had consonants that were acquired late in development (late-emerging consonants, LEC), and the length of words measured by number of syllables (MS). Using these measures, they indexed the difficulty of words that speakers produced and examined whether difficult words were stuttered at higher rates than easy ones. Their initial work with

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

this metric failed to find an association between difficulty and the likelihood of stuttering. However, this appears to be because the scoring metric was not designed to examine the parts of words that are particularly problematic for PWS (i.e., the onsets of words, see Wingate, 2002, for a review of this issue). Adaptation of the scheme to examine the same characteristics as Throneburg et al. (1994) for CS and LEC in initial word position has shown higher stuttering rates associated with these phonetic factors (Howell, Au-Yeung, & Sackin, 2000). The influences occur mainly on content rather than function words. (Function words are pronouns, articles, prepositions, conjunctions, and auxiliary verbs; content words are nouns, main verbs, adverbs, and adjectives.) The effects of phonetic difficulty are more apparent in older speakers who stutter than in younger speakers who stutter. The main finding is that phonetic difficulty affects stuttering rate when execution occurs concurrent with planning.

Motor timing of articulatory movements has been examined in PWS. Timing measures are sensitive and have provided evidence for motor deficits in PWS (Caruso, Abbs, & Gracco, 1988; Max, Caruso & Gracco, 2003; Smith & Kleinow, 2000; van Lieshout, Hulstijn, & Peters, 1996). Until recently, most studies have controlled linguistic demand by keeping the tasks simple by looking, for instance, at the details of motor timing responses for utterances like “sapapple” (Caruso et al., 1988) or a single syllable that has to be repeated at a regular rate (Howell, Au-Yeung, & Rustin, 1997). The results suggest a motor timing problem in PWS. In addition, the motor processing problem could be affected to a greater extent when linguistic complexity is increased (i.e., these findings do not rule out a role for planning–execution interactions). Indeed, Smith and Kleinow (2000) reported that motor response timing varies when linguistic complexity changes, suggesting just such an interaction. Studies such as the latter indicate (again) that the appropriate way of examining stuttering is to examine planning in interaction with motor output. The earlier studies on motor measures show that timing provides a sensitive measure of stuttering problems.

MODELS OF THE INTERACTION BETWEEN LINGUISTIC AND MOTOR SYSTEMS THAT HAVE BEEN APPLIED TO STUTTERING

Two models that have been developed to account for stuttering will be considered, both of which include linguistic and motor processes and their interaction. The models are representative versions where either (a) there is a link between perception and production (Kolk & Postma, 1997), or (b) the production process is modeled as autonomous from perception (Howell, 2002; Howell & Au-Yeung, 2002). The Kolk and Postma model is outlined only briefly, mainly to provide a contrasting view to the autonomous EXPLAN account.

The Linked Covert Repair Hypothesis Account of Stuttering

Kolk and Postma's (1997) account (called the covert repair hypothesis, CRH) approaches the issue of stuttering primarily from a psycholinguistic perspective. It was developed from Levelt's (1989) well-known model of speech breakdown in fluent speech control, in which speech errors are regarded as providing important evidence about speech control. A diagram of Levelt's model is presented as Figure 1.

The main features to note about the Levelt model follow:

- There is a hierarchical linguistic system, and errors can be made at different points in the hierarchy. Speech errors like “cuff” in “cuff of coffee” (Fromkin, 1971) are indisputably errors. However, Levelt (1983) regards repair events as results of errors in the psycholinguistic system. For instance, a speaker giving directions on how to go from one place to another might give an erroneous set of directions.

Europe PMC Funders Author Manuscripts

Assigning these a status of linguistic errors is disputable as they could stem from poor spatial, rather than linguistic, ability. To ensure that linguistic errors are involved, errors are defined here as a phoneme in the wrong position to complete an intended word. According to this view, Fromkin's “cuff” and the words “hissed” and “mistory” in the classic spoonerism “missed my history lesson” are all errors. “Turn left, no, turn right” in a repair and “m.missed” and “hhhistory,” on the other hand, are not errors. A further point to note for later is that errors according to this strict definition are infrequent in speech control; Garnham, Shillcock, Brown, Mill, and Cutler (1981) estimate that these occur in only approximately .01% of words.

- Errors are fed to a monitoring system via two routes—the external, and internal, loops. Both loops send information to the perception system, which decodes the speech. The decoded speech is transmitted to the monitor, which compares the intended message with that which perception indicates has actually been made. If the two correspond, speech has been produced as intended and output can continue. If there is a discrepancy, an error has occurred, speech is interrupted, and the correct message is reinitiated. Evidence from speech repairs has been used as support for both loops: Overt speech repairs such as “to the left of, no, to the right of the curtain” provide support for the external loop. The speaker has produced the word “left” when “right” was intended. The Levelt (1983) account assumed that the speaker hears this, interrupts speech (signified by the comma), and repeats some components that are not essential (“to the,” called a retrace) before making the correction. Using “left” instead of “right” may not be the result of an error in the linguistic system (for the same reason that such “errors” were excluded in the above definition, and, consequently, would have little relevance to the workings of psycholinguistic processes. Events that are taken as evidence for the internal loop are retraces like “to the, to the right of the curtain.” There is no overt error in these repairs; they are called covert repairs. It is possible that they are the results of internal monitoring where an error has been detected, interrupted, and corrected before the error is output.

- Some specification is given about the site of the language–speech interface in this model. A phonetic string is supplied by the linguistic system that is then output through the motor system. Subsequently, this output is detected by the peripheral auditory system and eventually transmitted to the monitor (that detects and interrupts production when an error occurs over the external loop).

Kolk and Postma's (1997) CRH was originally developed from Levelt's account to explain the frequent occurrence of pauses (interruptions) and word repetition (retraces) in stuttered speech. In Levelt's model, these events were seen as a result of errors detected over the internal loop that were repaired covertly. The early version of the CRH account did not explain why PWS should show a higher rate of covert repairs (and, by implication, a higher error rate in linguistic processing) than speakers who do not stutter. This was partly rectified by Kolk and Postma. They used Dell and O'Seaghdha's (1991) spreading activation model to explain how a slow phonological (or, more precisely, phonetic) system leads PWS to make errors (in particular, errors that lead to covert repairs). According to Kolk and Postma, when a speaker intends to say the word “cat” (the target unit), phonetically related competing units are also activated (e.g., “rat”). Dell and O'Seaghdha's model has steps involving lexical activation and phonetic encoding. Kolk and Postma focus on how activation patterns could lead to phonetic errors after lexical selection has taken place. The buildup of activation for the target and competing units follows similar trajectories in early epochs, but later they asymptote at different levels, as shown in Figure 2. The trajectories for target units asymptote at a higher level than do those for competing units. At asymptote, the higher activation level of the target unit always indicates what the appropriate word response was.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Operating under time pressure (such as when speech has to be produced rapidly) requires a speaker to generate words in the period where activation is still building up, for example, at points near "*S*-" in Figure 2. The word response at this point would still be the one with highest activation. However, as the target and competing options have similar activation trajectories during buildup, noise in the activation process can lead to the competing options having highest activation and be triggered (resulting in a speech error) if word selection is made in this time region. Kolk and Postma propose that PWS have slow phonetic systems. The result is, effectively, that the amount of time in the buildup phase is extended. A word response generated at the same time as that of a fluent speaker will be "early" for these speakers and lead to a heightened chance of speech error arising for the same reason as with speech produced by fluent speakers under time pressure. As speech output is made early, responses will mainly involve the internal loop, which explains why the proportion of overt errors is not high in these speakers (Melnick & Conture, 2000; Wolk, Edwards, & Conture, 1993). There will, however, be evidence that a repair is being made based on the occurrence of retraces and pauses.

Thus, the CRH explains whole word repetition and hesitation as due to errors that are detected over the internal loop. Using the terms presented in the introduction, the CRH can be characterized as a linked theory where errors occur in the linguistic system. The second theory, like the CRH, also addresses why whole word repetition and hesitation occur, but it does not include the questionable perceptual monitoring component (operating over either the internal or the external loop), nor does it assume that such repetitions and hesitations are the result of errors made during production. In this model, the principal site of fluency problems (for speakers in general, not just speakers who stutter) is the language–speech interface. The description begins, after some general background, with the account of word repetitions and hesitations, and then the application of the theory to other known characteristics of stuttering is given.

Introduction to the Autonomous EXPLAN Account of Stuttering

Characterization of the EXPLAN model—EXPLAN is an autonomous model of the production of spontaneous speech that applies to speakers who stutter and fluent speakers. Planning (PLAN) and execution (EX) are independent processes that reflect the linguistic and motor levels, respectively. Failures in the normal mode of interaction between the PLAN and EX processes can lead to fluency failures when plans are too late in being supplied to the motor system. Generically, "fluency failure" arises whenever there is this underlying problem but, according to EXPLAN, there are two distinct types of responses available to the speaker in these circumstances. These are (a) whole function word repetition and hesitation (referred to here as *disfluency*), and (b) problems on parts of content words (a feature mainly seen in persistent *stuttering*). The studies up to this point do not make this distinction between the two types of response (e.g., both types of response have been included in stuttering rate estimates in some of the studies reviewed above). In the remainder of this article, the terms disfluency and stuttering are used in the specific senses given here. The origin of these two types of response following fluency failure is outlined in detail below.

Planning and execution definitions—The planning process supplies output that the speech execution system uses. The time to generate a plan is the factor that determines whether the linguistic system operates fluently or not. Errors in planning can arise (and have to be accounted for), but they are rare in spontaneous speech (Garnham et al., 1981). The time to generate the plan is determined by the time taken in individual linguistic processes (such as those in a hierarchical system like Levelt's, 1989). The role of execution in EXPLAN is to organize the plan for output. Organization of output is fluent when planning

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

and execution operate so that the next segment of speech is ready when it is required for execution (that is, when the execution of the preceding segment has been completed). The emphasis on organization and timing of output suggests that the cerebellum has an important role in stuttering (see Stein & Glickstein, 1992, for a discussion of the view that the cerebellum plays the role of organizing motor plans for output).

Overview of operation of execution and planning processes during fluent speech and fluency failure—The independence of the planning and execution systems allows the plan for a future segment to be generated during the time the plan for the current segment is being executed. Here we use words as the segments, but the ideas apply to other segments. If execution time is long enough, the plan for the following word will be ready after execution of the present word has been completed, and is executed in sequence. A schematic representation of the process is provided in Figure 3, with time represented along the *x* axis. Planning is indicated for each word in turn in the top row, where the length of the line gives time taken to generate the plan of each successive word. Execution is indicated in the same manner below. Note that the execution of each word is offset relative to production (production cannot start until at least part of its plan is ready), and the plans of words are dealt with one at a time (series). Figure 3 shows the particular situation where planning of one word is completed during the time the preceding word is being executed, thus leading to fluent speech.

In secondary sources, the studies by Sternberg, Monsell, Knoll, and Wright (1978) are often reported as showing rapid planning, whereas EXPLAN requires planning to be of the same order as, or longer than, execution. Apart from the differences in paradigms that make comparison between Sternberg et al.'s work and EXPLAN difficult, the former authors were careful to indicate that their latency measures apply to “subprogram retrieval” rather than planning (see section VII-B of Sternberg et al., 1978). Also, the assumption about parallel processing during planning is not in conflict with Sternberg et al.'s claim that the retrieval process is serial.

Figure 3 represents the final planning stage that generates a word representation for the execution processes. Note, however, that any problem at prosodic, lexical, or other levels that increases planning time on the previous, present, or subsequent words will affect the time to generate the final output plan of the current word, provided that such planning happens before execution of the current word has been completed. On this assumption, anticipatory coarticulation that results at the planning level (Howell & Harvey, 1983) will be completed before phonetic output is generated. So, even when there is interruption of speech because the phonetic plan for a word takes a long time to generate, coarticulation will be appropriate. (See Howell & Vause, 1986, for a report that coarticulation of the consonant is appropriate for the subsequent vowel when there is a break on the initial consonant in a CV syllable).

The plan can be prepared for words in advance of the one currently being executed and speech will then be fluent. Problems only arise in the limiting case where there are no words that have been planned in advance so the next word is being planned as the current word is being produced (the situation shown in Figure 3). The latter situation is a critical conjunction point between planning and execution (conjunction point for short). These conjunction points are most likely to arise in spontaneous, rather than read, speech (part of the plan is effectively supplied by the text for read material, so conjunction points are less likely to occur).

Fluency fails when a speaker has finished execution of one plan and the next one is not ready for execution. There are two main reasons why this arises: (a) The inherent properties

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

of linguistic segments make their planning slow (difficulty), and (b) speech is executed at a high rate. Separating the two influences implies that they can occur independently of each other and, in particular, that speech rate can be adjusted independently of planning (see Guenther, 2001, who suggests that speech rate can be modulated voluntarily at the execution level, which is compatible with this assumption). The role of each of the factors can be appreciated by examining Figure 3. Difficulty increases the planning time of the following word beyond the time needed to execute the current word. Increasing execution rate has the effect of shortening the planning time allowed for the following word.

DETAILED ANALYSIS OF THE EXPLAN MODEL

EXPLAN is described in the following three parts of this article; the parts address planning, execution, and their interaction, respectively.

Part One: Planning

Overview—This part will first outline the EXPLAN account of word and phrase repetition around fluency failures (a topic addressed by CRH). The EXPLAN account of stuttering on parts of content words will then be presented.

EXPLAN account of word and phrase repetition around fluency failures—The phonetic properties that make words difficult for PWS (particularly adults) to produce were described (CS, LEC, and MS factors) earlier. Word initiation latency in picture naming tasks is a measure of planning time. Two of the factors that Throneburg et al. (1994) investigated are known to increase initiation latencies in fluent speakers. Thus, Santiago MacKay, Palma, and Rho (2000) found that words that began with a CS took longer to initiate than those that began with a singleton consonant. They also found that bi-syllabic words took longer to initiate than monosyllabic words. When a word contains one or both of these factors, planning time lengthens and there is an increased chance of fluency failure due to the plan not being ready in time for execution. Thus, according to EXPLAN, these two factors associated with phonetic difficulty should increase the likelihood of fluency failure (as reported by Howell, Au-Yeung, & Sackin, 2000). Initiation latency for words with the LEC factor has not been examined to date, but there is no compelling reason to suppose that this would not operate in a similar manner to CS and MS (i.e., that time would be longer in a word that contains an LEC).

It follows from the EXPLAN account that speakers need more time to prepare the next word when fluency fails. To overcome this problem, a speaker has to gain more time before attempting, or while attempting, a difficult word. One possible way of getting the extra time is to repeat the word before the one that is incomplete (that has already been planned and has just been executed). This assumes that the speaker still has the plan for this word available (see Blackmer & Mitton, 1991, for data that support this assumption). Related ways of gaining time would be the repetition of more than one word, or hesitation (using filled or unfilled pauses).

The EXPLAN account includes an explanation for two further factors that happen, relevant to a speaker gaining time by word repetition or hesitation. The first is that these disfluencies in children occur mainly on function words (Bloodstein & Gantwerk, 1967; Bloodstein & Grossman, 1981). The second is that such disfluencies involve whole words in children (Conture, 1990).

As background to understanding the EXPLAN account of these two phenomena, note that certain sequences of words increase the chance of fluency failure due to the plan not being ready in time. If there is a sequence of simple words followed by a complex one, planning

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

and execution of the simple words will both be rapid. At conjunction points, during the short time that the last simple word is being executed, the complex word that follows needs to be planned. The short execution time only allows a short planning time, but the complex word requires a long planning time. There will thus be an increased chance of fluency failure at this point.

One factor to consider when assessing whether or not a word is easy to plan is whether it is content or function in type. There are other ways of distinguishing easy and difficult words, for example, in terms of CS/LEC and MS properties, word frequency, or whether a word carries stress or not. One advantage of using content and function words is that these word classes are associated with fluency failure (disfluency and stuttering) at different ages (function words in young speakers, Bloodstein & Gantwerk, 1967; Bloodstein & Grossman, 1981, but content words in adults, Howell, Au-Yeung, & Sackin, 1999). A second advantage, for English at least, is that many other possible factors that affect difficulty correlate with these word classes (such as those listed at the beginning of this paragraph). Thus, using function/content words to specify difficulty engages other factors that could affect fluency. Subsequent studies should then estimate the effect of individual factors. A sequence of potentially easy words is likely to be function in type, and the difficult word that terminates this sequence is likely to be a content word. The speaker needs to gain extra time for planning the content word at conjunction points and can do this by pausing or repeating one or more prior words. Word repetition and hesitation should occur on the easy words before the content word (i.e., involving words at the point where an easy word is being executed and a word that takes a long time to plan is being prepared). Thus, the type of words that EXPLAN predicts should be repeated are function words, as observed to occur in children who stutter. As the simple words that are repeated reuse a previously completed plan, repetition of function words will involve the whole of the words.

Besides accounting for these previously known features observed in children who stutter, EXPLAN also makes the prediction that function words that follow a difficult word are unlikely to be repeated as these words could not gain time to prepare a preceding word (Au-Yeung, Howell, & Pilgrim, 1998). To test this prediction, Au-Yeung et al. developed the notion of phonological words (PWs) as a way of grouping words together. PWs consist of an obligatory content word as nucleus and an arbitrary number of function words preceding and following the content word (see Selkirk, 1984, for some background justification). The function words are associated with their content word by a set of sense unit rules (Au-Yeung et al., 1998). “I split it” is a PW consisting of a single initial function word, a content word, and a single final function word. Au-Yeung et al. confirmed that function word repetition occurs much more frequently when the function word precedes the content word than when it follows the content word (i.e., in this case, “I, I split it” is usual, but “I split it, it” is not).

This pattern of function word repetition is also observed in fluent speakers of all ages (Howell, Au-Yeung, & Sackin, 1999). The proposal that repetition serves the role of delaying the time at which the following word is produced has been made by several authors working on fluent speech (e.g., Blackmer & Mitton, 1991; Clark & Clark, 1977; Maclay & Osgood, 1959; MacWhinney & Osser, 1977). However, these accounts have not linked them to function words nor have they examined how word repetition depends on the position they occupy in PW contexts. PW contexts (as will be seen in the next section) also offer a different perspective for examining why patterns associated with fluency failure change in adult speakers who stutter, as well as providing a potential explanation for these changes.

To summarize, EXPLAN accounts for word repetition and hesitation and predicts why the easier function words are repeated by fluent speakers and young speakers who stutter. It also makes confirmed predictions about which function words are repeated (initial ones in a

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

PW). Finally, it is worth making two points about these repetition patterns explicit. First, according to EXPLAN, though repetition occurs on the function words, the locus of the fluency failure is actually the content word whose plan is not ready. Second, the explanation offered is solely in terms of plans not being ready in time for execution (not a result of a phonetic error the speaker makes).

EXPLAN account of the change to problems on parts of content words as speakers who stutter get older—Two changes occur when speakers who stutter get older: First, problems occur more on content words (Howell, Au-Yeung, & Sackin, 1999). Second, the problem is manifest as a change from whole words to parts of words (usually the first part) (Conture, 1990). If the interpretation of function word repetition at conjunction points is correct, then the older speakers who stutter have not used function word repetition or pausing to delay their attempt at the content word. In turn, this would mean that the content word is attempted before its plan is complete. Premature initiation could explain why problems on the first part of a content word arise, if the additional assumption is made that the plan for a word is built up left to right. Then, only the representation of the initial part of the content word would be ready when the speaker starts the utterance, and the rest would still need to be generated. When the later plan is not available, the initial section will lead to patterns of fluency failure that use up time until the remainder is generated. Thus, initial sounds may be prolonged (e.g., “ssssplit”), repeated (e.g., “s.s.split”), or a word break introduced (e.g., “s-plit”) at the start of the content word. Note, again, that these are examples of timing changes to speech, not errors. In these cases, when there is no function word repetition, fluency fails on the content word. Conversely, when there is function word repetition, this prevents problems on the content word. Thus, in any PW, either whole (disfluent) or part-word (stuttering) patterns of fluency failure will occur, not both.

The study by Howell, Au-Yeung, and Sackin (1999) was the first study that used PWs as a contextual unit to test whether stuttering on content words arises when speakers stop repeating function words. Their hypothesis predicted that a decrease in disfluency rate on function words would be mirrored by increased stuttering on subsequent content words within the PW if function word disfluency prevents content word stuttering. To test this prediction, PWs were selected that had initial function words. In these PWs, there is the opportunity for establishing when delaying occurs on the initial function words and when speakers have advanced to the content word. Howell, Au-Yeung, and Sackin confirmed that older speakers were disfluent less often on function words, but more frequently on content words, than younger speakers. They referred to the changing functional relationship between responses to fluency failure on function and content word over age groups as an exchange relation. According to this account, although the locus of the problem is the same over age groups, the way the problem presents changes with age in PWS. Since this initial report, exchange patterns have been reported for Spanish (Au-Yeung, Vallejo Gomez, & Howell, 2003) and for German (Dworzynski, Howell, Au-Yeung, & Rommel, in press). Fluent speakers mainly show the disfluent pattern over ages. Thus, it is only older speakers who stutter who depart from the disfluent pattern.

In summary, adults stop function word repetition at conjunction points. This requires them to attempt content words before they are fully prepared. In turn, this leads to two changes in the pattern of fluency failure over development: from whole to parts of words, and the locus changing from function, to content, words. Fluent speakers do not change the response they make to fluency failures. The latter observation may suggest that the content word response pattern is not conducive to developing fluent speech (otherwise, fluent speakers would not avoid it). Another implication of this position is that the pattern shown by children who stutter is more like that of fluent speakers, and these speakers have a good chance of recovery. This leads to the practical suggestion that, because function word repetition does

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

not seem to be problematic, maybe what should be done is to monitor whether and when the disfluent response pattern ceases to be used, not necessarily to admit the child immediately into a treatment program. The pattern change that pathologists need to be vigilant about is toward the adult pattern (part content word stutterings) if their occurrence makes the acquisition of fluent speech hard to achieve. The account offered does not imply that the child who stutters makes more errors in speech than fluent speakers, as does CRH. Also, no monitor is needed as neither type of fluency failure is a result of an error. The high frequency of function word repetitions in children who stutter (Howell, Au-Yeung, & Sackin, 1999) allows that adults who stutter may have had some underlying processing deficit in childhood (see the discussion for some reasons why this could be so). However, whether speakers who persist in stuttering had high rates of function word disfluency in childhood or not, the main characteristic that defines the underlying problem as they advance into adulthood is the change in the pattern of response to fluency failure.

Part Two: Execution

Overview—The timing of speech output can be changed by execution processes. Some timing variability will arise as a result of the relationship between the intrinsic timing of planning and execution processes (intrinsic timing), and this is not directly controlled by execution. It also seems necessary to postulate an external rate control mechanism that operates independently of planning processes (to deal with situations such as when execution is speeded up or slowed down). According to EXPLAN, execution is also disrupted when alterations are made to auditory feedback (AAF). Alterations that affect speech control include delaying the sound of a speaker's voice before it is heard (DAF) and shifting the frequency content of speech (FSF). These manipulations have marked effects on vocal control in fluent speakers and speakers who stutter. Traditionally, the effects of AAF have been explained on the basis that the alteration disrupts a process that monitors planning that, in turn, leads to speech control problems (and Levelt, 1989, uses this evidence as another line of support for his external loop). The EXPLAN proposal is that AAF disrupts execution, so it places the locus of these effects at more peripheral regions of the CNS.

Some data are reported that support the view that there is a distinction between local and global influences on rate control, which reflect intrinsic (involving planning and execution) and external (just execution) timing processes, respectively. Then some further topics concerning the execution process in EXPLAN are considered. These are (a) a proposal of how global rate is controlled during execution, (b) the implications of this view for whether AAF is used for perceptual monitoring, (c) the effects of AAF on fluent speakers, (d) the effects of AAF on PWS, and (e) the explanation is applied to account for why other techniques that operate at the execution level improve fluency.

Distinction between global and local influences on rate control—Generally speaking, if speech is slowed, there is less likelihood of planning getting ahead of execution. Such planning adjustments are only needed at the points where difficulty is high (local). Global changes appear to be necessary when speakers need to make a long-term adjustment to rate (as, for instance, when a speaker continues to make part-word stutterings).

Evidence that rate control operates at a local level in utterances (local rate change) has been obtained by dividing a sample of speech into sections according to their execution rate. Thus, Howell, Au-Yeung, and Pilgrim (1999) segmented the speech of adult speakers who stutter into tone units and separated those that were stuttered from those spoken fluently. Rate was measured in the fluent stretches in the tone units with and without stuttering in the section before the stuttering (the whole segment in the case of fluent tone units). The segments were divided into three rate categories based on the rate in the fluent section (fast,

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

medium, and slow), and stuttering rate was examined. The local segments that were spoken slowly led to a lower rate of stuttering than those that were spoken more rapidly. Assuming that phonetic difficulty is distributed evenly over the tone units divided according to rate, the Howell, Au-Yeung, and Pilgrim findings would support the idea that planning is taxed when speech rate is high in a local region of speech. Further work is needed to establish whether there is an interaction between local planning level effects and execution rate of the material. For instance, is material that is spoken fast and has content words with high phonetic difficulty more likely to be stuttered than material that is spoken fast that has content words with low phonetic difficulty?

Howell and Sackin (2000) addressed the issue of whether local rate change can occur independently of global rate change. They investigated timing variability in a simple utterance that was repeated several times by fluent speakers (“Cathy took some cocoa to the teletubbies”). Using the same utterance keeps planning and contextual influences constant. They had subjects do this in FSF and normal listening conditions, and when speaking and singing. They marked the plosives in the utterance and measured the duration of the intervals between the first and each of the subsequent plosives. The frequency distribution (for each interval type) that a speaker produced was then obtained. Global slowing between speaking conditions would be reflected in a shift in central tendency of the overall distribution to longer durations. Local slowing between conditions occurs when the intervals at the rapid (short duration) end of the distribution shift to longer durations, but there is no shift in the overall mean between conditions.

Looking at the means of the distributions first, comparisons were made between normal speech in ordinary and FSF listening, sung speech in ordinary and FSF listening, speech versus singing in ordinary listening, and speech versus singing in FSF listening. All revealed significant differences between overall means (global slowing) except speech versus singing in normal listening.

Though singing does not induce global slowing in the normal listening condition, does it lead to local slowing (as required if local changes are independent of global changes)? To test this, Howell and Sackin (2000) estimated the time at the point in the distribution where the 25th percentile occurred as a measure of whether the fast intervals shifted across the different conditions. When the same comparisons were made as with the means, all conditions showed a significant shift upwards of this percentile. This is not surprising when global shifts occur, as it may simply indicate a shift of the overall distribution. Of particular note, however, is the local slowing that occurred when speaking was compared with singing in a normal listening environment. This indicates significant local slowing where there was no global slowing, suggesting that these are two distinct modes of making a rate change. Although these results do not establish whether the local timing process is involved with the intrinsic planning–execution timing chain, or whether the global process is associated with an external timekeeper, they are consistent with this view, and further work will be conducted to test these notions in more detail. Thus, the tentative conclusion at this time is that timing variation under normal listening conditions in sung speech arises in the intrinsic relationship between planning and execution, whereas all the rest (all of which involve alteration to listening conditions) have associated global slowing that may reflect the operation of an external timekeeper.

EXPLAN proposal concerning global rate control at the execution level— EXPLAN maintains that global rate changes are made by an external timekeeping process (Wing & Kristofferson, 1973) that is located in the cerebellum (Ivry, 1997), and this timekeeper is disrupted (revealed by higher timing variance) when it receives asynchronous inputs. The timekeeper is also affected when the number of its inputs increases. The case is

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

made that DAF represents an example of this. Evidence is also presented that shows that DAF specifically affects the timekeeper (using a version of the Wing and Kristofferson task). The work with the Wing-Kristofferson task suggests a cerebellar location for the timekeeper, and other evidence that supports this position is presented.

Properties of the timekeeper: The timekeeper marks the rate of different events that are associated with speech tasks (or any other task, as the mechanism is general purpose). If the timekeeper cannot mark these times (because the events occur too quickly; are too numerous; or are not in a simple, synchronous, relationship to each other), the activities it regulates (motor actions) are slowed. Slowing could be achieved by reducing the rate of pacing signals supplied to the timekeeper via the basal ganglia that then attract the responses to this lower modulation rate. The alteration to a newly established rate represents a global change.

Considering the effects of global slowing on the timekeeper itself, the pacing signal and responses occur at a lower rate and responses are brought into synchrony with the pacing signal. Both of these changes would decrease load on the timekeeper and enable the timekeeper to mark time appropriately again.

Next, the effects on speech output of these changes to the timekeeper are considered. There appear to be two effects. First, speech rate will be slowed (see the next section, where it is argued that this is the sole effect in fluent speech). Second, when speech is stuttered (see the section, “effects of DAF and FSF on speakers who stutter”), slowing responses decreases the chance of planning getting ahead of execution (i.e., reduces the chance of fluency failure). These changes occur so the timekeeper can reduce its load. The load reduction brings the timekeeper's inputs back within its capacity and it is then able to mark the time of the inputs. The effect that a reduced response rate has on fluency is an indirect by-product of these changes.

What evidence is there that asynchrony increases timekeeper load: It seems reasonable to suppose that increasing the number or rate of inputs to a timekeeper would increase load, but there is no a priori reason to suppose that asynchrony between events increases load. Empirical observations suggest, however, that synchronous events are easy to perform concurrently, but that asynchronous events are not so easy. For instance, children are able to learn to sing canons in synchrony with other children, but find singing offbeat with other children (asynchronous) more difficult. Both types of singing have temporal (sound) inputs at the same rate; the difference is whether they are synchronous (canon) or not (offbeat).

Evidence that DAF affects the execution, not the planning, level: EXPLAN makes the radical proposal that DAF has its effect because it alters the inputs to the timekeeper, not to higher cognitive levels (as with a perceptual monitor). The perceptual monitoring account predicts that manipulations that destroy speech content should make the signal unusable for detecting errors and, by extension, remove the effects of DAF that are taken as support for this process. If DAF affects control because it creates disruptive rhythmic input to a timekeeper, noises with the same temporal structure as speech should disrupt the timekeeper as much as the speech itself. To test between these alternatives, Howell and Archer (1984) transformed speech into a noise that had the same temporal structure as speech, but none of the phonetic content. They then delayed the noise sound and compared performance of this with performance under standard DAF. The two conditions produced equivalent disruption over a range of delays. This suggests that the DAF signal does not need to be a speech sound to affect control in the way observed under DAF, and indicates that speech does not go through the speech comprehension system and then to a monitor. The disruption could,

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

however, arise if asynchronous inputs affect operation of the timekeeper in the execution system.

Evidence that DAF affects external timekeeper variability, not motor variability: In the previous section, it was shown that noise creates similar effects to speech under DAF conditions. Though these results rule out DAF affecting planning, they do not necessarily support the view that DAF affects the timekeeper. A modification of the Wing and Kristofferson (1973) task was developed to see whether DAF affects the timekeeper selectively (Howell & Sackin, 2002). Subjects in the standard Wing and Kristofferson procedure are required to produce a series of taps at a specified rate as accurately timed as possible. The modification here was that speech responses (the syllable /bae/) were used instead of a tap. With either form of the task, the timing variability in the sequence of responses can be decomposed into variance of motor processes (Mv) and variance of a timekeeper (Cv). The essence of the analysis procedure is that if the motor system leads to a tap being placed at the wrong point in time, it is compensated for in the next interval. Thus, Mv can be estimated from the lag one autocovariance. Taking Mv's contribution away from the total variance then leaves an estimate of Cv.

Howell and Sackin (2002) had subjects perform this task in conditions where they heard concurrent DAF. DAF led to a marked increase in Cv, the Cv increase being greater for longer DAF delays and more marked when the syllable had to be repeated at longer periods. Mv, on the other hand, stayed roughly constant across the DAF delay and repetition period. The fact that the timekeeping process rather than the motor processes are affected by DAF delay supports the idea (inherent in EXPLAN) that DAF has a direct influence on the timekeeper and has little effect on the motor processes. Taking this section and the previous one together, the results show that DAF is a technique that produces asynchronous input to the timekeeper, which then disrupts external timing control.

Cerebellar control and support from the Wing and Kristofferson task: The original Wing and Kristofferson (1973) task has been used with patients who have a lesion in the cerebellum to see whether (and if so, which) regions of the cerebellum are associated with Mv and which with Cv (Ivry, 1997). Lateral lesions of the cerebellum affect timing control, suggesting that the timekeeper mechanism is located in this part of the CNS (Ivry, 1997). The medial areas of the cerebellum appear to be involved with Mv, as lesions to this part affect this variance component. Generally speaking, the location of an area in the cerebellum responsible for Cv supports the idea, made in EXPLAN, that AAF disrupts a mechanism involved in organizing speech output for execution rather than higher level cognitive planning mechanisms.

There are data on the speech version of the Wing and Kristofferson task that indicate that children who stutter have problems in the Mv component (Howell, Au-Yeung, & Rustin, 1997). A recent study has failed to find any such deficits in adults (Max & Yudman, 2003). There are a number of procedural difficulties that could explain this failure (Howell, in press), or it could be that timing control differs between adults and children who stutter. Note also that there is other evidence for an execution–organizational problem in the cerebellum of adults who stutter. For instance, imaging studies that scan down to this level of the CNS invariably find differences between adults who stutter and controls (e.g., Fox et al., 1996).

In this section, the circumstances under which a timekeeper makes a global rate change have been presented. Evidence that the timekeeper is located in the cerebellum has been provided. Data have been reviewed that show that DAF affects the timekeeper and brings about a global rate change. The low level in the CNS at which DAF operates undermines the support

Europe PMC Funders Author Manuscripts

that procedures like DAF provide for feedback monitoring between speech output and planning. The effects of DAF are due to the changes they induce in speech timing, which has an indirect effect on fluency when more planning time is allowed.

Is auditory feedback used for perceptual monitoring?—In this section, other arguments against perceptual monitoring (in particular, arguments that feedback about executed output is transmitted via the perception system to the planning processes) are considered, as well as how EXPLAN answers each of them.

Howell (2002) argued that the amount of phonetic information a speaker can recover about vocal output is limited because bone-conducted sound masks a speaker's phonetic output (see Howell & Powell, 1984, for a study on this issue). This would limit the usefulness of the feedback that a speaker can recover by listening to his or her own voice, making it an unlikely source of information for use for feedback control. Other problems for the view that speakers use the external (auditory feedback) loop (issues first raised by Borden, 1979) for monitoring concern how quickly information can be recovered from the auditory signal and why speakers with hearing impairment, who have established language before the loss, can continue to speak. The former suggests that feedback would be too slow to use in feedback monitoring, and the latter that speech can proceed without this information (be under open loop control).

EXPLAN was developed specifically so speech control works when the speaker cannot retrieve phonetic information from auditory feedback (using the proposal that auditory information only affects low-level timekeeping mechanisms). Thus, veridical feedback is not required. The only processing that auditory input needs to undergo in the EXPLAN account is to a level where its synchrony with other events can be determined. Thus, processing time is not likely to be as long as when phonetic information has to be recovered. The timekeeping process in EXPLAN easily lends itself to an explanation of why people who adventitiously lose their hearing do not lose their ability to speak fluently, as auditory feedback does not have an essential role in ensuring that speech is controlled accurately.

Arguments that DAF affects the timekeeper rather than a monitoring process were given in the previous section (Howell & Archer, 1984; Howell & Sackin, 2002). Another argument against DAF supporting feedback control of the voice is based on whether this alteration gives rise to a Fletcher, or a Lombard, effect. A Fletcher effect occurs when speech level is altered (the speaker raises voice level when speech level is experimentally decreased and decreases voice level when speech level is experimentally increased). The opposite happens when noise level is changed (Lombard effect), where speakers raise voice level when noise level is increased and decrease voice level when noise level is decreased (Lane & Tranel, 1971). If the delayed speech under DAF is treated like a person's own speech after processing by the speech comprehension system, it should lead to a Fletcher, rather than a Lombard, effect. However, Howell (1990) reported that amplifying the delayed sound during DAF produces a Lombard effect, again suggesting that the delayed sound is treated as a noise rather than speech. Thus, DAF led to speakers responding to the sounds as nonspeech noises, suggesting that speech is not processed through to a full perceptual representation. To summarize, the evidence on the effects of DAF overall do not provide unequivocal support for perceptual monitoring.

Effects of DAF and FSF on fluent speakers—The effects of DAF and FSF alterations on fluent speakers could arise because the extra inputs under DAF and FSF lead to global slowing by increasing load on the timekeeper. Howell (2001) described how FSF could lead to local slowing in the same mechanism. However, as Howell and Sackin (2000) (see above) have found evidence for global slowing under this auditory perturbation, it is possible that

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

the timekeeper only needs to make global rate changes. This would allow global slowing to arise when auditory feedback is altered and leave local slowing as a feature associated with operation of the planning–execution chain.

Effects of DAF and FSF on speakers who stutter—EXPLAN explains the problems in speech control experienced by adults who stutter in normal listening conditions by proposing that these speakers are trying to execute speech at too rapid a rate. As rate is slowed under either form of altered feedback, these procedures allow the speakers more planning time, thus reducing the chance of stuttering. Consistent with this, FSF (Howell, El-Yaniv, & Powell, 1987) and DAF (Ryan & van Kirk Ryan, 1995), both of which affect speech rate, have been reported to improve control of the speech of speakers who stutter. It should also be noted that the effects of DAF or FSF are immediate, and very effective, ways of controlling speech rate while the alterations occur.

Accounts for why some other techniques that operate at the execution level improve fluency—EXPLAN predicts that all serial signals will be input to the timekeeper and lead to a globally slower speech rate that should improve the fluency of PWS. One example of this is interruption to vowels where speakers hear the voice with no delay but where the first part of each sound is omitted, creating the effect of a delayed, or asynchronous, onset (Howell, Powell, & Khan, 1983). Kalinowski, Dayalu, Stuart, Rastatter, and Rami (2000) reported that this manipulation improves speech control in PWS. The bizarre effect that the speech of PWS improves when they see a concurrent flashing light (Kuniszyk-Jozkowiak, Smolka, & Adamczyk, 1996) may also have its effect because the flashing light inputs to the timekeeper and induces a global rate change in the speakers that, in turn, leads to improved speech control.

Part Three: Interaction Between Planning and Execution (the Language–Speech Interface)

Overview—The goal of this section is to look at how planning and execution work together in the EXPLAN account of stuttering. The first issue addressed is to phrase the EXPLAN view about execution and planning in a spreading activation framework and show how the activation profiles could signal (a) lexical errors (on the rare occasions these occur), (b) function word repetition and hesitation, and (c) part-word stuttering. This is achieved without routing internal or external speech representations through a perceptual monitor. The second issue that is considered is how part-word stuttering can be signalled to the timekeeper so that it can make a global rate change. In the third section, the issue addressed is how the planning–execution cycle can be dissociated from the timekeeper during fluent speech, while at the same time, the timekeeper can be automatically coupled when a rate change is needed. In the fourth section, reasons why adult speakers might lose the ability to respond to indications that a speech rate change is needed are explored.

Spreading activation version of EXPLAN—Figure 2 shows Kolk and Postma's way of representing how activation profiles build up for two alternative words that are candidates for occupying a slot in an utterance. As discussed earlier, Kolk and Postma (1997) proposed that PWS respond at an early point in the buildup phase (*S*-), and this increases the chance that the speaker will make an error. The perceptual system establishes whether such errors have occurred (as described earlier). This proposal effectively discards (or ignores) information once word selection has been made. However, it is apparent from Figure 2 that if activation for both words had continued to build up beyond the early point at which the response was made (*S*-), the target word would have had higher activation than the competing word. That is, if buildup of activation was allowed to continue beyond *S*-, the speaker would have known automatically (without calling on the perception system) whether the correct word was selected, based on which word candidate had highest activation.

To apply the latter idea to EXPLAN, the buildup of activation over time can be equated with planning time (which is truncated at *S*- in CRH). EXPLAN allows responses to be initiated before planning is complete (before full activation). However, in the EXPLAN account, planning continues (implying that activation would continue to build up) after word initiation. As indicated in the preceding paragraph, this alone provides an account of how speakers are aware that an error has been made autonomously within the production system. If the plan is complete at the time the word is initiated (activation is at maximum), the plan will then decay (activation level will decrease) during the time it takes for a word to be executed. In the remainder of this section, a spreading activation version of EXPLAN based on these ideas is developed that accounts for the remaining issues raised above, namely, (a) word repetition and hesitation around function words, and (b) part-word stuttering on content words.

To adapt EXPLAN to a spreading activation account, it is assumed that (a) activation of word candidates for a slot in a PW takes place in parallel, (b) the activation onsets of successive words is offset according to their order of appearance in the utterance, (c) activation builds up at different rates for words of different complexity, and (d) activation begins to decay once a plan is completed. Only the first assumption is added relative to the earlier EXPLAN account (represented diagrammatically in Figure 3), and this assumption is included to account for how speakers are aware that they make errors.

Applying (b) to the planning phases of the words in the PW “in the spring,” activation starts to build up for “in” first, then “the,” and finally “spring.” Using (c), activation of words of different complexity builds up at different rates. This arises, to some extent, because of the complexity of the phonetic makeup of words. “In” and “the” build up rapidly as they have a simple structure, whereas “spring” has a complex onset (a consonant string that includes later emerging consonants, Throneburg et al., 1994) so its activation builds up more slowly. The phonetic output for a word builds up progressively as activation increases over time.

The situation for fluent speech is considered now. Assuming that execution of the first word starts immediately after it is fully activated, the first word will start to decay when execution starts (d). The buildup in activation for the subsequent words continues during the execution time of this word and, given the decay in activation of the first word and offsets for activation of successive words (b), the next word in the sequence will be the one with maximum activation. The process continues, assuming that the activation of successive words is complete after execution of the current one is completed (as would be the case if an appropriate execution rate was used).

The way in which whole-word disfluencies on function words arise is as follows. Figure 4 shows the activation pattern at the time that “the” has been spoken in the center panel (completed at time *Ty*). Activation patterns of the other two words over the interval of time *Tx* to *Ty* are shown in successive panels. “In” (left panel) had built up to maximum previously, but activation by the time *Ty* has dropped off. “The” (middle panel) has also been at maximum and is showing decay during time for its execution (far less than for “in”). Rates of activation buildup for “in” and “the” were more rapid than for “spring” (right panel), which has the complex onset (property c above), and this is shown as having a gentler rising slope. At *Ty* (i.e., at the time “the” has been executed), the plan for “spring” is not complete, though some activation has occurred. A threshold rule (produce the word whose activation is above *T* in Figure 4) would lead the speaker to repeat “the” in this case as this word has highest activation of all words. A lower threshold that is still above that

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

achieved at *Ty* for “spring” or a more rapid execution rate (that allows less time for decay of “in”) could leave both “in” and “the” above threshold. In this situation, both words would be above threshold, and the speaker would produce “in the, in the.” Activation for “spring” can continue during either of these examples of repetitions and can result in sufficient time for the plan for “spring” to be completed, its threshold to be above *T* and the word to be produced. Essentially, the overlapping activation patterns permit word repetition when they precede a word with a complex onset (usually a content word in English). Pauses would arise when “in” and “the” have decayed below threshold (due to threshold and rate parameters again), and “spring” has not reached *T.* Thus, such word repetition and hesitation that Levelt (1989) and Kolk and Postma (1997) took as evidence for corrections to errors detected over the internal perceptual loop arise in the proposed model from overlapping activation patterns and the decay and threshold parameters that apply to these activations in production.

The situation can arise, depending on threshold value, speech rate, or rate at which activation builds up (phonetic complexity), where the two initial words in the phrase have decayed to values lower than *T,* and the third word is at or above *T,* but its plan is not complete (activation below maximum). Such a situation is shown in Figure 5. Execution of this word can commence at the requisite time (*Ty*) (the later phones will not be available at this point in time), so the plan still needs to be completed. The plan can be completed in the time taken to execute the first part. However, if there is insufficient time during execution of the first part to complete the word, the plan runs out, only the first part of the word can be produced, and this leads to part-word stutterings at onset (the characteristics of persistent stuttering).

The spreading activation version of EXPLAN shows that (a) errors, (b) word repetition and hesitation, and (c) part-word stutterings can arise without a perceptual monitor. The current model is based on some reasonable assumptions about activation buildup and decay in phrases and how these relate to planning and execution time. As a perceptual monitor has been discarded, this poses a challenge to CRH. Although the current work uses Kolk and Postma's (1997) phonological activation profiles, it needs to be stressed that this does not commit EXPLAN to (a) a view that phonological activation is all that is important in leading to disfluency and stuttering, or (b) the Kolk and Postma proposal of the way phonological and lexical activation builds up. As indicated earlier, anything that varies the time-course of activation patterns (e.g., syntax as well as lexical influences) will affect whether disfluencies and stutterings occur.

How are part-word stutterings signaled to a speaker?—To determine whether speech timing needs to be altered, the speaker needs to know whether speech is being produced before the plan is complete. To determine whether speech is being produced prematurely in this manner, all that needs to be done is to subtract the plan at the point in time that execution is commenced from the plan at the point in time that execution is completed (Howell, 2002). If the whole plan is supplied before start of execution, the two plans will be identical, they will cancel, and speech is fluent. If the speaker initiates speech prematurely, more of the plan will be generated in the time taken to execute the first part and the two will differ; this indicates that speech needs to be slowed to allow more time for planning. The points in time that (a) execution starts and (b) execution is completed are markers that are used in the EXPLAN spreading activation account (the times delimited at onset and offset of execution). Given that the timing of these points is needed to account for the types of fluency failure, they would be available to be used as pointers to access the plan at these times which determine, as indicated above, whether execution started with a complete plan or not.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Coupling in the external timekeeper when speech is stuttered and decoupling the timekeeper when speech is fluent—The result of differencing the copy of the plan taken at the time execution commences and the plan at the time execution is completed is cancellation (result is zero) when speech is fluent or disfluent, but noncancellation (a nonzero result, termed an alert in EXPLAN) when speech is stuttered. Cancellation indicates that speech is fluent and as long as it holds, speech rate is appropriate for plans to be supplied in time to execution and the external timekeeper does not need to adjust timing. The intermittent alert pulse (nonzero result) occurs whenever part-word stutterings arise. If these alerts continue (speech rate is too high), they would constitute a pulse train indicating where part-word stutterings are happening, which is also the sort of task-associated input that is accessed by the timekeeper. That is, a train of pulses like this is of the type that the timekeeper would take as input (similar to the serial activity from the signal in a DAF task, as described above). Consequently, this extra input would affect timekeeper operation like the extra input created by the DAF procedure. The timekeeper would change its operation (slow the responses it controls) as described for DAF above, and slowing speech would restore speech to fluency by allowing more time for planning to go to completion.

Controlling speech that only involved chaining planning and execution together during fluent speech allows this cycle to operate independently of the timekeeper. The operation of the chain can be characterized in terms of an oscillator that has a preferred rate determined by the execution time of different words. The timekeeper also oscillates at the same rate as serial activity associated with speech responses and auditory activity associated with these speech responses will dominate input to the timekeeper. If these oscillators are coupled together, during fluent speech, they will not affect each other's operation because they operate at the same frequency. In effect, this allows speech to operate independently of the timekeeper, but the timekeeper and planning–execution cycle are coupled, allowing the oscillators to affect each other's operation if the rate of one of them changes.

When rate drifts, speech changes from being fluent to having part-word stutterings (see above). Then the planning–execution cycle generates alerts that input to the timekeeper. This extra input, in turn, leads the timekeeper to reduce its rate so that it is then oscillating at a lower frequency than the planning–execution cycle. The planning–execution cycle changes to the new frequency (an inherent property of coupled oscillators). This results in a lower speech execution rate that stops the stutterings as, effectively, more planning time is allowed. Speech becomes fluent, the alerts stop, and the coupled oscillators now operate independently, as before, at the newly established (slower) rate. The reestablished equilibrium between the planning–execution cycle and the timekeeper and the reinstated independence of the two oscillators essentially allows speech to return to an open-loop mode of control. All of these changes are built-in compensations and do not rely on inputs from outside (e.g., perceptual) processes.

Some reasons why adult speakers might lose the ability to respond to indications that speech rate has changed—Children who stutter or fluent speakers of any age make few part-word stutterings. This has been explained on the basis that function word repetition and hesitation prevent this type of stuttering. The timekeeper mechanism could account for recovery from occasional part-word stutterings. The increase in part-word stutterings in adults who stutter would then imply that the timekeeper-recovery mechanism functions less effectively in these speakers. Three possible reasons why this may be so are that (a) there could be changes in motor development at puberty that specifically affect children who stutter (van Lieshout, Rutjens, & Spauwen, 2002); (b) there is a change in style of interaction with peers (Howell, Au-Yeung, & Sackin, 1999) that has a disproportionate influence on children who stutter compared with fluent speakers of the same age; and (c) once the tendency to produce part-word stutterings has started, it is

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

progressive. For instance, if these stutterings lead to alerts, the speakers who persist in their stuttering ignore them, which in turn leads to a reduction in their sensitivity to the alerts, making it easier for them to ignore future alerts (Howell, Rosen, Hannigan & Rustin, 2000).

DISCUSSION

It is useful to have proposals available that address as many topics associated with stuttering as possible, even if, ultimately, those proposals are found wanting (see Smith & Kleinow, 2000, for another attempt at modeling a wide range of phenomena associated with stuttering). EXPLAN addresses a wide (though not comprehensive) range of topics in stuttering. Some of the main ones are (a) why stuttering is intermittent, (b) why function word repetition and hesitation occur frequently in the speech of young children who stutter and fluent speakers of all ages, (c) why stuttering patterns involving parts of content words emerge in adults, (d) how part-word stutterings relates to persistence of the disorder, and (e) in what sense is there continuity between stuttered and fluent speech.

There is sparse evidence relating to some of these topics. A notable example concerns why there should be a change in speech patterns at around 12 years of age. Relevant to this issue, Howell et al. (in press) showed that a bilingual speaker uses function word repetition in his second language (English) and shows content word stutterings in his first language (Spanish). This result shows two important things: (a) that the particular fluency failure pattern is associated with proficiency in a language, and (b) that the fluency failure patterns are not “hard-wired” responses that are triggered by particular linguistic structures. For instance, they do not appear to be the result of difficulty with particular phonemes; otherwise, the patterns would be similar in the different languages a bilingual speaker speaks (Wingate, 2002, has other grounds for dismissing “phoneme difficulty”). There are no data, at present, that indicate whether pragmatic, motor-maturational or neural changes occur around the age at which the change to the persistent patterns arises. This is one issue that merits further work.

Another such issue concerns the proposal about timing control. The EXPLAN account maintains that speakers shift between a mode of control not involving the timekeeper to a mode of control that does involve the timekeeper. The latter alerts the speaker that a speech rate change is needed. Other models consider that error information is computed continuously and a monitor changes speech control when one occurs or it is anticipated that one is likely to occur. For instance, feedforward models such as that of Kawato, Furakawa, and Suzuki's (1987) maintain that movement errors are continuously computed and used (when they arise) as correction signals. The difference between EXPLAN and Kawato et al.'s model is that in the latter, a target trajectory is computed for comparison with an achieved action. The differencing operation in EXPLAN can be regarded as a much simpler solution that uses “targets” as they are generated (not in advance). Furthermore, there is no error correction in EXPLAN (the motor system does not make continuous fine correction-adjustments to errors). These highlight some differences between the models. It is also possible, however, that the EXPLAN model develops from a feedforward system like Kawato et al.'s when skills are highly practiced (a possibility first raised by Borden, 1979, and included in Guenther's, 2001 DIVA model).

A final issue that needs more detailed work is the spreading activation model. An aspect of this is that more research needs to be conducted to establish whether different stages in the linguistic hierarchy have an effect on overall planning time (currently, most attention has been given to the phonetic level). A start has been made on improving metrics for analysis of phonological and phonetic structure (with John Harris, Chloe Marshall, Heather van der Lely, and Rachel Diment) that Diment and Howell plan to use to improve specification of

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

difficulty and how this affects timing. In particular, it is intended to use these improved specifications of difficulty to get a more precise indication of activation rate buildup. The spreading activation model is being implemented in an analytical version (parameterizing the profiles in Figures 4 and 5), and automatic learning algorithms are being applied to the problem (artificial neural networks, Howell, Sackin, & Glenn, 1997a, 1997b). The first step is to simulate the different fluency patterns by supplying information about phonetic difficulty and speech rate, and examining what variables need adjusting to produce a selected type of fluency failure patterns. For instance, if PWS have slower phonological systems than fluent speakers (Kolk & Postma, 1997), then activation rate profiles would need to be adjusted to lower rates to produce part content word stutterings. It might then be possible, depending on the success of this enterprise, to develop procedures that would automatically locate fluency failures. Applications of these procedures would include (a) objective and speedier speech assessment, and (b) introduction into fluency aids so alterations are given only in episodes where the speaker is stuttering.

Further to the second of these points, my group researches into possible treatment procedures, but is not in the business of delivering treatments or promoting any particular treatment. An outcome of our investigations on alterations to auditory feedback was the fluency-enhancing effects of FSF (Howell et al., 1987), which has been one basis of the SpeechEasy™ prosthetic device. Though some success has been reported with this device, there are some observations that it seems appropriate to make, concerning (a) what FSF does and why it works, (b) the ways in which our research suggest that FSF should be used, and (c) how speech performance with FSF should be compared with other techniques (see Howell, in press, for an extended discussion of auditory feedback and stuttering).

- When FSF is played to an adult speaker who stutters, speech becomes highly fluent immediately. The effects are restricted to the interval the device is on (fluency failure reoccurs once the device is switched off). There are no obvious side effects with FSF (for instance, speakers do not raise their voice level, Howell, 1990). In the part of this article that dealt with execution phenomena, it was argued that FSF (and AAF techniques in general) works by affecting timekeeper operation that allows more planning time. What FSF provides is a very effective way of controlling speech fluency over the short term. It is possible to have the FSF alteration on continuously, but there are at least two drawbacks to this: (a) It would be desirable to be able to cease using the device if speech fluency returns: Who would want to continue wearing a device like a hearing-aid if it is not necessary? (b) Having the alteration on continuously does not necessarily train the user to speak fluently.
- Exactly how users employ the device is likely to be a matter of personal choice and experience. It might be appropriate to use the device continuously when on the phone or when in anxiety-inducing situations where the risk of stuttering is high. However, if the research question is raised about how FSF compares with other techniques, an appropriate design would be to compare the techniques in comparable situations. If FSF is to be compared with operant procedures, then FSF should be used in a comparable way to the reinforcement given by the verbal contingencies in the operant procedures. Onslow, Andrews, and Lincoln (1994) described their operant technique (the Lidcombe program) as follows: It “is an operant treatment that incorporates parental verbal contingencies for stuttered speech and stutter-free speech. The contingencies for stutter-free speech are praise and tangible reinforcement, and the contingencies for stuttering are that the parents identify a stuttered utterance and request the child to correct the utterance.” Note that stutterings are located that then receive a particular verbal contingency in the operant procedure. To compare the effects that this operant procedure has with

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

> those that FSF has, in an FSF treatment, stutterings would also need to be located that then receive their appropriate contingency (e.g., switch the FSF on when a stuttering occurs). It does not seem appropriate to present FSF continuously, giving this contingency to fluent speech as well as stuttered speech (Ingham, Moglia, Frank, Ingham, & Cordes, 1997).

Another implication of EXPLAN for operant treatment of stuttering concerns whether all "stuttered utterances," irrespective of their type, should be corrected. The case has been made above that function word repetition and hesitation are associated with fluent speech control, suggesting that these might be appropriate events to receive "praise and tangible reinforcement." Conversely, part content word stutterings might be the appropriate events to target for correction (appropriate only for children older than those typically involved in the Lidcombe Program). To this end, our research has investigated techniques that could be used to increase the use of function word repetition as a form of 'praise' (Howell & Sackin, 2001). We have also investigated procedures involving direct reinforcement of function word disfluency and stopping speech when content word stutterings occur, with some success (Howell et al., 2001). (See Reed & Howell, 2001, for further discussion of some of these issues.)

Finally, there are some topics that are receiving attention in the stuttering literature on which EXPLAN currently takes a neutral view. One concerns whether PWS show brain activity differences relative to fluent control speakers and, if so, whether these reflect a structural or functional problem. EXPLAN could incorporate structural or functional problems in brain processes (in a similar way to that discussed when considering Kolk and Postma's proposal that speakers who stutter have a slow phonological system), providing they affect processing time. At present, there are only limited data on brain processes in stuttering and there are issues that are being addressed that will (hopefully) eventually be solved. At present, though, whether to introduce them into models needs to be approached cautiously. The problems include (a) motor artefacts, (b) (related to the previous point) high noise levels in recordings, (c) some imaging procedures that require speech in odd postural conditions and in noisy environments (due to equipment noise), and (d) in the case of children, there is no standard map to locate the exact area of the brain affected. One finding from the imaging studies that has direct relevance to EXPLAN is that there is an emerging consensus that the cerebellum is functionally involved in stuttering (Fox et al., 1996). EXPLAN offers a specific proposal as to what that function is (concerning organizing speech output).

A second area is whether genetics plays a part in predisposing an individual to stutter. As with brain differences, though the consensus among other research groups seems to be that there is a case for genetic involvement in the disorder, the evidence is equivocal. Studies that look at the families of PWS usually have to rely on past memory for information about family members who stutter. Besides the potential for selectivity (for instance, PWS might be more inclined to recall something odd about the speech of family members than people who do not stutter), lay persons might not be capable of making an accurate judgement about stuttering, as even professionals find such decisions hard (Kully & Boberg, 1988). Future work may firmly establish that genetic factors predispose certain people to stutter. As with brain imaging, this could be incorporated in EXPLAN providing that the genetic influence is manifest in terms of timing problems.

Acknowledgments

This work was supported by the Wellcome Trust.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

REFERENCES

Au-Yeung J, Howell P, Pilgrim L. Phonological words and stuttering on function words. Journal of Speech, Language, and Hearing Research. 1998; 41:1019–1030.

Au-Yeung J, Vallejo Gomez I, Howell P. Exchange of disfluency from function words to content words with age in Spanish speakers who stutter. Journal of Speech, Language, and Hearing Research. 2003; 46:754–765. [PubMed: 14697001]

Bishop, DVM. The Test for Reception of Grammar. Author; 1983.

Blackmer ER, Mitton JL. Theories of monitoring and the timing of repairs in spontaneous speech. Cognition. 1991; 39:173–194. [PubMed: 1841032]

Bloodstein O, Gantwerk BF. Grammatical function in relation to stuttering in young children. Journal of Speech and Hearing Research. 1967; 10:786–789. [PubMed: 5586944]

Bloodstein O, Grossman M. Early stutterings: Some aspects of their form and distribution. Journal of Speech and Hearing Research. 1981; 24:298–302. [PubMed: 7265947]

Borden GJ. An interpretation of research on feedback interruption in speech. Brain & Language. 1979; 7:307–319. [PubMed: 455050]

Caruso AJ, Abbs JH, Gracco VL. Kinematic analysis of multiple movement coordination during speech in stutterers. Brain. 1988; 111:439–455. [PubMed: 3378144]

Clark, HH.; Clark, E. Psychology and language. An introduction to psycholinguistics. New York: Harcourt; 1977.

Conture, EG. Stuttering. Englewood Cliffs, NJ: Prentice-Hall; 1990.

Dell GS, O'Seaghdha P. Mediated and convergent lexical priming in language production: A comment to Levelt et al. Psychological Review. 1991; 98:604–614. [PubMed: 1961775]

Dworzynski K, Howell P, Au-Yeung J, Rommel D. Stuttering on function and content words across age groups of German speakers who stutter. Journal of Multilingual Communication Disorders. in press.

Fox PT, Ingham RJ, Ingham JC, Hirsch T, Downs H, Martin C, et al. A PET study of the neural systems of stuttering. Nature. 1996; 382:158–162. [PubMed: 8700204]

Fromkin VA. The non-anomalous nature of anomalous utterances. Language. 1971; 47:27–52.

Garnham A, R. C. Shillcock GDA, Brown AI, Mill D, Cutler A. Slips of the tongue in the London-Lund corpus of spontaneous conversation. Linguistics. 1981; 19:805–817.

Guenther. Neural modeling of speech production. In: Maassen, B.; Hulstijn, W.; Kent, R.; Peters, HFM.; van Lieshout, PHMM., editors. Speech motor control in normal and disordered speech. Nijmegen: Uttgeverij Vantilt; 2001. p. 12-15.

Howell P. Changes in voice level caused by several forms of altered feedback in normal speakers and stutterers. Language and Speech. 1990; 33:325–338. [PubMed: 2133911]

Howell, P. Producing and perceiving speech. In: Green, D., et al., editors. Cognitive science. Blackwell; 1996. p. 120-147.

Howell, P. A model of timing interference to speech control in normal and altered listening conditions applied to the treatment of stuttering. In: Maassen, B.; Hulstijn, W.; Kent, R.; Peters, HFM.; van Lieshout, PHMM., editors. Speech motor control in normal and disordered speech. Nijmegen: Uttgeverij Vantilt; 2001. p. 291-294.

Howell, P. The EXPLAN theory of fluency control applied to the treatment of stuttering by altered feedback and operant procedures. In: Fava, E., editor. Pathology and therapy of speech disorders. Amsterdam: John Benjamins; 2002.

Howell P. Cerebellar activity and stuttering: Comments on Max and Yudman (2003). Journal of Speech, Language and Hearing Research. in press. [PubMed: 15072531]

Howell P, Archer A. Susceptibility to the effects of delayed auditory feedback. Perception & Psychophysics. 1984; 36:296–302. [PubMed: 6522222]

Howell, P.; Au-Yeung, J. The EXPLAN theory of fluency control and the diagnosis of stuttering. In: Fava, E., editor. Pathology and therapy of speech disorders. Amsterdam: John Benjamins; 2002.

Howell, P.; Au-Yeung, J.; Charles, N.; Davis, S.; Thomas, C.; Reed, P., et al. Operant procedures that increase function word repetition used with children whose speech had not improved during

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

previous treatment. In: Bosshardt, H-G.; Yaruss, JS.; Peters, HFM., editors. Fluency disorders: Theory, research, treatment and self-help. Proceedings of the Third World Congress of Fluency Disorders. Nijmegen: Nijmegen University Press; 2001. p. 133-137.

Howell P, Au-Yeung J, Pilgrim L. Utterance rate and linguistic properties as determinants of speech dysfluency in children who stutter. Journal of the Acoustical Society of America. 1999; 105:481–490. [PubMed: 9921672]

Howell, P.; Au-Yeung, J.; Rustin, L. Clock and motor variance in lip tracking: A comparison between children who stutter and those who do not. In: Hulstijn, W.; Peters, HFM.; van Lieshout, PHHM., editors. Speech production: Motor control, brain research and fluency disorders. Elsevier; 1997. p. 573-578.city, state

Howell P, Au-Yeung J, Sackin S. Exchange of stuttering from function words to content words with age. Journal of Speech, Language, and Hearing Research. 1999; 42:345–354.

Howell P, Au-Yeung J, Sackin S. Internal structure of content words leading to lifespan differences in phonological difficulty in stuttering. Journal of Fluency Disorders. 2000; 25:1–20. [PubMed: 18259599]

Howell P, Davis S, Au-Yeung J. Syntactic development in fluent children, children who stutter, and children who have English as an additional language. Child Language Teaching and Therapy. 2003; 19:311–337. [PubMed: 18259597]

Howell, P.; El-Yaniv, N.; Powell, DJ. Factors affecting fluency in stutterers when speaking under altered auditory feedback. In: Peters, HFM.; Hulstijn, W., editors. Speech motor dynamics in stuttering. New York: Springer Press; 1987. p. 361-369.

Howell, P.; Harvey, N. Perceptual equivalence and motor equivalence in speech. In: Butterworth, B., editor. Language production. Vol. 2. London: Academic Press; 1983. p. 203-224.

Howell P, Powell DJ. Hearing your voice through bone and air: Implications for explanations of stuttering behaviour from studies of normal speakers. Journal of Fluency Disorders. 1984; 9:247–264.

Howell P, Powell DJ, Khan I. Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. Journal of Experimental Psychology: Human Perception and Performance. 1983; 9:772–784.

Howell P, Rosen S, Hannigan G, Rustin L. Auditory backward masking performance by children who stutter and its relation to dysfluency rate. Perceptual and Motor Skills. 2000; 90:355–363. [PubMed: 10833723]

Howell, P.; Ruffle, L.; Fernández-Zúñiga, A.; Gutiérrez, R.; Fernández, AH.; O'Brien, ML., et al. Comparison of exchange patterns of stuttering in Spanish and English monolingual speakers and a bilingual Spanish-English speaker. IFA Montreal; in press

Howell P, Sackin S. Speech rate manipulation and its effects on fluency reversal in children who stutter. Journal of Developmental and Physical Disabilities. 2000; 12:291–315. [PubMed: 18259598]

Howell P, Sackin S. Function word repetitions emerge when speakers are operantly conditioned to reduce frequency of silent pauses. Journal of Psycholinguistic Research. 2001; 30:457–474. [PubMed: 11529422]

Howell P, Sackin S. Timing interference to speech in altered listening conditions. Journal of the Acoustical Society of America. 2002; 111:2842–2852. [PubMed: 12083218]

Howell P, Sackin S, Glenn K. Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: I. Psychometric procedures appropriate for selection of training material for lexical dysfluency classifiers. Journal of Speech, Language, and Hearing Research. 1997a; 40:1073–1084. [PubMed: 9328878]

Howell P, Sackin S, Glenn K. Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN Recognition of repetitions and prolongations with supplied word segment markers. Journal of Speech, Language, and Hearing Research. 1997b; 40:1085–1096. [PubMed: 9328879]

Howell P, Vause L. Acoustic analysis and perception of vowels in stuttered speech. Journal of the Acoustical Society of America. 1986; 79:1571–1579. [PubMed: 3711457]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Ingham RJ, Moglia RA, Frank P, Ingham JC, Cordes AK. Experimental investigation of the effects of frequency-altered auditory feedback on the speech of adults who stutter. Journal of Speech, Language, and Hearing Research. 1997; 40:361–372.

Ivry, R. Cerebellar timing systems. In: Schmahmann, J., editor. The cerebellum and cognition. San Diego, CA: Academic Press; 1997. p. 555-573.

Kalinowski J, Dayalu VN, Stuart A, Rastatter MP, Rami KM. Stutter-free and stutter-filled speech signals and their role in stuttering amelioration for English speaking adults. Neuroscience Letters. 2000; 293:115–118. [PubMed: 11027847]

Kawato M, Furakawa K, Suzuki R. A hierarchical neural-network model for control and learning of voluntary movement. Biological Cybernetics. 1987; 57:169–185. [PubMed: 3676355]

Kolk, H.; Postma, A. Stuttering as a covert repairs phenomenon. In: Curlee, RF.; Siegel, GM., editors. Nature and treatments of stuttering: New directions. Needham Heights, MA: Allyn & Bacon; 1997. p. 182-203.

Kully D, Boberg E. An investigation of interclinic agreement in the identification of fluent and stuttered syllables. Journal of Fluency Disorders. 1988; 13:309–318.

Kuniszyk-Jozkowiak W, Smolka E, Adamczyk B. Effect of acoustical, visual and tactile reverberation on speech fluency of stutterers. Folia Phoniatrica & Logopedics. 1996; 48:193–200.

Lane HL, Tranel B. The Lombard sign and the role of hearing in speech. Journal of Speech and Hearing Research. 1971; 14:677–709.

Levelt W. Monitoring and self-repair in speech. Cognition. 1983; 14:41–104. [PubMed: 6685011]

Levelt, WJM. Speaking: From intention to articulation. Cambridge, MA: Bradford Books; 1989.

Maclay H, Osgood CE. Hesitation phenomena in spontaneous English speech. Word. 1959; 15:169–182.

MacWhinney B, Osser H. Verbal planning functions in children's speech. Child Development. 1977; 48:97–985. [PubMed: 844366]

Max L, Caruso AJ, Gracco VL. Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering individuals. Journal of Speech, Language, and Hearing Research. 2003; 46:215–232.

Max L, Yudman EM. Accuracy and variability of isochronous rhythmic timing access motor systems in stuttering versus nonstuttering individuals. Journal of Speech, Language, and Hearing Research. 2003; 46:146–163.

Melnick KS, Conture EG. Relationship of length and grammatical complexity to the systematic and nonsystematic speech errors and stuttering of children who stutter. Journal of Fluency Disorders. 2000; 25:21–45.

Nippold MA. Concomitant speech and language disorders in stuttering children: A critique of the literature. Journal of Speech and Hearing Disorders. 1990; 55:51–60. [PubMed: 2405212]

Nippold MA. Phonological disorders and stuttering in children: What is the frequency of co-occurrence? Clinical Linguistics and Phonetics. 2001; 15:219–228.

Onslow M, Andrews C, Lincoln M. A control/experimental trial of operant treatment for early stuttering. Journal of Speech and Hearing Research. 1994; 37:1244–1259. [PubMed: 7877284]

Reed P, Howell P. Suggestions for improving the long-term effects of treatment for stuttering: A review and synthesis of frequency-shifted feedback and operant techniques. European Journal of Analysis of Behaviour. 2001; 1:89–106.

Ryan BP, van Kirk Ryan B. Programmed stuttering treatment for children: Comparison of two establishment programs through transfer, maintenance, and follow-up. Journal of Speech and Hearing Research. 1995; 38:61–75. [PubMed: 7731220]

Santiago J, MacKay DG, Palma A, Rho C. Sequential activation processes in producing words and syllables: Evidence from picture naming. Language and Cognitive Processes. 2000; 15:1–44.

Selkirk, E. Phonology and syntax: The relation between sound and structure. Cambridge, MA: MIT Press; 1984.

Smith A, Kleinow J. Kinematic correlates of speaking rate changes in stuttering and normally fluent adults. Journal of Speech, Language, and Hearing Research. 2000; 43:521–536.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Stein JF, Glickstein M. Role of the cerebellum in visual guidance of movement. Physiological Reviews. 1992; 72:972–1017. [PubMed: 1438583]

Sternberg, S.; Monsell, S.; Knoll, RL.; Wright, CE. The latency and duration of rapid movement sequences: Comparison of speech and typing. In: Stelmach, GE., editor. Information control in motor control and learning. New York: Academic Press; 1978. p. xx-xx.

Throneburg NR, Yairi E, Paden EP. The relation between phonological difficulty and the occurrence of disfluencies in the early stage of stuttering. Journal of Speech and Hearing Research. 1994; 37:504–509. [PubMed: 8084182]

van der Lely HKJ, Stollwerck L. Binding theory and grammatical specific language impairment in children. Cognition. 1997; 62:245–290. [PubMed: 9187060]

van Lieshout PHHM, Hulstijn W, Peters HFM. From planning to articulation in speech production: What differentiates a person who stutters from a person who does not stutter? Journal of Speech and Hearing Research. 1996; 39:546–564. [PubMed: 8783133]

van Lieshout PHHM, Rutjens CAW, Spauwen PHM. The dynamics of interlip coupling in speakers with a repaired unilateral cleft-lip history. Journal of Speech, Language, and Hearing Research. 2002; 45:5–19.

Wing AM, Kristofferson AB. Response delays and the timing of discrete motor responses. Perception & Psychophysics. 1973; 14:5–12.

Wingate, M. Foundations of stuttering. New York: Academic Press; 2002.

Wolk L, Edwards ML, Conture EG. Coexistence of stuttering and disordered phonology in young children. Journal of Speech and Hearing Research. 1993; 36:906–917. [PubMed: 8246479]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

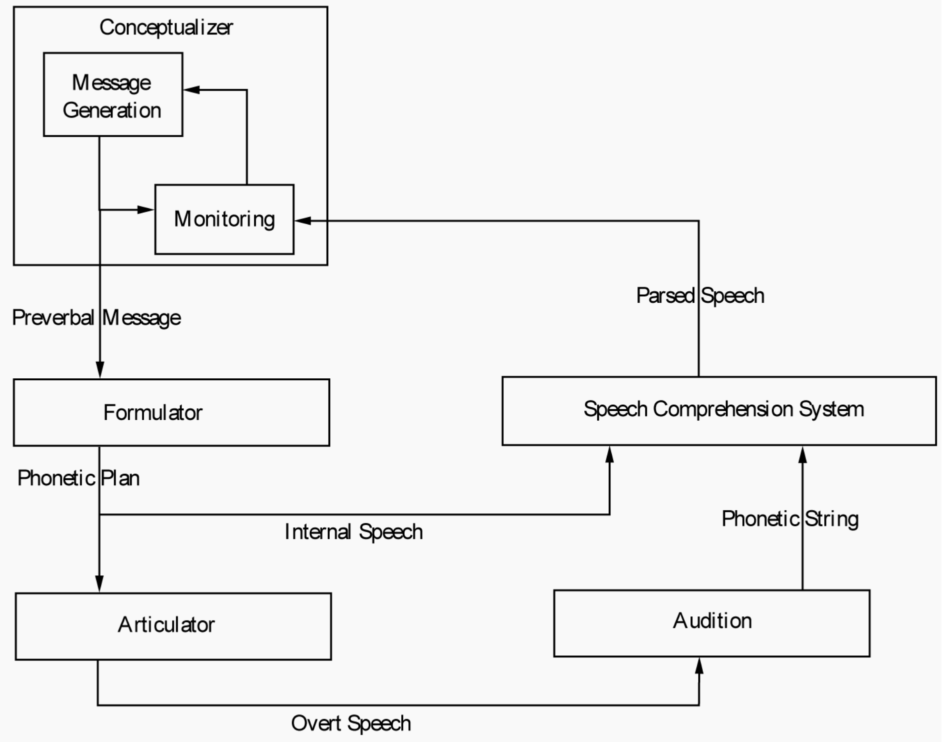

Figure 1.
Levelt's model of the processes involved in speech production. Production is on the left and perception on the right.

Europe PMC Funders Author Manuscripts

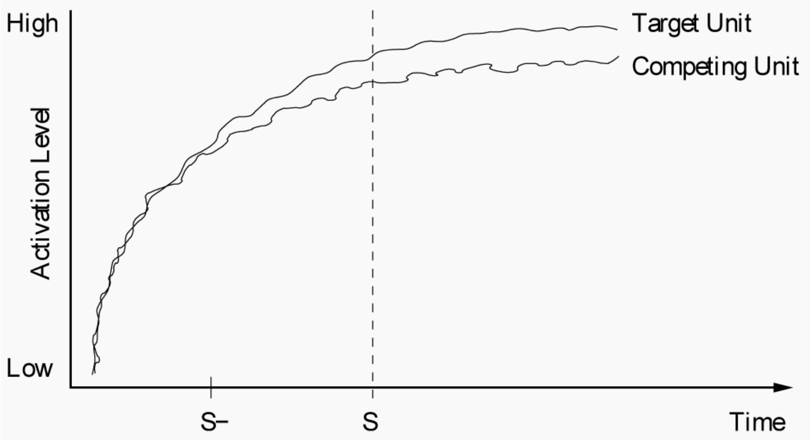

Figure 2.
Activation versus time for target, and competing, word candidates. Two selection points are shown, normal (*S*) and early (*S*–).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

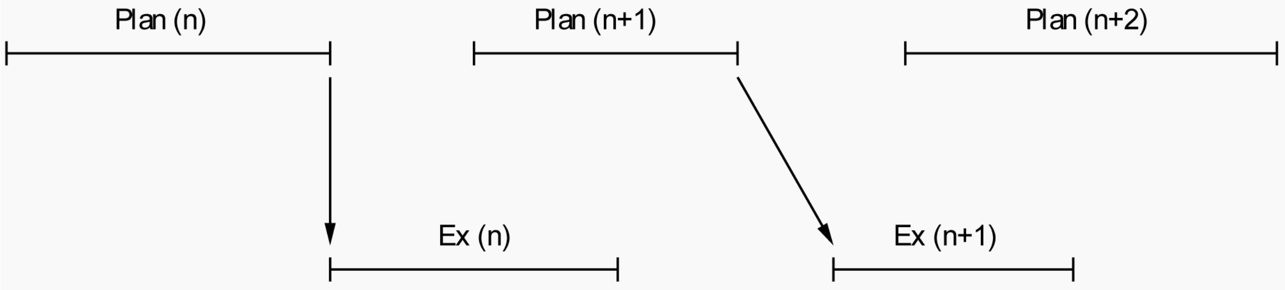

Figure 3.
Diagrammatic representation of the temporal relationship between planning and execution for three words (n, n+1, n+2) when speech is proceeding fluently. Time is along the abscissa. The epoch during which planning (PLAN) and execution (EX) occur is shown as bars in the top and middle rows respectively. Planning of adjacent words is shown in series for simplicity.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

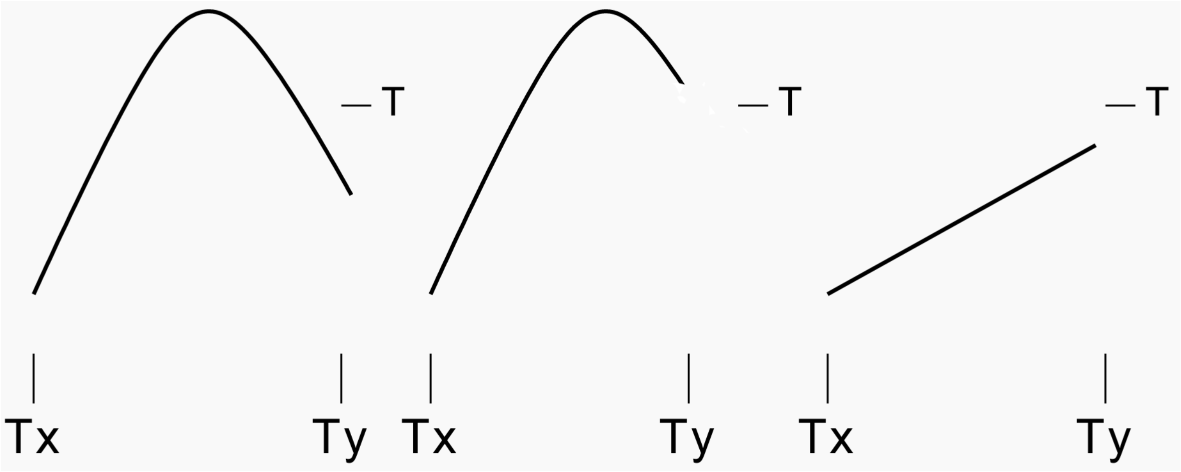

Figure 4.
Activation patterns for the three words in the test utterance each shown for the interval of time *Tx* to *Ty*. *Ty* is after execution of the second word and represents a situation that will lead to word repetition.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

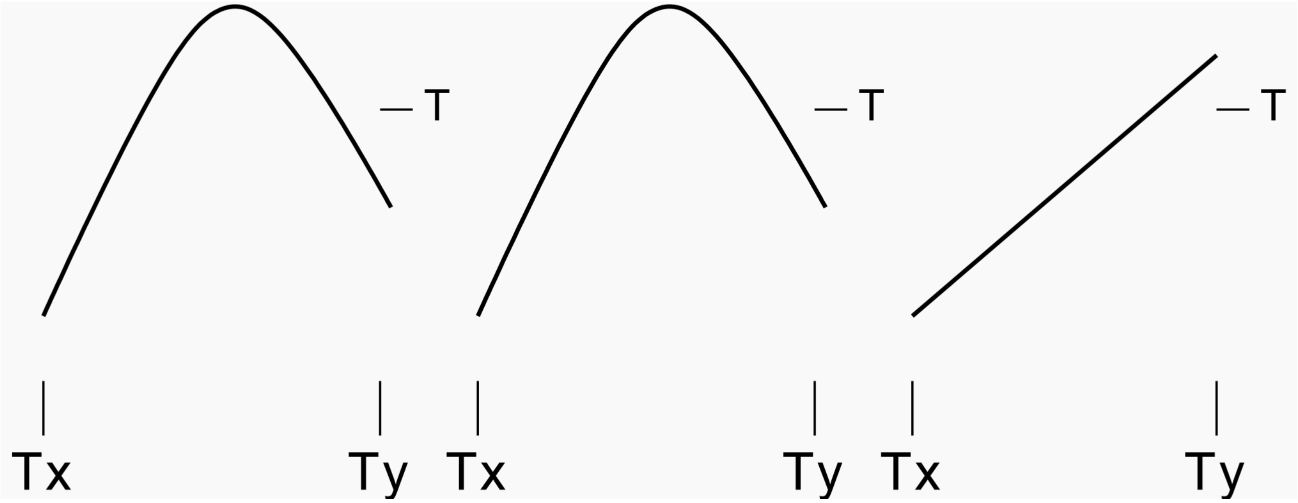

Figure 5.

Activation patterns for the three words in the test utterance each shown for the interval of time *Tx* to *Ty*. *Ty* is after execution of the second word and represents a situation that will lead to part-word stuttering involving the onset of the third word.