Simulations of Feedback and Feedforward Control in Stuttering

Oren Civier[1] and Frank H. Guenther[1,2,3]

[1]Department of Cognitive and Neural Systems, Boston University.
[2]Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology.
[3]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital.

Abstract

Simulation of speech production using a “neurologically impaired” model reveals patterns similar to stuttering: 1) High frequency of stutters on initial sounds; 2) enhancement of fluency by exposure to white noise; 3) enhancement of fluency by reducing the rate of speech. The results support the notion that stuttering may be in part due to weakening of the pathways involved in feedforward control of well practiced speech sounds and the consequent dominance of auditory feedback control.

Introduction

Dysfunction of auditory feedback has long been suspected as a source of stuttering (Fairbanks, 1954), mainly due to the apparent effect of delayed auditory feedback in alleviating stuttering (Kalinowski, Armson, Roland-Mieszkowski, Stuart, & Gracco, 1993) and the low incidence of stuttering among the deaf (Van Riper, 1971). Neilson & Neilson (1987) modeled feedback control and concluded that speech production is too fast to be controlled solely by feedback. The intrinsic delay between motor command and its auditory consequences will render the system unstable, especially during rapid transitions (e.g. consonants). Therefore, fluent speakers must rely primarily on feedforward control independent of auditory feedback.

The hypothesis investigated here is that, due to weakened feedforward control projections, some people who stutter may rely too heavily on feedback control (Max, Guenther, Gracco, Ghosh, & Wallace, 2004). When the feedback instabilities create too large an auditory error (the difference between the expected and produced auditory signal), the system performs a "reset" of the current syllable production, resulting in a part-word repetition. Our hypothesis is consistent with neurological evidence that stuttering adults show abnormalities in the white matter pathways underlying the orofacial area of the left hemisphere primary motor cortex (Sommer, Koch, Paulus, Weiller, & Buchel, 2002). Damage to these pathways may compromise the feedforward command from premotor to primary motor areas.

DIVA: A neural model of speech production

DIVA is a biologically plausible neural network model capable of simulating production and development of fluent speech (e.g., Guenther, Hampson, & Johnson, 1998). It combines mathematical descriptions of underlying commands, cerebral and cerebellar neural substrates corresponding to the model’s components, and computer simulations controlling an articulatory synthesizer. In the model (schematically represented in Figure 1), cells in the motor cortex generate the overall motor command, $M(t)$, for producing a speech sound. $M(t)$ is a combination of a feedforward command (i.e., a command that was previously learned for producing the sound and does not rely on auditory feedback for its execution) and a feedback command (created by comparing actual auditory feedback to a learned auditory target, then correcting for any mismatch)[1]:

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

$$\alpha_{ff} + \alpha_{fb} = 1$$

with α_{ff} and α_{fb} represent the amount of weighting toward feedforward and feedback control. $\dot{M}_{feedforward}(t)$ and $\dot{M}_{feedback}(t)$ are the feedforward and feedback commands, respectively.

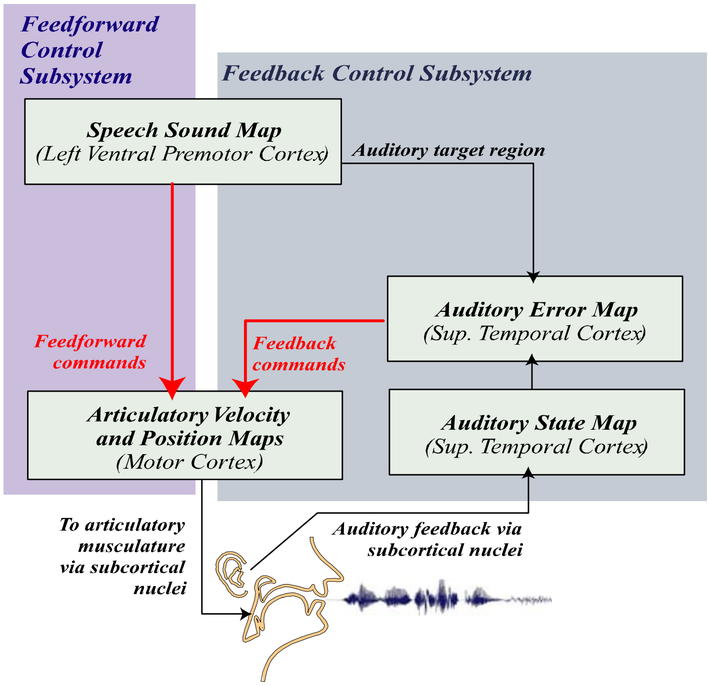

Figure 1 - Schematic of the DIVA model.

The stages of learning in the model are as follows:
1. Tune feedback control subsystem during babbling (self generated speech sounds),
2. Learn an auditory target (formant trajectory ranges) when a new sound sample is presented,
3. Learn a feedforward command for the sound by practicing its production.

In effect, the feedback control subsystem "tunes" the feedforward command with practice. Once an accurate feedforward command is learned, there will be little or no auditory error, and thus the feedback control subsystem will no longer play a major role. However, if the feedforward projections are weak (e.g, in some people who stutter), the feedback controller will always be engaged and can cause instabilities, particularly during rapid speech. Our hypothesis is in keeping with the fact that the onset of stuttering occurs in early childhood, when this transition from feedback to feedforward control takes place.

---

[1] The full DIVA model also includes somatosensory feedback control. It is not addressed here for the sake of simplicity. For more information regarding the full DIVA model, see Guenther, Ghosh, & Nieto-Castanon (2003).

Simulations

Normal vs. stuttering versions of the model

We simulated speech production of the utterance “good dog” with weakened feedforward projections by setting α_{ff} = 0.25 and α_{fb} = 0.75 (the stuttering version), and compared it with a simulation with normal feedforward projections by setting α_{ff} = 0.75 and α_{fb} = 0.25 (the normal version). From the simulation results (Figure 2), it is evident that dominance of feedback control in the stuttering version causes instabilities (large auditory errors due to delayed feedback processing) that increase the chance for stutters.

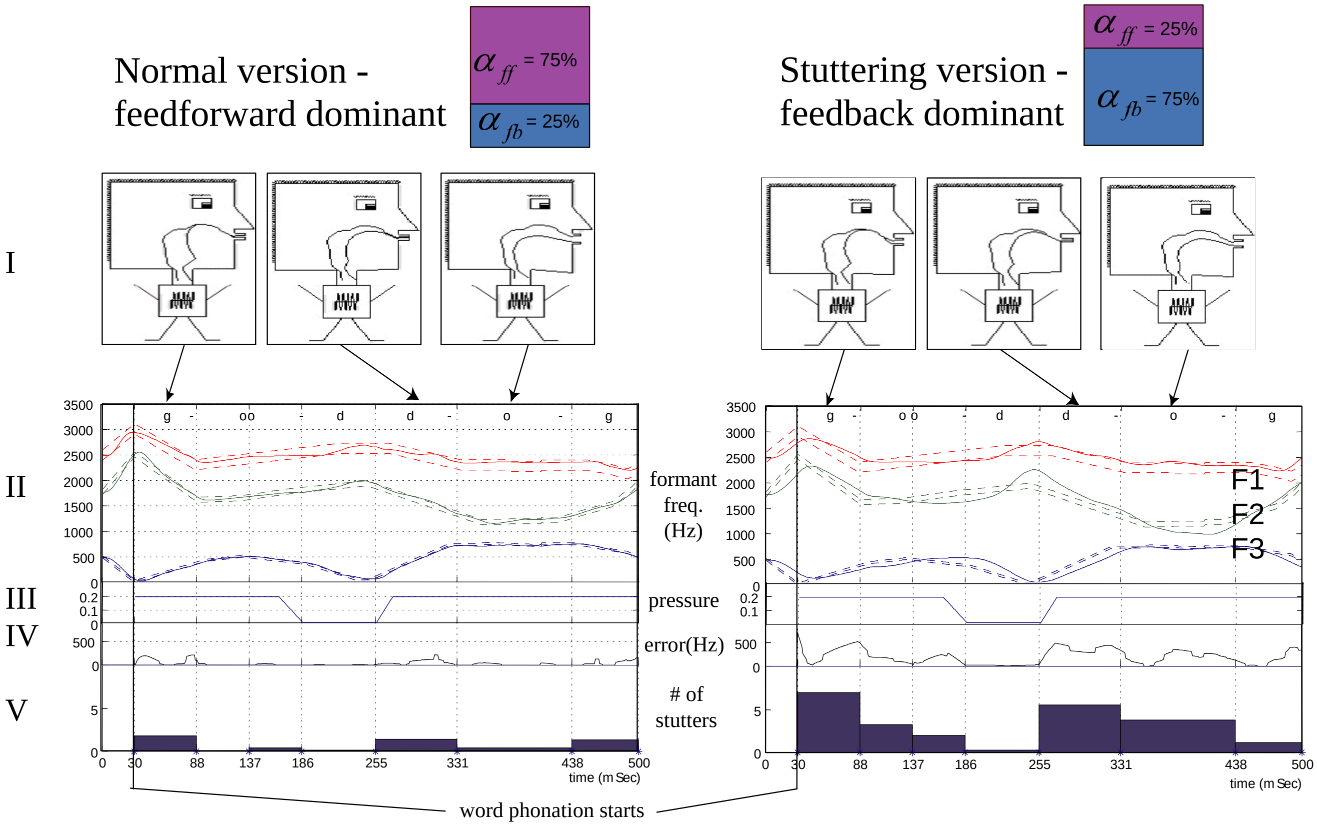

Figure 2 - Normal vs. Stuttering simulations. I. Vocal tract configurations. The images were created from the Maeda articulator model which is used by DIVA to synthesize the acoustic signal (Maeda, 1990). The Maeda model has 7 degrees of freedom: Jaw (1), tongue (3), lips (2) and larynx height (1). Each frame shows the vocal tract configuration at a specific moment in time (corresponding approximately to the phonemes “d”, “g”, and “aa”). **II. Auditory target region and actual trajectory.** The solid lines follow the 3 formant frequencies of the produced waveform. Dashed lines represent the 3 auditory target regions. When one of the formant frequencies is out of the target region (i.e., solid line is not between dashed lines), an auditory error is generated. **III. Glottal pressure.** The glottal pressure is the main indicator for vocal intensity, which modulates the auditory feedback. **IV. Auditory error.** The sum of the absolute auditory errors for the 3 formants. Notice that when glottal pressure is 0, no auditory error exists since there is no voicing. **V. Stutter distribution.** We assume that, due to noise in the system, the exact stutter position is not deterministic. For every 1/100 of a second, t, $P(stutter(t)) = \varepsilon * Error(t)$, where ε is a parameter set to $1x10^{-5}$ in the simulations. The histogram shows how many stutters on average occur on each sound in 37 repetitions of “good dog”. Sounds where the auditory error is greater, particularly at the beginning of words, are more likely to be stuttered.

Since the movement of the articulators into their initial positions normally takes place before word phonation starts (0-30ms), auditory feedback is not available at that time. For individuals with weak feedforward projections, the articulators will not reach their appropriate positions without auditory feedback. Thus, when phonation starts, the system will detect an auditory error and try to correct it using auditory feedback. Unfortunately, the correction can lead to instabilities (as described above), especially during sharp formant transitions. This can explain the higher frequency of stuttering on the initial sound or syllable of the word (where over 90% of stutters occur), especially if it is a consonant (Bloodstein, 1995). Figure 3 demonstrates a part-word repetition on the initial sound of the word "good" produced by the stuttering version.

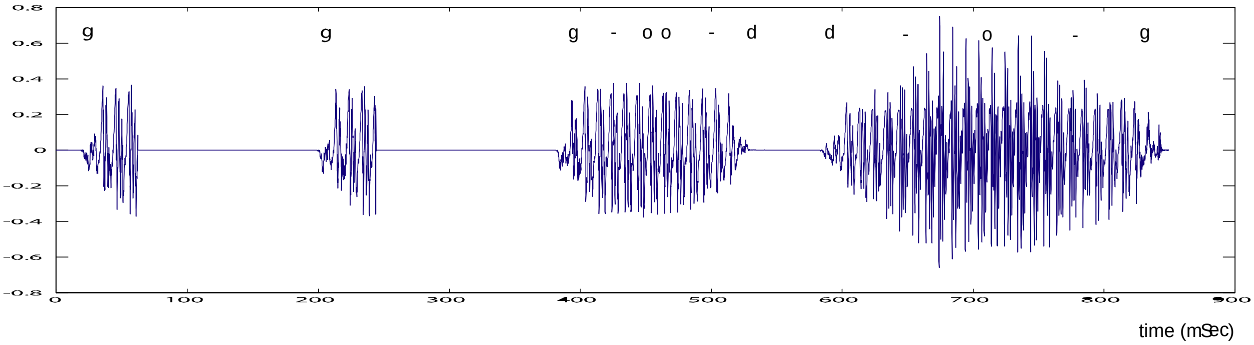

Figure 3 - Acoustic waveforms produced by the stuttering version of the model. Notice 3 repetitions of the initial sound "g" in the stuttering version.

Stuttering version with white noise

Since most subjects in white noise experiments report hearing themselves (even with loud noise presented binaurally through earphones), Bloodstein (1995) concluded that the noise acts as a distraction and not as a mere masker. Here we simulate increased noise level by reducing feedback dominance (Figure 4). We assume that the distraction of noisy auditory feedback prevents focusing on feedback control and forces control using primarily feedforward commands. Our assumption follows Van Riper's (1971) suggestion that distraction to the auditory feedback shifts attention to other forms of control (due to competition between control channels).

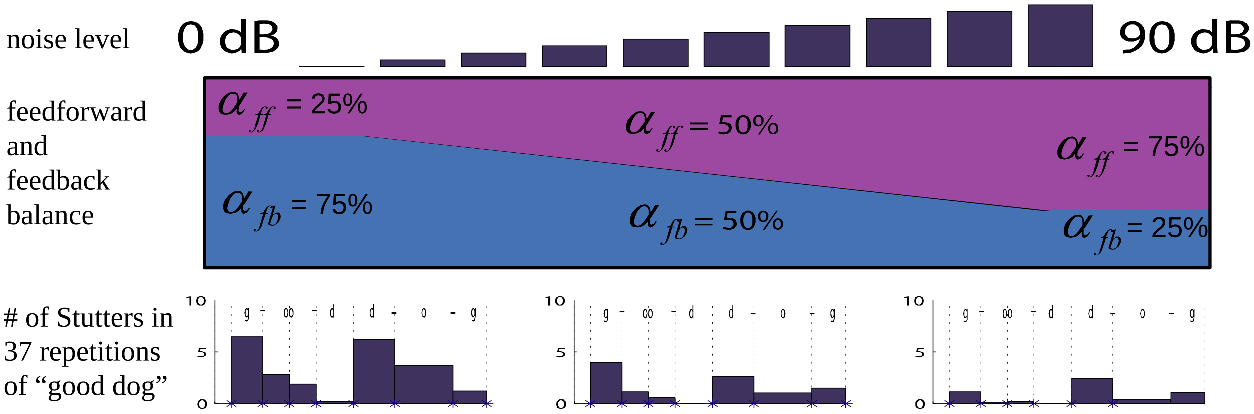

Figure 4 – Stuttering version with white noise.

In Figure 5, the simulation results (in blue) are compared with results of two very similar white noise experiments (in red) by Maraist & Hutton (1957) and Adams & Hutchinson (1974). Both involved ~15 subjects, and increasing noise intensities. In both simulation and experiments, there is an approximately linear increase in fluency with level of noise.

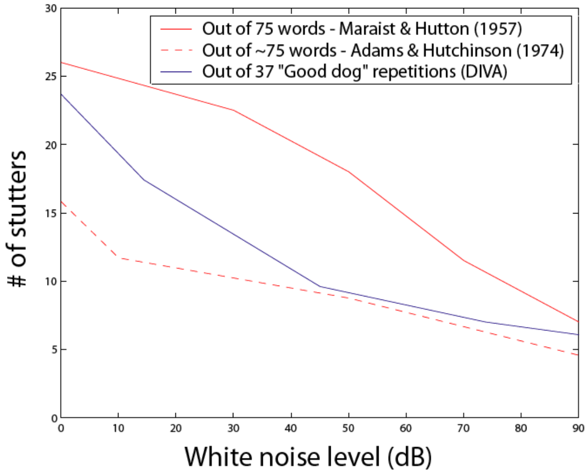

Figure 5 - Stuttering with white noise. Simulation compared with experiments.

In the white noise simulation, louder white noise enforces a feedforward/feedback balance more similar to the normal version. Consequently, less stutters are generated. We propose that the same mechanisms may act to enhance fluency in the white noise experiments described above.

Stuttering version at a slower rate

When slowing down the speaking rate, people who stutter tend to either reduce articulatory rate or insert pauses between words (depending on experiment design). Here we simulate reduced articulatory rate (induced by timed word production) by stretching of the formant trajectories in time (Figure 6).

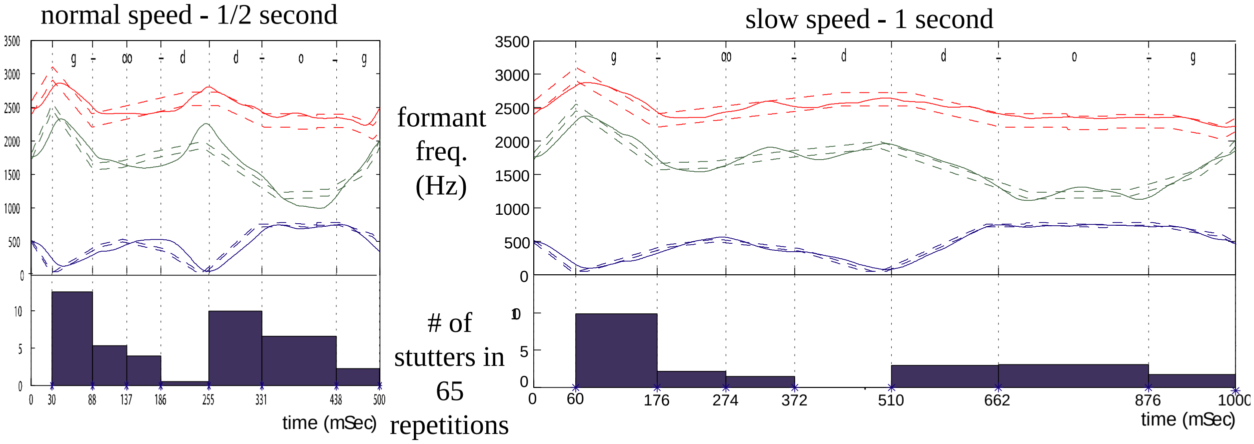

Figure 6 - Stuttering version at normal vs. slow speech rate.

When simulating the stuttering version at a reduced rate, transitions are not so sharp. Consequently, the feedback delays are less of a problem and fewer stutters are generated. We predict that the same mechanisms act to enhance fluency in the two comparable timed word production experiments: Perkins, Bell, Johnson, & Stocks (1979) who had 19 subjects read aloud 1 word per 2 seconds, and Adams, Lewis, & Besozzi (1973) who instructed 15 subjects to read 1 word per second. In both simulation and experiments, slowing down cuts stuttering by approximately half.

Conclusions

The simulation results account for several experimental findings regarding stuttering, therefore supporting the hypothesis that dominance of feedback control due to weakened feedforward projections is a possible source of stuttering. In the simulations of the normal vs. stuttering version we showed that an overemphasis on feedback control results in stuttering-like behavior. Moreover, since auditory feedback control is useless before phonation starts, stuttering is more likely to occur on the initial sound of a word. In the stuttering with white noise simulation, shifting emphasis from feedback to feedforward control by using white noise enhances fluency. Finally, we demonstrated that slowing down articulation can also enhance fluency, by creating better conditions for feedback control.

Acknowledgements

This study was supported by NIH/NIDCD grants R01 DC02852 (P.I. Frank Guenther) and RO1 DC03007 (J. Perkell, PI). We would like to thank Satrajit S. Ghosh and Jonathan Brumberg for the development and maintenance of the DIVA model code.

References

Adams, M.R., & Hutchinson, J. (1974) The effects of three levels of auditory masking on selected vocal characteristics and the frequency of disfluency of adult stutterers. J Speech Hear Res. 17(4):682-8

Adams, M.R., Lewis, J.I., & Besozzi, T.E. (1973) The effect of reduced reading rate on stuttering frequency. J Speech Hear Res. 16(4):671-5

Bloodstein, O. (1995) A handbook on stuttering / Oliver Bloodstein. San Diego, Calif. : Singular Pub. Group

Fairbanks, G. (1954) Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. J. Speech Hear. Disord., Vol 19, pp. 133-139

Guenther, F.H., Ghosh, S.S., & Nieto-Castanon, A. (2003) A neural model of speech production. Proceedings of the 6th International Seminar on Speech Production, Sydney, Australia.

Guenther, F.H., Hampson, M., & Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements. Psychological Review, 105, pp. 611-633.

Kalinowski, J., Armson, J., Roland-Mieszkowski, M., Stuart, A., & Gracco, V.L. (1993) Effects of alterations in auditory feedback and speech rate on stuttering frequency. Lang Speech. 36 (Pt 1):1-16.

Maraist, J.A., & Hutton, C. (1957) Effects of auditory masking upon the speech of stutterers. J Speech Hear Disord. 22(3):385-9

Max, L., Guenther, F.H., Gracco, V.L., Ghosh, S.S., & Wallace ME. (2004) Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. Contemporary Issues in Communication Science and Disorders, 31, pp. 105-122.

Maeda, S. (1990) Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. InW.J. Hardcastle and A. Marchal (Eds.), Speech production and speech modelling (pp. 131-149). Boston: Kluwer Academic Publishers.

Neilson, M.D., & Neilson, P.D. (1987) Speech motor control and stuttering: a computational model of adaptive sensory-motor processing. Speech communications 6. 325-333

Perkins, W.H., Bell, J., Johnson, L., & Stocks, J. (1979) Phone rate and the effective planning time hypothesis of stuttering. J Speech Hear Res. 22(4):747-55

Sommer, M., Koch, M.A., Paulus, W., Weiller, C., & Buchel, C. (2002) Disconnection of speech-relevant brain areas in persistent developmental stuttering. Lancet. 360(9330):380-3

Van Riper, C. (1971) The nature of stuttering. Englewood Cliffs, N.J.,: Prentice-Hall.