Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production

Oren Civier[a], **Stephen M. Tasko**[b], and **Frank H. Guenther**[a,c,d]

[a] Department of Cognitive and Neural Systems, Boston University, Boston, MA

[b] Department of Speech Pathology and Audiology, Western Michigan University, Kalamazoo, MI

[c] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA

[d] Division of Health Sciences and Technology, Harvard University – Massachusetts Institute of Technology, Cambridge, MA

Abstract

This paper investigates the hypothesis that stuttering may result in part from impaired readout of feedforward control of speech, which forces persons who stutter (PWS) to produce speech with a motor strategy that is weighted too much toward auditory feedback control. Over-reliance on feedback control leads to production errors which, if they grow large enough, can cause the motor system to "reset" and repeat the current syllable. This hypothesis is investigated using computer simulations of a "neurally impaired" version of the DIVA model, a neural network model of speech acquisition and production. The model's outputs are compared to published acoustic data from PWS' fluent speech, and to combined acoustic and articulatory movement data collected from the dysfluent speech of one PWS. The simulations mimic the errors observed in the PWS subject's speech, as well as the repairs of these errors. Additional simulations were able to account for enhancements of fluency gained by slowed/prolonged speech and masking noise. Together these results support the hypothesis that many dysfluencies in stuttering are due to a bias away from feedforward control and toward feedback control.

Keywords

stuttering; auditory feedback; fluency enhancement; error monitoring

Corresponding author: Oren Civier, orenciv@gmail.com; phone: +1-617-353-6181; fax: +1-617-353-7755, Postal address: Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215.

EDUCATIONAL OBJECTIVES

The reader will be able to (a) describe the contribution of auditory feedback control and feedforward control to normal and stuttered speech production, (b) understand the neural modeling approach to speech production and its application to stuttering, and (c) explain how the DIVA model accounts for enhancements of fluency gained by slowed/prolonged speech and masking noise.

1 Introduction

In his classic book "The Nature of Stuttering," Van Riper (1982) describes a person that stopped stuttering "after an incident in which he became completely deafened. The cessation of stuttering occurred within three hours of the trauma and shortly after he began to speak" (p. 383). What is the relationship between sensory feedback (in this case, hearing oneself speak) and stuttering? Although many investigations have attempted to account for the effect altered sensory feedback has in enhancing fluency (e.g., Fairbanks, 1954; Hutchinson & Ringel, 1975; Loucks & De Nil, 2006a; Mysak, 1960; Neilson & Neilson, 1987, 1991; Webster & Lubker, 1968; Wingate, 1970), no account has received overwhelming support. The obstacle may lie in the classic experimental devices typically used to study stuttering rather than the theoretical models themselves.

Most researchers have used psychophysical experiments to investigate sensory feedback in speech, but the results are often difficult to interpret because several feedback channels are simultaneously active during a typical experiment. In addition, experimental blockage of proprioceptive feedback channels is unfeasible (Scott & Ringel, 1971, p. 806), and auditory blockage is usually incomplete (e.g., Adams & Moore, 1972; but cf. Namasivayam, van Lieshout, McIlroy, & De Nil, 2009). Alternatively, sensory feedback can be studied by using computational models of speech production where feedback channels can be systematically blocked or altered.

In this paper we use the DIVA model (see Guenther, Ghosh, & Tourville, 2006), which is a biologically plausible model of speech production that mimics the computational and time constraints of the central nervous system (CNS), to test our hypothesis regarding the involvement of sensory feedback in stuttering. The DIVA model differs from other computational models applied to stuttering (e.g., Kalveram, 1991, 1993; Neilson & Neilson, 1987; Toyomura & Omori, 2004) in its ability to simulate the articulatory kinematic and acoustic features of both normal and disordered speech (for DIVA simulations of normal speech see Guenther, 1995; Guenther, 2006; Guenther et al., 2006; Guenther, Hampson, & Johnson, 1998; Nieto-Castanon, Guenther, Perkell, & Curtin, 2005; for simulations of childhood apraxia of speech, or CAS, see Terband & Maassen, in press; Terband, Maassen, Guenther, & Brumberg, 2009), which makes it possible to compare the simulations to a much larger set of data than permitted by other models.

Many authors have suggested that stuttering may be due in part or whole to an aberrant sensory feedback system. Some have hypothesized that persons who stutter (PWS) differ from those who do not stutter (PNS) by relying too heavily on sensory feedback (Hutchinson & Ringel, 1975; Tourville, Reilly, & Guenther, 2008, p. 1441; van Lieshout, Peters, Starkweather, & Hulstijn, 1993), while others have claimed that PWS actually benefit from reliance on sensory feedback (Max, 2004; Max, Guenther, Gracco, Ghosh, & Wallace, 2004; Namasivayam, van Lieshout, & De Nil, 2008; van Lieshout, Hulstijn, & Peters, 1996, 1996b; Zebrowski, Moon, & Robin, 1997). Our hypothesis is that due to an impaired feedforward (open-loop) control system[1], PWS rely more heavily on a feedback-based (closed-loop) motor control strategy (cf. De Nil, Kroll, & Houle, 2001; Jäncke, 1991; Kalveram, 1991; Kalveram & Jäncke, 1989; Lane & Tranel, 1971; Stromsta, 1972; Stromsta, 1986, p. 204; Toyomura & Omori, 2004; Van Riper, 1982, p. 387; Zimmermann, 1980c). Such an impairment in the feedforward control system is compatible with the general notion of stuttering arising out of diminished motor skill (Peters, Hulstijn, & van Lieshout, 2000; van Lieshout, Hulstijn, & Peters, 2004) because motor skill,

---

[1]The current investigation does not speak directly to the cause of the feedforward impairment (possibly, a dysfunction of the basal ganglia, see Alm, 2004; Alm, 2005, p. 30; Smits-Bandstra & De Nil, 2007) or to its exact nature (Civier, 2010; Civier, Bullock, Max, & Guenther, 2009); instead the focus is on the consequences of biasing away from feedforward control and toward feedback control.

in part, involves a transition from feedback to feedforward control (see Schmidt & Lee, 2005).

The hypothesized impairment in feedforward control and the resulting over-reliance on feedback control increase the frequency of production errors. The feedforward commands-- stored detailed instructions of how to move the articulators--are read out directly from memory, which in our model occurs via projections from the premotor to the motor cortex. Feedback control, on the other hand, requires the detection and correction of production errors (e.g., incorrect tongue position or unexpected formant pattern). Since a feedback-based strategy is relatively slow to detect and correct errors (e.g., Guenther et al., 2006), it is our contention that over-reliance on feedback control leads to error accumulation, and eventually, to a motor "reset" in which the system attempts to repair the error by restarting the current syllable. Such a reset would constitute a sound/syllable repetition (or simply *repetition*), a term we use to refer to any audible repetition of a syllable or part of it, without regard to phonemic boundaries, i.e., the cut-off can be within or between phonemes (cf. Conture, 2001, p. 6; Wingate, 1964). We also contend that an impaired feedforward control system may lead to other common types of dysfluency (Civier, 2010; Civier et al., 2009). However, to limit the scope of the paper, we will focus solely on sound/syllable repetitions (for further discussion see Section 7.2).

The idea that production errors can lead to stuttering is not new. Forty years ago, Wingate (1969) suggested that moments of stuttering result from errors in the speech movements required to transition between phonemes; this hypothesis found support in previous (Stromsta, 1965; summarized in Stromsta, 1986, pp. 64–67) and follow-up (Adams, 1978; Agnello, 1975; Harrington, 1987; Howell & Vause, 1986; Stromsta, 1986, p. 94; Stromsta & Fibiger, 1981; Zimmermann, 1980a) experimental work. In an attempt to clarify this mechanism, Van Riper (1982, pp. 117, 435) and Stromsta (1986, pp. 14, 28, 91, 111) reasoned that, after an error, the CNS detects the problem and searches for the response capable of repairing it; the response is often a "reset" in which the current syllable is attempted again. In their discussion of research approaches of stuttering, Postma and Kolk (1993) went further, combining the speech-motor-control approach with Levelt's (1983) self-repair theory. By assuming that speech-motor execution can be monitored via sensory feedback, they offered a theory in which "motor [production] errors are supposed to be detectable, and when reacted to, would cause stuttering" (Postma & Kolk, 1993, p. 482). To emphasize that production errors can be detected via sensory feedback (Guenther, 2006; Postma, 2000), we use the term *sensorimotor errors* (or simply *errors*). Wingate (1976, p. 338) claimed that PWS have mostly phonatory-movement errors which manifest as abnormal transitions in vocal pitch and intensity. We focus instead on PWS's articulatory movement errors which manifest as abnormal formant transitions (see Stromsta, 1986; Yaruss & Conture, 1993).

Here we investigate the hypothesis that, due their impaired ability to read out feedforward commands[2], PWS are forced to rely too heavily on auditory feedback control, which as described above, can lead to errors, and subsequent repetitions. The idea was previously presented in a limited form by our group (Civier & Guenther, 2005;Max et al., 2004) and inspired a series of simulations aimed at showing that when the DIVA model is biased away from feedforward control and toward feedback control, it behaves similarly to PWS, both in the type of errors produced, and in the way the errors are repaired.

This paper is organized as follows. After a description of the DIVA model, we report a series of modeling experiments in which the model is biased away from feedforward control and

---

[2]We acknowledge that in addition to auditory feedback control, the biasing may be toward somatosensory feedback control as well. For the sake of tractability, however, the simulations in this paper are limited to the auditory feedback channel. The possible limitations of this approach are discussed in Section 7.1.

toward auditory feedback control. The first modeling experiment mimics errors (observable in the acoustic domain) produced by PWS during fluent speech production. The second experiment tests whether the DIVA model's attempt to repair these errors resembles the acoustic and kinematic properties of repetitions produced by a PWS. The third and fourth experiments examine the model's ability to account for the reduction in the frequency of repetitions in two fluency-inducing conditions: slowed/prolonged speech and masking noise. To conclude we suggest how to further test our hypothesis using the model.

2 Computational modeling of speech kinematics and acoustics

The DIVA model is a biologically plausible neural network model capable of simulating the production and development of fluent speech (Callan, Kent, Guenther, & Vorperian, 2000; Guenther, 1994; Guenther, 1995; Guenther et al., 1999; Guenther et al., 2006; Guenther et al., 1998; Nieto-Castanon et al., 2005; Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004). The model focuses on speech control at the syllable and lower motor levels. Its name, DIVA (Directions Into Velocities of Articulators), derives from the sensory-motor transformations acquired during the first learning stage (roughly corresponding to infants' early babbling phase), in which the model uses semi-random articulatory movements to explore the auditory, somatosensory, and articulatory spaces, and the relations between them. The acquired transformations are assumed to be critical for human speech, and in the model they are involved in the acquisition and production of all speech sounds that will be learned later. More information on these sensory-motor transformations and how they bias the model away from awkward articulatory configurations not observed in actual speakers is available in Guenther et al. (1998). Guenther et al. (1998) also discuss the rationale behind another assumption we take: that acoustic information is the input for speech production. The inputs to the version of the model used here are continuous formant frequencies, but other auditory representations are feasible as well (see Guenther et al., 2006, p. 285; e.g., Nieto-Castanon, 2004). While the DIVA model depends on the yet-unproved assumptions mentioned above, and is far from an explanation of all aspects of human speech, it does provide a unified framework that accounts for many properties of fluent (and as we show here, also dysfluent) speech production.

The most recent version of the DIVA model combines mathematical descriptions of underlying commands with hypothesized neural substrates corresponding to the model's components, which are schematically represented in Fig. 1 (each map, or set of neurons, is represented by a box). The model is implemented in computer simulations in which it controls the movements of articulators in an articulatory synthesizer (Maeda, 1990), represented in the figure by a cartoon of a vocal tract. The mechanisms of the model are too complex to be described here in full. A detailed description of the model is available in Guenther et al. (2006), and a less technical description in Guenther (2006). The model's equations and parameter values used for this study are listed in the Appendix.

In the DIVA model, cells in the motor cortex generate the overall motor command, M(t), to produce a speech sound. M(t) is a combination of a *feedforward command* and a *feedback command* as indicated by the following formula:

$$\dot{M}(t) = \alpha_{ff}\dot{M}_{feedforward}(t) + \alpha_{fb}\dot{M}_{feedback}(t)$$

where α_{ff} and α_{fb} are gain parameters representing the amount of feedforward and feedback control weighting, respectively. The default parameter values are $\alpha_{ff} = 0.85$ and $\alpha_{fb} = 0.15$, settings that can account for normal speech, as well as for speech produced during auditory feedback perturbation (Tourville et al., 2008). Next we describe the *feedback control* and *feedfoward control subsystems* of the DIVA model (which are responsible to generate the

feedback and feedforward commands, respectively), as well as the *monitoring subsystem*, developed especially for this study. Although only the feedforward control subsystem is assumed to be impaired in PWS, all three subsystems are involved in the generation of repetitions.

2.1 Feedback control subsystem

Feedback, or closed-loop control (Borden, 1979; Max et al., 2004; Smith, 1992), relies on sensory feedback. Because of this, it can adapt to error-inducing perturbations of the speech-production system, such as the introduction of a bite block (e.g., Baum, McFarland, & Diab, 1996; Hoole, 1987), unexpected jaw perturbation (Guenther, 2006), and altering of auditory feedback of one's own speech (e.g., Tourville et al., 2008). For unperturbed speech, PNS rely very little on sensory-based feedback control (e.g., Gammon, Smith, Daniloff, & Kim, 1971), but as suggested in this paper, PWS rely heavily upon such feedback control to correct sensorimotor errors. In the *feedback control subsystem* of the DIVA model, auditory and somatosensory feedback loops are constantly comparing speech output with the desired auditory target (how the utterance should sound) and somatosensory target (how the utterance should feel). These targets consist of time-varying regions that encode the allowable variability of the acoustic and somatosensory signal throughout the utterance[3]. If the speech output exceeds the allowable variability of the target regions (which we refer to as sensorimotor error), sensory mismatches between speech output (e.g., formant values registered by the *Auditory State Map* cells in Fig. 1) and the desired target regions (e.g., lower and upper bounds for each formant over time, coded by the projections labeled *Auditory target region*) are detected in the feedback loops. The mismatch detected by the auditory feedback loop is an *auditory error* (calculated by the *Auditory Error Map* cells) and the mismatch detected by the somatosensory feedback loop is a *somatosensory error*. To correct the errors, the feedback loops issue appropriate corrective motor commands (*Feedback commands* in Fig. 1), based on the sensory-motor transformations acquired in the babbling phase (specifically, the transformation from auditory and somatosensory errors into articulatory movements that reduce these errors). As further discussed in Section 7.1, only the auditory feedback loop is implemented in this study so as to maintain tractability.

According to the DIVA model, infants initially speak using auditory feedback control alone. For each sound in the language, they first tune the boundaries of the sound's auditory target region by listening to examples spoken by others. Then, to produce the sound, they read out the auditory target from memory and try to reproduce it using the auditory feedback loop. Speaking in this way is like trying to sing along to a tune as one hears it playing on the radio for the first time--there will always be a delay between what the listener hears and what he or she produces. At each point of time, therefore, there will be a mismatch between the played tune (analogous to the auditory target in the DIVA model) and the tune the listener hears himself producing (analogous to the auditory feedback in the model). This time-lag problem is probably one of the reasons why feedback control is not the dominant mechanism used by adults to control their speech.

2.2 Feedforward control subsystem

To prevent errors that inevitably occur when using feedback control alone, speakers also utilize open-loop or feedforward control mechanisms (Borden, 1979; Kent, 1976; Kent & Moll, 1975; Lashley, 1951; Max et al., 2004; Smith, 1992). The *feedforward control subsystem* of DIVA issues preprogrammed motor commands (*Feedforward commands* in Fig. 1). These

---

[3]The use of target regions (rather than point targets) is an important aspect of the DIVA model that provides a unified explanation for a wide range of speech production phenomena, including motor equivalence, contextual variability, anticipatory coarticulation, and carryover coarticulation (for details see Guenther, 1995).

commands do not rely on error detection via sensory feedback and therefore avoid the time-lag problems of feedback control. Speaking with feedforward control is analogous to singing a song from memory. Ideally, there is no mismatch between desired and actual sensory consequences, and thus, feedback control never comes into play (but see Scott & Ringel, 1971). The strongest evidence for feedforward control during speech production comes from a variety of reports that illustrate that speech motor behavior is quite resistant to temporary loss in somatosensory information (Abbs, 1973; Gammon et al., 1971), to temporary (Gammon et al., 1971) and permanent (Cowie & Douglas-Cowie, 1983; Goehl & Kaufman, 1984) loss in auditory information, as well as to combined loss in both modalities (Gammon et al., 1971). Finally, feedforward control is also responsible for the after-effects observed after the sudden removal of speech perturbations that were applied for extended periods of time (e.g., Purcell & Munhall, 2006; Villacorta, Perkell, & Guenther, 2007).

The feedforward commands are formed by projections from the speech sound map (SSM) cells to the DIVA model's motor cortex (both directly and via the cerebellum, see Fig. 1). The SSM cells can be interpreted as forming a "mental syllabary" (e.g., Levelt & Wheeldon, 1994) which may be described as a "repository of gestural scores for the frequently used syllables of the language" (Levelt, Roelofs, & Meyer, 1999, p. 5). Support for this interpretation comes from a recent functional imaging study (Peeva et al., 2010) showing that activity in the left ventral premotor cortex, where the neurons corresponding to the SSM cells are hypothesized to reside, is indeed related to syllable processing. Following this interpretation, each SSM cell in the model codes the feedforward commands required to produce one specific well-learned syllable. According to our account, higher-level brain regions involved in phonological encoding of an intended utterance (e.g., anterior Broca's area) sequentially activate the SSM cells corresponding to the syllables to be produced, with the timing of syllable activation being controlled by the basal ganglia. This account is formalized by the GODIVA model, which is an extension of the DIVA model, and is described elsewhere (Bohland, Bullock, & Guenther, in press;Civier, 2010).

Feedforward commands are acquired in the DIVA model's learning stage corresponding to early word imitation. To learn the feedforward commands for the syllables used in the simulations, the model first attempts to produce each syllable using auditory feedback control. On each attempt, the model updates the feedforward commands to incorporate the corrective commands generated by the auditory feedback control subsystem on that attempt. This process results in more accurate feedforward commands for the next attempt. After 3 to 5 attempts, the feedforward commands become fully tuned, i.e., they are sufficient to produce the syllable with very few errors.

2.3 Monitoring subsystem

Earlier we mentioned that both auditory and somatosensory errors are used to generate corrective feedback commands. Here we hypothesize that the auditory errors are also inspected by a *monitoring subsystem*, which was not previously described by the DIVA model (depicted at the top of Fig. 1). The monitoring subsystem is responsible for detecting and repairing errors that are too large to be corrected by feedback loops. A *repair* (by the monitoring subsystem) is triggered only as a last resort because it requires a motor "reset," thus, interruption of speech flow; a *correction* (by the feedback control subsystem), on the other hand, usually passes unnoticed by listeners because corrective commands do not interrupt speech flow. The monitoring subsystem detects errors by inspecting auditory feedback alone (without consulting somatosensory feedback) following Levelt's (1983,1989) implementation of a speech monitor (perceptual loop theory); a theoretical design consideration that has some empirical support (Postma, 2000). Levelt (1989) limited the scope of his model to language formulation errors, but it may be extended to deal with non-phonetic errors as well (Postma, 2000, p. 102).

Unlike the feedforward and feedback subsystems, the monitoring subsystem is not yet associated with specific neural substrates. For this reason, its components are algorithmically implemented, i.e., a computer algorithm performs the computation without using numerical integration of differential equations. The *self-repair controller* component, whose function is to execute the error repair, is described in Section 4.1. It is engaged by the *excessive error detector,* the component that detects when an error too large to be corrected by feedback loops occurs. The likelihood that the excessive error detector will trigger an error repair, R_{error}, at time *t*, is calculated by the following probabilistic formula:

$$P((R_{error}(t)) = \varepsilon * f_{masking}(f_{size}(\Delta Au(t)), I_{masking})$$

where the parameter ε is set to $2*10^{-3}$, *t* is time in milliseconds, $I_{masking}$ is masking noise intensity in decibels, and $\Delta Au(t)$ is the auditory error at time *t*, calculated by auditory error cells in the DIVA model's superior temporal cortex. The f_{size} function quantifies the size of the auditory error (in hertz), as further described in section 3.1. While a perception-based scale may be more appropriate to measure the size of the auditory error, this study uses the hertz scale as a first approximation (cf. Villacorta et al., 2007). Lastly, the $f_{masking}$ function is described in the fourth modeling experiment, where it is used to calculate the effect of masking noise. When noise is absent ($I_{masking}$ = 0 dB) as in the first three modeling experiments, this function simply returns the auditory error size it gets as input, i.e., $f_{masking}(x, 0) = x$.

We developed the repair-likelihood formula based on several design considerations. First, errors do not need to reach some threshold size in order to be repaired; instead, error repair is more likely to occur for large errors as compared to small errors (cf. Zimmermann, 1980c, p. 131). This design consideration is biologically plausible, and is supported by data showing that occurrence of dysfluency can be predicted by the extent of sensorimotor error, measured as lack of anticipatory coarticulation (Stromsta, 1986, p. 79; Stromsta & Fibiger, 1981). Second, the formula is probabilistic to account for the fact that not all speech errors are repaired (e.g., Levelt, 1983, p. 59). Nondeterministic repair may explain the variability of stuttering due to emotions, muscle tension, and other perceptual and physiological factors not currently addressed in the DIVA model (Alm, 2004, pp. 332, 344, 359; Smith, 1999; Van Riper, 1982; Zimmermann, 1980c). While a probabilistic formula cannot capture the exact effects these factors have on stuttering, it may at least imitate the unpredictable nature of the disorder.

The only constant in the repair-likelihood formula is the parameter ε which stands for the relation between a given auditory error size and the likelihood of the error being repaired. We initially considered letting this parameter fluctuate to reflect changes in the mental state of the speaker under fluency-inducing conditions (PWS may feel more comfortable and confident when speaking slower, or feel less anxious, believing their errors pass unnoticed, when speaking under masking noise, see Shane, 1955), but past studies actually reject this scenario (for masking noise see Adams & Moore, 1972; for prolonged speech following treatment see Ingham & Andrews, 1971). Modification of the ε parameter can also represent an abnormal sensitivity to errors, but since we assume that the error sensitivity of PWS is normal (for an extended discussion see Section 7.1), no modification was required here either. As stated in the next section, the only modification required to induce stuttering in the model was in the gain parameters of feedforward and feedback control.

3 Modeling experiment 1: Bias toward feedback control leads to sensorimotor errors during fluent speech

There is evidence that PWS have errors not only in their dysfluent speech preceding repetitions, but in their fluent speech as well (see Bloodstein, 1995, p. 35; Postma & Kolk, 1993, p. 482;

Stromsta, 1986, pp. 189, 191; Van Riper, 1982, pp. 396, 402). As noted earlier, a possible cause of these errors is PWS's impaired ability to read out feedforward commands resulting in an over-reliance on auditory feedback, which can lead to delays due to time lags associated with afferent information reaching the central nervous system (Guenther et al., 2006; Neilson & Neilson, 1987, 1991; Van Riper, 1982). The DIVA model can be biased away from feedforward control and toward feedback control by utilizing a relatively low gain on the feedforward control subsystem (low α_{ff}) in conjunction with relatively high gain on the feedback control subsystem (high α_{fb}).

Without sufficient feedforward control, sensorimotor errors arise, and the auditory feedback loop detects them as auditory errors. Consistent with CNS physiological delays, the DIVA model takes 20 ms to perceive auditory feedback, and then an additional 45 ms to execute the corrective feedback commands (Guenther et al., 2006); but since some consonant sequences contain articulatory events separated by as little as 10 ms (Kent, 1976; Kent & Moll, 1975), the DIVA model is expected to be too slow in issuing the feedback commands and thus unable to correct the errors quickly. The errors will then grow large, increasing the likelihood of a motor "reset." Nevertheless, in this modeling experiment none of the errors triggered a repetition. Simulations that do not produce stuttering can be compared to data on PWS's fluent speech, which are more abundant than data on their dysfluent speech.

Errors due to reliance on auditory feedback are expected to be the greatest for phonetic events with rapid formant transitions since the rate of acoustic change will exceed the feedback controller's ability to make timely adjustments (also see Borden, 1979). The result will be auditory error. The prominence of errors in rapid transitions can also be observed in auditory tracking experiments which force subjects to rely heavily on auditory feedback--the largest errors follow rapid shifts in the auditory target (e.g., Nudelman, Herbrich, Hoyt, & Rosenfield, 1987).

3.1 Methods

During the performance stage, we biased the DIVA model away from feedforward control and toward feedback control by modifying the feedforward and feedback gain parameter values to $\alpha_{ff} = 0.25$ and $\alpha_{fb} = 0.75$ (from the default settings of $\alpha_{ff} = 0.85$ and $\alpha_{fb} = 0.15$). Like the default gain parameter values, the modified parameters also sum to one. This modeling constraint formalizes the assumption that the increase in reliance on feedback control is proportional to the severity of impairment in feedforward control. Simulations were performed with both the modified settings (*stuttering DIVA version*) and the default settings (*non-stuttering DIVA version*).

The performance stage was preceded by the learning stages (described in Section 2), in which both versions of the model first learned the desired auditory target regions, and then the corresponding feedforward commands. This learning procedure was employed for all utterances used in this study. It is important to note that the aforementioned biasing of the model took place only after learning was complete (i.e., both versions of the model used the same default parameter settings for the learning stages). To investigate developmental aspects of stuttering, we could have possibly modified the gain parameters earlier on (already during the learning stages), but due to time and space constraints, we decided against it. Despite this focus on the mature system, we acknowledge that the development of the disorder is an important area for future studies, and discuss it further in Section 7.4.

To assess performance on different formant transition rates, we used two auditory targets for the first modeling experiment: the utterance /bid/ which requires a rapid second (F2) and third (F3) formant transitions, and the utterance /bad/ which requires slower transitions. The difference between the two targets can be seen by observing the lines indicating the F2 and F3

formants in Fig. 2. The slopes of these lines are steep in the transition from /b/ to /i/ in Fig. 2 (a) and almost flat in the transitions from /b/ to /a/ in Fig. 2(c).

In all experiments, auditory errors were based on formant values. An auditory error is a three-dimensional vector whose components are *formant errors* corresponding to the three first formants. For each time interval within the utterance, a *formant error* is defined as the distance (in hertz) of the actual produced formant value from the target region (bounded by upper and lower limits) for that formant. To help visualize auditory errors (as well as calculating the likelihood of repair for a given error, see Section 2.3), we use the f_{size} function to calculate *auditory error size*, that is the sum of the three formant errors that constitute the auditory error. If, within a given production, all three formants are within their corresponding target regions, auditory error size at that time point is 0 Hz (e.g., Fig. 2(a) at 158 ms).

3.2 Results

Fluent speech was produced in all simulations. Fig. 2 shows the target region (low and high boundaries of the three formants), the auditory feedback (values of the three formants), the formant errors, and the auditory error size (all shown in hertz) of the non-stuttering (left plots) and stuttering (right plots) versions of the DIVA model during a fluent production of both /bid/ (upper plots) and /bad/ (lower plots). To illustrate the performance of the model during the voiced (white background) as well as the silent (shaded background) periods of the utterance, the target region and auditory feedback data are plotted as if voicing is always on. However, to emphasize that auditory feedback is not available when voicing is off, auditory error size and formant errors are plotted in the voiced periods only. Notice that formants with higher frequencies have wider target region. This represents the reduced sensitivity of speakers to frequency differences in the higher frequency ranges, which translates to greater allowed variability.

Note that the stuttering version, which relies more on feedback control, had larger auditory errors than the non-stuttering DIVA version, and that the largest auditory error occurs during formant transitions. Moreover, at one point the size of the auditory error made by the stuttering DIVA version on /bid/ (rapid transition) exceeded 1000 Hz--much larger than the maximum size of auditory error made by the stuttering DIVA version on /bad/ (slow transition), which was below 500 Hz.

3.3 Discussion

The simulation results demonstrate that a bias toward feedback control leads to increased auditory error, especially on utterances with rapid formant transitions, as the utterance /bid/.

Do PWS show similar acoustic patterns during fluent speech production? Although the effect is not always statistically significant (Caruso, Chodzko-Zajko, Bidinger, & Sommers, 1994; Subramanian, Yairi, & Amir, 2003; Zebrowski, Conture, & Cudahy, 1985), there are several reports that both adults and children who stutter exhibit abnormal F2 transitions during fluent speech (Chang, Ohde, & Conture, 2002; Robb & Blomgren, 1997). PWS may have abnormal F1 transitions as well, but since the bulk of data links stuttering to abnormalities in F2 rather than F1 transitions, we here focus on F2 transitions only (e.g., Agnello, 1975; also see Subramanian et al., 2003; Yairi & Ambrose, 2005, p. 259; Yaruss & Conture, 1993). The data most relevant to our simulation comes from a pair of studies (Blomgren, Robb, & Chen, 1998; Robb & Blomgren, 1997) which examined F2 transitions and steady states in CVC words produced by PWS and PNS. Blomgren et al. found that PWS had greater vowel centralization than PNS. Like our simulations, this finding suggests that PWS have larger auditory errors. Moreover, the studies resemble our simulations in another aspect: the performance of PWS depended on the vowel they produced.

Compared to PNS, Blomgren et al. (1998) found that PWS produced /Cit/ and /Cat/ syllables differently. Specifically, PWS had significantly lower average F2 (defined as the average of F2 values measured in five equally spaced points spanning the vowel production, which was predetermined to start 40ms after voicing onset, see Blomgren et al., 1998, p. 1044) in the /Cit/ syllables but not the /Cat/ syllable. To compare these results with the model simulations, we approximated the average F2 formant error for Blomgren et al. data by assuming that the PNS reproduced the desired auditory target without auditory errors. The average F2 formant error of the PWS is roughly equal then to the difference in average F2 between the two groups. For the study in question, the group differences in average F2 values were 125 Hz for /bit/, and 94 Hz for /bat/. These results are qualitatively similar to the stuttering DIVA version's average F2 formant errors: 127 Hz on /bid/, and 68 Hz on /bad/[4] (Fig. 2(b) and (d)). Although the stuttering version did not have errors toward the end of the /i/ vowel in /bid/, the large vowel-initial errors were significant enough to increase the average F2 formant error (notice that error measurement started 40 ms post-voicing, thus keeping with the method used by Blomgren et al., 1998). There were also some vowel-initial F2 formant errors in the stuttering DIVA version's production of /bad/, but they were much smaller.

Other studies examining formant values are consistent with our simulations as well. For example, several studies suggest, like Blomgren et al. (1998), that PWS have significant F2 formant errors during the vowel of /CiC/ syllables (Hirsch et al., 2007; Klich & May, 1982; Prosek, Montgomery, Walden, & Hawkins, 1987). Furthermore, one of the studies (Prosek et al., 1987) also supports the Blomgren et al. finding of no significant average F2 difference between PWS and PNS when producing /CaC/. In conclusion, the model simulations account for several published results. According to our hypothesis, /CiC/ syllables are more likely to have errors because /i/ has the highest F2 of all vowels (~2250 Hz), requiring larger, faster F2 transition rates in many consonant-vowel environments. By contrast, the vowel /a/, with a much lower F2 (~1150 Hz), typically requires smaller, slower, F2 transition rates in similar consonant environments[5].

Close inspection of Fig. 2(b) reveals that the vowel-initial errors made by the stuttering version of DIVA on /bid/ are in large part due to a delayed initiation of the F2 transition. The delay is a direct consequence of the bias toward feedback control--a control mechanism with inherent time lags. This pattern is similar to that reported by Robb and Blomgren (1997). Using the same subjects and utterances from Blomgren et al. (1998), the 1997 study provided evidence that the PWS group exhibited delayed F2 transitions into the vowel as compared to the PNS group. For a comparison between our simulations and the Robb and Blomgren results, we have plotted the authors' data in Fig. 2. Specifically, the median F2 at voicing onset, as well as 30 ms and 60 ms later, are plotted in Fig. 2(a) for the PNS, and in Fig. 2(b) for the PWS. Although the F2 values of the DIVA simulations are reduced in comparison to the median data points plotted, they do fall within the range of the Robb and Blomgren (1997) data. Note that the errors made by PWS appear to be associated with delays, as was the case for the errors made by the stuttering DIVA version.

---

[4]The syllables from the Blomgren et al. (1998) and Robb and Blomgren (1997) studies terminate with /t/, while the syllables uttered by the DIVA model terminate by /d/. However, this should not make a difference on the transition from the initial /b/ to the middle vowel. We did not use /t/ for our simulations because the DIVA model cannot produce that sound. It is also noteworthy that in the above studies the slow transition from the /b/ to /a/ in /bat/ is rising rather than falling as in the DIVA simulations of /bad/. This is probably because the DIVA model uses a relatively high F2 locus for /b/ in that context (compare to the inter-subject differences in transition direction noted by Kewley-Port, 1982, p. 358). We believe the comparison of the simulations with the studies remains valid because in both cases the transitions from /b/ to /a/ are slow and errors are minimal.

[5]An additional factor contributing to the difference in auditory error size between the /i/ and /a/ vowels may be their different tactile contact patterns leading to different somatosensory feedback processing. Since we did not simulate the somatosensory feedback loop, this hypothesis cannot be tested here. Yet, the apparent correlation across all the acoustic spectrum between the F2 formant errors PWS make on vowels, and the vowels' F2 values (Hirsch et al., 2007; Prosek et al., 1987) suggests that error size depends more on the acoustics than on the tactile pattern of vowels.

The similarity of our simulations to the behavioral data from both Robb and Blomgren (1997) and Blomgren et al. (1998) suggests that the gain parameter values of the stuttering DIVA version ($\alpha_{ff} = 0.25$, $\alpha_{fb} = 0.75$) adequately reflect the relative contributions of feedforward and feedback to the speech motor control of PWS. Simulations with other combinations of low α_{ff} and high α_{fb} generated abnormal acoustics as well, but no parameter combination resulted in a better fit to the data than reported here. Nevertheless, auditory error size increased as α_{ff} got lower (and α_{fb} got higher), which may indicate that the gain parameters are related to the severity of stuttering (for quantitative comparison between different gain parameter values see Terband et al., 2009).

Fig. 2(b) indicates that the errors made by the stuttering DIVA version on /bid/ are due to a delay in the onset of the F2 transition as well as a slower rate of transition as compared to the auditory target region. This slow transition rate can be attributed to the fact that the feedback control subsystem is limited in the speed of movement it can handle effectively (van Lieshout et al., 1996a, p. 89); faster transition rates may be achieved by increasing the gain on the feedback control subsystem even more, but this has the risk of increasing the undesirable overshoots and oscillations generated when feedback commands over-correct errors, leading to new errors of the opposite sign (Ghosh, 2005; Stromsta, 1986, p. 203). There is recent empirical evidence that PWS exhibit reduced formant transition rates. Chang et al. (2002) found that children who stutter exhibited reduced F2 transition rates compared to normally fluent children for the syllables /beC/ and /meC/, tokens with F2 transitions[6].

The observation that PWS exhibit F2 transitions that are delayed (Robb & Blomgren, 1997) and slow (Chang et al., 2002) specifically on tokens that have bilabial consonants preceding high-F2 (or front) vowels is not unexpected. The contrast between the low F2 locus of the bilabials and the high F2 frequency of the vowels forces PWS to produce rapid transitions, which are difficult for a feedback-based control system to track accurately. Reports of PWS having errors on high-F2 vowels preceded by consonants of other classes as well (Hirsch et al., 2007; Klich & May, 1982; Prosek et al., 1987) suggest that the problem is not limited to bilabials, yet, for consonants with very high F2 loci (e.g., alveolars) the error pattern may be quite different. Owing to the high F2 locus, PWS are expected to experience errors when combining such consonants with vowels whose F2 is low rather than high. The data on the older children who stutter in Chang et al. provide some support for this prediction.

Because the DIVA model controls an artificial vocal tract (Maeda, 1990), we could have potentially compared the model simulation to the tongue kinematics of PWS's fluent speech. In bilabial contexts, the tongue is the articulator whose position correlates with F2 the most (Stevens, 1998), and should reflect the slowness PWS experience in rapid F2 transitions. Unfortunately, such comparison is not feasible at this time because most studies measuring the tongue movement speeds of PWS either did not report data (e.g., Alfonso, 1991; McClean, Tasko, & Runyan, 2004) or only reported qualitative data (McClean & Runyan, 2000) from fluent speech segments with rapid transitions of the second formant (cf. Wood, 1995). Nevertheless, data we collected on tongue (and other articulators) kinematics during dysfluent speech segments will be compared with the simulations of the second modeling experiment.

---

[6]Based on their data points (plotted in Fig. 2), Robb and Blomgren (1997) calculated F2 transition rates as well, but in contrast with Chang et al. (2002), they did not find PWS to be slower. We believe this may be sourced to their measurement method, which was based on the slope of a straight line connecting the 0 ms and the 30 ms post-voicing data points. Since PNS's formant transition for /b/ may not span up to the 30 ms data point (Lieberman & Blumstein, 1988, p. 224; Miller & Baer, 1983), the slope of the line may be flatter than the slope of the actual transition (as suggested in Fig. 2(a)). PNS will appear then slower than they really are, artificially changing the experiment results. The same argument applies to similar measurement Robb and Blomgren (1997) took using the 60 ms post-voicing data point.

Even though no repetitions occurred in this experiment, when we repeated each simulation multiple times, repetitions did occur. The most likely simulation to produce a stutter was the /bid/ production by the stuttering DIVA version, which is also the simulation where the largest auditory errors occur. This apparent relation between auditory error size and frequency of repetitions is investigated in the third modeling experiment. The third experiment builds on the second modeling experiment that investigates how repetitions are generated.

4 Modeling experiment 2: Sensorimotor errors lead to sound/syllable repetitions

The correspondence between published data and the DIVA simulations in the first modeling experiment supports the view that abnormal speech motor patterns observed in the fluent speech of PWS could be the result of auditory errors stemming from a motor control strategy that relies too heavily on feedback. We also contend that auditory errors, especially if large, can cause repetitions commonly observed in PWS. Recall from Fig. 1 that the DIVA model has a monitoring subsystem that serves to repair auditory errors that are too large to be corrected by the feedback control subsystem. The monitoring subsystem repairs an error by repeatedly restarting the current syllable until the auditory error is small enough to allow the forward flow of speech. Therefore, repetitions result from the attempts to repair large sensorimotor errors.

This general view that repetitions arise as a *self-repair* strategy is not new (Postma & Kolk, 1993). However, previous attempts to apply the self-repair hypothesis to stuttering have largely assumed linguistic (phonologic) formulation errors (Postma & Kolk, 1993; Van Riper, 1971, 1982) rather than sensorimotor errors. Unfortunately, this assumption has not been borne out experimentally (Burger & Wijnen, 1999; Howell & Vause, 1986; Stromsta, 1986; Wingate, 1976) and clear evidence for a strong link between stuttering and phonological processing is lacking (Hennessey, Nang, & Beilby, 2008; Melnick & Conture, 2000; Nippold, 2002), although further research is required (Conture, Zackheim, Anderson, & Pellowski, 2004; Sasisekaran, De Nil, Smyth, & Johnson, 2006). An added advantage of the sensorimotor self-repair model is that it could account for stuttering-like behavior observed in healthy speakers when exposed to delayed (Black, 1951; Lee, 1951; Van Riper, 1982; Venkatagiri, 1980; Yates, 1963) or phase shifted (Stromsta, 1959) auditory feedback. Presumably, these speakers were making repairs to perceived auditory error.

This view also suggests that errors in the fluent and dysfluent speech of PWS have a similar origin. There appears to be some empirical support for such a contention. Studies that have focused on the acoustic characteristics of repetitions indicate that F2 variations are observed (Agnello, 1975; Harrington, 1987; Howell & Vause, 1986; Stromsta, 1986; Yaruss & Conture, 1993) that are similar to the errors in the fluent speech of PWS. Furthermore, just as PWS are more likely to make errors in their fluent speech when uttering syllables with high-F2 vowels, so in their dysfluent speech these syllables seem more likely to be stuttered. Although data on the whole speech corpus are lacking, if we are only to consider syllables that start with a vowel (reported by Johnson & Brown, 1935; reproduced in Wingate, 2002, p. 309), those starting with the /ε/, /I/, or /i/ vowels are among the most likely to be stuttered[7] (compared to syllables that start with other vowels, a greater proportion of these syllables are stuttered). This information, combined with the well-known fact that most repetitions occur on syllables that start with consonants (Bloodstein, 1995; Brown, 1938; Wingate, 1976), suggests that repetitions may be related to speech elements that require rapid formant transitions and are therefore most prone to auditory error.

---

[7] Only one group of syllables – those starting with the vowel /ɝ/ – is more likely to be stuttered than syllables with high-F2 vowels. The fact that /ɝ/ does not have high F2 may seem to contradict our view, but given the general difficulty in pronouncing this vowel, challenging pronunciation (rather than rapid F2 transitions) is the more likely reason for the frequent stuttering.

The goal of this modeling experiment is to use the same stuttering DIVA version used in the first modeling experiment to simulate a self-repair of sensorimotor errors by the monitoring subsystem, and compare these simulations to the acoustic and kinematic characteristics of sound repetitions made by a single PWS.

4.1 Methods

We used the stuttering DIVA version to simulate productions of the nonsense phrase "a bad daba" (pronounced as /ebæd dæbə/). The *DIVA dysfluent production* started from the vocal tract configuration at rest. The DIVA model had to perform a rapid transition toward the initial configuration of the utterance (the configuration for the first vowel of "a bad daba"), and thus large errors, leading to self-repair, were likely. The *DIVA fluent production* served as a control condition. In order to generate a fluent, error-free production, we used DIVA simulations at half the normal movement speed, a fluency-inducing technique that will be further investigated in the third modeling experiment.

To simulate repetitions, we algorithmically implemented the self-repair controller introduced in Section 2.3. When the self-repair controller receives a repair trigger from the excessive error detector, it first disrupts the initial attempt (or the *1st attempt*) to produce the syllable by suspending phonation as soon as possible, which given the DIVA model time constraints, takes 65 ms from the time of error (Guenther et al., 2006)[8]. Phonation is temporarily suspended by updating the cells that correspond to voicing control parameters in the DIVA model (specifically, parameters AGP and AGO of the Maeda articulatory synthesizer) and which reside in the voicing control map (not explicitly defined in previous versions of the model, this map encompasses the motor larynx and motor respiration components defined in Guenther et al., 2006). While voicing is off, the feedback control subsystem issues corrective motor commands to reduce the error that triggered the repair. These commands reposition the articulators in order to place them in an appropriate position for the upcoming attempt to produce the syllable correctly. When voicing turns back on, the self-repair controller performs a motor "reset" by initiating the next attempt (*2nd attempt*) to produce the syllable. If the 2nd attempt still contains an auditory error, the excessive error detector will once again instruct the self-repair controller to initiate a self-repair. The cycle repeats until speech output is either error-free or the error is small enough to be handled by the sensory feedback loops. At this point no further disruption occurs. Each repetition consists therefore of one or more disrupted attempts followed by a *final complete production* of the utterance. This account of error repair is drawn from previous theoretical work (Postma & Kolk, 1993; Stromsta, 1986; Van Riper, 1982).

The simulations were compared with kinematic and acoustic data based on a single speaker drawn from the Walter Reed-Western Michigan University Stuttering Database, a large speech acoustic and physiological dataset of adults who do and do not stutter (McClean et al., 2004; Tasko, McClean, & Runyan, 2007). The speaker was 19 year old male (denoted CXX) with no reported history of stuttering therapy. A behavioral analysis of videotaped samples of CXX's monologue and oral reading was performed by expert clinicians. His score on the Stuttering Severity Instrument-3 (Riley, 1994) was 26, which placed him in the range of a moderate fluency impairment. The analysis reported here is based on CXX's repeated production of a nonsense phrase "a bad daba" (/ebæd dæbə/) performed once at habitual rate and loudness, and once at half the normal speaking rate. For the reduced rate condition, the instructions were "Now I want you to speak in half rate. But don't insert pauses, instead, stretch the sound."

---

[8]This estimation is based on the minimum time it takes to accelerate the tongue muscle, which includes 30 ms from EMG to onset of acceleration (Guenther et al., 2006). The laryngeal muscles, which are involved in interruption of phonation, have a similar delay relative to EMG, as demonstrated by Poletto, Verdun, Strominger, and Ludlow (2004).

Subject CXX was selected because he usually stuttered on the initial vowel of the test phrase in the habitual rate condition, but never in the reduced rate condition.

The participant was seated in a sound-treated room while chest wall motion, orofacial movement and speech acoustics were recorded. Recordings were obtained of the two-dimensional positions of the lips, tongue blade, jaw, and nose within the midsagittal plane by means of a Carstens AG100 Articulograph (Carstens Medizinelektronik GmbH, Lenglern, Germany), an electromagnetic movement analysis system commonly known as EMA. Three sensor coils (3 mm × 2 mm × 2 mm), or pellets, were attached with biomedical tape to the bridge of the nose and the vermilion borders of the upper (UL) and lower lip (LL). Two more sensor coils were attached with surgical adhesive (Isodent) to the tongue blade (TB) approximately 1 cm from the tip and at the base of the mandibular incisor (MI). Sensor locations are schematically presented in Fig. 3(c). The acoustic speech signal was transduced with a Shure M93 miniature condenser microphone (Shure, Inc., Niles, IL) positioned 7.5 cm from the mouth and the microphone-amplifier setup was calibrated to permit measurement of absolute sound pressure levels. The orofacial kinematic and acoustic signals were digitized to a computer at 250 Hz per channel and 16 kHz respectively. The subject performed approximately 12–15 speech tasks, each 30 seconds in length.

After data acquisition, the lip, jaw, and tongue movement signals were low-pass filtered (zero phase distortion fifth-order Butterworth filter) at 8 Hz and the nose signal at 3 Hz. While head movements during recording were slight (<1.0 mm), nose sensor movements were subtracted from the lip, tongue, and jaw movement signals in the X and Y dimensions in order to minimize any head movement contributions. The kinematic signal was downsampled to 146 Hz and the acoustic signal was upsampled to 22.050 kHz to fit the file format supported by the TF32 (Time-Frequency analysis for 32-bit windows) software (Copyright 2000 Paul Milenkovic. Revised May 7, 2004). The TF32 software was then used for data analysis and presentation. To compare the data to the model simulations, we calculated the approximate positions of the sensors on DIVA's artificial vocal tract (Maeda, 1990) and sampled their X and Y coordinates during the simulations. Since EMA data do not provide detailed information regarding specific vocal tract dimensions, scaling the model simulation for comparison with the empirical data was rough at best. As a result, comparisons between the model simulations and subject data were necessarily qualitative. Acoustic analysis was carried using the SpeechStation 2 software (Copyright 1997–2000 Semantic Corporation. Version 1.1.2). Formants were identified by the first author using linear predictive coding (LPC) technique (12 coefficients) in combination with a wideband spectrogram. Measurements were taken in the first glottal pulse of each disrupted attempt or fluent production with the condition that both the first and second formants be visible on the spectrogram.

4.2 Results

Fig. 3(a) is an F1–F2 plot showing the formant values of both fluent and dysfluent productions of the initial vowel in the phrase "a bad daba" (/ebæd dæbə/) made by the stuttering DIVA version (in blue) and the test subject CXX (in other colors). As expected, a self-repair was triggered in the beginning of the DIVA dysfluent production. The 1st attempt to produce the /e/ is marked by the big square (its location specifies the F1 and F2 of the auditory feedback at the beginning of the attempt), which is connected by a dashed line to the small square, marking the 2nd attempt to produce this sound. Another line connects that square to a circle that marks the final, complete production of /e/. Four separate sound repetition sequences made by CXX are depicted in the same manner, with each repetition plotted in a different color. A blue star marks the F1–F2 position of the initial vowel for the DIVA fluent production where no self-repairs occurred. A black star similarly marks the average F1–F2 position of the initial vowel for CXX's fluent productions of the utterance (F1 mean = 466 Hz, SD = 38 Hz, F2 mean =

1660 Hz, SD = 36 Hz). As mentioned in the Methods section, in both cases fluency was induced by reduced speaking rate. The DIVA model and CXX used slightly different frequencies to produce /e/ because they have different vocal tract dimensions, a factor that modulates the formant values produced by a speaker (Peterson & Barney, 1952).

Note that, although there are generally changes in F1 and F2 for production attempts of a given replicate, the most consistent finding for both the simulation and the speaker data is an upward shift in F2. Shifts in F1 are generally smaller and less consistent with regard to the direction of the shift. The F2 values of the 1st and 2nd attempts by the DIVA model are 1755 Hz and 1808 Hz, respectively. The final complete production has an F2 of 1928 Hz. Second formant values for the DIVA simulation and the eleven repetitions made by CXX are listed in Table 1 (the four repetitions appearing in Fig. 3 are marked with an asterisk). For each repetition, the F2 values at the starting points of the disrupted attempts, as well as of the final complete production, are listed. For repetitions that included only a single disrupted attempt, the columns for the second attempt are left blank. An unpaired samples t-test ($p < 0.0004$) showed that there was an increase in F2 from the disrupted attempts to the final complete production. Also listed in the table are the durations of CXX's and DIVA's disrupted attempts. The durations of the 1st and 2nd attempts in the DIVA model simulation were 66 ms and 76 ms respectively, indicating that the errors were quickly detected. According to the time constants mentioned in Section 3, the model can detect an error no sooner than 20 ms following voicing onset, and in the simulations errors were detected once 21 ms post-voicing (1st attempt), and once 31 ms post-voicing (2nd attempt).

Fig. 3(b) shows the articulatory movements for the repetitions made by the DIVA model and CXX. For each repetition (depicted in the same color as in Fig. 3(a)), four continuous lines trace the positions of the 4 sensors (TB, UL, MI, and LL) from the beginning of the first attempt (big squares) to that of the second attempt (small squares), and from there to the beginning of the final complete production (circles). While the simulation generally agrees with CXX data regarding the direction to which the articulators are moving, there are some observable differences. First, the tongue blade of DIVA does not seem to move backwards similarly to CXX's tongue blade (even though in both cases the tongue blade moves upwards). Second, the DIVA simulation produce tongue blade movements that are smaller in magnitude. This difference in movement magnitude and orientation is not that surprising given that there are differences in vocal tract morphology (and correspondingly, differences in articulatory-acoustic relations) between CXX and the simplified Maeda (1990) articulatory synthesizer used in the DIVA simulations, the previously mentioned difficulties with scaling the model simulations, and some differences in the details of the reference frame used for each.

4.3 Discussion

The speaker data show that within a single repetition, the disrupted attempts exhibit a formant pattern converging toward the F2 value for the fluent production of the vowel. This suggests there may be some corrective action occurring through the repetition. The DIVA simulation, which is based upon the self-repair hypothesis, also shows such a pattern. This suggests that it is plausible that repetitions may result from a self-repair of sensorimotor error. It is also noteworthy that the self-repair investigated here primarily affects F2. This can be explained by the second formant being the primary locus of auditory error in stuttering (see Section 3.3).

Patterns similar to our empirical and simulation results have been previously observed. Stromsta (1986) reported on a child who repeated part of the initial vowel in the word "apple". The disrupted attempts did not show any evidence of acoustic transitions one would expect for a vowel followed by a /p/, suggesting an error. In line with our simulation, the error was limited to the disrupted attempts, and was resolved when the utterance was finally produced. A variety of studies have described F2 abnormalities during repetitions. In Howell and Vause's (1986)

data for repetitions on /CiC/ syllables, the F2 values were abnormally low in the disrupted attempts of the repetitions, but shifted to higher values in the final, complete production. Harrington (1987) demonstrated repetitions where abnormal F2 transitions were present in the disrupted attempts but not in the final complete production. Similar abnormal F2 transitions were also observed in 16% to 45% of the disrupted attempts in the Yaruss and Conture (1993) study of repetitions by children who stutter. Unfortunately, the aforementioned findings cannot be directly compared to our DIVA simulation since none of the studies could generalize the abnormalities to all repetitions (see also Subramanian et al., 2003), for one of several reasons: (a) relatively few repetitions were analyzed (Howell & Vause, 1986; Yaruss & Conture, 1993); (b) no quantitative measures were used (Harrington, 1987); or (c) each participant repeated on different words (Yaruss & Conture, 1993).

The acoustic data cannot reveal the speech motor behavior during the silent periods between the disrupted attempts of the repetition. Therefore, we also evaluated articulatory kinematic data. Fig. 3(b) shows that CXX repositions the articulators between the disrupted attempts. Between the 1st and 2nd attempts, and between the 2nd attempt and the final complete production, the tongue blade moves upwards and does not return to its initial position. The same pattern of repositioning is also observed in the repetition made by the stuttering DIVA version. This suggests that both the speaker and the DIVA model may be repositioning articulators by issuing corrective feedback commands to reduce the error that triggered the self-repair. These results are in line with reports by Zimmermann (1980a) who reported "consistent repositioning of articulators occurs during oscillatory behaviors [i.e., repetitions] …" (p. 117).

The acoustic data in Table 1 and Fig. 3(a) suggest that the repositioning of articulators by CXX (primarily the elevation of his tongue) affected the formants in a consistent way. This can be best observed when CXX made two attempts before producing the utterance fluently. In all but one of these repetitions (#2, #6–#9) there was a gradual increase in F2, from the 1st to the 2nd attempt, and from there to the final complete production (in the context of front vowels, increase in F2 is an expected consequence of narrowed constriction due to tongue elevation, see Kent & Read, 1992, pp. 26–27; Stevens & House, 1955). The same gradual increase in F2 was observed during the DIVA model simulated repetition (from 1755 Hz to 1808 Hz to 1928 Hz). Moreover, this gradual increase in F2 seems to bring it toward the F2 of the model's fluent production (1976 Hz), which by itself is close to the F2 of the target (according to Section 2.3, the absence of repetitions from the fluent production implies a small auditory error, i.e., proximity to the auditory target). These acoustic observations, combined with the aforementioned kinematic observations, suggest that articulatory repositioning during repetitions brings the auditory feedback closer to the desired auditory target[9]. Most studies that have investigated repetitions in stuttering have not used simultaneous analysis of acoustic and articulatory kinematic data (Howell & Vause, 1986;Wood, 1995;Yaruss & Conture, 1993;Zimmermann, 1980c) and as a result, may have led to conclusions different from the present study (cf. Harrington, 1987). Some previous studies may also be limited by the measurement methods used. Wood (1995) and Harrington (1987), for example, performed electropalatographic analysis which requires subjects to wear an artificial palate, thus affecting sensory feedback from the region of the hard palate (see Wood, 1995, p. 232).

The hypothesis that sensorimotor errors lead to self-repairs might be rejected on the grounds that such errors are not prominent enough to cause a "reset" of the motor system, as other

[9]Getting closer to the auditory target is equivalent to reducing the auditory error. Indirect evidence for error reduction during the repetition is found in the durations of CXX's disrupted attempts (reported in Table 1). In all but one of the multiple-attempt repetitions by CXX, the 2nd attempt is longer in duration than the 1st attempt, which suggests that of the two, the 2nd attempt has a smaller error. This is because smaller errors have reduced likelihood of repair, and thus require on the average more time to be repaired (hence, longer attempt duration). Similar pattern is observed in the model simulation: the 2nd attempt by the DIVA model is 10 ms longer than the 1st attempt.

speech errors do. Small errors, as those observed in the fluent speech of PWS, may indeed be too small to affect intelligibility, but the larger errors, which often trigger repetitions, are more likely to do so. Examples of such large errors preceding repetitions are brought by Howell and Vause (1986); in accordance with our claim, some of the errors were large enough to cause incorrect classification of vowels based on their formants (when employing discriminant function analysis). The possible effect of sensorimotor errors on a listener's perception is demonstrated by a recent study showing that untrained listeners judge the fluent speech of PWS as less fluent than the speech of PNS (Lickley, Hartsuiker, Corley, Russell, & Nelson, 2005; Russell, Corley, & Lickley, 2005). Even without growing large enough to trigger a repetition, the errors in the fluent speech of PWS may have influenced listeners' judgment.

The assumption that PWS use an auditory-based speech monitor has been challenged as well. Kolk and Postma (1997), for example, argued that auditory feedback is simply too slow to allow detection and repair of errors and therefore cannot account for short duration speech disruptions. Our simulation suggests the contrary: even the longer of the two disrupted attempts by the stuttering DIVA version, an attempt that lasted 76 ms, was fast enough to account for the disrupted attempts by CXX that were 99 ms long on the average (Table 1) as well as durations of disrupted attempts reported in the literature, which range anywhere from 80–150 ms (Stromsta, 1986; Yaruss & Conture, 1993). We are aware that because DIVA is a theoretical model, its performance may be faster than that of the CNS. However, even if we were to adjust the time constants of the model according to the results of a recent experiment from our lab (Tourville et al., 2008), the model would still be able to account for short duration speech disruptions. In the experiment, subjects performed same-trial compensation for formant shifting, i.e., reacted to an auditory error, in an average of 107.7 or 164.8 ms, depending on the direction of the shift. These figures are in the same range as those from subject CXX and the other reports mentioned above.

5 Modeling experiment 3: Slowed/prolonged speech reduces the frequency of sound/syllable repetitions

It has been frequently reported that PWS stutter less when speaking at a slower rate (for review see Andrews et al., 1983, p. 232; Starkweather, 1987, p. 192; Van Riper, 1982, p. 401) and/or prolong their words (Davidow, Bothe, Andreatta, & Ye, 2009; Ingham et al., 2001; Packman, Onslow, & van Doorn, 1994). These two conditions may reduce the speed that PWS move their articulators. We hypothesize that slower articulatory movements may reduce the extent and frequency of errors made by PWS (Max et al., 2004); as a result, fewer repetitions are triggered. This hypothesis can explain, for example, why PWS who speak more slowly after fluency-shaping treatment have less vowel centralization compared to PWS who speak at a normal rate (Blomgren et al., 1998). The current simulation was designed to demonstrate that the frequency of errors, and as a consequence also repetitions, drops significantly when the DIVA model is required to move the articulators at a rate that is 50% of the normal speaking rate.

5.1 Methods

Before simulating slowed/prolonged speech, we first verified that the DIVA model can account for the difference in the frequency of repetitions made by PWS and PNS. To that end, we ran simulations with both the stuttering and the non-stuttering versions of DIVA producing "good dog" at a normal speech rate (*normal speed condition*). Then, for the slow/prolonged speech simulation of the stuttering version, the target region of the utterance "good dog" was linearly stretched in time so as to double its duration (*slow speed condition*). The stretching in time reduces the rate of formant transitions. Thus, the correct production of the utterance does not require as fast articulatory movements as before. This manipulation also makes the steady-

state portion of vowels longer and increases the pause between words. We recognize that the stretching in time of the target region is an oversimplification of how speech rate modification occurs. However, it does serve the purpose of determining how changing target space transitions can affect the frequency of repetitions. To analyze how the repetitions distribute over the phonemes of the utterance, in each one of the simulations, the DIVA model produced the utterance 500 times.

5.2 Results

Fig. 4 shows the simulation results for the stuttering DIVA version in the normal speed condition (plot a) and the slow speed condition (plot b), and that of the non-stuttering DIVA version at normal speed (plot c). For each simulation, the auditory feedback during one of the fluent productions of the utterance is displayed in the top of the corresponding plot. The distribution of repetitions over the phonemes of the utterance is given by the frequency histogram in the bottom of the plot.

Consistent with the first modeling experiment, at the normal speed condition the auditory errors (mismatch between auditory feedback and target region) of the stuttering DIVA version were larger than those of the non-stuttering version. Some of these errors are due to delayed and slow formant transitions as those simulated in the first modeling experiment, while others should be attributed instead to overshoots of the target (see Section 3.3). The difference in the auditory error size resulted in a much greater frequency of repetitions for the stuttering (36/500) vs. the non-stuttering DIVA version (9/500). The stuttering DIVA version differs from the non-stuttering version also in having the greatest number of repetitions on word-initial positions. In the slow speed condition the stuttering DIVA version had both smaller errors and fewer repetitions (21/500) than in the normal speed condition.

5.3 Discussion

The simulation confirmed that in the slow speed condition, halving movement speed reduced the size of the auditory errors, thus also reducing the frequency of repetitions. According to the simulation, the reduction was an outcome of the fact that the target region had slower formant transitions, while the system time lags remained unchanged. Given the same feedback control delays, the reduced rates of the target region transitions were much easier for the stuttering DIVA version to follow (cf. Max, 2004; Max et al., 2004; van Lieshout et al., 1996a). The agreement of the simulation with the reduction in stuttering observed when PWS speak half as fast--when instructed to speak slower (Andrews, Howie, Dozsa, & Guitar, 1982; Ingham, Martin, & Kuhl, 1974), or when instructed to prolong their speech (Davidow et al., 2009; Perkins, Bell, Johnson, & Stocks, 1979)--provides further evidence for the applicability of our theory to PWS. However, while the effect of slowed/prolonged speech generally agrees with our simulations, it is difficult to conduct a quantitative comparison due to the paucity of the data for these procedures in a controlled setting (Ingham et al., 2001, p. 1230). Specifically, there is little available data regarding how slowed speech affects the distribution of specific types of dysfluency.

Further complications arise from the way PWS interpret the instruction to slow down (Andrews et al., 1982; Healey & Adams, 1981) or prolong (Ingham et al., 2001; Packman et al., 1994) their speech. The difficulty is more acute for slowed speech where PWS prefer prolonging pauses between words to prolonging the words themselves (Healey & Adams, 1981). For example, in one reduced speech rate study (Andrews et al., 1982), all three PWS at least doubled the duration of pauses, while only one of the PWS significantly increased the duration of words. In order to overcome this problem, many stuttering treatment programs train PWS to prolong their speech by stretching sounds and not pauses (see Curlee, 1999; cf. O'Brian, Onslow, Cream, & Packman, 2003); a method that has the effect of reducing the speed and increasing

the duration of articulatory movements associated with transitions (McClean, Kroll, & Loftus, 1990; Tasko et al., 2007). Ingham et al. (2001), for example, developed the MPI (Modifying Phonation Intervals) treatment program which trains PWS to specifically reduce the frequency of short phonation intervals, thus encouraging prolongation of words. The results of this treatment are encouraging--all the participants in the program achieved stutter-free and natural-sounding speech.

This paper does not reject the possibility that other methods for slowing down speech rate--ones which do not involve reduced movement speed--may improve fluency as well (see Andrews et al., 1982). For example, using longer vowel durations may facilitate readout of feedforward commands based on sensory information (Civier, 2010, p. 98). Similarly, longer vowel or pause durations may normalize intraoral pressure patterns (cf. van Lieshout et al., 2004, p. 323) and/or permit more time for speech motor planning (see Perkins et al., 1979; Riley & Ingham, 2000; Van Riper, 1982, p. 398); both are suggested to be abnormal in PWS (e.g., Peters & Boves, 1988; Peters, Hulstijn, & Starkweather, 1989; but see van Lieshout et al., 1996b). Studies where subjects are explicitly instructed how to reduce the rate of their speech (e.g., Davidow et al., 2009; Perkins et al., 1979) should help clarify the relative contribution to fluency of the various rate reduction methods.

6 Modeling experiment 4: Masking noise reduces the frequency of sound/ syllable repetitions

Masking noise (constant binaural white noise) significantly reduces the average frequency of stuttering (for review see Andrews et al., 1983, p. 233; Bloodstein, 1995, p. 345; R. R. Martin, Johnson, Siegel, & Haroldson, 1985, p. 492; Van Riper, 1982, p. 380; Wingate, 1970), with the reduction being the greatest for sound/syllable repetitions (Altrows & Bryden, 1977; Conture & Brayton, 1975; Hutchinson & Norris, 1977). Some have argued that masking noise completely blocks auditory feedback (Andrews et al., 1982; Sherrard, 1975; Stromsta, 1972; Van Riper, 1982, p. 382), but this is inconsistent with PWS's frequent reports that they keep hearing themselves above the noise (Adams & Moore, 1972; Shane, 1955). We suggest, instead, that masking noise reduces the quality of auditory feedback (cf. Garber & Martin, 1977, p. 239; Starkweather, 1987, p. 184) by decreasing the signal-to-noise ratio, or SNR (Liu & Kewley-Port, 2004). The decreased SNR may deteriorate the precision with which PWS perceive formant values, thus preventing them from detecting small formant errors (since error detection requires the speaker to perceive small mismatches between the formants of the auditory feedback and those of the target region; see Section 3.1). Reduced error detection will then lead to a lower probability of triggering a repetition, resulting in less stuttering. Our treatment of masking noise is supported by a study from Postma and Kolk (1992) reporting that noise caused PWS to detect a smaller fraction of the speech errors they made while producing tongue-twisters. Relative to the number of errors, PWS who spoke in noise also had fewer dysfluencies.

The current account makes predictions regarding changes in masking noise intensity as well. Since louder noise means a lower SNR (and detection of fewer errors), the frequency of repetitions should decrease as noise rises. Indeed, in a series of experiments that modulated masking noise intensity (Adams & Hutchinson, 1974; Maraist & Hutton, 1957; R. R. Martin, Siegel, Johnson, & Haroldson, 1984; Shane, 1955), it was found that the frequency of stuttering is reduced to a greater extent under louder noises. Here we investigate whether the stuttering version of the DIVA model makes fewer repetitions as it is exposed to higher levels of masking noise.

6.1 Methods

Liu and Kewley-Port (2004) demonstrated that masking noise increases the threshold for discriminating between two perceived formants. To approximate this property in the DIVA model, in the simulations the masking noise increases the threshold for detection of an excessive auditory error by the monitoring subsystem (remember that an auditory error consists of three formant errors; see Section 3.1). In light of Liu and Kewley-Port's report of a close-to-linear relation between masking noise and formant discrimination threshold, $T_{excessive}$, the threshold for detection of excessive error (in hertz), was defined as:

$$T_{excessive}(I_{masking}) = \xi * I_{masking}$$

where $I_{masking}$ is measured in decibels, and ξ is set to 2.

The $T_{excessive}$ threshold forms an integral part of the $f_{masking}$ function which calculates the effect of masking noise on auditory error size. $f_{masking}$ is used by the repair-likelihood formula (see Section 2.3) and is given by:

$$f_{masking}(x, I_{masking}) = \max(0, x - T_{excessive}(I_{masking}))$$

where x is auditory error size in hertz, $I_{masking}$ is measured in decibels, and $T_{excessive}$ is defined above. According to this function, masking noise of $I_{masking}$ dB "masks" (i.e., turns into 0 dB) all auditory errors smaller than $T_{excessive}(I_{masking})$ Hz in size (hence, a linear relationship exists between masking intensity and the minimum error size that can lead to a repair). Moreover, errors that are not completely masked are still being perceived as smaller than they really are; an error perceived as x Hz under normal conditions is perceived under masking noise of $I_{masking}$ dB as only $x - T_{excessive}(I_{masking})$ Hz in size.

We conducted 10 simulations of the DIVA model repeating the utterance "good dog", with each simulation having different noise intensity $I_{masking}$ ranging from 0 dB to 90 dB in steps of 10 dB.

6.2 Results

Fig. 5 shows, for each noise intensity $I_{masking}$, the number of repetitions made by the stuttering DIVA version (dashed line). The excessive-error detection thresholds for masking noise intensities ranging from 0 to 90 dB are depicted at the bottom of Fig. 5. In silence (0 dB), the DIVA model made on the average 3.7 repetitions per 100 syllables; the model made fewer repetitions at higher intensities, with only 1.1 repetitions on average with a masking noise of 90 dB.

6.3 Discussion

The simulations confirm that the DIVA model reduces the frequency of repetitions under masking noise and that the reduction is greater for louder noise. Moreover, as can be seen in Fig. 5, the simulation results resemble the estimated reduction in repetitions in the data reported by Adams and Hutchinson (1974)[10]. In that study, the estimated number of repetitions in the no-noise condition was much higher than in the 10 dB condition because headphones were

---

[10]Adams and Hutchinson (1974) reported the number of all dysfluencies combined. Therefore, we based our estimation of how many of these dysfluencies are sound/syllable repetitions on Hutchinson and Norris (1977). They reported the fraction of part-syllable repetitions out of the total number of dysfluencies, a fraction that seems to vary with noise intensity. Hutchinson and Norris also reported the fraction of whole-syllable repetitions, but because the frequency of these dysfluencies barely decreased in the noise condition, we did not take them into account (for the relation between sound/syllable repetitions and auditory-based monitoring see Section 7.2).

only worn in the noise conditions. Speaking with headphones is a novel situation that reduces stuttering in its own right (e.g., Brown, Sambrooks, & MacCulloch, 1975). The simulations are also consistent with the results of a study that directly measured the frequency of repetitions in silence and at 80 dB (Hutchinson & Norris, 1977). Other studies have reported results that are similar to the simulations (Conture & Brayton, 1975;Maraist & Hutton, 1957). In conclusion, the agreement between the simulations and the experimental data suggests that the effect of masking noise on PWS is similar to its effect on the stuttering DIVA version: the masking noise lowers SNR, which prevents detection of errors and subsequently reduces the frequency of repetitions.

To demonstrate that masking noise affects the signal-to-noise ratio, repetitions should be also affected by modulation of the "signal" part of the ratio--the auditory feedback (Starkweather, 1987, pp. 184, 239). Garber and Martin (1977) indeed found that, on average, PWS had fewer dysfluencies when using a lower vocal intensity as compared to a higher vocal intensity in conjunction with masking noise. Another explanation is possible however: the PWS had to keep their vocal intensity low albeit the noise, thus resisting the Lombard effect (the tendency to increase vocal intensity in noisy conditions, see Lane & Tranel, 1971). This conscious control of speech style may reduce stuttering on its own right (Alm, 2004; Bloodstein, 1995). Hearing loss and whispering are additional conditions that affect the intensity (as well as quality) of auditory feedback, and according to the current account, should improve fluency as well. Indeed, the incidence of stuttering in the hearing impaired is low, and especially so in the completely deaf (Montgomery & Fitch, 1988; Starkweather, 1987, p. 243; Van Riper, 1982, p. 47; Wingate, 1976, p. 216). Similarly, several reports demonstrated enhancement of fluency during whispered speech (Cherry & Sayers, 1956; Johnson & Rosen, 1937; Perkins, Rudas, Johnson, & Bell, 1976). Whispering, however, may also improve fluency due to simplified motor control. Because no voicing is required, the task of coordinating the vocal folds with the articulators is evaded (as suggested by Perkins et al., 1976; Van Riper, 1982, p. 401).

It is noteworthy that not all studies of stuttering in noise can be accounted for by a reduced signal-to-noise ratio view. The suggestion that noise reduces stuttering through distraction (Bloodstein, 1995, p. 350; Stephen & Haggard, 1980) can explain the effect of some noise forms that do not significantly reduce SNR[11]: noise that is played for less than 20% of the time (Murray, 1969); noise that is played monaurally (Yairi, 1976); narrow-band noise played at 4 or 6 kHz, much higher than the range of the first 3 formants (Barr & Carmel, 1969); or pure tones of various frequencies (Parker & Christopherson, 1963; Saltuklaroglu & Kalinowski, 2006; Stephen & Haggard, 1980). Yet, the effect of masking noise on stuttering probably has a more profound effect than merely distracting PWS because, for a given intensity, it seems to be more effective than distracting noise forms (Murray, 1969; Yairi, 1976). Moreover, the effectiveness of masking noise after prolonged use (e.g., Altrows & Bryden, 1977; MacCulloch, Eaton, & Long, 1970) cannot be explained if masking noise is only a distractor, which presumably loses its effect with time.

Other than affecting PWS's perception of their own speech, masking noise may also modify the speech movements themselves, either due to increased vocal intensity (Dromey & Ramig, 1998; Tasko & McClean, 2004) associated with the Lombard effect mentioned earlier in this

---

[11]Some noise forms may not appear to reduce SNR (see Bloodstein, 1995, pp. 347–349), but close inspection suggests that they do. Binaural white noise that is played only during silent periods (Sutton & Chase, 1961; Webster & Dorman, 1970) does not overlap with phonation, but it may still reduce the SNR in the initial tens of milliseconds of words (where errors are most likely to occur, see Section 7) via forward masking (Moore, 1995). Similarly, binaural white noise that is turned on shortly after a word begins and up to its end (see Chase & Sutton, 1968) may mask the initial portion of the word via backward masking (Moore, 1995). Very loud pure tones (approaching pain level) at low frequencies ranging from 120 Hz to 300 Hz (individual ranges vary, see Parker & Christopherson, 1963, p. 122) may "set the basilar membrane into vibration and so obliterate all hearing, over the whole range of pitch of human voice" (Cherry & Sayers, 1956, p. 241), thus drastically reducing SNR. Similar effect can be achieved using loud 500 Hz low-pass white noise (Cherry & Sayers, 1956; Conture, 1974; May & Hackwood, 1968).

section, or due to the reduced quality of auditory feedback, which may result in the corrective commands being less accurate. Such articulatory changes may modify the size of auditory errors, and thus, the frequency of stuttering. However, this is unlikely in light of experimental data showing that the features of formants and frequency of speech errors (including vowel distortion), which in our model are both indicators of auditory error size, are minimally affected by masking noise for PWS (Klich & May, 1982; Postma & Kolk, 1992)[12]. While noise-induced articulatory changes affect the acoustics of PWS in some other ways (Brayton & Conture, 1978; Stager, Denman, & Ludlow, 1997), the apparent limited effect of such changes on auditory error size did not justify simulating them for the current study. In the DIVA model therefore, masking noise only affected the monitoring subsystem, and did not affect the feedforward and feedback control subsystems which drive the articulators. Future studies should test whether the noise-induced articulatory changes in PWS are similar to those observed in PNS (Forrest, Abbas, & Zimmermann, 1986; Garnier, Bailly, Dohen, Welby, & Loevenbruck, 2006; Postma & Kolk, 1992; cf. Stager et al., 1997), and if they contribute to stuttering (see Brayton & Conture, 1978; Wingate, 1970; Wingate, 1976, pp. 225–227).

7 Conclusions and General discussion

Based on recent insights into the neural control of speech production gained from neurocomputational modeling, we presented a specific hypothesis about a possible source of the sound/syllable repetitions of people who stutter. The hypothesis suggests that due to faulty feedforward control, PWS rely more heavily on auditory feedback control. The time lags inherent in feedback-based motor control cause a sensorimotor error, detected as a mismatch between the desired speech output and actual sensory feedback. When the error becomes too large (usually in rapid formant transitions), the speech monitor issues a self-repair by immediate disruption of phonation, repositioning of the articulators, and an attempt to reproduce the erroneous syllable. The repeated disrupted attempts to produce the syllable, until it is produced without errors, are what clinicians formally label as sound/syllable repetition dysfluency.

In light of how difficult it is to collect direct behavioral evidence for a bias away from feedforward control and toward feedback-based motor control (also see Introduction), in this paper the hypothesis is supported by simulations of the DIVA speech production model. The first modeling experiment replicated published work on sensorimotor errors in the fluent speech of PWS. In the second experiment, simulation of a self-repair that is being initiated in order to eliminate such errors resembled our own data on PWS dysfluent speech. Additional experiments showed that slowed/prolonged speech minimizes errors, while masking noise prevents PWS from repairing them. In both cases, the frequency of sound/syllable repetitions is reduced. As a whole, the experiments demonstrate that the overreliance-on-feedback hypothesis has some merit in explaining stuttering and induced fluency. Moreover, by showing that slowed/prolonged speech and masking noise enhance fluency through different mechanisms, this study predicts that the combination of both conditions will reduce stuttering more than each condition alone.

7.1 Auditory-based feedback motor control and monitoring for errors

The hypothesis of over-reliance on auditory feedback control (feedback control subsystem of the DIVA model) explains why sound/syllable repetitions are most likely to occur on, or immediately after, consonants (because consonants require rapid formant transitions). It can also explain why repetitions are usually on word-initial positions (Andrews et al., 1983; Bloodstein, 1995; Wingate, 2002). We argue that because auditory feedback is not available

---

[12]Klich & May (1982) reported that F1 transition rate in the context of /æ/ was the only formant feature that changed significantly when PWS spoke in noise.

in the silent periods of speech, PWS must depend then more heavily on the impaired feedforward commands that are not capable of bringing the articulators to their correct positions. This will lead to a large auditory error at voicing onset (cf. Terband et al., 2009) and an increased likelihood of a repetition. The first and third modeling experiments support this argument by showing that the stuttering DIVA version makes the biggest auditory errors on word-initial phonemes; that these phonemes are also the most likely to be stuttered was confirmed by the third modeling experiment. These results should be taken with caution though, because the somatosensory feedback loop which can be used before voicing onset was not simulated.

It is important to note that, in addition to auditory feedback control, the speech production system also uses somatosensory feedback in the control of speech movements, especially for the production of consonants (Brancazio & Fowler, 1998; Fowler, 1994; for further discussion see Guenther et al., 1998). For the sake of tractability, we focus here on the auditory feedback system and therapeutic manipulations that affect auditory feedback. Overreliance on somatosensory feedback control has been suggested by several studies (De Nil et al., 2001; Hutchinson & Ringel, 1975; van Lieshout et al., 1996a, 1996b; van Lieshout et al., 1993). Such findings provide motivation for future modeling studies of somatosensory feedback's possible role in stuttering. Somatosensory feedback may be important, for example, to explain why unlike monitoring for errors, feedback-based motor control seems to be relatively unaffected by masking noise (see evidence presented in Section 6.3). When speaking under noise, somatosensory feedback control may substitute auditory feedback control (see Namasivayam et al., 2009, p. 702), enabling PWS to sustain proper articulation regardless of the feedforward control impairment. The same substitution cannot take place in error monitoring because somatosensory feedback is presumably not used there (see Section 2.3).

Because auditory-based monitoring for errors (monitoring subsystem of the DIVA model) can explain the effect of masking noise, it might also account for other fluency enhancers that perturb or interfere with auditory feedback, such as DAF (Delayed Auditory Feedback), FAF (Frequency-shifted Auditory Feedback), and choral speech (Andrews et al., 1983; Antipova, Purdy, Blakeley, & Williams, 2008; Armson & Kiefte, 2008; Bloodstein, 1995; Howell, El-Yaniv, & Powell, 1987; Lincoln, Packman, & Onslow, 2006; MacLeod, Kalinowski, Stuart, & Armson, 1995). In all of these conditions speakers keep hearing the original auditory feedback (for example, through bone-conducted feedback), and thus, the extra auditory input may be regarded a form of noise (which may explain, for example, why PWS under DAF raise their vocal intensity, see Howell, 1990). This noise has the potential to reduce SNR in the same way masking noise does (resulting in fewer repetitions, see Section 6). Yet, repeated findings that masking noise is less effective than DAF or choral speech (Andrews et al., 1982; Stager et al., 1997), as well as FAF (Howell et al., 1987; Kalinowski, Armson, Roland-Mieszkowski, Stuart, & Gracco, 1993), suggest that reduction in SNR cannot be the sole mechanism behind auditory-based fluency enhancers. Some additional mechanisms that might come into play are the engagement of mirror neuron networks (Kalinowski & Saltuklaroglu, 2003), de-automatization (Alm, 2004) which may be carried by activation of the lateral premotor system (Alm, 2005, p. 50; Snyder, Hough, Blanchet, Ivy, & Waddell, 2009), and reduction of speech rate (Wingate, 1970; Wingate, 1976, p. 239), possibly as a consequence of the timekeeper, which regulates motor actions, being overloaded by the extra auditory input (an effect that is based on the rhythmic structure rather than auditory content of the signal, see Howell, 2004). It should be also noted that DAF in particular has additional effects (Yates, 1963), some of them reduce rather than enhance fluency (for an example see Section 4). By simulating the effects of DAF on the stuttering DIVA version, future studies may be able to explain why in PWS the upsides of DAF ultimately overcome its downsides.

Notwithstanding the important roles of feedback-based motor control (DIVA's feedback control subsystem) and monitoring for errors (DIVA's monitoring subsystem) in stuttering, it should be emphasized that neither mechanism was assumed to be impaired in this study (the only impairment is in the readout of feedforward commands, i.e., DIVA's feedforward control subsystem). Nevertheless, the hypothesis of defective speech monitoring--due to hyper-sensitivity to errors (Maraist & Hutton, 1957; J. E. Martin, 1970; Russell et al., 2005; Sherrard, 1975), or due to inaccuracy in predicting the sensory consequences of movements (Max, 2004; Max et al., 2004)--can be investigated using the DIVA model by modifying the ε parameter which modulates the likelihood of error repair (described in Section 2.3). Similarly, combined impairment in both feedforward and feedback control (as suggested, for example, by Loucks & De Nil, 2006b) may be simulated by using low values for both the α_{ff} and α_{fb} gain parameters, such that the parameters would not sum to one (compare to the gain parameter values used in the first modeling experiment). Such combined impairment is likely to occur if not only the feedforward commands for an utterance, but also its auditory target, cannot be read out from memory.

7.2 Sound/syllable repetitions

Although the DIVA model can simulate different types of dysfluency (Civier, 2010; Civier et al., 2009), we focused this study on sound/syllable repetitions since these are the only dysfluencies that arise directly from the actions of the auditory-based monitoring subsystem. There is limited evidence that masking noise has a specific effect of reducing sound/syllable repetitions (Altrows & Bryden, 1977; Conture & Brayton, 1975; Hutchinson & Norris, 1977) and that sound/syllable repetitions were absent in a deaf child that showed stuttering (subject 12 in Montgomery & Fitch, 1988, p. 133). This view would suggest, for example, that silent repetitions (repeated silent articulatory movements that sometimes precede prolongations and may be associated with laryngeal blocks, see van Lieshout, Peters, & Bakker, 1997; Van Riper, 1982, p. 119), which in many aspects resemble sound/syllable repetitions (Zimmermann, 1980a), do not depend on auditory-based detection of errors, and therefore, should not be affected by masking noise.

The above discussion must not imply that sound/syllable repetitions have nothing in common with other dysfluencies. On the contrary, we hypothesize that all types of dysfluency have a common underlying cause: an impairment in the ability to read out feedforward commands (Civier, 2010; Civier et al., 2009). In the second modeling experiment we simulated one consequence of this impairment, namely, a bias away from feedforward control (and toward feedback control) which resulted in increased error size, and ultimately lead to repetitions; but, other scenarios are possible as well. For example, if the consequence of the impairment is that no feedforward commands are read out at all, speech production will halt and the outcome will be a pause (or initiation problem). Simulations of pauses are available elsewhere (Civier, 2010; Civier et al., 2009), and together with the simulations in this paper, they support our hypothesis that a single motor control problem can explain the range of dysfluencies. If correct, this hypothesis can also account for the observation that multiple dysfluency types often co-exist in the same moment of stuttering (e.g., repetitions and prolongations, see Van Riper, 1982, p. 118).

7.3 Slow movement speed in PWS

This paper speaks to the ongoing debate (see van Lieshout et al., 1996a, pp. 559–560; van Lieshout et al., 2004, p. 318) about the origin of the slower-than-normal articulatory movement speeds in PWS (e.g., Max, Caruso, & Gracco, 2003; McClean et al., 2004; Smith & Kleinow, 2000; van Lieshout et al., 1996a, 1996b; van Lieshout et al., 1993; Zimmermann, 1980b). It has been suggested (see the aforementioned papers, as well as Max, 2004; Max et al., 2004, p. 113; McClean et al., 1990; Peters et al., 2000, p. 113; van Lieshout et al., 2004; Zimmermann,

1980c) that to improve fluency, PWS may exercise planned prolongation of speech movements (permanently or more strategically, see van Lieshout et al., 1993), which presumably is similar to their response when explicitly instructed to prolong their speech (reduce the desired movement speed, see Section 5). While acknowledging that some PWS may indeed prolong their speech intentionally, the model simulations presented here suggest that at least some aspects of slowing down may be an unconscious rather than planned reaction to the motor control deficit. PWS may intend to speak at their normal rate, but due to the time lags inherent in feedback control, they are not able to[13]. According to the model, this discrepancy between desired and actual movement speed is what leads to large errors, and then repetitions.

The view that the slowness of PWS might not be conscious is suggested by data reviewed in Section 4.3, showing that the movements of PWS at the onset of dysfluency (as reflected by their abnormal formant transitions and reduced vowel formants) are even slower than during the fluent production that follows[14]. Such data suggest that the slowness is not a feature of intentional dysfluency-preventing adaptation (it does not fit with dysfluency occurring at the slower part of the utterance) but rather an indication of unavoidable dysfluency-inducing limitation (see van Lieshout et al., 2004, p. 331). This paper suggests that the limiting factors are the impaired feedforward control and the motor control strategy (heavy reliance on feedback control) that fails to cope with the impairment. To summarize, we claim that for many untreated PWS, slowness of articulation may be part of the speech problem, and not a technique to resolve it (similar views are expressed in Chang et al., 2002; Van Riper, 1982, p. 417). Treated PWS are excluded from the above statement because many of them do often use slowed speech techniques learned in treatment programs (See Section 5). We also exclude those untreated PWS who, based on favorable experience, voluntary adopted a slower speech rate.

The hypothesis of a bias away from feedforward control and toward feedback control accounts well for the articulatory behavior of more severe PWS, but it fails to explain why, compared to PNS, mild PWS seem to have faster, rather than slower, opening movements of the tongue, and possibly also the lower lip (McClean, Kroll, & Loftus, 1991; McClean et al., 2004; cf. Smith & Kleinow, 2000). The behavior of mild PWS can be explained, however, by assuming (see Namasivayam et al., 2008; van Lieshout et al., 2004) that they opt for a control strategy that involves faster movements, where the coupling strength between coordinated elements becomes smaller, and the system "regresses to the most stable and basic pattern of coordination" (van Lieshout et al., 2004, p. 321). If mild PWS indeed choose this strategy over the maladaptive strategy of heavy reliance on feedback control, this can potentially explain both their enhanced fluency, and faster movement speeds.

7.4 Challenges to the hypothesis of overreliance on auditory feedback

We believe we have addressed many of Postma and Kolk's (1993, p. 482) objections to a speech motor control theory of stuttering. Our overreliance-on-feedback hypothesis causally links sensorimotor error with repetition-based dysfluency. As error size grows, so does the likelihood of dysfluency (see Section 2.3). Therefore, we assume that errors are not a secondary consequence of a lifetime of stuttering or some sort of compensation strategy (Armson &

---

[13]Some have argued that if PWS do indeed rely heavily on feedback control, one must conclude that they also intentionally prolong their speech (Max, 2004; Max et al., 2004; van Lieshout et al., 1996a). Although we agree that PWS are better off using prolonged movements (as we showed in the third modeling experiment, feedback control can be utilized more effectively this way), we still consider the possibility that they try to maintain their original speech rate (see Wingate, 1976, p. 340). After all, even our worst habits are difficult to change.

[14]Due to their proximity to the moment of stuttering, the complete productions following repetitions may not be representative of fluent utterances. A study that tested this argument (Stromsta, 1986, p. 79; Stromsta & Fibiger, 1981) suggests against it, however. All fluent utterances, whether following a repetition or not, showed anticipatory labial activity (as measured by EMG) less deviant than that of the dysfluent utterances. Similar results regarding delayed onset of phonation were reported as well (Peters, Hietkamp, & Boves, 1995; Peters et al., 2000, pp. 113–115).

Kalinowski, 1994). This view is supported by the observation that errors are evident already in young age (see Sections 3.3 and 4.3, cf. Conture, 1991). However, we do not currently have sufficient data to reject the possibility that errors result from a gross change in the speech apparatus as suggested by Prosek et al. (1987).

Namasivayam et al. (2009) do not object to a speech motor control theory of stuttering, but rather to hypotheses that PWS heavily rely on sensory feedback. To challenge our specific hypothesis of a bias toward auditory feedback, they showed that complete auditory masking did not significantly increase the variability of speech motor coordination in PWS. Unfortunately, because the dependent measures used were different from our study, direct comparison between the two studies is rather complicated. Nevertheless, our hypothesis may be maintained by assuming that when auditory feedback is perturbed (as in masking), PWS rely more on jaw proprioceptive or lip tactile feedback, or when jaw proprioception is perturbed as well and lip tactile feedback is limited, on lip proprioceptive feedback. This assumption might be supported by Namasivayam et al.'s observation that PWS under combined perturbation had greater upper lip movement amplitude in utterances with limited tactile lip contact; the extended upper lip movement may have increased the gain of proprioceptive input from the lips (see Namasivayam et al., 2009, pp. 703–704), presumably facilitating feedback channel substitution. That said, using extended upper lip movements probably has additional benefits for PWS (Namasivayam & van Lieshout, 2008; Namasivayam et al., 2008; Namasivayam et al., 2009) since their movement coordination in such conditions is not only equal, but even superior to that of PNS (Namasivayam et al., 2009).

In contrast with the Namasivayam et al. (2009) study which challenges the hypothesis of overreliance on sensory feedback, other studies that measured speech motor variability in PWS may actually support that hypothesis. Several kinematic studies (e.g., Kleinow & Smith, 2000; Ward, 1997) have shown that PWS are more variable than controls at normal (unperturbed) speaking conditions (but see Smith & Kleinow, 2000)[15]. The overreliance-on-feedback hypothesis indeed predicts somewhat elevated within-subject variation in motor output. A feedback-based control system will be subject to inherent variations in sensory feedback. Fluctuation of feedback will result in movement delays of variable duration, and consequently, higher articulatory variability. On the other hand, a feedforward strategy, which relies instead on stable internal commands, is likely to exhibit less motor variation. However, this prediction has not been tested with simulations yet and will be the topic of future work.

The overreliance-on-feedback hypothesis may be further tested by its ability to explain developmental aspects of stuttering. For example, in the framework of the DIVA model, changes in articulatory complexity and speech rate (as well as in sensitivity to errors) during childhood may explain why the onset of stuttering usually occurs between two to five years of age (see Max et al., 2004, p. 115). Moreover, improvement of initially impaired feedforward control through sensorimotor learning and neuroanatomical/neurochemical maturation may potentially explain the spontaneous recovery of most children who stutter (see Max et al., 2004, p. 116).

It is possible to simulate the effect of feedback over-reliance on speech development by introducing the bias away from feedforward control and toward feedback control in the learning stages of the DIVA model (unlike the current study in which the bias was introduced only in the mature system, see Section 3.1) as was recently done by Terband and colleagues (Terband

---

[15]Variability in speech production has been formally examined also in the acoustic domain (Wieneke, Eijken, Janssen, & Brutten, 2001) and in treated PWS (McClean, Levandowski, & Cord, 1994). These and other relevant studies (e.g., Caruso, Abbs, & Gracco, 1988) are not discussed here since their results might have been confounded with the variations in speech rate mentioned in Section 7.3 (see McClean et al., 1990; Namasivayam et al., 2009, p. 692).

& Maassen, in press; Terband et al., 2009). These authors studied childhood apraxia of speech rather than stuttering, but some of their results can still be compared to ours. For example, Terband et al. (2009) reported that the speech was unintelligible when the feedback gain parameter, α_{fb}, was raised above 0.45. In contrast, the productions in our study stayed intelligible even with α_{fb} being set to 0.75. This suggests that extreme bias toward feedback ($\alpha_{fb} > 0.45$) deteriorates speech intelligibility when applied during both learning and performance stages (Terband et al., 2009), but not when applied during the performance stage alone (our study). The reason for the difference, we believe, is that feedforward commands are crucial for intelligible speech production, and they cannot be acquired properly when the model is abnormally biased during learning.

The correspondence of the DIVA model components to anatomical locations in the brain (Golfinopoulos, Tourville, & Guenther, 2009; Guenther et al., 2006) provides us with yet an additional way to test our hypothesis: by comparing the activities of the model components to activities of brain regions measured using functional brain imaging during the fluent and dysfluent speech of PWS (Civier, 2010; Civier et al., 2009; for use of this process of comparing model simulations and experimental results for the speech of PNS, see Golfinopoulos et al., 2009; also see Guenther, 2006; Guenther et al., 2006; Tourville et al., 2008). Based on our recent association of the motor mechanisms for error correction with right hemisphere inferior frontal cortex (Tourville et al., 2008), we can account for the over-activation of that region often reported in PWS's speech production as compared to that of PNS (Brown, Ingham, Ingham, Laird, & Fox, 2005). The association of auditory error cells with the posterior superior temporal gyrus (Tourville et al., 2008) predicts that PWS have differential activation in that region as well, but further simulations are needed to ensure that the sign and size of the model's activity are comparable with physiological measurements. Lastly, mapping the anatomical location of the newly-defined monitoring subsystem's components may be possible when comparing model activities to imaging data collected during sound/syllable repetitions. Since this is the subsystem that initiates repetitions, its components are expected to be extremely active in these instances.

Acknowledgments

References

Abbs JH. The influence of the gamma motor system on jaw movements during speech: A theoretical framework and some preliminary observations. Journal of Speech and Hearing Research 1973;16(2): 175–200. [PubMed: 4269271]

Adams MR. Further analysis of stuttering as a phonetic transition defect. Journal of Fluency Disorders 1978;3:265–271.

Adams MR, Hutchinson J. The effects of three levels of auditory masking on selected vocal characteristics and the frequency of disfluency of adult stutterers. Journal of Speech and Hearing Research 1974;17 (4):682–688. [PubMed: 4444288]

Adams MR, Moore WH Jr. The effects of auditory masking on the anxiety level, frequency of dysfluency, and selected vocal characteristics of stutterers. Journal of Speech and Hearing Research 1972;15(3): 572–578. [PubMed: 5080050]

Agnello, JG. Voice onset and voice termination features of stutterers. In: Webster, LM.; Furst, LC., editors. Vocal tract dynamics and dysfluency; Proceedings of the First Annual Hayes Martin Conference on Vocal Tract Dynamics; New York: Speech and Hearing Institute; 1975. p. 40-70.

Alfonso, PJ. Implications of the concepts underlying task-dynamic modeling on kinematic studies of stuttering. In: Peters, HFM.; Hulstijn, W.; Starkweather, CW., editors. Speech motor control and stuttering; Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering; 1990; Nijmegen, Netherlands. Amsterdam: Excerpta Medica; 1991. p. 79-100.

Alm PA. Stuttering and the basal ganglia circuits: A critical review of possible relations. Journal of Communication Disorders 2004;37(4):325–369. [PubMed: 15159193]

Alm, PA. On the causal mechanisms of stuttering. Lund, Sweden: Lund University; 2005.

Altrows IF, Bryden MP. Temporal factors in the effects of masking noise on fluency of stutterers. Journal of Communication Disorders 1977;10(4):315–329. [PubMed: 925194]

Andrews G, Craig A, Feyer AM, Hoddinott S, Howie P, Neilson M. Stuttering: A review of research findings and theories circa 1982. Journal of Speech and Hearing Disorders 1983;48(3):226–246. [PubMed: 6353066]

Andrews G, Howie PM, Dozsa M, Guitar BE. Stuttering: Speech pattern characteristics under fluency-inducing conditions. Journal of Speech and Hearing Research 1982;25(2):208–216. [PubMed: 7120960]

Antipova EA, Purdy SC, Blakeley M, Williams S. Effects of altered auditory feedback (AAF) on stuttering frequency during monologue speech production. Journal of Fluency Disorders 2008;33(4): 274–290. [PubMed: 19328980]

Armson J, Kalinowski J. Interpreting results of the fluent speech paradigm in stuttering research: Difficulties in separating cause from effect. Journal of Speech and Hearing Research 1994;37(1):69–82. [PubMed: 8170133]

Armson J, Kiefte M. The effect of SpeechEasy on stuttering frequency, speech rate, and speech naturalness. Journal of Fluency Disorders 2008;33(2):120–134. [PubMed: 18617052]

Barr DF, Carmel NR. Stuttering inhibition with high frequency narrow-band masking noise. The journal of Auditory Research 1969;9:40–44.

Baum SR, McFarland DH, Diab M. Compensation to articulatory perturbation: Perceptual data. Journal of the Acoustical Society of America 1996;99(6):3791–3794. [PubMed: 8655810]

Black JW. The effect of delayed side-tone upon vocal rate and intensity. Journal of Speech and Hearing Disorders 1951;16(1):56–60.

Blomgren M, Robb M, Chen Y. A note on vowel centralization in stuttering and nonstuttering individuals. Journal of Speech, Language, and Hearing Research 1998;41:1042–1051.

Bloodstein, O. A handbook on stuttering. 5. San Diego, CA: Singular; 1995.

Bohland JW, Bullock D, Guenther FH. Neural representations and mechanisms for the performance of simple speech sequences. Journal of Cognitive Neuroscience. (in press).

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

Brancazio L, Fowler CA. On the relevance of locus equations for production and perception of stop consonants. Perception and Psychophysics 1998;60(1):24–50. [PubMed: 9503910]

Brayton ER, Conture EG. Effects of noise and rhythmic stimulation on the speech of stutterers. Journal of Speech and Hearing Research 1978;21(2):285–294.

Brown S, Ingham RJ, Ingham JC, Laird AR, Fox PT. Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. Human Brain Mapping 2005;25(1):105–117. [PubMed: 15846815]

Brown SF. A further study of stuttering in relation to various speech sounds. Quarterly Journal of Speech 1938;24:390–397.

Brown T, Sambrooks JE, MacCulloch MJ. Auditory thresholds and the effect of reduced auditory feedback on stuttering. Acta Psychiatrica Scandinavica 1975;51:297–311. [PubMed: 1146589]

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

Callan DE, Kent RD, Guenther FH, Vorperian HK. An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. Journal of Speech, Language, and Hearing Research 2000;43(3):721–736.

Caruso AJ, Abbs JH, Gracco VL. Kinematic analysis of multiple movement coordination during speech in stutterers. Brain 1988;111(Pt 2):439–456. [PubMed: 3378144]

Caruso AJ, Chodzko-Zajko WJ, Bidinger DA, Sommers RK. Adults who stutter: Responses to cognitive stress. Journal of Speech and Hearing Research 1994;37(4):746–754. [PubMed: 7967559]

Chang SE, Ohde RN, Conture EG. Coarticulation and formant transition rate in young children who stutter. Journal of Speech, Language, and Hearing Research 2002;45(4):676–688.

Chase RA, Sutton S. Reply to: "masking of auditory feedback in stutterers' speech". Journal of Speech and Hearing Research 1968;11:222–223.

Cherry C, Sayers BM. Experiments upon the total inhibition of stammering by external control, and some clinical results. Journal of Psychosomatic Research 1956;1(4):233–246. [PubMed: 13377328]

Civier O. *Computational modeling of the neural substrates of stuttering and induced fluency* (Doctoral dissertation, Boston University, 2010). Dissertation Abstracts International 2010;70:10.

Civier, O.; Bullock, D.; Max, L.; Guenther, FH. Simulating neural impairments to syllable-level command generation in stuttering. Poster session presented at the 6th World Congress on Fluency Disorders; Rio de Janeiro, Brazil. 2009 Aug.

Civier, O.; Guenther, FH. Simulations of feedback and feedforward control in stuttering. Paper presented at the 7th Oxford Dysfluency Conference; St. Catherine's College, Oxford University; 2005 June–July.

Conture EG. Some effects of noise on the speaking behavior of stutterers. Journal of Speech and Hearing Research 1974;17(4):714–723. [PubMed: 4444291]

Conture, EG. Young stutterers' speech production: A critical review. In: Peters, HFM.; Hulstijn, W.; Starkweather, CW., editors. Speech motor control and stuttering; Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering; 1990; Nijmegen, Netherlands. Amsterdam: Excerpta Medica; 1991. p. 365-384.

Conture, EG. Stuttering: Its nature, diagnosis, and treatment. Boston: Allyn and Bacon; 2001.

Conture EG, Brayton ER. The influence of noise on stutterers' different disfluency types. Journal of Speech and Hearing Research 1975;18:381–384.

Conture, EG.; Zackheim, CT.; Anderson, JD.; Pellowski, MW. Linguistic processes and childhood stuttering: Many's a slip between intention and lip. In: Maassen, B.; Kent, RD.; Peters, HFM.; van Lieshout, PHHM.; Hulstijn, W., editors. Speech motor control in normal and disordered speech. Oxford, UK: Oxford University Press; 2004. p. 253-281.

Cowie, RI.; Douglas-Cowie, E. Speech production in profound post-lingual deafness. In: Lutman, ME.; Haggard, MP., editors. Hearing science and hearing disorders. New York: Academic Press; 1983. p. 183-231.

Curlee, RF. Stuttering and related disorders of fluency. 2. New York: Thieme Medical; 1999.

Davidow JH, Bothe AK, Andreatta RD, Ye J. Measurement of phonated intervals during four fluency-inducing conditions. Journal of Speech, Language, and Hearing Research 2009;52(1):188–205.

De Nil LF, Kroll RM, Houle S. Functional neuroimaging of cerebellar activation during single word reading and verb generation in stuttering and nonstuttering adults. Neuroscience Letters 2001;302 (2–3):77–80. [PubMed: 11290391]

Dromey C, Ramig LO. Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation. Journal of Speech, Language, and Hearing Research 1998;41(5):1003–1018.

Fairbanks G. Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. Journal of Speech and Hearing Disorders 1954;19(2):133–139. [PubMed: 13212816]

Forrest K, Abbas PJ, Zimmermann GN. Effects of white noise masking and low pass filtering on speech kinematics. Journal of Speech and Hearing Research 1986;29(4):549–562. [PubMed: 3795897]

Fowler CA. Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. Perception and Psychophysics 1994;55(6):597–610. [PubMed: 8058448]

Gammon SA, Smith PJ, Daniloff RG, Kim CW. Articulation and stress-juncture production under oral anesthetization and masking. Journal of Speech and Hearing Research 1971;14(2):271–282. [PubMed: 5558079]

Garber SF, Martin RR. Effects of noise and increased vocal intensity on stuttering. Journal of Speech and Hearing Research 1977;20(2):233–240. [PubMed: 895095]

Garnier, M.; Bailly, L.; Dohen, M.; Welby, P.; Loevenbruck, H. An acoustic and articulatory study of Lombard speech: Global effects on the utterance. INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing; Curran Associates; 2006. p. 2246-2249.

Ghosh SS. *Understanding cortical and cerebellar contributions to speech production through modeling and functional imaging* (Doctoral dissertation, Boston University, 2005). Dissertation Abstracts International 2005;66:11.

Goehl H, Kaufman DK. Do the effects of adventitious deafness include disordered speech? Journal of Speech and Hearing Disorders 1984;49(1):58–64. [PubMed: 6700203]

Golfinopoulos E, Tourville JA, Guenther FH. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage. 200910.1016/j.neuroimage.2009.10.023

Guenther FH. A neural network model of speech acquisition and motor equivalent speech production. Biological Cybernetics 1994;72(1):43–53. [PubMed: 7880914]

Guenther FH. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychological Review 1995;102(3):594–621. [PubMed: 7624456]

Guenther FH. Cortical interactions underlying the production of speech sounds. Journal of Communication Disorders 2006;39(5):350–365. [PubMed: 16887139]

Guenther FH, Espy-Wilson CY, Boyce SE, Matthies ML, Zandipour M, Perkell JS. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. Journal of the Acoustical Society of America 1999;105(5):2854–2865. [PubMed: 10335635]

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain and Language 2006;96(3):280–301. [PubMed: 16040108]

Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. Psychological Review 1998;105(4):611–633. [PubMed: 9830375]

Harrington, J. Coarticulation and stuttering: An acoustic and electropalatographic study. In: Peters, HFM.; Hulstijn, W., editors. Speech motor dynamics in stuttering. New York: Springer-Verlag; 1987. p. 381-392.

Healey EC, Adams MR. Rate reduction strategies used by normally fluent and stuttering children and adults. Journal of Fluency Disorders 1981;6:1–14.

Healey EC, Gutkin B. Analysis of stutterers' voice onset times and fundamental frequency contours during fluency. Journal of Speech and Hearing Research 1984;27(2):219–225. [PubMed: 6738033]

Hennessey NW, Nang CY, Beilby JM. Speeded verbal responding in adults who stutter: Are there deficits in linguistic encoding? Journal of Fluency Disorders 2008;33(3):180–202. [PubMed: 18762061]

Hirsch, F.; Fauvet, F.; Ferbach-Hecker, V.; Béchet, M.; Bouarourou, F.; Sturm, J. In: Trouvain, J., editor. Formant structures of vowels produced by stutterers at normal and fast speech rates; Proceedings of the 16th International Congress of Phonetic Sciences; Saarbrücken, Germany. Saarbrücken: Universität des Saarlandes; 2007. p. 1345-1348.

Hoole, P. In: Bolla, K., editor. Bite-block speech in the absence of oral sensibility; Proceedings of the 11th International Congress on Phonetic Sciences; Tallin. Budapest, Hungary: Academy of Sciences of the Estonian SSR; 1987. p. 16-19.

Howell P. Changes in voice level caused by several forms of altered feedback in fluent speakers and stutterers. Language and Speech 1990;33(Pt 4):325–338. [PubMed: 2133911]

Howell P. Assessment of some contemporary theories of stuttering that apply to spontaneous speech. Contemporary Issues in Communication Science and Disorders 2004;31:122–139. [PubMed: 18259590]

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-Verlag; 1987. p. 361-369.

Howell P, Vause L. Acoustic analysis and perception of vowels in stuttered speech. Journal of the Acoustical Society of America 1986;79(5):1571–1579. [PubMed: 3711457]

Hutchinson JM, Norris GM. The differential effect of three auditory stimuli on the frequency of stuttering behaviors. Journal of Fluency Disorders 1977;2:283–293.

Hutchinson JM, Ringel RL. The effect of oral sensory deprivation on stuttering behavior. Journal of Communication Disorders 1975;8(3):249–258. [PubMed: 802975]

Ingham RJ, Andrews G. The relation between anxiety reduction and treatment. Journal of Communication Disorders 1971;4:289–301.

Ingham RJ, Kilgo M, Ingham JC, Moglia R, Belknap H, Sanchez T. Evaluation of a stuttering treatment based on reduction of short phonation intervals. Journal of Speech, Language, and Hearing Research 2001;44(6):1229–1244.

Ingham RJ, Martin RR, Kuhl P. Modification and control of rate of speaking by stutterers. Journal of Speech and Hearing Research 1974;17(3):489–496. [PubMed: 4420465]

Jäncke, L. The 'audio-phonatoric' coupling in stuttering and nonstuttering adults: Experimental contributions. In: Peters, HFM.; Hulstijn, W.; Starkweather, CW., editors. Speech motor control and stuttering; Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering; 1990; Nijmegen, Netherlands. Amsterdam: Excerpta Medica; 1991. p. 171-180.

Johnson W, Brown SF. Stuttering in relation to various speech sounds. The Quarterly Journal of Speech 1935;31:481–496.

Johnson W, Rosen L. Studies in the psychology of stuttering: VII. Effect of certain changes in speech pattern upon frequency of stuttering. Journal of Speech Disorders 1937;2:105–110.

Kalinowski J, Armson J, Roland-Mieszkowski M, Stuart A, Gracco VL. Effects of alterations in auditory feedback and speech rate on stuttering frequency. Language and Speech 1993;36(1):1–16. [PubMed: 8345771]

Kalinowski J, Saltuklaroglu T. Choral speech: The amelioration of stuttering via imitation and the mirror neuronal system. Neuroscience and Biobehavioral Reviews 2003;27(4):339–347. [PubMed: 12946686]

Kalveram, KT. How pathological audio-phonatoric coupling induces stuttering: A model of speech flow control. In: Peters, HFM.; Hulstijn, W.; Starkweather, CW., editors. Speech motor control and stuttering; Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering; 1990; Nijmegen, Netherlands. Amsterdam: Excerpta Medica; 1991. p. 163-169.

Kalveram KT. A neural-network model enabling sensorimotor learning: Application to the control of arm movements and some implications for speech-motor control and stuttering. Psychological Research. Psychologiche Forschung 1993;55:299–314.

Kalveram KT, Jäncke L. Vowel duration and voice onset time for stressed and nonstressed syllables in stutterers under delayed auditory feedback condition. Folia Phoniatrica 1989;41(1):30–42. [PubMed: 2759514]

Kent, RD. Models of speech production. In: Lass, NJ., editor. Contemporary issues in experimental phonetics. New York: Academic Press; 1976. p. 79-104.

Kent RD, Moll KL. Articulatory timing in selected consonant sequences. Brain and Language 1975;2:304–323.

Kent, RD.; Read, C. The acoustic analysis of speech. San Diego, CA: Singular; 1992.

Kewley-Port D. Measurement of formant transitions in naturally produced stop consonant-vowel syllables. Journal of the Acoustical Society of America 1982;72(2):379–389. [PubMed: 7119280]

Kleinow J, Smith A. Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. Journal of Speech, Language, and Hearing Research 2000;43(2):548–559.

Klich RJ, May GM. Spectrographic study of vowels in stutterers' fluent speech. Journal of Speech and Hearing Research 1982;25(3):364–370. [PubMed: 7176608]

Kolk, H.; Postma, A. Stuttering as a covert repair phenomenon. In: Curlee, RF.; Siegel, GM., editors. Nature and treatment of stuttering: New directions. 2. Boston: Allyn and Bacon; 1997. p. 182-203.

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

Lashley, KS. The problem of serial order in behavior. In: Jeffress, LA., editor. Cerebral mechanisms in behavior. New York: Wiley; 1951. p. 507-528.

Lee BS. Artificial stutter. Journal of Speech and Hearing Disorders 1951;16(1):53–55.

Levelt WJM. Monitoring and self-repair in speech. Cognition 1983;14(1):41–104. [PubMed: 6685011]

Levelt, WJM. Speaking: From intention to articulation. Cambridge, MA: MIT Press; 1989.

Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behavioral and Brain Sciences 1999;22(1):1–38. [PubMed: 11301520]

Levelt WJM, Wheeldon L. Do speakers have access to a mental syllabary? Cognition 1994;50(1–3):239–269. [PubMed: 8039363]

Lickley RJ, Hartsuiker RJ, Corley M, Russell M, Nelson R. Judgment of disfluency in people who stutter and people who do not stutter: Results from magnitude estimation. Language and Speech 2005;48(3):299–312. [PubMed: 16416939]

Lieberman, P.; Blumstein, S. Speech physiology, speech perception, and acoustic phonetics. Cambridge, UK: Cambridge University Press; 1988.

Lincoln M, Packman A, Onslow M. Altered auditory feedback and the treatment of stuttering: A review. Journal of Fluency Disorders 2006;31(2):71–89. [PubMed: 16750562]

Liu C, Kewley-Port D. Formant discrimination in noise for isolated vowels. Journal of the Acoustical Society of America 2004;116(5):3119–3129. [PubMed: 15603157]

Loucks TM, De Nil LF. Anomalous sensorimotor integration in adults who stutter: A tendon vibration study. Neuroscience Letters 2006a;402(1–2):195–200. [PubMed: 16698179]

Loucks TM, De Nil LF. Oral kinesthetic deficit in adults who stutter: A target-accuracy study. Journal of motor behavior 2006b;38(3):238–246. [PubMed: 16709563]

MacCulloch MJ, Eaton R, Long E. The long term effect of auditory masking on young stutterers. The British Journal of Disorders of Communication 1970;5(2):165–173. [PubMed: 5495168]

MacLeod J, Kalinowski J, Stuart A, Armson J. Effect of single and combined altered auditory feedback on stuttering frequency at two speech rates. Journal of Communication Disorders 1995;28(3):217–228. [PubMed: 8530718]

Maeda, S. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In: Hardcastle, WJ.; Marchal, A., editors. Speech production and speech modelling. Boston: Kluwer Academic; 1990. p. 131-149.

Maraist JA, Hutton C. Effects of auditory masking upon the speech of stutterers. Journal of Speech and Hearing Disorders 1957;22(3):385–389. [PubMed: 13463921]

Martin JE. The signal detection hypothesis and perceptual defect theory of stuttering. Journal of Speech and Hearing Research 1970;35:352–355.

Martin RR, Johnson LJ, Siegel GM, Haroldson SK. Auditory stimulation, rhythm, and stuttering. Journal of Speech and Hearing Research 1985;28(4):487–495. [PubMed: 4087883]

Martin RR, Siegel GM, Johnson LJ, Haroldson SK. Sidetone amplification, noise, and stuttering. Journal of Speech and Hearing Research 1984;27(4):518–527. [PubMed: 6521457]

Max, L. Stuttering and internal models for sensorimotor control: A theoretical perspective to generate testable hypotheses. In: Maassen, B.; Kent, RD.; Peters, HFM.; van Lieshout, PHHM.; Hulstijn, W., editors. Speech motor control in normal and disordered speech. Oxford, UK: Oxford University Press; 2004. p. 357-387.

Max L, Caruso AJ, Gracco VL. Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. Journal of Speech, Language, and Hearing Research 2003;46(1):215–232.

Max L, Guenther FH, Gracco VL, Ghosh SS, Wallace ME. Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. Contemporary Issues in Communication Science and Disorders 2004;31:105–122.

May AE, Hackwood A. Some effects of masking and eliminating low frequency feedback on the speech of stammerers. Behaviour Research and Therapy 1968;6(2):219–223. [PubMed: 5734538]

McClean MD, Kroll RM, Loftus NS. Kinematic analysis of lip closure in stutterers' fluent speech. Journal of Speech and Hearing Research 1990;33(4):755–760. [PubMed: 2273888]

McClean, MD.; Kroll, RM.; Loftus, NS. Correlation of stuttering severity and kinematics of lip closure. In: Peters, HFM.; Hulstijn, W.; Starkweather, CW., editors. Speech motor control and stuttering; Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering; 1990; Nijmegen, Netherlands. Amsterdam: Excerpta Medica; 1991. p. 117-122.

McClean MD, Levandowski DR, Cord MT. Intersyllabic movement timing in the fluent speech of stutterers with different disfluency levels. Journal of Speech and Hearing Research 1994;37(5): 1060–1066. [PubMed: 7823552]

McClean MD, Runyan CM. Variations in the relative speeds of orofacial structures with stuttering severity. Journal of Speech, Language, and Hearing Research 2000;43(6):1524–1531.

McClean MD, Tasko SM, Runyan CM. Orofacial movements associated with fluent speech in persons who stutter. Journal of Speech, Language, and Hearing Research 2004;47(2):294–303.

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(1):21–45.

Miller JL, Baer T. Some effects of speaking rate on the production of /b/ and /w/. Journal of the Acoustical Society of America 1983;73(5):1751–1755. [PubMed: 6863754]

Montgomery BM, Fitch JL. The prevalence of stuttering in the hearing-impaired school age population. Journal of Speech and Hearing Disorders 1988;53(2):131–135. [PubMed: 3361855]

Moore, BCJ. An introduction to the psychology of hearing. 3. London: Academic Press; 1995.

Murray FP. An investigation of variably induced white noise upon moments of stuttering. Journal of Communications Disorders 1969;2:109–114.

Mysak ED. Servo theory and stuttering. Journal of Speech and Hearing Disorders 1960;25:188–195. [PubMed: 14425655]

Namasivayam AK, van Lieshout PHHM. Investigating speech motor practice and learning in people who stutter. Journal of Fluency Disorders 2008;33(1):32–51. [PubMed: 18280868]

Namasivayam AK, van Lieshout PHHM, De Nil LF. Bite-block perturbation in people who stutter: Immediate compensatory and delayed adaptive processes. Journal of Communication Disorders 2008;41(4):372–394. [PubMed: 18405914]

Namasivayam AK, van Lieshout PHHM, McIlroy WE, De Nil L. Sensory feedback dependence hypothesis in persons who stutter. Human Movement Science 2009;28(6):688–707. [PubMed: 19692132]

Neilson MD, Neilson PD. Speech motor control and stuttering: A computational model of adaptive sensory-motor processing. Speech Communication 1987;6(4):283–373.

Neilson, MD.; Neilson, PD. Adaptive model theory of speech motor control and stuttering. In: Peters, HFM.; Hulstijn, W.; Starkweather, CW., editors. Speech motor control and stuttering; Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering; 1990; Nijmegen, Netherlands. Amsterdam: Excerpta Medica; 1991. p. 149-156.

Nieto-Castanon A. *An investigation of articulatory-acoustic relationships in speech production* (Doctoral dissertation, Boston University, 2004). Dissertation Abstracts International 2004;65:05.

Nieto-Castanon A, Guenther FH, Perkell JS, Curtin HD. A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. Journal of the Acoustical Society of America 2005;117(5):3196–3212. [PubMed: 15957787]

Nippold MA. Stuttering and phonology: Is there an interaction? American Journal of Speech-Language Pathology 2002;11:99–110.

Nudelman, HB.; Herbrich, KE.; Hoyt, BD.; Rosenfield, DB. Dynamic characteristics of vocal frequency tracking in stutterers and nonstutterers. In: Peters, HFM.; Hulstijn, W., editors. Speech motor dynamics in stuttering. New York: Springer-Verlag; 1987. p. 161-169.

O'Brian S, Onslow M, Cream A, Packman A. The Camperdown program: Outcomes of a new prolonged-speech treatment model. Journal of Speech, Language, and Hearing Research 2003;46(4):933–946.

Packman A, Onslow M, van Doorn J. Prolonged speech and modification of stuttering: Perceptual, acoustic, and electroglottographic data. Journal of Speech and Hearing Research 1994;37(4):724–737. [PubMed: 7967557]

Parker CS, Christopherson F. Electronic aid in the treatment of stammer. Medical Electronics and Biological Engineering 1963;1:121–125.

Peeva MG, Guenther FH, Tourville JA, Nieto-Castanon A, Anton JL, Nazarian B, et al. Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. Neuroimage 2010;50(2):626–638. [PubMed: 20035884]

Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, et al. The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. Journal of the Acoustical Society of America 2004;116(4):2338–2344. [PubMed: 15532664]

Perkell JS, Matthies ML, Tiede M, Lane H, Zandipour M, Marrone N, et al. The distinctness of speakers' /s/-/s/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. Journal of Speech, Language, and Hearing Research 2004;47(6):1259–1269.

Perkins WH, Bell J, Johnson L, Stocks J. Phone rate and the effective planning time hypothesis of stuttering. Journal of Speech and Hearing Research 1979;22(4):747–755. [PubMed: 513683]

Perkins WH, Rudas J, Johnson L, Bell J. Stuttering: Discoordination of phonation with articulation and respiration. Journal of Speech and Hearing Research 1976;19(3):509–522. [PubMed: 979213]

Peters HFM, Boves L. Coordination of aerodynamic and phonatory processes in fluent speech utterances of stutterers. Journal of Speech and Hearing Research 1988;31(3):352–361. [PubMed: 3172752]

Peters, HFM.; Hietkamp, RK.; Boves, L. Aerodynamic and phonatory processes in disfluent speech utterances in stutterers. In: Starkweather, CW.; Peters, HFM., editors. Stuttering; Proceedings of the First World Congress on Fluency Disorders; Munich, Germany. Nijmegen, Netherlands: Nijmegen University Press; 1995. p. 76-81.

Peters HFM, Hulstijn W, Starkweather CW. Acoustic and physiological reaction times of stutterers and nonstutterers. Journal of Speech and Hearing Research 1989;32(3):668–680. [PubMed: 2779210]

Peters HFM, Hulstijn W, van Lieshout PHHM. Recent developments in speech motor research into stuttering. Folia Phoniatrica et Logopaedica 2000;52(1–3):103–119. [PubMed: 10474010]

Peterson GE, Barney HL. Control methods used in a study of the vowels. The Journal of the acoustical society of America 1952;24(2):175–184.

Poletto CJ, Verdun LP, Strominger R, Ludlow CL. Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures. Journal of Applied Physiology 2004;97(3):858–866. [PubMed: 15133000]

Postma A. Detection of errors during speech production: A review of speech monitoring models. Cognition 2000;77(2):97–132. [PubMed: 10986364]

Postma A, Kolk H. Error monitoring in people who stutter: Evidence against auditory feedback defect theories. Journal of Speech and Hearing Research 1992;35(5):1024–1032. [PubMed: 1447913]

Postma A, Kolk H. The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. Journal of Speech and Hearing Research 1993;36(3):472–487. [PubMed: 8331905]

Prosek RA, Montgomery AA, Walden BE, Hawkins DB. Formant frequencies of stuttered and fluent vowels. Journal of Speech and Hearing Research 1987;30(3):301–305. [PubMed: 3669636]

Purcell DW, Munhall KG. Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. Journal of the Acoustical Society of America 2006;120(2):966–977. [PubMed: 16938984]

Riley, GD. Stuttering severity instrument for children and adults. 3. Austin, TX: PRO-ED; 1994.

Riley GD, Ingham JC. Acoustic duration changes associated with two types of treatment for children who stutter. Journal of Speech, Language, and Hearing Research 2000;43(4):965–978.

Robb M, Blomgren M. Analysis of F2 transitions in the speech of stutterers and nonstutterers. Journal of Fluency Disorders 1997;22(1):1–16.

Russell, M.; Corley, M.; Lickley, R. Magnitude estimation of disfluency by stutterers and nonstutterers. In: Hartsuiker, RJ.; Bastiaanse, R.; Postma, A.; Wijnen, F., editors. Phonological encoding and monitoring in normal and pathological speech. Hove, UK: Psychology Press; 2005. p. 248-260.

Saltuklaroglu T, Kalinowski J. The inhibition of stuttering via the presentation of natural speech and sinusoidal speech analogs. Neuroscience Letters 2006;404(1–2):196–201. [PubMed: 16806702]

Sasisekaran J, De Nil LF, Smyth R, Johnson C. Phonological encoding in the silent speech of persons who stutter. Journal of Fluency Disorders 2006;31(1):1–21. [PubMed: 16405985]

Schmidt, RA.; Lee, TD. Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics; 2005.

Scott CM, Ringel RL. Articulation without oral sensory control. Journal of Speech and Hearing Research 1971;14(4):804–818. [PubMed: 5163911]

Shane, MLS. Effects on stuttering of alteration in auditory feedback. In: Johnson, WS.; Leutenegger, RR., editors. Stuttering in children and adults: Thirty years of research at the university of Iowa. Minneapolis: University of Minnesota Press; 1955. p. 286-297.

Sherrard CA. Stuttering as "false alarm" responding. British Journal of Disorders of Communication 1975;10(2):83–91. [PubMed: 1191509]

Smith A. The control of orofacial movements in speech. Critical Reviews in Oral Biology and Medicine 1992;3:233–267. [PubMed: 1571473]

Smith, A. Stuttering: A unified approach to a multifactorial, dynamic disorder. In: Ratner, NB.; Healey, EC., editors. Stuttering research and practice: Bridging the gap. Mahwah, NJ: Erlbaum; 1999. p. 27-43.

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(2):521–536.

Smits-Bandstra S, De Nil LF. Sequence skill learning in persons who stutter: Implications for cortico-striato-thalamo-cortical dysfunction. Journal of Fluency Disorders 2007;32(4):251–278. [PubMed: 17963936]

Snyder GJ, Hough MS, Blanchet P, Ivy LJ, Waddell D. The effects of self-generated synchronous and asynchronous visual speech feedback on overt stuttering frequency. Journal of Communication Disorders 2009;42(3):235–244. [PubMed: 19304293]

Stager SV, Denman DW, Ludlow CL. Modifications in aerodynamic variables by persons who stutter under fluency-evoking conditions. Journal of Speech, Language, and Hearing Research 1997;40(4):832–847.

Starkweather, CW. Fluency and stuttering. Englewood Cliffs, NJ: Prentice-Hall; 1987.

Stephen SCG, Haggard MP. Acoustic properties of masking/delayed feedback in the fluency of stutterers and controls. Journal of Speech and Hearing Research 1980;23:527–538. [PubMed: 7421156]

Stevens, KN. Acoustic phonetics. Cambridge, MA: The MIT Press; 1998.

Stevens KN, House AS. Development of a quantitative description of vowel articulation. Journal of the Acoustical Society of America 1955;27(3):484–493.

Stromsta C. Experimental blockage of phonation by distorted sidetone. Journal of Speech and Hearing Research 1959;2:286–301. [PubMed: 13835296]

Stromsta, C. A spectrographic study of dysfluencies labeled as stuttering by parents. Proceedings of the 13th Congress of the International Association of Logopedics and Phoniatrics, Vienna. De therapia vocis et loquelae; 1965. p. 317-320.

Stromsta C. Interaural phase disparity of stutterers and nonstutterers. Journal of Speech and Hearing Research 1972;15(4):771–780. [PubMed: 4652398]

Stromsta, C. Elements of stuttering. Oshtemo, MI: Atsmorts; 1986.

Stromsta, C.; Fibiger, S. In: Urban, BJ., editor. Physiological correlates of the core behavior of stuttering; Proceedings of the 18th Congress of the International Association of Logopedics and Phoniatrics; Washington, DC. 1980; Rockville, MD: American Speech-Language-Hearing Association; 1981. p. 335-340.

Subramanian A, Yairi E, Amir O. Second formant transitions in fluent speech of persistent and recovered preschool children who stutter. Journal of Communications Disorders 2003;36:59–75.

Sutton S, Chase RA. White noise and stuttering. Journal of Speech and Hearing Research 1961;4(1):72.

Tasko SM, McClean MD. Variations in articulatory movement with changes in speech task. Journal of Speech, Language, and Hearing Research 2004;47(1):85–100.

Tasko SM, McClean MD, Runyan CM. Speech motor correlates of treatment-related changes in stuttering severity and speech naturalness. Journal of Communication Disorders 2007;40(1):42–65. [PubMed: 16765980]

Terband H, Maassen B. Speech motor development in childhood apraxia of speech (CAS): Generating testable hypotheses by neurocomputational modeling. Folia Phoniatrica et Logopaedica. (in press).

Terband H, Maassen B, Guenther FH, Brumberg J. Computational neural modeling of speech motor control in childhood apraxia of speech (CAS). Journal of Speech, Language, and Hearing Research 2009;52(6):1595–1609.

Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. Neuroimage 2008;39(3):1429–1443. [PubMed: 18035557]

Toyomura, A.; Omori, T. Stochastic control system of basal ganglia and relation with stuttering. In: Hamza, MH., editor. Neural networks and computational intelligence; Proceedings of the 2nd IASTED International Conference; Grindelwald, Switzerland. Anaheim, CA: Acta Press; 2004. p. 31-36.

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 1996a;39(3):546–564. [PubMed: 8783133]

van Lieshout PHHM, Hulstijn W, Peters HFM. Speech production in people who stutter: Testing the motor plan assembly hypothesis. Journal of Speech and Hearing Research 1996b;39(1):76–92. [PubMed: 8820700]

van Lieshout, PHHM.; Hulstijn, W.; Peters, HFM. Searching for the weak link in the speech production chain of people who stutter: A motor skill approach. In: Maassen, B.; Kent, RD.; Peters, HFM.; van Lieshout, PHHM.; Hulstijn, W., editors. Speech motor control in normal and disordered speech. Oxford, UK: Oxford University Press; 2004. p. 313-355.

van Lieshout, PHHM.; Peters, HFM.; Bakker, A. En route to a speech motor test: A first halt. In: Hulstijn, W.; Peters, HFM.; van Lieshout, PHHM., editors. Speech production: Motor control, brain research, and fluency disorders; Proceedings of the 3rd International Conference on Speech Motor Production and Fluency Disorders; 1996; Nijmegen, Netherlands. Amsterdam: Elsevier; 1997. p. 463-471.

van Lieshout PHHM, Peters HFM, Starkweather CW, Hulstijn W. Physiological differences between stutterers and nonstutterers in perceptually fluent speech: EMG amplitude and duration. Journal of Speech and Hearing Research 1993;36:55–63. [PubMed: 8450665]

Van Riper, C. The nature of stuttering. Englewood Cliffs, NJ: Prentice-Hall; 1971.

Van Riper, C. The nature of stuttering. Englewood Cliffs, NJ: Prentice-Hall; 1982.

Venkatagiri HS. The relevance of DAF-induced speech disruption to the understanding of stuttering. Journal of Fluency Disorders 1980;5(2):87–98.

Villacorta VM, Perkell JS, Guenther FH. Sensorimotor adaptation to feedback perturbations on vowel acoustics and its relation to perception. Journal of the Acoustical Society of America 2007;122(4): 2306–2319. [PubMed: 17902866]

Ward D. Intrinsic and extrinsic timing in stutterers' speech: Data and implications. Language and Speech 1997;40(3):289–310. [PubMed: 9509581]

Webster RL, Dorman MF. Decreases in stuttering frequency as a function of continuous and contingent forms of auditory masking. Journal of Speech and Hearing Research 1970;13(1):82–86. [PubMed: 5421144]

Webster RL, Lubker BB. Masking of auditory feedback in stutterers' speech. Journal of Speech and Hearing Research 1968;11(1):221–223. [PubMed: 5648537]

Wieneke GH, Eijken E, Janssen P, Brutten GJ. Durational variability in the fluent speech of stutterers and nonstutterers. Journal of Fluency Disorders 2001;26(1):43–53.

Wingate ME. A standard definition of stuttering. Journal of Speech and Hearing Disorders 1964;29:484–489. [PubMed: 14257050]

Wingate ME. Stuttering as phonetic transition defect. Journal of Speech and Hearing Disorders 1969;34:107–108.

Wingate ME. Effect on stuttering of changes in audition. Journal of Speech and Hearing Research 1970;13(4):861–873. [PubMed: 5491360]

Wingate, ME. Stuttering: Theory and treatment. New York: Irvington; 1976.

Wingate, ME. Foundations of stuttering. San Diego, CA: Academic Press; 2002.

Wood SE. An electropalatographic analysis of stutterers' speech. European Journal of Disorders of Communication 1995;30(2):226–236. [PubMed: 7492853]

Yairi E. Effects of binaural and monaural noise on stuttering. The journal of Auditory Research 1976;16:114–119.

Yairi, E.; Ambrose, NG. Early childhood stuttering for clinicians by clinicians. Austin, TX: PRO-ED; 2005.

Yaruss JS, Conture EG. F2 transitions during sound/syllable repetitions of children who stutter and predictions of stuttering chronicity. Journal of Speech and Hearing Research 1993;36(5):883–896. [PubMed: 8246477]

Yates AJ. Delayed auditory feedback. Psychological Bulletin 1963;60:213–232. [PubMed: 14002534]

Zebrowski PM, Conture EG, Cudahy EA. Acoustic analysis of young stutterers fluency: Preliminary observations. Journal of Fluency Disorders 1985;10(3):173–192.

Zebrowski, PM.; Moon, JB.; Robin, DA. Visuomotor tracking in children who stutter: A preliminary view. In: Hulstijn, W.; Peters, HFM.; van Lieshout, PHHM., editors. Speech production: Motor control, brain research, and fluency disorders; Proceedings of the 3rd International Conference on Speech Motor Production and Fluency Disorders; 1996; Nijmegen, Netherlands. Amsterdam: Elsevier; 1997. p. 579-584.

Zimmermann G. Articulatory behaviors associated with stuttering: A cinefluorographic analysis. Journal of Speech and Hearing Research 1980a;23(1):108–121. [PubMed: 7442171]

Zimmermann G. Articulatory dynamics of fluent utterances of stutterers and nonstutterers. Journal of Speech and Hearing Research 1980b;23(1):95–107. [PubMed: 7442187]

Zimmermann G. Stuttering: A disorder of movement. Journal of Speech and Hearing Research 1980c; 23(1):122–136. [PubMed: 6777605]

Appendix

The DIVA model parameters are given in Table A1. The values of the time constant parameters are taken from Guenther et al. (2006), and those of the α_{ff} and α_{fb} parameters in the non-stuttering DIVA version from Tourville et al. (2008). The ε and ξ parameters, not previously used in the DIVA model, were chosen such as to fit best published behavioral data. This extra degree of modeling flexibility does not weaken the results of the study since the model can mimic the trends in the behavioral data also when using other values of these parameters. Table A2 lists the model's equations; all but the first three are adapted from Guenther et al. (2006). The equation for the feedback motor command was modified to use the auditory feedback loop only (see Section 2.1).

Biographies

Oren Civier received his PhD from Boston University in 2010. Under the supervision of Prof. Guenther, Oren used computational modeling to investigate the neural substrates of stuttering and induced fluency. His research interests include the involvement of the medial wall and basal ganglia in motor control, sequencing, and dopamine-related disorders.

Stephen Tasko is an associate professor of speech pathology at Western Michigan University. His research interests include the speech motor characteristics of stuttering, normal speech motor control and voice disorders.

Frank Guenther is a computational and cognitive neuroscientist specializing in speech and sensorimotor control. He received his MS from Princeton University in 1987 and PhD from Boston University in 1993. His research combines theoretical modeling with behavioral and neuroimaging experiments to characterize the neural computations underlying speech and language.
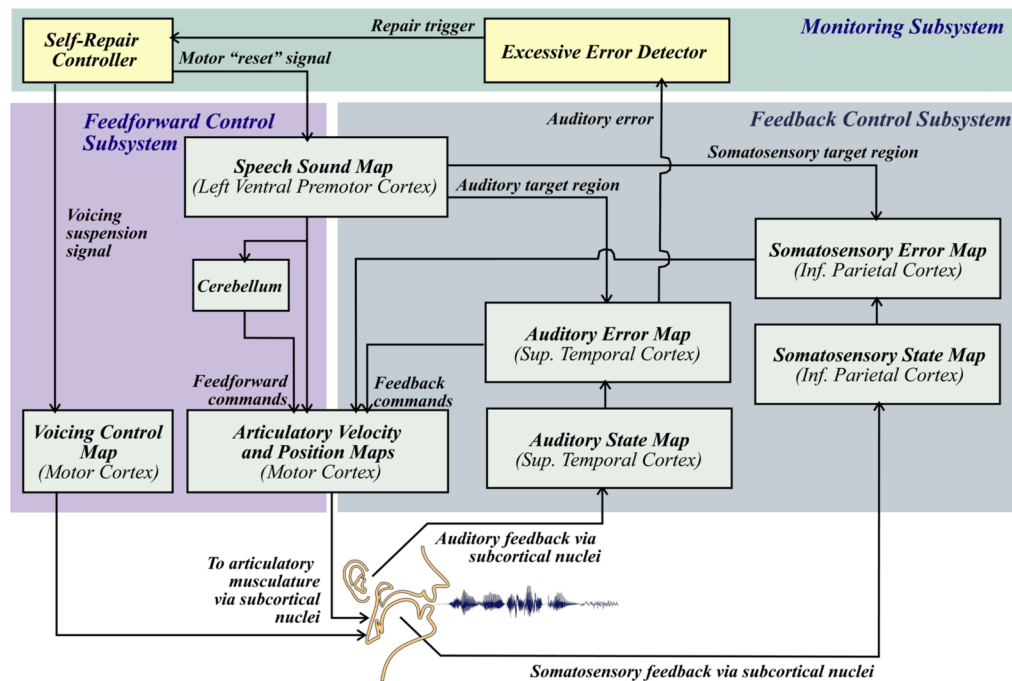
Fig. 1.
Schematic of the DIVA model. The cyan boxes correspond to sets of neurons (or maps) in the model, and the yellow boxes correspond to model's components whose anatomical locations are not mapped yet. Arrows correspond to synaptic projections that transform one type of neural representation into another, or transmit information between model's components. The model is divided into three distinct systems: the Feedforward control subsystem on the left, the Feedback control subsystem on the right, and the Monitoring subsystem on the top. The feedforward control and feedback control subsystem are reproduced from Guenther et al. (2006) except for the Voicing control map, which is described in Section 4.1. The monitoring control subsystem, not previously described by the DIVA model, was developed for this study.
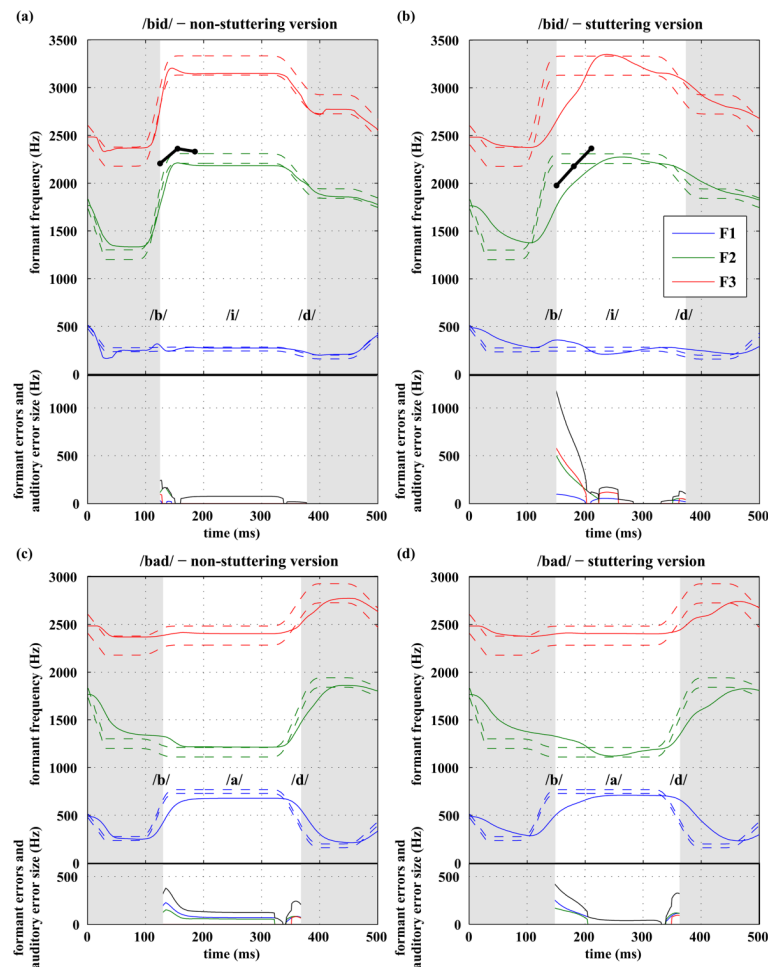
Fig. 2.
Auditory feedback, target region, and errors for the non-stuttering (left plots) and stuttering (right plots) versions of the DIVA model, producing /bid/ (top plots) and /bad/ (bottom plots). The top of each plot contains the target region lower and upper bound frequencies (dashed lines), with a pair of bounding lines for each formant (different formants coded by different colors). For each formant, the auditory feedback frequency during the simulation is indicated by solid colored line (including in the shaded periods were voicing was off, see Section 3.2 for details). The auditory error size (black) and the formant errors (other colors) are plotted on the bottom of each plot. Indicated on the top plots with filled circles connected with straight lines are the results from Robb & Blomgren (1997) who sampled F2 during productions of /bid/ at 0 ms, 30 ms, and 60 ms after voicing onset. Voicing onset follows initial articulatory movement by 28.88 ms for PWS and 18.10 ms for PNS (Healey & Gutkin, 1984).
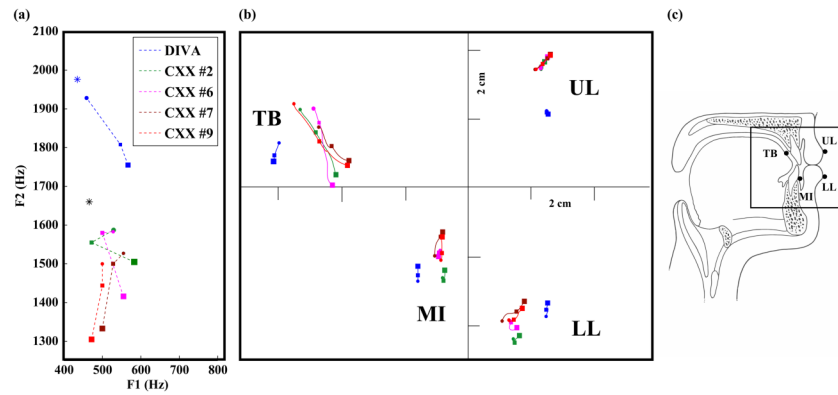
Fig. 3.
One simulated (blue) and four recorded (other colors) productions of the first vowel
(pronounced as /e/) in "a bad daba". Simulations were performed with the stuttering version
of the DIVA model. Recordings are of the stuttering speaker CXX. For each replicate, large
and small squares mark the beginning points of the 1st and 2nd attempts respectively. Filled
circles mark the beginning points of the final complete production that follows the repetition.
(a) Acoustic data. F2 is plotted against F1. Formant values were measured at the first glottal
pulse of each attempt. A dashed line connects the markers of each repetition. The locations of
the replicates in the series listed in Table 1 are specified in the legend. Stars mark the formant
locations for fluent productions of /e/ (blue for the simulation and black for CXX). (b)
Articulatory data. For each production, the trajectories (X and Y coordinates) of four sensors
positioned on different articulators are plotted. UL--upper lip. LL--lower lip. MI--Mandibular
incisor. TB--tongue blade. For the simulation, the trajectories are of the approximate positions
of the sensors on DIVA's artificial vocal tract. For subject CXX, sensor trajectories were
measured by means of an electromagnetic movement analysis (EMA) system. The trajectories
follow sensor locations starting at the beginning of the 1st attempt and ending at the beginning
of the final complete production. (c) A schematic diagram (subject identity unknown) showing
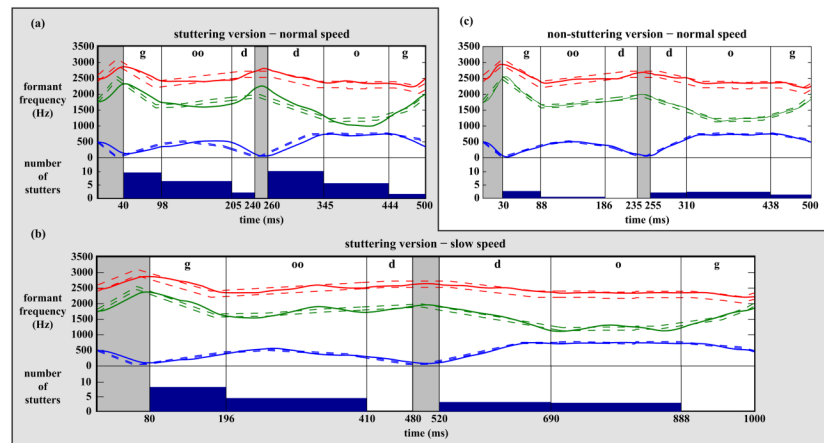sensor approximate locations in the mid-sagittal plane of the vocal tract.

Fig. 4.
Acoustics and frequency of sound/syllable repetitions made by the stuttering version of the DIVA model at normal (a) and slow speeds (b), and repetitions made by the non-stuttering version of the DIVA model, at normal speed only (c). The top parts of the plots use the same conventions as in Fig. 2. In the bottom of the plots, histograms show how many repetitions occurred on each phoneme of the utterance in 500 productions of "good dog" (silent periods are shaded).
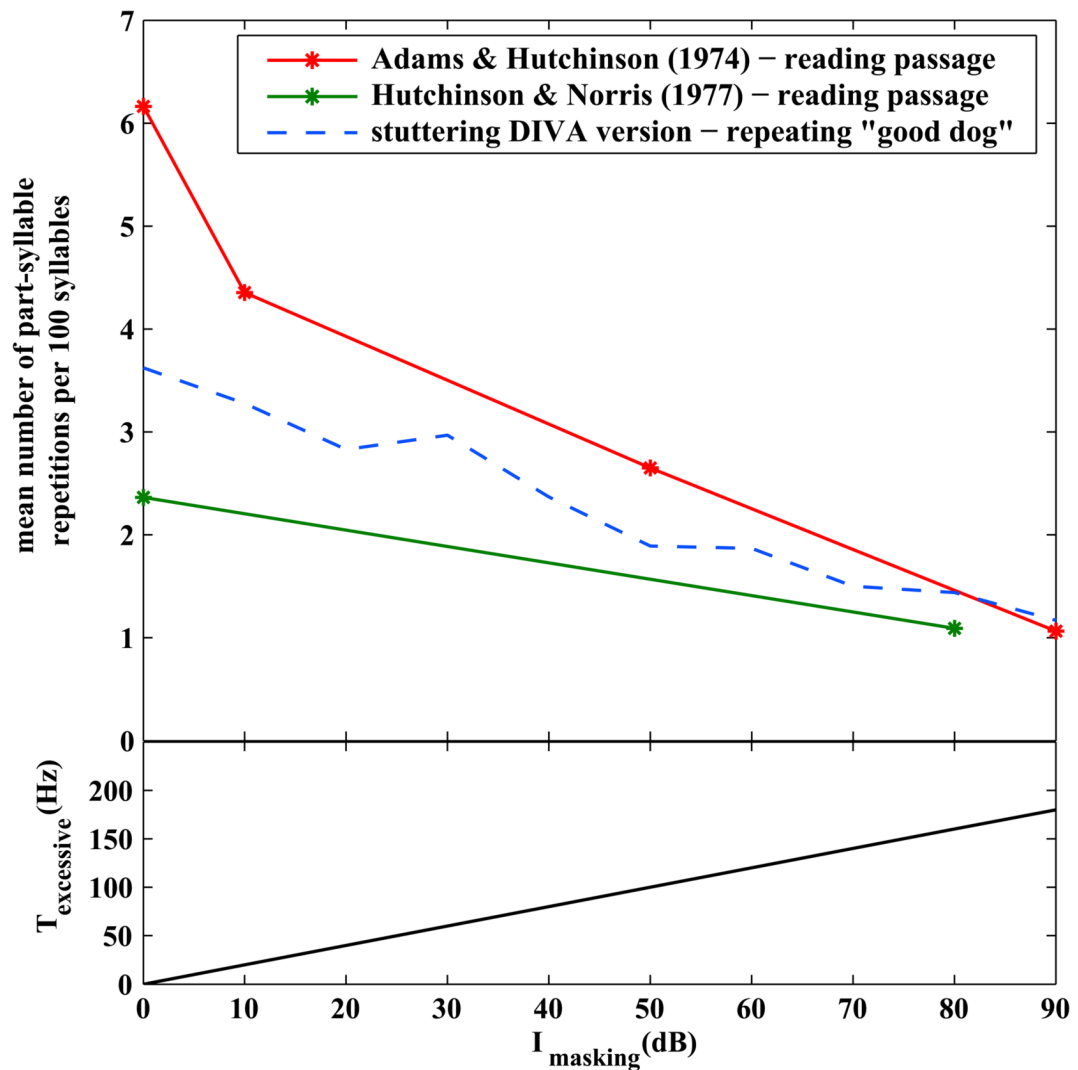
Fig. 5.
Frequency of sound/syllable repetitions made by the stuttering version of the DIVA model and selected studies under masking noise of different intensities. The number of repetitions (upper y-axis) under different noise intensities (x-axis) contrasts the DIVA model simulation and the experimental results. The threshold for detecting excessive errors, $T_{excessive}$ (lower y-axis), is indicated by the black line. Study results are expressed as repetitions per 100 syllables (see Section 6.3 for details).

Table 1

F2 values during repetitions made on the first vowel (pronounced as /e/) of "a bad daba" by the stuttering speaker CXX and the stuttering version of the DIVA model

#	1st disrupted attempt		2nd disrupted attempt		Final complete production
	F2 (Hz)	Duration (ms)	F2 (Hz)	Duration (ms)	F2 (Hz)
Stuttering speaker CXX					
1	1465	130			1709
2*	1506	80	1555	120	1588
3	1465	130			1627
4	1465	130			1627
5	1546	60	1587	70	1543
6*	1415	180	1581	20	1584
7*	1329	90	1498	120	1526
8	1302	10	1546	140	1627
9*	1303	100	1448	110	1494
10	1383	70			1505
11	1343	120			1546
Mean (SD)	1411 (85)	100 (45)	1535 (53)	97 (44)	1580 (65)
Stuttering version of the DIVA model					
	1755	66	1808	76	1928

*
Repetition is also plotted in Fig. 3.

Table A1

DIVA model parameters used in the simulations

Name	Value	Description
α_{ff}	0.85 (non-stuttering version) 0.25 (stuttering version)	Contribution of feedforward command to total command; feedforward gain
α_{fb}	0.15 (non-stuttering version) 0.75 (stuttering version)	Contribution of the feedback command; feedback gain
E	$2 * 10^{-3}$	Relation between auditory error size (in hertz) and likelihood of error repair
Ξ	2	Relation between masking noise intensity (in decibels) and excessive-error detection threshold (in hertz)
τ_{MAr}	42 ms	Transmission delay from motor cortex cell activity to physical movement of articulators
τ_{AcAu}	20 ms	The time it takes an acoustic signal transduced by the cochlea to make its way to the auditory cortical areas
τ_{PAu}	3 ms	Transmission delay from premotor to auditory cortex
τ_{AuM}	3 ms	Transmission delay from auditory to motor cortex

Table A2

DIVA model equations used in the simulations

Equation	Description
$P(R_{error}(t)) = \varepsilon * f_{masking}(f_{size}(\Delta Au(t)), I_{masking})$	Likelihood of repair trigger
$f_{masking}(x, I_{masking}) = \max(0, x - T_{excessive}(I_{masking}))$	Masking noise effect on auditory error size
$T_{excessive}(I_{masking}) = \xi * I_{masking}$	Error detection threshold
$\dot{M}(t) = \alpha_{ff} \dot{M}_{feedforward}(t) + \alpha_{fb} \dot{M}_{feedback}(t)$	Overall motor command
$g(t)$	Go signal
$\dot{M}_{feedforward}(t) = P(t)z_{PM}(t) - M(t)$	Feedforward motor command
$P_i(t) = \begin{cases} 0 & \text{If sound is being produced or perceived} \\ 1 & \text{Otherwise} \end{cases}$	Speech sound map activity
z_{PM}	Synaptic weights encoding feedforward commands for a speech sound
$\dot{M}_{feedback}(t) = \Delta Au(t - \tau_{AuM})z_{AuM}$	Feedback motor command
z_{AuM}	Synaptic weights that transform auditory error into corrective motor velocities for a speech sound
$\Delta Au(t) = Au(t) - P(t - \tau_{PAu})z_{PAu}(t)$	Auditory error map activity
z_{PAu}	Synaptic weights encoding auditory expectation for a speech sound
$Au(t) = f_{AcAu}(Acoust(t - \tau_{AcAu}))$	Auditory state activity
$Acoust(t) = f_{ArAc}(Artic(t))$	Physical acoustic signal resulting from the current articulator configuration
$Artic(t) = f_{MAr}(M(t - \tau_{MAr}))$	Position of model articulators (Maeda parameter values)