Sensorimotor learning and control in individuals who stutter

Kwang S. Kim

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Ludo Max, Chair

Fred Minifie

David J. Perkel

Kristie Spencer

Program Authorized to Offer Degree:

Speech and Hearing Sciences

University of Washington

ABSTRACT

Sensorimotor learning and control in individuals who stutter

Kwang S. Kim

Chair of the Supervisory Committee:

Professor Ludo Max

Speech and Hearing Sciences

Despite accumulating evidence that stuttering is associated with deficiencies in sensorimotor integration, the exact mechanisms underlying this disorder of speech fluency remain not only unknown but controversial. Since our research group's first formulation of a broad theoretical framework and a set of testable hypotheses (Max, 2004; Max, Guenther, Gracco, Ghosh, & Wallace, 2004), numerous studies have produced mostly supporting but also some seemingly incompatible results. Here, I present new results from a series of closely related studies designed to fill several critical gaps in our understanding of the sensorimotor mechanisms underlying stuttering. Specifically, I examined in both children and adults who stutter several aspects of speech and limb sensorimotor *learning* by means of adaptation paradigms in which

participants learn to update the planning of future movements in response to an experimentally introduced feedback perturbation. In addition, in adults who stutter, I also examined aspects of sensorimotor *control* by recording kinematic data for unperturbed articulatory movements and applying statistical methods that yield estimates of the relative contributions of feedforward and feedback control mechanisms.

The experiments presented in Chapter I were directly motivated by recent work, from our own lab and others, demonstrating poor performance in auditory-motor learning among adults who stutter (AWS) as compared with adults who do not stutter (AWNS) (Daliri, Wieland, Cai, Guenther, & Chang, 2018; Daliri & Max, 2018; Sengupta, Shah, Gore, Loucks, & Nasir, 2016). However, it remained controversial whether such poor sensorimotor learning can be causally related to the mechanisms that are responsible for breakdowns in speech fluency. For example, in adult speakers, multiple years of coping with stuttering may have affected the functioning of neuromotor control processes, and this, in turn, may affect assessments of sensorimotor learning. In fact, Daliri et al. (2018) specifically reported not finding a difference in auditory-motor learning between children who stutter (CWS) and children who do not stutter (CWNS), suggesting that no learning deficit was present during the childhood years closer to the onset of stuttering. In addition, it remains unclear whether sensorimotor learning limitations in individuals who stutter are speech-specific or whether they also affecting nonspeech effector systems such as the upper limb during reaching and grasping. sensorimotor system.

Therefore, the first study (Chapter I) in the series aimed to examine both speech auditory-motor and reach visuomotor learning in adults and children who stutter. Results suggest that both children and adults who stutter showed statistically significant limitations in auditory-motor adaptation to formant-shifted auditory feedback. In fact, this limitation was even more profound in children than in adults and in younger children (3-6 years of age) versus older children (7-9 years of age). Between-group differences in the adaptation of reaching movements performed with rotated visual feedback were subtle, but still statistically significant, for adult participants. In children, even nonstuttering children showed limited visuomotor adaptation, and, as a result, there was no difference between the stuttering and nonstuttering groups. From these results, I conclude that sensorimotor learning is deficient in individuals who stutter, and that at least substantial speech auditory-motor learning problems are already present at a young age near the onset of stuttering. Thus, atypical motor learning may play a critical role in the fundamental mechanisms underlying stuttering.

In the first study described above, participants' auditory feedback was manipulated with a commercially available vocal effects processor. Prior to the adoption of a different methodological approach that implements the feedback perturbation in user-adjustable Matlab (The Mathworks) software (i.e., the Audapter package, Cai, Boucek, Ghosh, Guenther, & Perkell, 2008; Cai, Ghosh, Guenther, & Perkell, 2010; Cai, Ghosh, Guenther, & Perkell, 2011), it became clear that the literature contained much implausible information about the minimum time latencies that can be achieved when combining this software with a personal computer and professional or consumer-grade audio interfaces. It is crucial, however, for researchers to be cognizant of the fact that the reproducibility of results from speech auditory-motor adaptation studies depends on accurately measuring and reporting the overall feedback loop latency that was in effect during the experiment. Therefore, in a second study (Chapter II), I measured the total feedback loop latencies (including both hardware and software latency) for various hardware and software combinations. Results showed that hardware-specific latencies were overlooked in several published reports, but that these latencies are not at all negligible for some of the tested audio interfaces (adding up to 15 ms delay). In addition, the measured total latencies were also generally larger than claimed in the literature. Therefore, the manuscript included as Chapter II emphasizes that the actual total latency (hardware plus software) needs to be correctly measured and described in all published reports. Furthermore, the study also demonstrated that the use of non-default parameter values can improve Audapter's own processing latency without negative impact on formant tracking. The paper concludes with several recommendations to improve feedback latency in speech auditory-motor adaptation paradigms.

The third study (Chapter III) examined which sub-processes of learning may be implicated in stuttering individuals' auditory-motor learning difficulties. Recent studies in upper-limb motor control indicate that sensorimotor learning involves at least two distinct components: (a) an explicit component that includes intentional strategy use and presumably is driven by target error, and (b) an implicit component that updates an internal model without awareness of the learner and presumably is driven by sensory prediction error. The presented study constitutes an initial attempt at dissociating these components for speech auditory-motor learning in AWS versus adults who do not stutter (AWNS). I developed a novel paradigm to obtain information on participants' awareness and intent related to adaptive articulatory behavior when speaking with formant-shifted feedback. First, results replicated previous findings that such auditory-motor learning is indeed limited in AWS. Second, in neither group did participants report any awareness at all of changing their productions in response to the perceived auditory feedback. These findings suggest that speech auditory-motor adaptation relies exclusively on implicit learning, and, therefore, that the limited adaptation found in AWS is likely due to poor implicit learning mechanisms. This conclusion is in direct agreement with other recent lines of evidence implying sensory prediction deficits in stuttering.

Chapter IV includes a study investigating sensorimotor *control* rather than *learning* in AWS versus AWNS. The foundation for this work lies in the well-established observation that the central nervous system (CNS) has the remarkable ability to both pre-plan a desired movement prior to movement onset (i.e., feedforward control) and quickly and efficiently adjust the ongoing movement based on sensory input (i.e., feedback control). Our laboratory's theoretical perspective on stuttering suggests that an atypically large weight on online feedback-based control may render the sensorimotor system unstable—potentially leading to sustained or repetitive articulatory movements and, thus, speech dysfluencies (Max, 2004; Max et al., 2004). My previous work already demonstrated the feasibility of using kinematic data to examine feedforward versus feedback mechanisms in the control of unperturbed speech movements in nonstuttering adults (Kim & Max, 2014). For the new study presented here, I used the same approach to analyze the kinematic data of jaw and tongue movements in AWS and AWNS. Specifically, the study determined how well initial kinematic landmarks (peak acceleration, peak velocity) could predict final movement kinematics (movement extent). Results show generally large correspondence between initial and final kinematics in both groups, suggesting that speech movements are mostly under feedforward control. However, this relationship was statistically significantly weaker for stuttering speakers. In addition, estimates of feedback-driven adjustments in movement duration were significantly larger for AWS than for AWNS in the second half of the movement, suggesting that stuttering individuals' regulate duration online in order to compensate for less efficient planning of the final kinematics. Thus, the findings add support for the hypothesis that stuttering individuals may over-rely on online feedback control and less on feedforward control (Max et al., 2004).

Taken together, the findings from these studies provide novel insights into the sensorimotor integration impairments underlying stuttering. The sensorimotor learning studies demonstrated that sensory prediction errors may not be correctly integrated for subsequent movement planning in both CWS and AWS, and that this limitation reflects less than optimal implicit learning processes. The sensorimotor control study confirmed that AWS are indeed more dependent on online feedback for immediate within-movement corrections. This necessary but inefficient control strategy may ultimately lead to the repetitive corrections or postural fixations that are perceived as stuttering moments during speech production (Max et al., 2004; Max & Daliri, 2019).

References

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/. *Proceedings of the 8th Intl. Seminar on Speech Production,* , 65-68.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America, 128*(4), 2033-2048. 10.1121/1.3479539 [doi]

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 31*(45), 16483-16490. 10.1523/JNEUROSCI.3653-11.2011 [doi]

Daliri, A., & Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 99*, 55-68. 10.1016/j.cortex.2017.10.019

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science, 21*(2)10.1111/desc.12521

Kim, K. S., & Max, L. (2014). Estimating feedforward vs. feedback control of speech production through kinematic analyses of unperturbed articulatory movements. *Frontiers in Human Neuroscience, 8*, 911. 10.3389/fnhum.2014.00911

Max, L. (2004). Stuttering and internal models for sensorimotor control: A theoretical perspective to generate testable hypotheses. *Speech motor control in normal and disordered speech* (pp. 357–388) Oxford University Press.

Max, L., & Daliri, A. (2019). Limited Pre-Speech Auditory Modulation in Individuals Who Stutter: Data and Hypotheses. *Journal of Speech, Language, and Hearing Research: JSLHR, 62*(8S), 3071-3084. 10.1044/2019_JSLHR-S-CSMC7-18-0358

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (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 Sciences and Disorders, 31*, 105-122.

Sengupta, R., Shah, S., Gore, K., Loucks, T. M., & Nasir, S. M. (2016). Anomaly in neural phase coherence accompanies reduced sensorimotor integration in adults who stutter. *Neuropsychologia, 93*(Pt A), 242-250. 10.1016/j.neuropsychologia.2016.11.004

ACKNOWELDGEMENTS

First and foremost, I am extremely grateful to my advisor, Dr. Ludo Max who has helped me grow both professionally as a scientist and as a person. Without the tremendous support and constant nurturing of Dr. Max since the first day I met him in my undergraduate program, I definitely would not have made it this far. At my wedding, my father personally thanked Dr. Max for being like a father figure to me. Those words were surely not an exaggeration. I am truly thankful for his investment in me.

I would also like to extend my deepest gratitude to Dr. David Perkel, who has been very generous with his time and wisdom throughout my time in this program. I would like to express my sincere appreciation to my other esteemed committee members, Dr. Fred Minifie, and Dr. Kristie Spencer, who have always been patient with my progress and a constant source of encouragement. I must also extend my gratitude to Dr. Richard Folsom who has been very supportive of me from my first day in this department.

I also want to thank previous and current laboratory members Churan Li, Annie Zhu, Alisson Stewart, Chrystal Chau, Ayoub Daliri, Ph.D., Selena Galaviz, Robert Hermosillo, Ph.D., Frances Ingram-Bate, Nick Kitchen, Ph.D., Yusuf Ali, Elise LeBovidge, Santosh Maruthy, Ph.D., Takashi Mitsuya, Ph.D., Nadia Siddiqui, Justin Walker, Hantao Wang, and Feiran Yang. Special thanks to Tiffani Kittilstved who always cared for me and my research. Much appreciation also goes to all the participants who endured many hours of numerous experiments. For the funding provided, I am to grateful to the American Speech-Language-Hearing Foundation, the Olswang Travel Fund, and The Graduate School’s Fund for Excellence and Innovation, and the lab’s grants from the National Institute on Deafness and Communication Disorders.

Throughout my time here, I was so fortunate to have so many who loved and cared for me. I am thankful for my wife who never fails to make me smile. I also want to acknowledge my father-in-law, Chan-Young Choi, and mother-in-law, Bok-Jin Kim, who always have made me feel at home with their kindness and care. I am also deeply indebted to Pastor Bryan Staab for his guidance and building my character. I also want to thank my church family, especially Prisca, Brian, and Elice for always being there for me. Last but not the least, my success would not have been possible without my parents' endless love, sacrifice, and support.

DEDICATION

To people who stutter

TABLE OF CONTENTS

Chapter I

Dissociated development of speech and limb sensorimotor learning in stuttering: speech auditory-motor learning is impaired in both children and adults who stutter

Kwang S. Kim
University of Washington, Seattle, WA
Ludo Max
University of Washington, Seattle, WA
Haskins Laboratories, New Haven, CT

Correspondence concerning this work should be addressed to Ludo Max, Ph.D., Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd St., Seattle, WA 98105-6246. E-mail: LudoMax@uw.edu. Phone: 206 543 2674.

Abstract

Stuttering is a neurodevelopmental disorder in which speech fluency is disrupted by repetitions and prolongations of orofacial or laryngeal movements. Various experimental paradigms have demonstrated that affected individuals show limitations in sensorimotor control and learning. However, there is continuing controversy regarding two core aspects of this perspective. First, it has been claimed that sensorimotor learning limitations are detectable only in adults who stutter (after many years of coping with the disorder) but not during childhood close to the onset of stuttering. Second, it remains unclear whether stuttering individuals' sensorimotor learning limitations affect only speech movements or also unrelated effector systems involved in nonspeech movements. Here, we report new data from separate experiments investigating speech auditory-motor learning ($N = 60$) and limb visuomotor learning ($N = 84$) in both children and adults who stutter versus matched nonstuttering individuals. Both children and adults who stutter showed statistically significant limitations in auditory-motor adaptation to formant-shifted feedback. In fact, this limitation was even more profound in children than in adults and in younger children versus older children. Between-group differences in the adaptation of reach movements performed with rotated visual feedback were subtle but statistically significant for adult participants. In children, even the nonstuttering groups showed limited visuomotor adaptation just like their stuttering peers. We conclude that sensorimotor learning is deficient in individuals who stutter, and that the ability for speech auditory-motor learning—which was already adult-like in 3-6 year-old typically developing children—is severely compromised in young children near the onset of stuttering. Thus, motor learning limitations may play an important role in the fundamental mechanisms contributing to the onset of this speech disorder.

Keywords: stuttering, sensorimotor learning, feedback, adaptation, internal model

Introduction

Reaching a thorough understanding of the biological foundations of stuttering requires parallel investigations of the “distal” cause(s) (What causes an individual to have the disorder?) as well as the “proximal” cause(s) (What are the mechanisms that cause specific instances of stuttering moments in the individual's speech?) (Packman & Attanasio, 2004). Several research groups' recent accomplishments in the areas of neuroimaging (e.g., Beal, Gracco, Brettschneider, Kroll, & De Nil, 2013; Cai et al., 2014; Chang, Horwitz, Ostuni, Reynolds, & Ludlow, 2011; Choo, Chang, Zengin-Bolatkale, Ambrose, & Loucks, 2012; Sitek et al., 2016; Wymbs, Ingham, Ingham, Paolini, & Grafton, 2013) and gene discovery (e.g., Benito-Aragón et al., 2020; Frigerio-Domingues & Drayna, 2017; Frigerio-Domingues et al., 2019; Kang et al., 2010; Raza, Amjad, Riazuddin, & Drayna, 2012) indicate substantial progress regarding the former of these two causality questions. Progress in addressing the latter question continues to lag as very few specific hypotheses have been formulated to suggest biologically plausible mechanisms that may explain the individual occurrences of fluency breakdowns in the form of sound/syllable repetitions and sound prolongations. In fact, there is not even general agreement about whether or not these observable disruptions in speech motor behavior originate within the sensorimotor system itself (Daliri & Max, 2015a; Max, 2004; Max, Guenther, Gracco, Ghosh, & Wallace, 2004) or at higher levels of linguistic processing such as those involved in phonological encoding, retrieving items from a lexicon, or encoding grammatical structures (Buhr & Zebrowski, 2009). For at least some of the linguistically-oriented views, however, it is not evident how the proposed hypotheses are compatible with the clinical facts that stuttering also occurs on non-words for which no lexical representation exists (Dayalu, Kalinowski, & Stuart, 2005; Onslow & Packman, 2002; Packman, Onslow, Coombes, & Goodwin, 2001) or on single words for which no sentence planning occurs (Blomgren & Goberman, 2008; Dayalu et al., 2005) and that individuals who stutter also differ from individuals who do not stutter in many aspects of nonspeech movements performed with entirely unrelated effector systems (e.g. Daliri, Prokopenko, Flanagan, & Max, 2014; Max, Caruso, & Gracco, 2003).

Although many linguistic and cognitive-emotional factors may play a *contributing* role, it is clear that—during stuttering moments—efferent signals to the orofacial and laryngeal musculature lead to articulatory and/or phonatory actions that are repeated or sustained. In previous theoretical work, we have therefore formulated a biologically plausible framework for one potential proximal cause that may occur within the sensorimotor system itself as the direct result of a dysfunction in well-documented sensorimotor mechanisms (Max, 2004; Max et al., 2004; Neilson & Neilson, 1987). In particular, the proposed framework was based on psychophysical data and computational modeling suggesting that, for the planning and execution of voluntary movements, the central nervous system (CNS) relies on both (a) a feedforward controller that determines which motor commands will achieve a desired movement outcome and (b) a feedback controller that uses efference copy and adaptable forward internal models in combination with afferent information and prior knowledge to predict future system states (e.g., Desmurget & Grafton, 2000; Kawato, 1999; Krakauer & Mazzoni, 2011; Shadmehr, Smith, & Krakauer, 2010; Wolpert & Miall, 1996; Wolpert & Flanagan, 2009). In this framework, internal models are stored neural representations of effector- and environment-dependent mappings between central motor commands and their sensory consequences. An individual's ability to learn or update these internal models can be tested with sensorimotor adaptation paradigms in which the input-output mapping is experimentally manipulated through perturbations of either the movement or the resulting feedback signals: the adaptive learning and after-effects that are observed in neurologically healthy subjects reflect internal model updates that serve to minimize performance errors and sensory prediction errors (e.g., Hadjiosif, Krakauer, & Haith, 2020 PREPRINT; Mazzoni & Krakauer, 2006; Taylor, Krakauer, & Ivry, 2014).

Within this framework, our model of stuttering proposed that individuals who stutter may have difficulty learning accurate internal models of the complex transformations from neuromuscular activation to vocal tract configurations and from those articulatory movements and postures to acoustic speech output (Max, 2004; Max et al., 2004; Neilson & Neilson, 1987). Developmental stuttering has its onset during early childhood when substantial anatomical changes take place in the CNS as well as in the peripheral structures of the vocal tract and, consequently, also in the resulting speech kinematics and acoustics (Callan, Kent, Guenther, & Vorperian, 2000; Kent, 1976; Kent, 1997; Vorperian, Kent, Gentry, & Yandell, 1999; Vorperian et al., 2005; Vorperian & Kent, 2007; Vorperian et al., 2009). If individuals who stutter are impaired in their ability to update internal models of these intricate relationships, it would be problematic for the feedforward controller to derive accurate motor commands for achieving the desired sensory outcomes and for the feedback controller to precisely predict the sensory consequences of the planned motor commands. Hypothetically, the resulting sensory prediction errors could trigger maladaptive system corrections that interfere with fluent speech production.

Consistent with this theoretical perspective, recent studies have found that adults who stutter (AWS) exhibit a reduced amount of auditory-motor adaptation as compared with adults who do not stutter (AWNS) (Daliri, Wieland, Cai, Guenther, & Chang, 2018; Daliri & Max, 2018; Sengupta, Shah, Gore, Loucks, & Nasir, 2016). However, it remains controversial whether such poor sensorimotor learning can be directly related to the mechanisms that are responsible for stuttering given that Daliri et al. (2018) recently reported that adaptation limitations are not present in *children* who stutter (CWS). Hence, those authors argued that the learning limitations found in AWS reflect a compensatory effect that arises as a result of years of experiencing stuttering.

A second controversy relates to the question whether sensorimotor learning problems in individuals who stutter are specific to the speech system. Differences in movement *control* between stuttering and nonstuttering individuals certainly are not limited to speech tasks, and have been found in behavioral measures of finger movement sequencing accuracy, initiation and execution time (Webster, 1997), manual reaction time (Bishop, Williams, & Cooper, 1991; Webster & Ryan, 1991), bimanual coordination (Forster & Webster, 2001; Zelaznik, Smith, Franz, & Ho, 1997), manual responses to visual feedback (Jones, White, Lawson, & Anderson, 2002), and directional and target accuracy of reaching movements (Daliri et al., 2014). With regard to motor *learning*, several studies have found finger-tapping sequence learning difficulties in AWS (Smits-Bandstra, De Nil, & Rochon, 2006; Smits-Bandstra, De Nil, & Saint-Cyr, 2006; Smits-Bandstra & De Nil, 2007; Smits-Bandstra & De Nil, 2009). Yet, the process of nonspeech sensorimotor adaptation in response to feedback perturbations has not been explored in stuttering. This is surprising because the upper-limb sensorimotor learning literature offers various paradigms that have been extensively tested and that are well established. One common paradigm investigates sensorimotor learning by experimentally applying a visuomotor rotation such that the location of a cursor representing hand position is rotated around the center of the workspace during reaching movements (while the actual hand remains invisible). Participants adapt by reaching in the direction opposite to that induced by the visual feedback (e.g., Krakauer & Mazzoni, 2011; Mazzoni & Krakauer, 2006; Shadmehr et al., 2010).

The studies presented here were designed to directly investigate both speech auditory-motor learning and reach visuomotor learning in AWS as well as CWS. We tested age-, sex-, and handedness-matched stuttering and nonstuttering adults and children in each of the two paradigms. Experiment 1 (*N* = 60) involved a speech auditory-motor adaptation task in which formant frequencies in the real-time auditory feedback were shifted up or down either gradually over many trials or suddenly between two successive trials. Experiment 2 (*N* = 84) involved a reach visuomotor adaptation task in which real-time visual feedback was suddenly rotated counterclockwise around the center of the workspace.

EXPERIMENT 1: SPEECH AUDITORY-MOTOR ADAPTATION IN ADULTS AND CHILDREN WHO STUTTER

Method

Adult participants

We recruited 18 AWS and 18 age- (± 3 years), sex-, and handedness-matched AWNS who reported that they were not taking any medications with a possible effect on sensorimotor functioning and that they had not been diagnosed with any neurological, psychological, emotional, or speech-language-hearing problems (other than stuttering in case of AWS). Each stuttering participant was individually matched with a nonstuttering control participant. We included only native speakers of American English (or individuals who were bilingual but had started speaking English before the age of 5 years). In addition, for AWS, the onset of stuttering had to have occurred during childhood (age 8 years or younger). Prior to participation, all participants provided written informed consent in accordance with procedures approved by the Institutional Review Board.

An American Speech-Language-Hearing Association certified speech-language pathologist administered the Stuttering Severity Instrument (SSI-3 or SSI-4, Riley, 1994; Riley, 2009) to all recruited participants. Based on these SSI results, we excluded three AWS because their speech samples failed to reach the minimum required overall SSI score of 10 or more (i.e., this participant's SSI score did not reach the ”very mild” category). On the other hand, one additional AWS was excluded due to being unable to produce a sufficient number of fluent trials during the actual speech adaptation task. Consequently, the final data set included 14 AWS (age *M* = 24.9 years, *SD* = 7.8, range 18-49 years) and 14 AWNS (*M* = 25.0 years, *SD* = 7.9, range 19-48 years).

Stuttering severity for the remaining AWS ranged from very mild to severe (3 very mild, 3 mild, 5 moderate, 3 very severe). In each group, there were 10 men and 4 women. The groups were matched to include 12 right-handed and 2 left-handed individuals (per self-report). Pure tone air-conduction hearing screenings were completed for all octave frequencies from 250-4000 Hz. Most of the participants had thresholds at or below 20 dB HL for all tested frequencies. One stuttering participant had a threshold of 40 dB HL at 4 kHz for the left ear whereas another stuttering participant also had a threshold of 40 dB HL at 4 kHz but for the right ear; one other stuttering participant had a threshold of 25 dB HL at 4 kHz for the left ear.

Child participants

Data were successfully collected for 16 CWS and 16 individually age- (± 3 months)-, sex-, and handedness-matched children who do not stutter (CWNS). The children were further divided into two subgroups based on age: 8 CWS and 8 CWNS in a younger age group from 3-6 years old (CWS: *M* = 5.41 years, *SD* = 1.36, range = 3.50-6.83 years; CWNS: *M* = 5.49 years, *SD* =

1.29, range = 3.75-6.83 years), and 8 CWS and 8 CWNS in an older age group from 7-9 years old (CWS: *M* = 8.02 years, *SD* = 0.69, range 7.08-9.33 years; CWNS: *M* = 8.05 years, *SD* = 0.60, range = 7.25-9.33 years). Generally consistent with early childhood changes in the distribution of affected girls versus boys, the younger age groups each included 3 girls and 5 boys whereas the older age groups included only boys.

Children's handedness was determined with a custom-modified version of the Edinburgh Handedness Inventory (Coren, 1993; Oldfield, 1971). Each child was asked to perform 5 tasks (drawing, throwing, using children's scissors, using a spoon, opening the lid of a box) selected from the 10-items normally included in the Edinburgh inventory. In the younger age groups, seven CWS were right-handed (defined here as completing 4 or 5 of the tasks with the right hand) and one was ambidextrous (defined here as completing 2 or 3 items with the right hand and the other 3 or 2 items with the left hand) whereas all CWNS were right-handed. In the older age groups, all CWS and CWNS were right-handed, based on the same criteria. A guardian or parent of each child provided written informed consent following procedures approved by the Institutional Review Board.

All child participants (a) spoke American English as their native language, (b) had normal pure tone air-conduction hearing thresholds (20 dB HL or better for all octave frequencies from 250-4000 Hz in both ears), and (c) scored above the 20th percentile on each of the following speech and language tests: Peabody Picture Vocabulary Test (PPVT-4, Dunn & Dunn, 2007), Expressive Vocabulary Test (EVT-2, Williams & Williams, 2007), Goldman-Fristoe Test of Articulation (GFTA-2, Goldman & Fristoe, 2000; GFTA-3, Goldman & Fristoe, 2015), and either Test of Early Language Development (TELD-3, Hresko, Reid, & Hammill, 1999) or (for children 8 or 9 years old) Clinical Evaluation of Language Fundamentals (CELF-4, Semel, Wiig,

& Secord, 2004). Specific eligibility criteria for children who stutter included (a) confirmation of the diagnosis of stuttering by an American Speech-Language-Hearing Association certified speech-language pathologist, (b) a minimum overall score of 6 on the Stuttering Severity Instrument (SSI-3 or SSI-4, Riley, 1994; Riley, 2009), and (c) stuttering onset prior to age 5. Based on SSI scores, the stuttering severity classification was severe for 1 child, moderate for 7 children, mild for 5 children, and very mild for 3 children.

Experimental set-up

All participants completed the speech adaptation task in a large sound booth. For each trial, adults spoke one of three monosyllabic words ("*tech*," *"tuck*," "*talk*") into a microphone (SM58, Shure). The words were presented on a computer monitor in randomized order per block of three trials. Children named pictures that were used in a board game: each trial elicited the production of a monosyllabic word (*"buck", "bus", "puck", "pup", "cut", "cup", "gut", "duck"*) while their speech was recorded with a wireless lapel microphone (WL185 with transmitter ULX1-M1 and receiver ULXP4, Shure). The pictures were presented in random order per block of 8 trials.

As illustrated in Figure 1A, the speech signal from the microphones was amplified (DI/O Preamp System II, ART) and digitally recorded on one channel of a CD recorder (CD-RW901SL, Tascam). The same signal was also routed to a vocal effects processor (VoiceOne, TC Helicon) that can manipulate all formant frequencies (i.e., resonance frequencies of the vocal tract) in real-time with a total feedback loop delay of only 10 ms.[1] The processor's output signal was amplified by a headphones amplifier (S-phone, Samson) and sent to insert earphones (ER-

---

[1] Due to a communication malfunction between the control computer and vocal effects processor, one adult and one child from the stuttering groups experienced a delay of 35 ms instead of 10 ms. As the amount of learning observed for these two participants was typical for that seen in their respective groups, we included the obtained data sets and also applied an identical 35 ms delay during the recording sessions of the two matched nonstuttering participants.

3A, Etymotic Research) worn by the participant. Immediately before each individual participant's recording session, this feedback loop was carefully calibrated in such a manner that a speech signal with an intensity of 75 dB SPL at the microphone (placed 15 cm from a loudspeaker through which a production of "tuck" was played) resulted in an output of 72 dB SPL in the earphones (measured in a 2 cc coupler Type 4946 connected to a sound level meter Type 2250A with Type 4947 1⁄2" pressure field microphone, Bruel & Kjaer). This input-output ratio was based on simultaneous recordings of a speech signal at a microphone in front of a speaker and at the entrance to the speaker's ear (Cornelisse, Gagné, & Seewald, 1991).

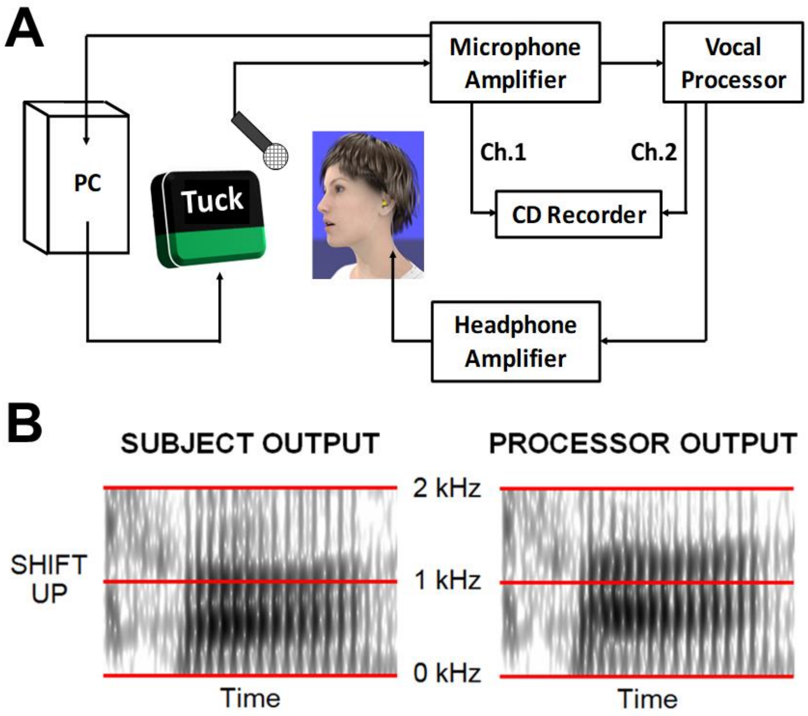


Figure 1. *Experimental set-up for speech auditory-motor adaptation. (A) The speech signal from the microphone is routed to a vocal processor that applies a formant shift perturbation to the signal that is presented through insert earphones as real-time auditory feedback. (B) Effect of an upward formant shift illustrated with separate spectrographic displays of the original acoustic speech signal produced by the participant (left) and the altered signal that is output by the vocal processor and heard by the subject (right).*

Speech auditory-motor adaptation task

Each session started with a baseline phase without any formant perturbation (i.e., veridical feedback). After the baseline phase, all formant frequencies in the speech signal were shifted either up or down by the VoiceOne processor which was controlled by a computer running custom MATLAB (The Mathworks) code. The processor selectively shifted the formant frequencies while leaving the fundamental frequency and consonant-related noise components completely unaltered (Figure 1B). We opted here to have all formant frequencies shifted by the same proportional amount and in the same direction. This experimental manipulation does not induce the perception of phonemic errors (a manipulation that we applied in Feng, Gracco, & Max, 2011), but implements a different motor-to-sensory transformation corresponding to speech output from a vocal tract with increased or decreased geometrical dimensions (as we have also applied in Daliri & Max, 2018; Max, Wallace, & Vincent, 2003; Max & Maffett, 2015). Thus, a participant's articulatory movements resulted in real-time auditory feedback with resonance frequencies that were globally increased or decreased relative to those associated with the participant's own motor-to-sensory mapping.

For adults, the processor implemented upward or downward formant shifts of 250 cents (the difference in cents between two frequencies fa and fb is 1200 × log2(fa/fb)). For each of the two shift directions there was one condition in which the formant perturbation was introduced suddenly (i.e., the maximum perturbation was introduced in full as a single step between two successive trials) and one condition in which the formant perturbation was introduced gradually (i.e., incrementally ramped up or down to its maximum value across many trials). Thus, there were four conditions: sudden shift up, gradual shift up, sudden shift down, and gradual shift down. For children, to make the perturbation sufficiently salient and limit the testing duration, we selected a shift of 335 cents but only in the upward direction. Thus, children completed two conditions: sudden shift up and gradual shift up.

The decision to include both sudden (or step) and gradual (or ramp) introductions of the auditory perturbation was made based on others' suggestions that intact cerebellar functioning is critical specifically for adapting to gradual perturbations (Doya, 2000; Robertson & Miall, 1999) whereas intact basal ganglia functioning is critical specifically for adapting to sudden perturbations (Contreras-Vidal & Buch, 2003; Doya, 2000; Venkatakrishnan, Banquet, Burnod, & Contreras-vidal, 2011)—suggestions that were deemed potentially informative for the present study given that dysfunction of both these neural substrates has been implicated in stuttering (e.g., Civier, Bullock, Max, & Guenther, 2013; Ingham et al., 2004). It should be noted, however, that the relationship between neural substrates and adaptation is complex and poorly understood: other empirical results suggest that some basal ganglia patients are unimpaired when adapting to sudden visuomotor distortions or mechanical force fields during reaching (Smith & Shadmehr, 2005; Weiner, Hallett, & Funkenstein, 1983) whereas cerebellar patients may be (a) impaired in adapting to sudden visuomotor or force field perturbations during reaching (e.g., Criscimagna-Hemminger, Bastian, & Shadmehr, 2010; Maschke, Gomez, Ebner, & Konczak, 2004; Rabe et al., 2009; Smith & Shadmehr, 2005; Weiner et al., 1983; Werner, Bock, & Timmann, 2009; Werner, Bock, Gizewski, Schoch, & Timmann, 2010) and (b) relatively unimpaired when adapting to a gradual force field perturbations during such reaching movements (e.g., Criscimagna-Hemminger et al., 2010). Thus, the relationship between cortico-striato-thalamo-cortical or cortico-cerebellar-thalamo-cortical circuits and impaired adaptation to sudden versus gradual perturbations is clearly not as straightforward as had been suggested. In addition, there is also evidence that both neural circuits may contribute to different stages of the same adaptation process (Doyon, Penhune, & Ungerleider, 2003) and that the cerebellum may process performance-related information (e.g., an error signal) that is a pre-requisite for adaptation but that is not directly related to learning *per se* (Werner et al., 2009).

The order of completion of the conditions was pseudo-balanced (or balanced in children) across stuttering individuals, and each nonstuttering individual then performed the four conditions (or two conditions for children) in the same order that had been used for the matched stuttering participant. Adult participants completed for each condition 60 blocks of the three test words (180 words) at a rate of 5 blocks (15 words) per minute. In the adults' gradual conditions, there were 10 blocks of *baseline*, 20 blocks of *gradual ramp*, 20 blocks of *full exposure* (i.e., 250 cents up or down formant shift), and 10 blocks of *after-effects* or wash-out (i.e., veridical feedback was restored). In the adults' sudden conditions, there were 20 blocks of *baseline*, 30 blocks of *full exposure*, and 10 blocks of *after-effects*. Intensity of the subject's speech was kept relatively constant across trials by means of color-coded visual feedback on the computer monitor. Words were presented on the monitor at a consistent pace so that subjects read 15 words per minute.

Children completed for each condition 13 blocks of the eight test words. Even though the use of a board game prevented us from controlling the rate of trial presentation as strictly as we did for adults, the staff member interacting with the children always attempted to keep the pace at approximately 10-12 trials per minute. In the children's gradual condition, there were 2 blocks of *baseline*, 4 blocks of *gradual ramp*, 5 blocks of *full exposure* (i.e., 335 cents upward formant shift), and 2 blocks of *after-effects*. In their sudden condition, there were 4 blocks of *baseline*, 7 blocks of *full exposure*, and 2 blocks of *after-effects*.

Data extraction and analyses

Prior to analyzing formant frequency changes to quantify auditory-motor adaptation, median vowel durations were extracted for each participant in each condition. Between-group comparisons of vowel duration can rule out the possibility that one group of speakers may disproportionately benefit from a slower rate of speech which would allow more time for *within-trial* compensation based on online sensory feedback (as opposed to sensorimotor adaptation which affects moving planning based on feedback from previous trials).

The acoustic recordings were then resampled at 10 kHz, and the frequencies of the first (F1) and second (F2) formant were measured in the middle of the vowel using a custom MATLAB program. The window defining vowel midpoint had its onset 40% into the vowel and its offset 60% into the vowel. Within the MATLAB program, we extracted children's formants using custom code implementing an order-optimized Linear Predictive Coding method (Feng et al., 2011; Vallabha & Tuller, 2002) and adults' formants using calls to Praat scripts (Boersma & Weenink, 2008). F1 and F2 in each trial were converted to cents normalized to the participant's baseline values (calculated from blocks 6-10 for adults and block 2 for children). F1 and F2 were then averaged for all analyses reported here. Trials in which stuttering occurred, the word was mispronounced, or pronunciation was affected by yawning, etc., were excluded.

Both the final extent of adaptation and initial or early adaptation were determined for each participant in each condition. Measures of the final extent of adaptation were based on the last five blocks from the full perturbation phase for adults and the last two blocks from the full perturbation phase for children. Measures of initial adaptation were based on early trials in the full perturbation phase of the conditions with sudden onset of the perturbation, specifically the initial five blocks for adults and the first block for children. We were not able to estimate initial *rate* of learning by fitting exponential functions adaptation because the auditory-motor adaptation profiles of many AWS and CWS did not demonstrate exponential learning.

Using the ezANOVA function in the ez package (Lawrence, 2016) in R (R Core Team, 2018), we conducted repeated measures analyses of variance (ANOVA) separately for final extent of adaptation and initial adaptation. Applied to the adult data, we tested the between-groups effect (stuttering vs. nonstuttering) and within-groups effects related to perturbation type (sudden vs. gradual) and perturbation direction (i.e., up-shift vs. down-shift) as well as all interactions.[2] For the children, who had completed only up-shift conditions, we used two between-group factors (stuttering vs. nonstuttering and younger children 3-6 years of age vs. older children 7-9 years of age) in addition to the within-group effect of perturbation type (sudden vs. gradual). Effect sizes were calculated as partial omega-squared (ω_p^2). Welch's *t*-tests in R were used for post-hoc tests. In all cases where the ANOVA necessitated follow-up with pair-wise comparisons, we report *p* values that have been adjusted with the Holm-Bonferroni method (Holm, 1979), keeping the family-wise error rate at .05. Lastly, we conducted two sets of correlational analyses, calculating Pearson's correlation coefficient in R. First, for the groups of stuttering adults and children, we examined whether there was a relationship between either final extent of adaptation or initial adaptation and stuttering frequency (i.e., averaged percent stuttered syllables across reading and speaking samples for the SSI assessment). Second, for stuttering and nonstuttering children separately, we examined whether there was a relationship between either final extent of adaptation or initial adaptation and age.

---

[2] Adult participants' auditory-motor adaptation task included separate conditions with upward and downward perturbation directions, so we normalized the data by dividing by 250 for down-shift conditions and by -250 for up-shift conditions. This caused adaptive changes (i.e., moving in the opposite direction of the perturbation) to be represented as a positive number in the range from 0 to 1, with 1 corresponding to full compensation. Following the perturbation (i.e., moving in the same direction of the perturbation) resulted in a negative number.

Results

Vowel duration

Although AWS tended to produce longer vowels than AWNS during the auditory-motor learning task, the difference was not statistically significant, $F(1, 26) = 3.381$, $p = 0.062$, $\omega_p^2 = 0.049$. Vowel duration also did not differ between the conditions with upward vs. downward directions of formant perturbation, $F(1, 26) = 0.496$, $p = 0.487$, $\omega_p^2 < 0.001$, or with sudden vs. gradual types of perturbation, $F(1, 26) = 3.491$, $p = 0.073$, $\omega_p^2 < 0.001$). None of the interactions were statistically significant (Direction × Group: $F(1, 26) = 0.960$, $p = 0.336$, $\omega_p^2 < 0.001$; Type × Group: $F(1, 26) = 0.355$, $p = 0.557$, $\omega_p^2 < 0.001$; Direction × Type: $F(1, 26) = 0.533$, $p = 0.472$, $\omega_p^2 < 0.001$; Direction × Type × Group: $F(1, 26) = 0.002$, $p = 0.965$, $\omega_p^2 < 0.001$).

Similarly, CWS and CWNS did not differ in vowel duration, $F(1, 28) = 2.689$, $p = 0.608$, $\omega_p^2 < 0.001$. Vowel duration was also not significantly different between the two age-based subgroups of children, $F(1, 28) = 0.083$, $p = 0.775$, $\omega_p^2 < 0.001$, or between the two types of perturbation, $F(1, 28) = 2.800$, $p = 0.105$, $\omega_p^2 = 0.006$. There were also no significant interactions (Age × Group, $F(1, 28) = 0.416$, $p = 0.524$, $\omega_p^2 < 0.001$; Type × Group, $F(1, 28) = 1.289$, $p = 0.266$, $\omega_p^2 = 0.001$; Age × Type, $F(1, 28) = 0.007$, $p = 0.932$, $\omega_p^2 < 0.001$; Group × Age × Type, $F(1, 28) = 0.209$, $p = 0.651$, $\omega_p^2 < 0.001$).

Auditory-motor adaptation in AWS vs. AWNS

Figure 2 shows adult group data (means and standard errors of the mean) across all trials in the sudden and gradual shift-up conditions. Formant frequencies are represented in cents relative to baseline (i.e., average frequency for the blocks from which baseline is calculated is always 0 cents). Participants lowered their formant frequencies during the perturbation trials in both the sudden and gradual shift conditions. Figure 3 shows the corresponding data for the shift-down conditions. In this case, participants tended to increase their formant frequencies, albeit to a smaller extent. The ANOVA results for final extent of auditory-motor learning (per participant averaged across F1 and F2, the 3 test words, and trial blocks 46-50) revealed statistically significant main effects for Group, $F(1, 26) = 9.780$, $p = 0.004$, $\omega_p^2 = 0.140$, and Direction, $F(1, 26) = 20.738$, $p < 0.001$, $\omega_p^2 = 0.195$: AWS adapted significantly more than AWNS, and upward formant shift perturbations resulted in significantly more adaptation than downward formant shift perturbations. The main effect for perturbation Type was not significant, $F(1, 26) = 0.719$, $p = 0.404$, $\omega_p^2 < 0.001$, but there was a significant Group × Type interaction, $F(1, 26) = 4.696$, $p = 0.040$, $\omega_p^2 = 0.021$. This interaction was due to the AWS vs. AWNS between-group difference in final extent of adaptation being larger for the sudden perturbation than for the gradual perturbation. Nevertheless, post-hoc tests revealed that the between-group difference was statistically significant in both the sudden perturbation condition, $t(19.135) = -3.337$, $p = 0.007$, $d = 1.261$, and the gradual perturbation condition $t(20.687) = -2.298$, $p = 0.032$, $d = 0.869$. There were no other statistically significant interactions for the final extent of adaptation (Direction × Group: $F(1, 26) = 0.243$, $p = .626$, $\omega_p^2 < 0.001$; Type × Direction: $F(1, 26) = 0.282$, $p = .600$, $\omega_p^2 = -0.003$; Type × Direction × Group: $F(1, 26) = 0.177$, $p = .678$, $\omega_p^2 < 0.001$).

We examined adults' initial adaptation using the first five perturbation blocks of the sudden conditions. ANOVA revealed a significant between-group difference, $F(1, 26) = 6.267$, $p < 0.019$, $\omega_p^2 = 0.089$, with greater initial adaptation for AWNS than for AWS. There was also a significant main effect of perturbation direction, $F(1, 26) = 4.872$, $p = 0.036$, $\omega_p^2 = 0.057$, with greater initial learning for upward perturbations than for downward perturbations.

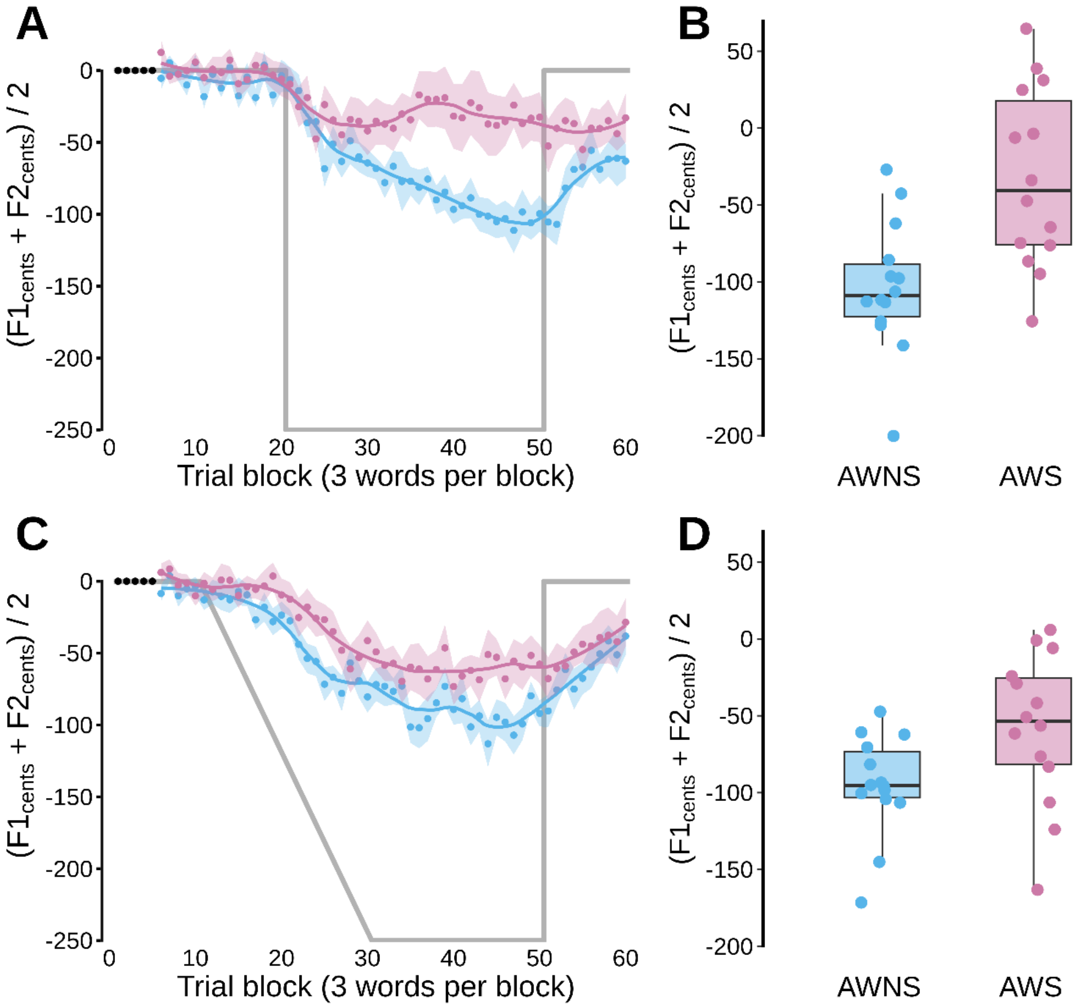


Figure 2. *Auditory-motor adaptation for adults who do not stutter (blue) and adults who stutter (pink) in the 250 cents upward formant perturbation conditions. Panels A (suddenly introduced perturbation) and C (gradually introduced perturbation) show group averaged formant frequencies in cents relative to baseline (averaged across F1 and F2 and across 3 test words after normalization of the individual formants and words). Shaded areas indicate standard errors of the mean. Gray lines plot the formant perturbation with reversed sign (thus indicating hypothetical full compensation). Panels B and D show corresponding individual participant data (averaged across the last 5 blocks of 3 words) overlaid on group boxplots.*

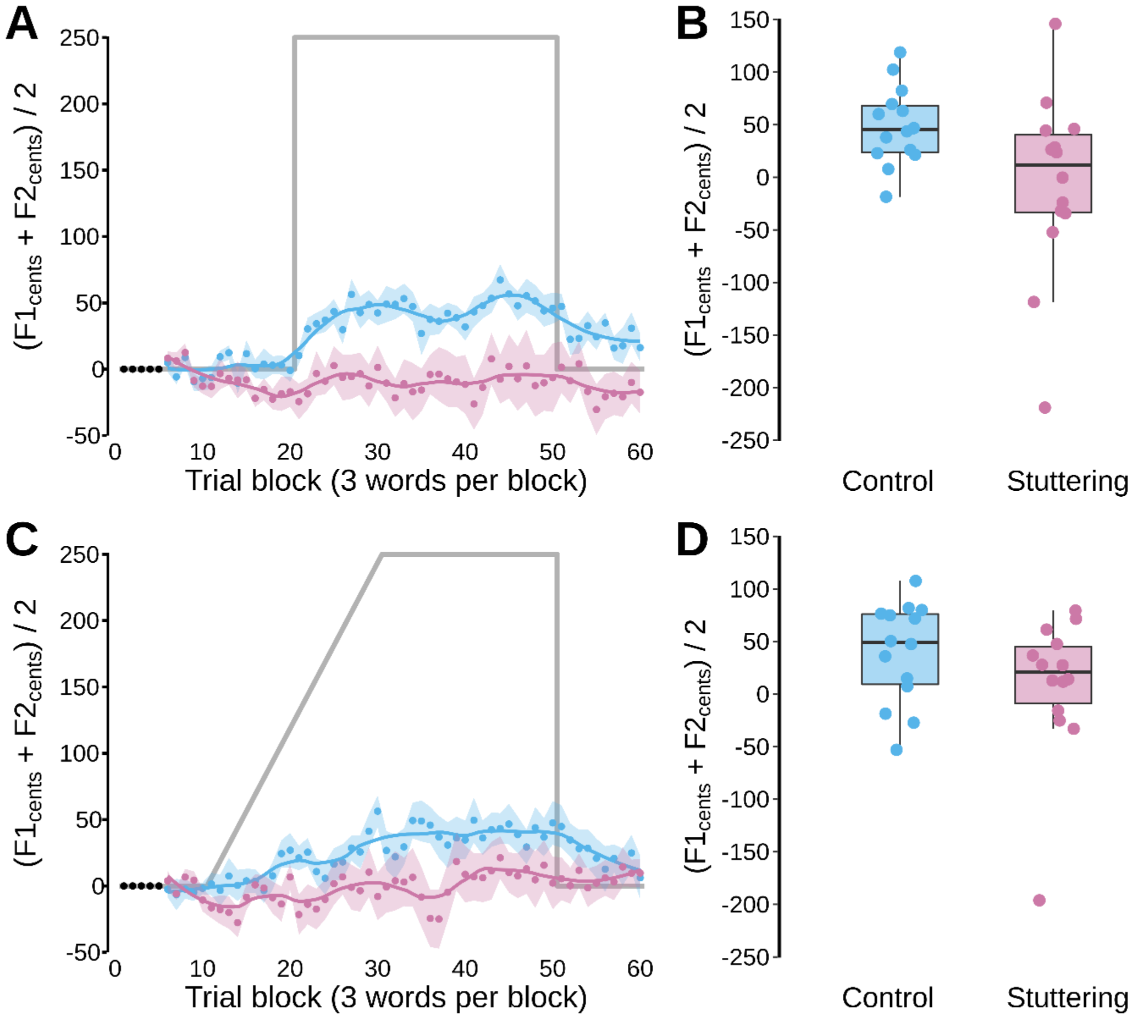


Figure 3. *Auditory-motor adaptation for adults who do not stutter (blue) and adults who stutter (pink) in the 250 cents downward formant perturbation conditions. Panels A (suddenly introduced perturbation) and C (gradually introduced perturbation) show group averaged formant frequencies in cents relative to baseline (averaged across F1 and F2 and across 3 test words after normalization of the individual formants and words). Shaded areas indicate standard errors of the mean. Gray lines plot the formant perturbation with reversed sign (thus indicating hypothetical full compensation). Panels B and D show corresponding individual participant data (averaged across the last 5 blocks of 3 words) overlaid on group boxplots.*

However, these two main effects were modified by a Group × Direction interaction, *F*(1, 26) = 4.352, p = 0.047, ω_p^2 = 0.049. Post hoc tests revealed that AWS, as compared with AWNS, learned less in the initial blocks of the sudden down condition, *t*(19.701) = -3.483, *p* = 0.005, *d* = 1.317, but not the sudden up condition, *t*(22.423) = -0.527, *p* = 0.604, *d* = 0.199.

There was no statistically significant correlation between stuttering frequency (averaged across the reading and speaking samples from the SSI evaluation) and either final adaptation extent (Sudden, *r*(12) = -0.041, *p* = 0.888, Gradual, *r*(12) = 0.171, *p* = 0.560) or initial adaptation (*r*(12) = -0.211, *p* = 0.470).

Auditory-motor adaptation in CWS vs. CWNS

In both the sudden and gradual perturbation conditions, CWNS showed a final extent of auditory-motor adaptation that was very similar to that seen in nonstuttering adults (Figure 4 for the younger subgroup 3-6 years of age, Figure 5 for the older subgroup 7-9 years of age). Even the youngest subgroup of CWNS adapted in both conditions to an extent that compensated for 30~40% of the perturbation, which is highly similar to the data described above for nonstuttering adults. On the other hand, both younger and older CWS made only minimal, if any, changes in their formant frequencies in response to the perturbation. In fact, based on an additional family of four pair-wise tests, neither the younger nor the older subgroup of CWS showed a final extent of adaptation (per participant averaged across F1 and F2, the different tests words, and blocks 10 and 11) that was significantly different from baseline in either of the two perturbation conditions (3-6 year-olds in Sudden, *t*(7) = 0.611, *p* = 1.000, *d* = 0.216, 3-6 year-olds in Gradual, *t*(7) = 0.135, *p* = 1.000, *d* = 0.048, 7-9 year-olds in Sudden, *t*(7) = -2.175, *p* = 0.198, *d* = 0.769, and 7-9 year-olds in Gradual, *t*(7) = -2.411, *p* = 0.189, *d* = 0.852.

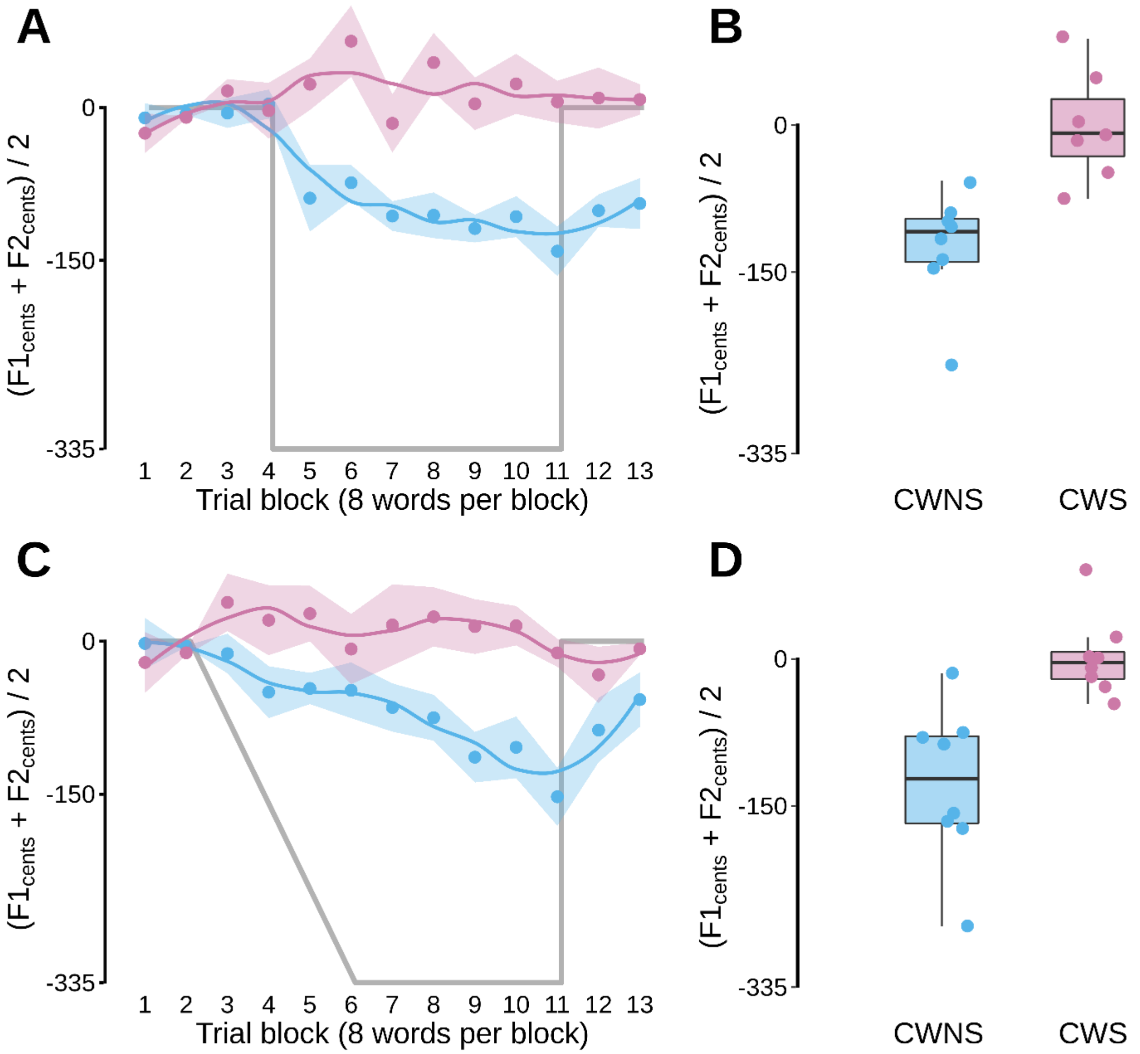


Figure 4. *Auditory-motor adaptation for 3-6 year-old children who do not stutter (blue) and age-matched children who stutter (pink) in 335 cents upward formant perturbation conditions. Panels A (suddenly introduced perturbation) and C (gradually introduced perturbation) show group averaged formant frequencies in cents relative to baseline (averaged across F1 and F2 and across 8 test words after normalization of the individual formants). Shaded areas indicate standard errors of the mean. Gray lines plot the formant perturbation with reversed sign (thus indicating hypothetical full compensation). Panels B and D show corresponding individual participant data (averaged across the last 2 blocks of 8 words) overlaid on group boxplots.*

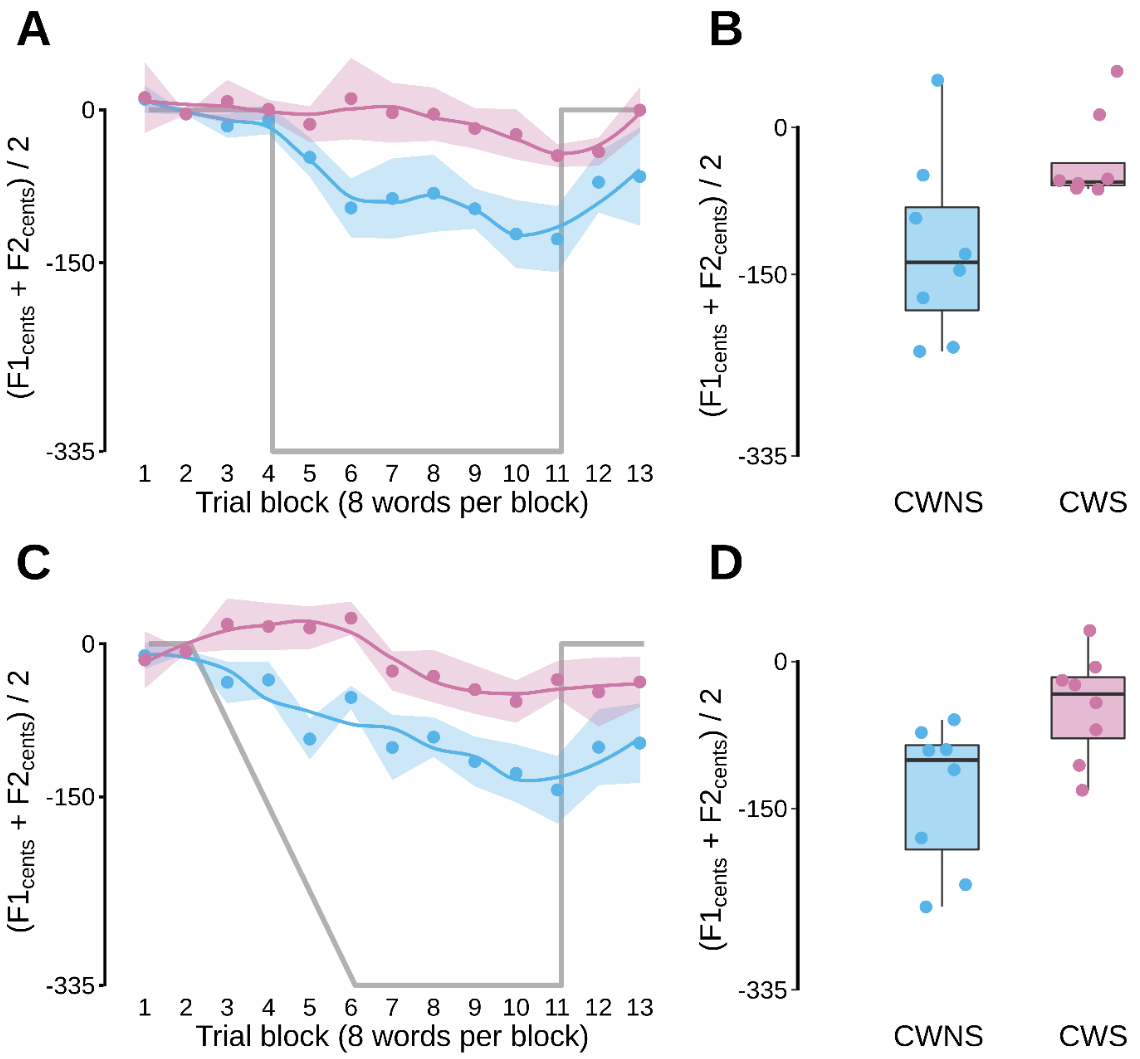


Figure 5. *Auditory-motor adaptation for 6-9 year-old children who do not stutter (blue) and age-matched children who stutter (pink) in 335 cents upward formant perturbation conditions. Panels A (suddenly introduced perturbation) and C (gradually introduced perturbation) show group averaged formant frequencies in cents relative to baseline (averaged across F1 and F2 and across 8 test words after normalization of the individual formants). Shaded areas indicate standard errors of the mean. Gray lines plot the formant perturbation with reversed sign (thus indicating hypothetical full compensation). Panels B and D show corresponding individual participant data (averaged across the last 2 blocks of 8 words) overlaid on group boxplots.*

Consequently, the ANOVA for the children's final extent of adaptation revealed a significant between-group effect, $F(1, 28) = 41.947$, $p < 0.001$, $\omega_p^2 = 0.414$, in the absence of main effects for age subgroup, $F(1, 28) = 2.257$, $p = 0.144$, $\omega_p^2 = 0.021$, or perturbation type $F(1, 28) = 0.385$, $p = 0.540$, $\omega_p^2 < 0.001$. There were also no significant interactions among any of the variables (Age × Group, $F(1, 28) = 1.684$, $p = 0.205$, $\omega_p^2 = 0.012$; Type × Group, $F(1, 28) = 0.019$, $p = 0.891$, $\omega_p^2 < 0.001$; Age × Type, $F(1, 28) = 0.007$, $p = 0.932$, $\omega_p^2 < 0.001$; Group × Age × Type, $F(1, 28) = 0.015$, $p = 0.903$, $\omega_p^2 < 0.001$).

Despite the absence of a statistically significant Group × Age interaction, inspection of the individual participant data clearly showed that final extent of adaptation increased with age in CWS but not CWNS (Figure 6). In both the sudden and gradual perturbation conditions, the final adaptation extent for CWS was significantly correlated with age, (Sudden, $r(14) = -0.571$, $p = 0.021$; Gradual, $r(14) = -0.508$, $p = 0.045$). For CWNS, on the other hand, there were no statistically significant correlations in either condition (Sudden, $r(14) = -0.031$, $p = 0.910$; Gradual, $r(14) = -0.109$, $p = 0.915$). Note that a greater adaptation extent corresponds to a more negative number, and, thus, that a negative correlation here indicates greater adaptation with increasing age.

A family of two comparisons focusing on initial adaptation in the sudden perturbation condition revealed that 3-6-year-old CWNS already learned more than matched CWS in the very first perturbation block (i.e., block 5), $t(13.155) = -2.702$, $p = 0.036$, $d = 1.351$, whereas the 7-9-year-old CWNS and CWS showed no group difference in this measure of initial adaptation, $t(13.910) = -1.241$, $p = 0.235$, $d = 0.621$). Furthermore, the initial amount of adaptation in this block of trials did not significantly correlate with age for either group (CWS, $r(14) = -0.225$, $p = 0.401$, CWNS, $r(14) = -0.163$, $p = 0.548$). Lastly, there was no statistically significant correlation between stuttering frequency from the SSI speaking samples and either final extent of adaptation (Sudden, *r*(14) = -0.139, *p* = 0.608, Gradual, *r*(14) = -0.346, *p* = 0.190) or initial adaptation (*r*(14) = 0.180, *p* = 0.504).

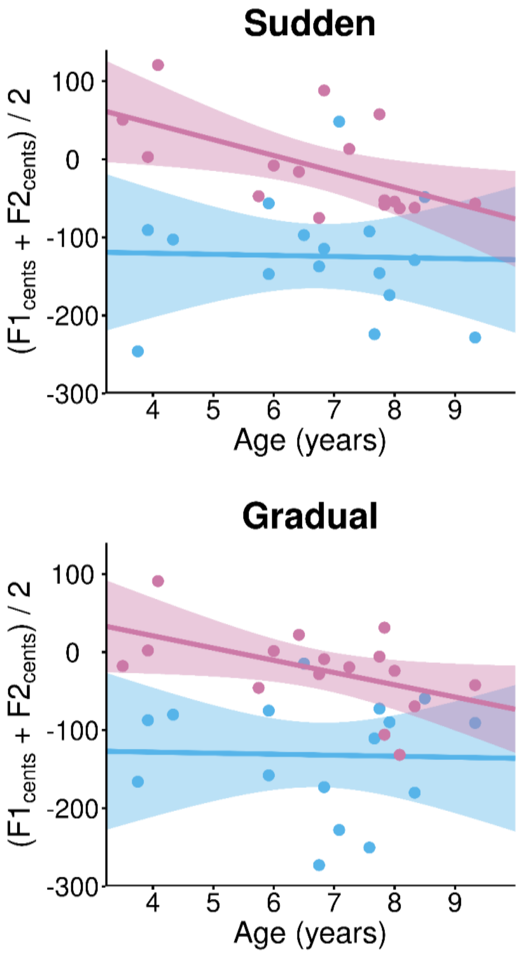


Figure 6. *Relationship between extent of speech auditory-motor adaptation and age for CWS (pink) and CWNS (blue). Data for the sudden condition shows a negative correlation for CWS whereas there is no such clear trend for CWNS (top). A similar finding in the gradual condition (bottom). The shaded areas indicate the confidence interval (95%).*

Discussion

In this first experiment, we examined adaptation to formant-shifted real-time auditory feedback in adults who stutter and children who stutter as compared with matched control participants who do not stutter. Specifically, the auditory perturbations consisted of global formant shifts (i.e., applying the same proportional amount of shift to all formants) such that the feedback signal corresponded to vowels produced with a vocal tract that is smaller (in the case of upward shifts) or larger (in the case of downward shifts) than the speaker's own vocal tract. These perturbations were applied continually while participants produced *CVC* words, where *C* was always a stop consonant and *V* was a front, central, or back mid vowel. The speaker's fundamental frequency and consonant bursts were not affected by the perturbation. In other words, the formant shifts specifically changed the transformation from motor commands for vowel-related vocal tract configurations to acoustic output (typically leading the CNS to update its internal models of these transformations), but without changing the phonemic category of the produced vowels or the primary characteristics of the surrounding consonants.

As a first finding, we replicated previous results indicating that AWS show reduced speech auditory-motor learning in response to formant perturbations that are introduced suddenly in-between two successive trials (Daliri & Max, 2018; Sengupta et al., 2016) as well as formant perturbations that are introduced gradually across multiple trials (Daliri et al., 2018). Second, we found that this between-group difference in auditory-motor learning for AWS vs. AWNS holds up for both up-shift and down-shift global formant perturbations. Third, although adaptation for AWS was statistically significantly smaller than that for AWNS in both the suddenly introduced and the gradually introduced perturbation conditions, the between-group difference was larger when the perturbation was introduced suddenly. Fourth, not only did we find a difference in speech auditory-motor learning also in CWS vs. CWNS, this between-group difference for children was even larger than that seen in adults: the youngest group of CWS (3-6 years of age) showed essentially no learning at all whereas the age-matched CWNS already showed adult-like learning. Descriptively, the older group of CWS (7-9 years of age) showed slightly more adaptation, but even for this group the change from baseline was still not statistically significant.

Fifth, like the situation for adults, stuttering and nonstuttering children differed in adaptation for both suddenly and gradually introduced formant perturbations. For children, however, this between-group difference was not affected by the sudden or gradual manner in which the perturbation was introduced as, statistically, neither the younger nor the older CWS showed any adaptation at all (i.e., no changes from baseline in either condition). Sixth, among CWS, the final extent of speech auditory-motor learning was correlated with age (more adaptation in older children); in contrast, for CWNS there was no correlation between these variables as even the younger children already adapted to the same extent as older children (and even as adults, but formant perturbation size as well as sample size differed between the child and adult groups so no direct statistical comparisons were made for children vs. adults).

Clearly, our findings that (a) speech auditory-motor learning problems are present not only in AWS but also in CWS, and (b) this learning impairment in CWS is even more profound than in AWS (statistically, our groups of stuttering children showed no adaptation at all) are not consistent with previous work by Daliri et al. (2018). Those authors reported that speech auditory-motor learning limitations occur only in adults who stutter (who have experienced stuttering for many years and whose performance may be affected by compensatory sensorimotor strategies) and not in children who stutter (who are closer to the onset of the disorder). Thus, our contrasting findings with regard to children constitute some of the most important outcomes of the present work because for any sensorimotor learning impairment to potentially play a role in the onset of stuttering (which typically occurs between 2 and 4 years of age), the impairment must of course already be influencing the development of speech motor behavior during early childhood.

One possible reason for the discrepancy between our own results and those of Daliri et al.

(2018) may lie in the specific formant perturbations that were used in the studies. In our work described here, we shifted all formants in the same direction such that the auditory feedback signal reflected an altered motor-to-sensory transformation that did not change the phonemic category of the produced vowels. In other words, when the children participating in our study produced the test words *bus, pup, cup, duck*, etc., they never perceived these words as having been produced with the wrong vowel. Instead, they perceived the words as having the correct vowel but produced by a vocal tract with different geometric properties (akin to developmental changes in vocal tract geometry during childhood, albeit on a shortvtime scale of only minutes). Daliri et al. (2018), on the other hand, shifted the first and second formants in different directions such that when participating children produced the words *bed, head,* and *Ted,* they heard their productions as similar to *bad, had,* and *Tad*. It is very well possible, and perhaps even expected, that CWS are indeed able to correct their productions when auditory feedback from previous trials indicates that they produced the words with a completely wrong sound (as in Daliri et al., 2018). For example, the fact that CWS are not more likely than other children to have speech sound disorders (Nippold, 2002; Unicomb, Kefalianos, Reilly, Cook, & Morgan, 2020) suggests that CWS do not have difficulty with adjusting movement planning to target the correct speech sounds. In addition, the symptoms of stuttering do not involve producing incorrect sounds, but certain movements or postures for sound production being repeated or sustained.

It is therefore important to recall that what had been hypothesized previously in a theoretical framework (e.g., Max, 2004; Max et al., 2004) is not that CWS would fail to correct for errors in target sound selection, but, rather, that there may be “an inability to learn stable or correct mappings between motor and sensory signals and to update these mappings in the presence of rapid neural and craniofacial maturation during speech development” (Max, 2004, p.

374). To be more specific, the crux of that particular hypothesis is that “After their initial acquisition during babbling and early speech, these internal models require continuous updating, due to the rapid developmental changes in neural, anatomical, and biomechanical characteristics during childhood. … If, for some reason, the CNS would fail to accurately update the internal models to match the currently applicable transformations, it would become unable to predict with great precision the sensory consequences of planned movements …” (Max, 2004, p. 374). The formant perturbations that we implemented in our work reported here were selected very specifically to test participants’ updating of internal models that represent such fine-grained and vocal tract geometry-related spectral features of the motor-to-auditory transformations involved in speech production. Our findings very strongly suggest that, at least for the time scale tested here, CWS did not update these internal models at all whereas age-matched CWS already showed adult-like learning. Moreover, our correlational analyses demonstrate that younger CWS—who are closer to the onset of stuttering—showed less learning than older CWS, thus rejecting the notion that impaired speech auditory-motor learning is something that develops as a consequence of stuttering rather than a factor that may contribute to the onset of stuttering.

EXPERIMENT 2: REACH VISUOMOTOR ADAPTATION IN ADULTS AND CHILDREN WHO STUTTER

Method

Adult participants

We recruited 16 AWS, but one participant was excluded due to the SSI score being too low to reach the “very mild” category. Thus, the final data set included 15 AWS (age *M* = 27.07

years, *SD* = 7.88, range 19-48 years) and 15 individually age- (±3 years), sex-, and handedness-matched AWNS (age *M* = 26.67 years, *SD* = 8.16, range = 20-48 years). The same inclusion criteria as in Experiment 1 were applied, except that participants did not need to be native speakers of American English given that the task involved upper limb reaching movements rather than speaking. All participants were right-handed based on self-report, and there were 12 males and 3 females in each group. According to the SSI scores, 2 individuals' stuttering was categorized as very severe, 1 as severe, 4 as moderate, 4 as mild, and 4 as very mild.

Child participants

The participating children were 27 CWS and 27 CWNS from 3 to 9 years of age. Inclusion criteria were as described for Experiment 1. Each stuttering child was individually matched with a nonstuttering child based on age (i.e., ±3 months) and sex. Based on the handedness assessment described above for Experiment 1, most children were right-handed except 4 CWS and 1 CWNS who were ambidextrous. We organized the data into three age-based subgroups. A 3-4 years old subgroup included 9 CWS (*M* = 3.80 years, *SD* = 0.57, range = 3.11-4.55 years) and 9 CWNS (*M* = 3.76 years, *SD* = 0.52, range = 3.26-4.66 years). A 5-6 years old subgroup included 8 CWS (*M* = 6.10 years, *SD* = 0.50, range =5.46-6.89 years) and 8 CWNS (*M* = 6.14 years, SD = 0.46, min = 5.50, max = 6.69). A 7-9 years old subgroup included 10 CWS (*M* = 8.13 years, *SD* = 0.68, range = 7.09-9.13 years) and 10 CWNS (*M* = 8.09 years, *SD* = 0.72, range = 7.06-9.09 years). In the 3-4 years-old range, there were 2 girls in each group. In the 5-6 years-old range, there were 3 girls in each group. In the 7-9 years-old range, each group included 1 girl.

With regard to severity based on SSI scores ((SSI-3 or SSI-4, Riley, 1994; Riley, 2009), stuttering in the youngest group of CWS (3-4 years of age) was rated as very severe in 1 case, severe in 1 case, moderate in 4 cases, and mild in 3 cases. For the group of CWS in the range 5-6 years old, stuttering was severe for 1 child, moderate for 3 children, mild for 2, and very mild for 2. The oldest group of CWS, 7-9 years of age, included 2 children rated as severe, 4 rated as moderate, 3 rated as mild, and 1 rated as very mild.

Experimental set-up

Seated on a height-adjustable chair, participants made fast out-and-back reaching movements with their right arm supported by an acrylic air sled on a glass table top (Figure 7A). The air sled was connected to a compressed-air supply so that friction across the glass surface was minimized. The position of a sensor attached to the participant's extended right index finger was recorded by an electromagnetic tracking system (Liberty, Polhemus; 240 samples/s). A cursor representing this fingertip position was projected, together with the trial's start and target positions, onto a back-projection screen in real-time (visual feedback delay ≤ 32 ms). The back-projection screen was mounted horizontally above a first-surface mirror in such a manner that, when participants viewed the image in the mirror, the start position, target, and feedback cursor all appeared to be located in the same plane as the participant's hand, even though the hand itself remained invisible below the mirror (Figure 7B).

Reach visuomotor adaptation task

Adult participants were instructed to make out-and-back reaching movements from a white center start position (radius 1 cm) to one of eight radially oriented green targets (radius 0.75 cm) located 10 cm from the start position at angles ranging from 0° to 315° in 45° increments. A custom MATLAB program presented all targets in random order within each block of 8 trials. The start position was fixed in the participant's midsagittal plane, with

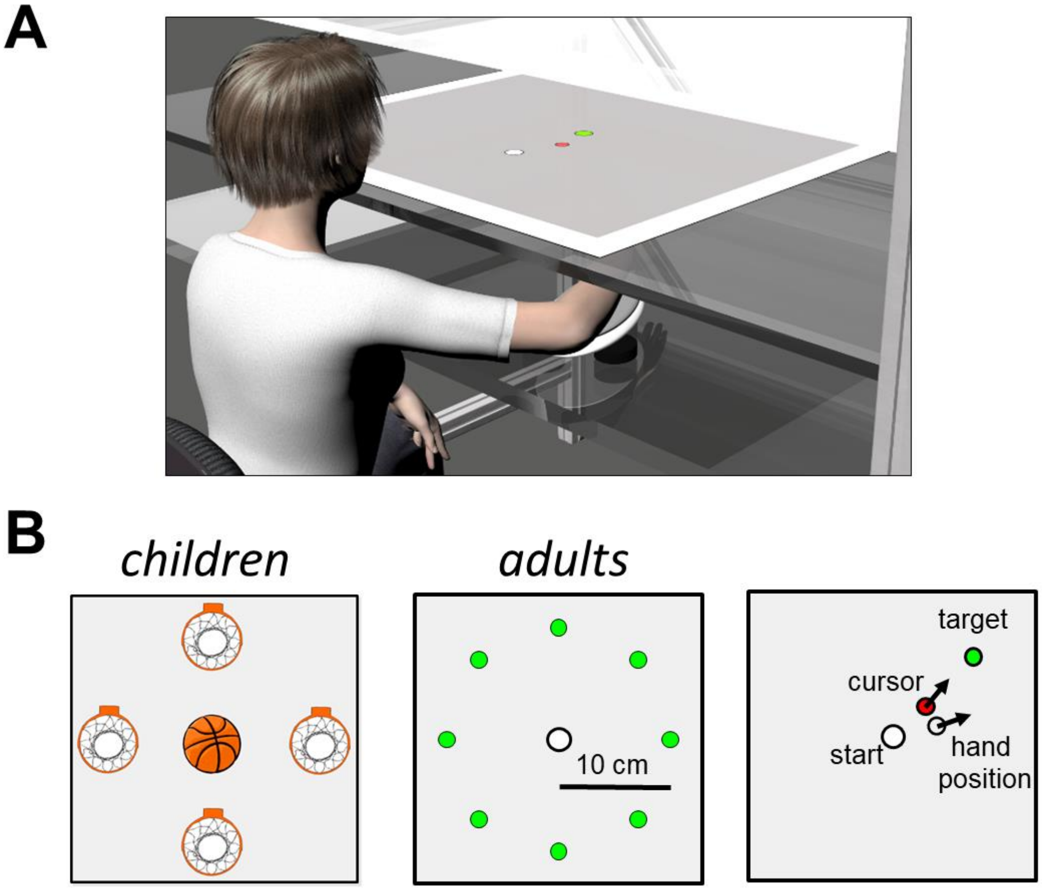


Figure 7. *Experimental set-up for reach visuomotor adaptation. (A) Participant viewing the workspace displayed in a mirror that is part of a virtual display system. Reaching movements are performed with the right arm which is supported by an air sled on a glass surface. (B) The workspace displayed reach targets as basketball hoops for children (left) and as green circles for adults (middle). All possible targets are shown here, but targets appeared one at a time in blocked randomized order. Fast out-and-back reach movements were performed by moving a basketball (children) or red cursor (adults) to a target and back to the start position. During trials with perturbed visual feedback, the cursor's position was rotated 30 degrees counter-clockwise around the center of the workspace (right).*

distance from the body adjusted such that the elbow was flexed approximately 90° when the fingertip rested on the start position. Continuous visual feedback of fingertip position was presented by a red cursor of the same size as the targets (radius 0.75 cm). Participants were asked to move the cursor out to the target and back to the start position without stopping at the target. They were told that it was important to move as fast as possible after initiating the movement, to not make any corrections during the movement, and that it was not necessary to initiate the movement quickly after the appearance of a new target.

Each session for adult participants consisted of 25 blocks of the 8 possible targets (200 trials) across three phases. During the *baseline* phase (initial 5 blocks, 40 trials), veridical feedback was presented (i.e., no perturbation). During the *full-exposure* phase (15 blocks, 120 trials), the position of the cursor representing fingertip position was rotated 30° counterclockwise (CCW) around the center of the work space (corresponding to the start position). During the wash-out or *after-effects* phase (5 blocks, 40 trials), veridical feedback was restored. Before data collection, each participant was given a set of familiarization trials with veridical feedback.

In the children's visuomotor adaptation task, participants were also instructed to make out-and-back reaching movements from a white center start position, but there were only four targets, located 10 cm from the start position at 0°, 90°, 180°, and 270°. The targets were represented by an image of a basketball hoop and the cursor was an image of a basketball (Figure 7B). The start position, cursor, and targets all had a radius of 2.5 cm. For each child, the start position location was determined in the same manner as described above for adults. Our custom MATLAB control program presented the targets in randomized order per block of 4 trials. We asked the children to move the ball to the hoop and back as fast as possible and without stopping at the hoop. They were encouraged throughout the task to keep moving fast and to not correct their movements within a trial even if the cursor missed the target.

Children completed 18 blocks of 4 targets (72 trials total). Like in the adults' task, there were three phases: *baseline* (4 blocks, 16 trials), *full-exposure* (10 blocks, 40 trials), and *after-effects* (4 blocks, 16 trials). During the full-exposure phase, the position of the cursor representing fingertip position was rotated 30° CCW around the center of the workspace. During the baseline and after-effects phases, no perturbation was applied.

Data extraction and analyses

Custom MATLAB code smoothed the motion sensor data using a butterworth low-pass filter with cutoff frequency 10 Hz. These filtered position data were differentiated to obtain tangential velocity signals. Measurements of movement duration were made given that a between-group difference in this variable may allow the slower group to implement more within-trial corrections (as opposed to updating internal models relied upon during movement planning). Movement onset was defined as the time point where tangential velocity first exceeded 5 cm/s. Movement offset was defined as the time point where tangential velocity dropped below 5 cm/s or, if that did not happen, where tangential velocity reached a local minimum.

For adult participants, initial reach angle was measured as the direction of a vector between start position and cursor location at the time point when peak tangential velocity was reached or 150 ms after movement onset, whichever occurred first (Tong & Flanagan, 2003). For each trial, this initial reach angle was then expressed relative to a vector from the start position to the specific target presented on that trial. Similar to the data processing steps in the speech auditory-motor adaptation task, all relative reach angles were normalized to baseline by subtracting, for each target direction separately, the median relative reach angle across baseline trials 3-5 for that same target. Lastly, the normalized reach angle for each block of 8 trials was obtained by averaging the normalized reach angles toward the 8 different targets in that block.

For children, similar analysis procedures were used, but the cut-off for tangential peak velocity to be reached was 200 ms (rather than 150 ms) after movement onset. Thus, if tangential peak velocity had not been reached 200 ms after onset, we used the cursor location at that time point to calculate initial movement direction. Further, we used each target's baseline trials 1-4 for normalizing the relative reach angles to baseline. Final data were obtained by averaging the normalized relative reach angles across the 4 different targets in each block of 4 trials.

We conducted all statistical analyses once for final adaptation extent and once for initial rate of adaptation. To analyze adaptation extent, we used each participant's last two blocks of the perturbation phase. For initial rate of adaptation, we fitted each participant's data with an exponential function as in common in studies of reach adaptation (e.g., Flanagan et al., 1999; Smith, Ghazizadeh, & Shadmehr, 2006). Specifically, we fitted the function

$$A = Be^{\lambda T} + C$$

where *A* represents normalized relative reach angle data (averaged across 8 directions per block) during the perturbation phase and *T* represents the trial block. The unknown parameters *B*, *λ*, and *C* were determined by a nonlinear least squares algorithm computed using the *nls* function in R (R Core Team, 2018), and the λ values were used to quantify each individual's initial rate of learning. However, the learning curves for most children were not exponential and, thus, not well fitted by the function. Hence, given that the function fitting method was only appropriate for adult participants, we also examined initial learning with an alternative method that determined how much adaptation had already occurred in the first two blocks of the perturbation phase.

For statistical analysis, all between-group comparisons were completed with Welch's *t*-tests in R (R Core Team, 2018). Where such tests are listed as a family of comparisons, the reported *p*-values are those obtained after adjustment with the Holm-Bonferroni method (Holm, 1979) to keep the family-wise error rate at .05. Cohen's *d* effect sizes were calculated using the *cohen.d* function in the *effsize* package for R (Torchiano, 2018).

Results

Movement duration

AWS and AWNS did not differ in movement duration, *t*(17.296) = -1.724, *p* = 0.103, *d* = 0.630. CWS also did not differ from CWNS in movement duration for any of the three age groups (Age 3-4, *t*(15.196) = 0.363, *p* = 1.000, *d* = 0.171; Age 5-6, *t*(10.985) = 1.073, *p* = 0.919, *d* = 0.537; Age 7-9, *t*(11.213) = 0.065, *p* = 1.000, *d* = 0.029).

Visuomotor adaptation in AWS vs. AWNS

Adult participants in both the stuttering and nonstuttering groups exhibited adaptation in response to the visual perturbation. Figure 8A includes group average relative reach angle data throughout the baseline, perturbation, and after-effects phase. By the last two blocks of the perturbation phase, AWS and AWNS showed a similar final extent of adaptation that was not statistically different, *t*(21.569) = -0.598, *p* = 0.556, *d* = 0.218.

Given that we examined both learning rate (by fitting an exponential function) and early learning (by analyzing adaptation in the first two blocks of the perturbation phase), the associated between-group comparisons were treated as a family of two tests. Fitting an exponential function to each participant's adaptation data (Figure 8B) revealed a statistically significant between-group difference in the initial rate of learning when the perturbation was first introduced: the exponential rate term was significantly more negative (i.e., a faster drop in the curve and, thus, faster learning) for AWNS as compared with AWNS, *t*(21.573) = -3.193, *p* = 0.013, *d* = 1.166 (Figure 8C). The same result was confirmed by the second method of examining initial learning: the extent of adaptation reached in the first two blocks after introduction of the perturbation was significantly greater for AWNS than for AWS, *t*(25.430) = -3.033, *p* = 0.013, *d* = 1.108.

There was no statistically significant correlation between stuttering frequency and extent of adaptation (*r*(13) = -0.008, *p* = 0.978). Stuttering frequency was also not significantly correlated with rate of adaptation determined by either the initial two perturbation blocks, *r*(13) = -0.085, *p* = 0.764, or the exponent rate terms, *r*(13) = -0.375, *p* = 0.168.

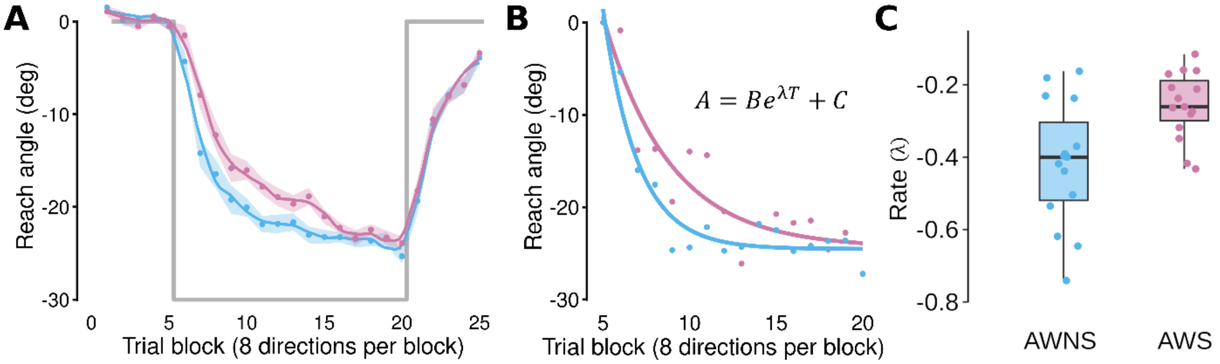


Figure 8. *Visuomotor adaptation to a 30° CCW rotation for adults who do not stutter (blue) and adults who stutter (pink). (A) Group averaged reach angles relative to baseline (averaged across 8 reach directions after normalization by direction). Shaded areas indicate standard errors of the mean. Gray lines plot the visual perturbation with reversed sign (thus indicating hypothetical full compensation). (B) Two individual participant illustrations of exponential fitting of reach angle data from the perturbation phase. (C) Individual participant data for the exponential rate term (λ), overlaid on group box plots.*

Visuomotor adaptation in CWS vs. CWNS

Children in both groups adapted to the perturbation (Figure 9), although the extent of adaptation was substantially smaller (11-16°) than that seen in adults (24-25°). One family of pair-wise between-group comparisons revealed that there was no statistically significant difference in final extent of adaptation between CWS and CWNS in any age subgroups (Age 3-4, *t*(15.786) = 1.133, *p* = 0.548, *d* = 0.534; Age 5-6, *t*(13.211) = 0.367, *p* = 0.719, *d* = 0.183; Age 7-9, *t*(17.993) = 1.680, *p* = 0.331, *d* = 0.751). The second family of pair-wise between-group comparisons revealed that there was also no statistically significant between-group difference in initial adaptation (Age 3-4, *t*(14.803) = -1.616, *p* = 0.331, *d* = 0.762; Age 5-6, *t*(9.482) = 1.430, *p* = 0.331, *d* = 0.715; Age 7-9, *t*(17.985) = -0.298, *p* = 0.331, *d* = 0.133).

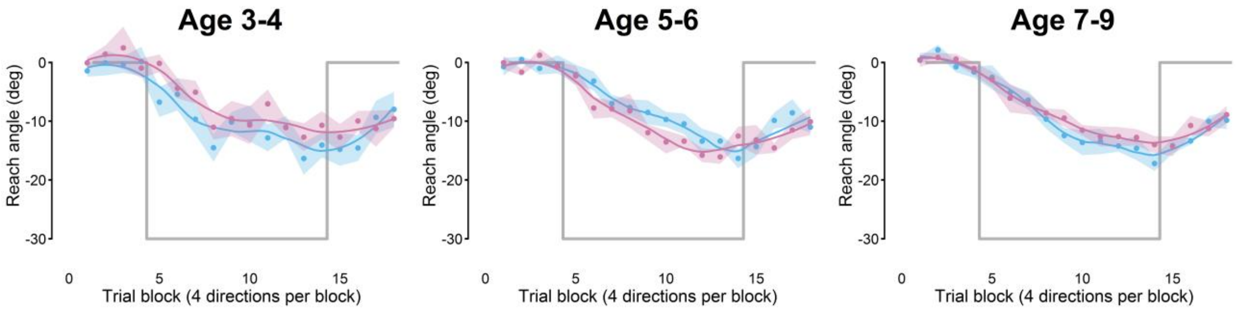


Figure 9. *Visuomotor adaptation to a 30° CCW rotation for children who do not stutter (blue) and children who stutter (pink). Data show group averaged reach angles relative to baseline (averaged across 4 reach directions after normalization by direction). Shaded areas indicate standard errors of the mean. Gray lines plot the visual perturbation with reversed sign (thus indicating hypothetical full compensation). the mean.*

The extent of adaptation in the final two blocks of the perturbation phase did not significantly correlate with age for either CWNS (*r*(25) = -0.049, *p* = 0.808) or CWS (*r*(25) = -0.091, *p* = 0.652).

Similarly, initial adaptation in the first two blocks of the perturbation phase also did not significantly correlate with age for either CWNS (*r*(25) = 0.142, *p* = 0.480) or CWS (*r*(25) = -0.109, *p* = 0.588). Lastly, for none of the three age groups of CWS was there a statistically significant correlation between the final extent of visuomotor adaptation and stuttering frequency from the clinical assessment (Age 3-4, *r*(7) = -0.040, *p* = 0.919; Age 5-6, *r*(6) = 0.028, *p* = 0.948; Age 7-9, *r*(8) = 0.426, *p* = 0.220), or between initial visuomotor adaptation and stuttering frequency from the clinical assessment (Age 3-4, *r*(7) = -0.364, *p* = 0.335; Age 5-6, *r*(6) = 0.162, *p* = 0.702; Age 7-9, *r*(8) = 0.465, *p* = 0.176).

Discussion

We investigated nonspeech sensorimotor learning in adults and children who stutter as compared with age-matched individuals who do not stutter. Specifically, participants were tested in a visuomotor rotation paradigm that involved fast out-and-back reaching movements while visual feedback was manipulated during the perturbation phase of the experiment. The perturbation involved a 30° CCW rotation (around the center of the workspace) of the location of a cursor representing fingertip position (the participant's hand and arm remained hidden below a screen throughout the experiment). We analyzed participants' adaptation to this perturbation (i.e., adjustments in reach direction in the opposite direction of the manipulation) in terms of both overall adaptation extent and initial rate of adaptation.

For adults, findings show that AWS did reach the same final level of adaptation as AWNS after more than 80 exposure trials, but that their initial rate of learning was statistically significantly slower. Thus, stuttering adults' limitations in limb visuomotor learning certainly are more subtle than those observed in the speech auditory-motor learning task in Experiment 1 where the overall extent of learning was also atypical. Nevertheless, this novel finding of a lower rate of learning for the updating of internal models that are used in the planning of upper limb reach movements suggests that stuttering individuals' sensorimotor integration difficulties are not limited to the orofacial effector system or speech motor control processes.

Unlike the situation for speech auditory-motor adaptation, however, the reach visuomotor task revealed a between-group difference only for the adult participant groups and not for children. Indeed, CWS did not differ from CWNS in either the final extent of learning or initial learning in any of the three age groups (3-4, 5-6, 7-9 years of age). This result could be interpreted as indicating that nonspeech visuomotor learning deficits develop only after the age of 9 years, although it is not obvious how stuttering itself, personal experiences related to stuttering, or techniques applied in stuttering treatment would contribute to the development of *upper limb* motor learning difficulties. Thus, the mechanisms behind such late-developing nonspeech effects would be entirely unknown.

Alternatively, it is also possible that the different outcome for speech auditory-motor learning (limitations observed in both AWS and CWS) and reach visuomotor learning (limitations observed in AWS but not CWS) stems from inherent differences in the typical developmental trajectories for these two tasks. In our speech auditory-motor adaptation task, even the youngest group of 3-4-year-old nonstuttering children already showed an extent of adaptation that was highly similar to that of the nonstuttering adult participants. Thus, speech auditory-motor learning may show a fast developmental trajectory, and an advanced learning ability in this domain may be necessary at an early age to continually update the relevant internal models that take account of how developmental changes in the biomechanical and neural systems for speech continually affect the transformations from motor commands to acoustic speech output (Callan et al., 2000; Kent, 1976; Kent, 1997; Vorperian et al., 1999; Vorperian et al., 2005; Vorperian & Kent, 2007; Vorperian et al., 2009). In contrast, in our reach visuomotor adaptation task, even the oldest group of 7-9-year-old nonstuttering children showed an extent of adaptation that was only 64% as large as that seen in nonstuttering adults. This finding is generally consistent with prior studies demonstrating that several aspects of reach sensorimotor learning continue to develop throughout the school-age years (Contreras-Vidal, Bo, Boudreau, & Clark, 2005)(Ferrel, Bard, & Fleury, 2001; Kagerer & Clark, 2014). Additionally, Kagerer and Clark (2015) reported that the learning of visuomotor maps develops gradually over many years whereas the learning of (nonspeech) auditory-motor maps may reach adult-like performance levels earlier around the age of 5-6 years. Hence, upper limb visuomotor learning shows a slower developmental trajectory, and even typical children in the age range tested here (3-9 years of age) still have limited learning abilities in this domain. We speculate that subtle differences between individuals who stutter and individuals who do not stutter are more likely to be detected once adult levels of performance are reached.

Despite of the lack of group difference found in children, visuomotor adaptation limitations in AWS suggests that the sensorimotor learning difficulties are not limited to the speech domain. Notably, this finding is supported by the fact that the neural circuits considered to be responsible for visuomotor learning are also consistently found to be abnormal in PWS. For example, abnormalities in the basal ganglia or its related circuit associated with stuttering have been found (Brown, Ingham, Ingham, Laird, & Fox, 2005; Chang & Zhu, 2013; Giraud et al., 2008; Wu et al., 1995; Wu et al., 1997) and hypothesized to cause stuttering moments (Alm, 2004; Civier et al., 2013; Max et al., 2004). Specifically, medications that reduce the hyperactivity of the dopaminergic neurotransmitter system in PWS can improve stuttering (e.g., Anderson, Hughes, Rothi, Crucian, & Heilman, 1999; Koller, 1983; Maguire, Yeh, & Ito, 2012; Shahed & Jankovic, 2001). In addition, abnormal dopaminergic system can reduce the rate of visuomotor adaptation as seen with Parkinson's disease patients (e.g., Contreras-Vidal & Buch, 2003). Another circuit widely thought to play a critical role in visuomotor learning is cocrtico-cerebellar. Both abnormal functional activation (Brown et al., 2005; De Nil, Kroll, & Houle, 2001; Fox et al., 1996; Lu et al., 2010) and atypical anatomy (e.g., Lu et al., 2010) in the cerebellum have been found in individuals who stutter. These abnormalities may be highly relevant to our finding in the current study as individuals with lower white matter integrity in the cerebellum exhibit slower rate of adaptation (Tomassini et al., 2011). Furthermore, a slower rate of visuomotor learning was found in cerebellar patients (e.g., Tseng, Diedrichsen, Krakauer, Shadmehr, & Bastian, 2007). Taken together, the visuomotor adaptation limitation in PWS is largely consistent with what has been reported in the literature, adding support to the idea that general sensorimotor learning difficulties are associated with stuttering.

General discussion

The aim of the present study was designed to examine whether there are sensorimotor learning difficulties, specifically speech auditory-motor learning and arm reaching visuomotor learning limitations, in both AWS and CWS. For the speech auditory-motor adaptation task (experiment 1), the formant frequencies of speech production were perturbed in real-time and played back over multiple trials. Acoustic analyses of the participants' speech production, specifically the changes their formant frequencies in response to the auditory perturbation, were conducted. For the arm visuomotor adaptation task (experiment 2), we applied the commonly used visuomotor rotation perturbation in which the visual feedback of the hand location (i.e., cursor) is rotated counter-clockwise around the center of the workspace. The recorded movements were analyzed to examine how the participants learned to change the movement directions for the cursor to reach closer to the target.

Experiment 1 demonstrated that auditory-motor adaptation was limited in AWS as well as CWS compared to their matched participants who do not stutter. Importantly, the limitations were severe in CWS, resulting in no adaptation in some cases. Additionally, younger CWS had the most difficulties in adapting compared to older stuttering individuals. These results together suggest that the auditory-motor learning deficits may already be present in childhood, and such deficits are actually more severe at early ages when there has not been much experiences of stuttering. Hence, the auditory-motor learning deficits may be closely related to the proximal cause of stuttering rather than being a by-product of years of experiences of stuttering. Experiment 2's results showed that AWS had a slower initial rate of visuomotor adaptation compared to AWNS, but the groups did not differ in the extent of adaptation suggesting only a subtle difference in visuomotor learning. CWS did not differ from CWNS in their visuomotor adaptation performances. Therefore, the subtle visuomotor adaptation limitation in individuals who stutter may not be detectable until the adaptation performances achieve adult-level. Taken together, the findings suggest that individuals who stutter have sensorimotor learning deficits in both speech and non-speech effector systems, albeit the deficits in non-speech system may be more subtle.

Overall, the results are consistent with the widely accepted theoretical framework for stuttering which hypothesizes that the neural representations of the involved motor-to-sensory transformations (i.e., internal models) are unstable, noisy, and/or incorrect in individuals who stutter (Cai et al., 2012; Daliri, Prokopenko, & Max, 2013; Hickok, Houde, & Rong, 2011; Max, 2004; Max et al., 2004; Neilson & Neilson, 1987). Importantly, our research group hypothesized that sensorimotor learning may be limited in stuttering individuals and that it may be especially challenging for children to acquire stable and/or correct internal models as the internal models must account for rapid neural and craniofacial developmental changes in childhood (Max et al., 2004). It is also critical to note that a vast amount of speech motor control is acquired in early childhood as most speech sounds are developed by age 5 and 6 (Sander, 1972; Smit, Hand, Freilinger, Bernthal, & Bird, 1990). Hence, under the intense pressure of urgent demand for acquiring the internal models for various speech movements, unstable and incorrect internal models may likely emerge, resulting in the severe auditory-motor learning limitations in CWS and leading to stuttering moments (Max et al., 2004).

Moreover, correlation data for CWS revealed that younger CWS who have the least amount of experiences of stuttering also exhibited the most severe auditory-motor adaptation limitations. Therefore, auditory-motor learning deficits found in PWS are likely not merely a compensatory effect as a result of years of experiencing stuttering (Daliri et al., 2018), but are rather likely associated with the development of stuttering in the early childhood years.

This finding, then, raises the question: which specific sensorimotor *mechanism(s)* in the unstable and/or incorrect internal models is responsible for limited sensorimotor learning in PWS? In the general motor control literature, it has been widely accepted that sensorimotor adaptation relies on sensory prediction error (Flanagan, Vetter, Johansson, & Wolpert, 2003; Haith & Krakauer, 2013), that is, the error signal that comes from forward models predicting the sensory consequences of planned movements (e.g., Wolpert & Miall, 1996; Wolpert & Flanagan, 2001). Thus, if sensory prediction errors are not accurate/appropriate in PWS, their adaptation performances may be affected (Max et al., 2004). In support of this view, a series of recent studies examined whether PWS can prepare their auditory systems in prediction of the upcoming self-generated speech sounds (Daliri & Max, 2015a; Daliri & Max, 2015b; Daliri & Max, 2018; Max & Daliri, 2019). The authors played auditory stimuli immediately before speech movements or silent reading conditions and measured the auditory evoked potentials in response to the stimuli. Compared to PWNS who modulated their auditory systems in preparation for upcoming speech sounds opposed to silent reading), PWS did not exhibit such pre-speech auditory modulation (i.e., no difference between before speaking and before silent reading), demonstrating abnormal sensory prediction. However, it should be noted that a separate study with upper-limb reaching movements found problems in control, rather than predictive component in PWS (Daliri et al., 2014).

Another—perhaps obvious and potentially interrelated with the problems in the forward models—problem involved may be poor sensory acuity. Specifically reduced oral somatosensory acuity in PWS has been found in numerous studies (Archibald & De Nil, 1999; De Nil & Abbs, 1991; Howell, Sackin, & Rustin, 1995; Loucks & De Nil, 2006a; Loucks & De Nil, 2006b) although Daliri et al. (2013) did not replicate the group difference. Regarding auditory acuity, there is evidence that auditory acuity to formant changes is not different between PWS and PWNS (Cai et al., 2012). Yet, this idea deserves more investigations because the vowel sounds tested for acuity in the study were much longer (300 ms) than vowel sounds in everyday speech, which may have allowed more opportunities for PWS to notice changes in the formant frequencies. Therefore, it remains plausible that sensory acuity may be responsible to sensorimotor learning difficulties.

An additional problem may be in updating the control policy (i.e., inverse model) for the subsequent trials during adaptation (Max et al., 2004; Max, 2004). In support of the view, a recent study has argued that at least a certain component of adaptation (i.e., implicit, see below) may be driven by inverse models rather than forward-model based learning (Hadjiosif et al., 2020 PREPRINT). However, it is difficult to draw a conclusion, as the specific sensorimotor mechanisms underlying auditory-motor adaptation remain largely speculative. One paradigm that can further shed light on these issues is examining different *components* of sensorimotor adaptation. Recent motor control literature has suggested that sensorimotor learning is comprised of both implicit component that is thought to be generated from prediction error and explicit component that is driven by performance error. Taylor and his colleagues succeeded to dissociate the two components in upper-limb visuomotor adaptation and found that each component had different time scales of learning (McDougle, Bond, & Taylor, 2015; Taylor et al., 2014). Examining which component(s) is affected in PWS may provide insights on specific sensorimotor mechanisms that are impaired.

We acknowledge several limitations in this study. We included the auditory-motor adaptation data from three stuttering adults who had hearing thresholds higher than 20 dB (two participants had 40 dB HL and one participant had 25 dB HL thresholds) at 4 kHz, even though hearing thresholds in all AWNS were 20 dB or lower. However, the effect of higher hearing thresholds on adaptation may have been minimal, because all of the three participants adapted more than their group average in all four conditions. Another limitation is that we did not find any significant correlation between stuttering frequency and extent/rate of adaptation. This is consistent with a previous study on auditory-motor learning in AWS by Daliri et al. (2018), but not with another study that found a significant negative relationship between stuttering frequency during a conversational sample and the amount of learning (Daliri & Max, 2018). The lack of correlation is not necessarily surprising, however, given that the relationship between movement measures and stuttering frequencies may be extremely complicated (see Daliri et al., 2014)

In conclusion, both AWS and CWS exhibited limited sensorimotor adaptation in speech auditory-motor domain. The auditory-motor adaptation limitations were not only present, but also more severe in CWS compared to AWS. Surprisingly, younger CWS had the most difficulties in adaptation, exhibiting no adaptation at all. Therefore, auditory-motor learning deficits in AWS are likely not by-products of years of experiencing stuttering, instead they may represent impairment in the acquisition of precise/stable internal models required for fluent speech. In non-speech visuomotor domain, even though the sensorimotor learning performance did not differ between CWS and CWNS, when visuomotor learning had reached advanced levels of learning in the adult years, subtle limitations (i.e., in the initial rate of adaptation) were again observable in individuals who stutter. These findings provide additional evidence on sensorimotor learning difficulties in PWS as well as new inquiries on their potential link to the development of stuttering.

Acknowledgements

The authors thank Roman Prokopenko, Ayoub Daliri, Caitlin Baldwin, Derek Maffett, Churan Li, and Patrick McCleary for extensive contributions to data collection and Elizabeth Rylance and Prince Z. Wang for help with data analysis. We also acknowledge our research speech-language pathologist Allison Stewart and graduate clinicians Annie Zhu and Chrystal Chau for conducting speech, language, and hearing evaluations for all children and stuttering assessments for all individuals who stutter. This research was supported by grants from the National Institute on Deafness and Other Communication Disorders (R01 DC007603, R01 DC014510, and R01 DC17444, L. Max PI) and the Canadian Institutes of Health Research (MOP-137001, L. Max Co-I). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, or the Canadian Institutes of Health Research.

References

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: A critical review of possible relations. *Journal of Communication Disorders, 37*(4), 325-369. doi:10.1016/j.jcomdis.2004.03.001

Anderson, J. M., Hughes, J. D., Rothi, L. J., Crucian, G. P., & Heilman, K. M. (1999). Developmental stuttering and parkinson's disease: The effects of levodopa treatment. *Journal of Neurology, Neurosurgery, and Psychiatry, 66*(6), 776-778. doi:10.1136/jnnp.66.6.776

Archibald, L., & De Nil, L.,F. (1999). The relationship between stuttering severity and kinesthetic acuity for jaw movements in adults who stutter. *Journal of Fluency Disorders, 24*(1), 25-42. doi:10.1016/S0094-730X(98)00023-0

Beal, D. S., Gracco, V. L., Brettschneider, J., Kroll, R. M., & De Nil, L. F. (2013). A voxel-based morphometry (VBM) analysis of regional grey and white matter volume abnormalities within the speech production network of children who stutter. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 49*(8), 2151-2161. doi:10.1016/j.cortex.2012.08.013

Benito-Aragón, C., Gonzalez-Sarmiento, R., Liddell, T., Diez, I., d'Oleire Uquillas, F., Ortiz-Terán, L., . . . Sepulcre, J. (2020). Neurofilament-lysosomal genetic intersections in the cortical network of stuttering. *Progress in Neurobiology, 184,* 101718. doi:10.1016/j.pneurobio.2019.101718

Bishop, J. H., Williams, H. G., & Cooper, W. A. (1991). Age and task complexity variables in motor performance of children with articulation-disordered, stuttering, and normal speech. *Journal of Fluency Disorders, 16*(4), 219-228. doi:10.1016/0094-730X(91)90004-V

Blomgren, M., & Goberman, A. M. (2008). Revisiting speech rate and utterance length manipulations in stuttering speakers. *Journal of Communication Disorders, 41*(2), 159-178. doi:10.1016/j.jcomdis.2007.10.001

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer [computer software]

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping, 25*(1), 105-117. doi:10.1002/hbm.20140

Buhr, A., & Zebrowski, P. (2009). Sentence position and syntactic complexity of stuttering in early childhood: A longitudinal study. *Journal of Fluency Disorders, 34*(3), 155-172. doi:10.1016/j.jfludis.2009.08.001

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter:

Evidence for abnormal auditory-motor transformation. *PloS One, 7*(7), e41830. doi:10.1371/journal.pone.0041830

Cai, S., Tourville, J. A., Beal, D. S., Perkell, J. S., Guenther, F. H., & Ghosh, S. S. (2014). Diffusion imaging of cerebral white matter in persons who stutter: Evidence for network-level anomalies. *Frontiers in Human Neuroscience, 8* doi:10.3389/fnhum.2014.00054

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). 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, 43*, 721-736. doi:10.1044/jslhr.4303.721

Chang, S., Horwitz, B., Ostuni, J., Reynolds, R., & Ludlow, C. L. (2011). Evidence of left inferior frontal-premotor structural and functional connectivity deficits in adults who stutter. *Cerebral Cortex (New York, N.Y.: 1991), 21*(11), 2507-2518. doi:10.1093/cercor/bhr028

Chang, S., & Zhu, D. C. (2013). Neural network connectivity differences in children who stutter. *Brain : A Journal of Neurology, 136*(Pt 12), 3709-3726. doi:10.1093/brain/awt275

Choo, A. L., Chang, S., Zengin-Bolatkale, H., Ambrose, N. G., & Loucks, T. M. (2012). Corpus callosum morphology in children who stutter. *Journal of Communication Disorders, 45*(4), 279-289. doi:10.1016/j.jcomdis.2012.03.004

Civier, O., Bullock, D., Max, L., & Guenther, F. H. (2013). Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. *Brain and Language, 126*(3), 263-278. doi:10.1016/j.bandl.2013.05.016

Contreras-Vidal, J. L., Bo, J., Boudreau, J. P., & Clark, J. E. (2005). Development of visuomotor representations for hand movement in young children. *Experimental Brain Research, 162*(2), 155-164. doi:10.1007/s00221-004-2123-7

Contreras-Vidal, J. L., & Buch, E. R. (2003). Effects of parkinson's disease on visuomotor adaptation. *Experimental Brain Research, 150*(1), 25-32. doi:10.1007/s00221-003-1403-y

Coren, S. (1993). The lateral preference inventory for measurement of handedness, footedness, eyedness, and earedness: Norms for young adults. *Bulletin of the Psychonomic Society, 31*(1), 1-3. doi:10.3758/BF03334122

Cornelisse, L. E., Gagné, J., & Seewald, R. C. (1991). Ear level recordings of the long-term average spectrum of speech*. *Ear and Hearing, 12*(1), 47.

Criscimagna-Hemminger, S. E., Bastian, A. J., & Shadmehr, R. (2010). Size of error affects cerebellar contributions to motor learning. *Journal of Neurophysiology,* doi:10.1152/jn.00822.2009

Daliri, A., & Max, L. (2015a). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain and Language, 150*, 37-44. doi:10.1016/j.bandl.2015.08.008

Daliri, A., & Max, L. (2015b). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language, 143*, 59-68. doi:10.1016/j.bandl.2015.03.002

Daliri, A., & Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 99*, 55-68. doi:10.1016/j.cortex.2017.10.019

Daliri, A., Prokopenko, R. A., Flanagan, J. R., & Max, L. (2014). Control and prediction components of movement planning in stuttering versus nonstuttering adults. *Journal of Speech, Language, and Hearing Research: JSLHR, 57*(6), 2131-2141. doi:10.1044/2014_JSLHR-S-13-0333

Daliri, A., Prokopenko, R. A., & Max, L. (2013). Afferent and efferent aspects of mandibular sensorimotor control in adults who stutter. *Journal of Speech, Language, and Hearing Research: JSLHR, 56*(6), 1774-1788. doi:10.1044/1092-4388(2013/12-0134)

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science, 21*(2) doi:10.1111/desc.12521

Dayalu, V. N., Kalinowski, J., & Stuart, A. (2005). Stuttering frequency on meaningful and nonmeaningful words in adults who stutter. *Folia Phoniatrica Et Logopaedica : Official Organ of the International Association of Logopedics and Phoniatrics (IALP), 57*(4), 193-201. doi:10.1159/000085187

De Nil, L. F., & Abbs, J. H. (1991). Kinaesthetic acuity of stutterers and non-stutterers for oral and non-oral movements. *Brain: A Journal of Neurology, 114 (Pt 5)*, 2145-2158. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1933239

De Nil, L. F., Kroll, R. M., & Houle, S. (2001). Functional neuroimaging of cerebellar activation during single word reading and verb generation in stuttering and nonstuttering adults. *Neuroscience Letters, 302*(2-3), 77-80. doi:10.1016/s0304-3940(01)01671-8

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences, 4*(11), 423-431. doi:10.1016/s1364-6613(00)01537-0

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology, 10*(6), 732-739. doi:10.1016/s0959-4388(00)00153-7

Doyon, J., Penhune, V., & Ungerleider, L. G. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia, 41*(3), 252-262. doi:10.1016/s0028-3932(02)00158-6

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test* Pearson Assessments.

Feng, Y., Gracco, V., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology, 106*(2), 667-679. doi:10.1152/jn.00638.2010

Ferrel, C., Bard, C., & Fleury, M. (2001). Coordination in childhood: Modifications of visuomotor representations in 6- to 11-year-old children. *Experimental Brain Research, 138*(3), 313-321. doi:10.1007/s002210100697

Flanagan, J. R., Nakano, E., Imamizu, H., Osu, R., Yoshioka, T., & Kawato, M. (1999). Composition and decomposition of internal models in motor learning under altered kinematic and dynamic environments. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 19*(20), RC34. doi:10.1523/JNEUROSCI.19-20-j0005.1999

Flanagan, J. R., Vetter, P., Johansson, R. S., & Wolpert, D. M. (2003). Prediction precedes control in motor learning. *Current Biology: CB, 13*(2), 146-150. doi:10.1016/s0960-9822(03)00007-1

Forster, D. C., & Webster, W. G. (2001). Speech-motor control and interhemispheric relations in recovered and persistent stuttering. *Developmental Neuropsychology, 19*(2), 125-145. doi:10.1207/S15326942DN1902_1

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., . . . Lancaster, J. L. (1996). A PET study of the neural systems of stuttering. *Nature, 382*(6587), 158-161. doi:10.1038/382158a0

Frigerio-Domingues, C., & Drayna, D. (2017). Genetic contributions to stuttering: The current evidence. *Molecular Genetics & Genomic Medicine*, *5*(2), 95-102. doi:10.1002/mgg3.276

Frigerio-Domingues, C., Gkalitsiou, Z., Zezinka, A., Sainz, E., Gutierrez, J., Byrd, C., . . . Drayna, D. (2019). Genetic factors and therapy outcomes in persistent developmental stuttering. *Journal of Communication Disorders, 80*, 11-17. doi:10.1016/j.jcomdis.2019.03.007

Giraud, A. L., Neumann, K., Bachoud-Levi, A. C., von Gudenberg, A. W., Euler, H. A., Lanfermann, H., & Preibisch, C. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. *Brain and Language, 104*(2), 190-199. doi:10.1016/j.bandl.2007.04.005

Goldman, R., & Fristoe, M. (2000). *Goldman fristoe test of articulation-2 (GFTA-2)* . Circle Pines, Minnesota: American Guidance Service. Inc.

Goldman, R., & Fristoe, M. (2015). *GFTA-3: Goldman fristoe 3 test of articulation* PsychCorp.

Hadjiosif, A. M., Krakauer, J. W., & Haith, A. M. (2020). Implicit adaptation is driven by direct policy updates rather than forward-model-based learning. *bioRxiv*, , 2020.01.22.914473. doi:10.1101/2020.01.22.914473

Haith, A. M., & Krakauer, J. W. (2013). Model-based and model-free mechanisms of human motor learning. *Advances in Experimental Medicine and Biology, 782*, 1-21. doi:10.1007/978-1-4614-5465-6_1

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69*(3), 407-422. doi:10.1016/j.neuron.2011.01.019

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70. Retrieved from http://www.jstor.org/stable/4615733

Howell, P., Sackin, S., & Rustin, L. (1995). Comparison of speech motor development in stutterers and fluent speakers between 7 and 12 years old. *Journal of Fluency Disorders, 20*(3), 243-255. doi:10.1016/0094-730X(94)00011-H

Hresko, W. P., Reid, D. K., & Hammill, D. D. (1999). *TELD-3: Test of early language development* Pro-ed.

Ingham, R. J., Fox, P. T., Ingham, J. C., Xiong, J., Zamarripa, F., Hardies, L. J., & Lancaster, J. L. (2004). Brain correlates of stuttering and syllable production: Gender comparison and replication. *Journal of Speech, Language, and Hearing Research: JSLHR, 47*(2), 321-341. doi:10.1044/1092-4388(2004/026)

Jones, R. D., White, A. J., Lawson, K. H. C., & Anderson, T. J. (2002). Visuoperceptual and visuomotor deficits in developmental stutterers: An exploratory study. *Human Movement Science, 21*(5), 603-619. doi:10.1016/S0167-9457(02)00165-3

Kagerer, F. A., & Clark, J. E. (2014). Development of interactions between sensorimotor representations in school-aged children. *Human Movement Science, 34*, 164-177. doi:10.1016/j.humov.2014.02.001

Kagerer, F. A., & Clark, J. E. (2015). Development of kinesthetic-motor and auditory-motor representations in school-aged children. *Experimental Brain Research, 233*(7), 2181-2194. doi:10.1007/s00221-015-4288-7

Kang, C., Riazuddin, S., Mundorff, J., Krasnewich, D., Friedman, P., Mullikin, J. C., & Drayna, D. (2010). Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. *The New England Journal of Medicine, 362*(8), 677-685. doi:10.1056/NEJMoa0902630

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology, 9*(6), 718-727. doi:10.1016/s0959-4388(99)00028-8

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research, 19*(3), 421-447.

Kent, R. D. (1997). *The speech sciences*. San Diego, CA: Singular.

Koller, W. C. (1983). Dysfluency (stuttering) in extrapyramidal disease. *Archives of Neurology, 40*(3), 175-177. doi:10.1001/archneur.1983.04050030069014

Krakauer, J. W., & Mazzoni, P. (2011). Human sensorimotor learning: Adaptation, skill, and beyond. *Current Opinion in Neurobiology, 21*(4), 636-644. doi:10.1016/j.conb.2011.06.012

Lawrence, M. A. (2016). *Ez: Easy analysis and visualization of factorial experiments*

Loucks, T. M., & De Nil, L. F. (2006a). Anomalous sensorimotor integration in adults who stutter: A tendon vibration study. *Neuroscience Letters, 402*(1-2), 195-200. doi:10.1016/j.neulet.2006.04.002

Loucks, T. M., & De Nil, L. F. (2006b). Oral kinesthetic deficit in adults who stutter: A target-accuracy study. *Journal of Motor Behavior, 38*(3), 238-246. doi:10.3200/JMBR.38.3.238-247

Lu, C., Chen, C., Ning, N., Ding, G., Guo, T., Peng, D., . . . Lin, C. (2010). The neural substrates for atypical planning and execution of word production in stuttering. *Experimental Neurology, 221*(1), 146-156. doi:10.1016/j.expneurol.2009.10.016

Maguire, G. A., Yeh, C. Y., & Ito, B. S. (2012). Overview of the diagnosis and treatment of stuttering. *Journal of Experimental & Clinical Medicine, 4*(2), 92-97. doi:10.1016/j.jecm.2012.02.001

Maschke, M., Gomez, C. M., Ebner, T. J., & Konczak, J. (2004). Hereditary cerebellar ataxia progressively impairs force adaptation during goal-directed arm movements. *Journal of Neurophysiology, 91*(1), 230-238. doi:10.1152/jn.00557.2003

Max, L. (2004). Stuttering and internal models for sensorimotor control: A theoretical perspective to generate testable hypotheses. *Speech motor control in normal and disordered speech* (pp. 357–388) Oxford University Press.

Max, L., Caruso, A. J., & Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech,* *Language, and Hearing Research: JSLHR, 46*(1), 215-232. doi:10.1044/1092-4388(2003/017)

Max, L., & Daliri, A. (2019). Limited pre-speech auditory modulation in individuals who stutter: Data and hypotheses. *Journal of Speech, Language, and Hearing Research: JSLHR, 62*(8S), 3071-3084. doi:10.1044/2019_JSLHR-S-CSMC7-18-0358

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (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 Sciences and Disorders, 31*, 105-122. doi:10.1044/cicsd_31_S_105

Max, L., & Maffett, D. G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience Letters, 591*, 25-29. doi:10.1016/j.neulet.2015.02.012

Max, L., Wallace, M. E., & Vincent, I.Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. Paper presented at the *International Congress of Phonetic Sciences*, 1053-1056.

Mazzoni, P., & Krakauer, J. W. (2006). An implicit plan overrides an explicit strategy during visuomotor adaptation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 26*(14), 3642-3645. doi:10.1523/JNEUROSCI.5317-05.2006

McDougle, S. D., Bond, K. M., & Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *The Journal of Neuroscience, 35*(26), 9568-9579. doi:10.1523/JNEUROSCI.5061-14.2015

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

Nippold, M. A. (2002). Stuttering and phonology: Is there an interaction? . *American Journal of Speech-Language Pathology, 11*(2), 99-110. doi:10.1044/1058-0360(2002/011)

Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia, 9*(1), 97-113. doi:10.1016/0028-3932(71)90067-4

Onslow, M., & Packman, A. (2002). Stuttering and lexical retrieval: Inconsistencies between theory and data. *Clinical Linguistics & Phonetics, 16*, 295-298. doi:10.1080/02699200210128972

Packman, A., & Attanasio, J. S. (2004). *Theoretical issues in stuttering*. New York, NY: Psychology Press.

Packman, A., Onslow, M., Coombes, T., & Goodwin, A. (2001). Stuttering and lexical retrieval. *Clinical Linguistics & Phonetics, 15*(6), 487-498. doi:10.1080/02699200110059114

R Core Team. (2018). R: A language and environment for statistical computing [computer software]. Vienna, Austria:

Rabe, K., Livne, O., Gizewski, E. R., Aurich, V., Beck, A., Timmann, D., & Donchin, O. (2009). Adaptation to visuomotor rotation and force field perturbation is correlated to different brain areas in patients with cerebellar degeneration. *Journal of Neurophysiology, 101*(4), 1961-1971. doi:10.1152/jn.91069.2008

Raza, M. H., Amjad, R., Riazuddin, S., & Drayna, D. (2012). Studies in a consanguineous family reveal a novel locus for stuttering on chromosome 16q. *Human Genetics, 131*(2), 311-313. doi:10.1007/s00439-011-1134-2

Riley, G. (1994). Stuttering severity instrument for children and adults. *Doi,*

Riley, G. (2009). *SSI-4: Stuttering severity instrument*

Robertson, E., & Miall, R. (1999). Visuomotor adaptation during inactivation of the dentate nucleus. *Neuroreport, 10*(5), 1029-1034. doi:10.1097/00001756-199904060-00025

Sander, E. K. (1972). When are speech sounds learned? *The Journal of Speech and Hearing Disorders, 37*(1), 55-63. doi:10.1044/jshd.3701.55

Semel, E. M., Wiig, E. H., & Secord, W. (2004). *Celf 4* PsychCorp.

Sengupta, R., Shah, S., Gore, K., Loucks, T. M., & Nasir, S. M. (2016). Anomaly in neural phase coherence accompanies reduced sensorimotor integration in adults who stutter. *Neuropsychologia, 93*(Pt A), 242-250. doi:10.1016/j.neuropsychologia.2016.11.004

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience, 33*(1), 89-108. doi:10.1146/annurev-neuro-060909-153135

Shahed, J., & Jankovic, J. (2001). Re-emergence of childhood stuttering in parkinson's disease: A hypothesis. *Movement Disorders: Official Journal of the Movement Disorder Society, 16*(1), 114-118. doi:10.1002/1531-8257(200101)16:13.0.co;2-2

Sitek, K. R., Cai, S., Beal, D. S., Perkell, J. S., Guenther, F. H., & Ghosh, S. S. (2016). Decreased cerebellar-orbitofrontal connectivity correlates with stuttering severity: Whole-brain functional and structural connectivity associations with persistent developmental stuttering. *Frontiers in Human Neuroscience, 10,* 190. doi:10.3389/fnhum.2016.00190

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The iowa articulation norms project and its nebraska replication. *The Journal of Speech and Hearing Disorders, 55*(4), 779-798. doi:10.1044/jshd.5504.779

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLOS Biology, 4*(6), e179. doi:10.1371/journal.pbio.0040179

Smith, M. A., & Shadmehr, R. (2005). Intact ability to learn internal models of arm dynamics in huntington's disease but not cerebellar degeneration. *Journal of Neurophysiology, 93*(5), 2809-2821. doi:10.1152/jn.00943.2004

Smits-Bandstra, S., & De Nil, L. (2009). Speech skill learning of persons who stutter and fluent speakers under single and dual task conditions. *Clinical Linguistics & Phonetics, 23*(1), 38-57. doi:10.1080/02699200802394914

Smits-Bandstra, S., & De Nil, L. F. (2007). Sequence skill learning in persons who stutter: Implications for cortico-striato-thalamo-cortical dysfunction. *Journal of Fluency Disorders, 32*(4), 251-278. doi:10.1016/j.jfludis.2007.06.001

Smits-Bandstra, S., De Nil, L. F., & Saint-Cyr, J. (2006). Speech and nonspeech sequence skill learning in adults who stutter. *Journal of Fluency Disorders, 31*(2), 116-136. doi:10.1016/j.jfludis.2006.04.003

Smits-Bandstra, S., De Nil, L., & Rochon, E. (2006). The transition to increased automaticity during finger sequence learning in adult males who stutter. *Journal of Fluency Disorders, 31*(1), 22-40. doi:10.1016/j.jfludis.2005.11.004

Taylor, J. A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 34*(8), 3023-3032. doi:10.1523/JNEUROSCI.3619-13.2014

Tomassini, V., Jbabdi, S., Kincses, Z. T., Bosnell, R., Douaud, G., Pozzilli, C., . . . Johansen-Berg, H. (2011). Structural and functional bases for individual differences in motor learning. *Human Brain Mapping, 32*(3), 494-508. doi:10.1002/hbm.21037

Tong, C., & Flanagan, J. R. (2003). Task-specific internal models for kinematic transformations. *Journal of Neurophysiology, 90*(2), 578-585. doi:10.1152/jn.01087.2002

Torchiano, M. (2018). *Effsize - A package for efficient effect size computation* doi:10.5281/zenodo.1480624

Tseng, Y., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of Neurophysiology*, *98*(1), 54-62. doi:10.1152/jn.00266.2007

Unicomb, R., Kefalianos, E., Reilly, S., Cook, F., & Morgan, A. (2020). Prevalence and features of comorbid stuttering and speech sound disorder at age 4 years. *Journal of Communication Disorders, 84,* 105976. doi:10.1016/j.jcomdis.2020.105976

Vallabha, G. K., & Tuller, B. (2002). Systematic errors in the formant analysis of steady-state vowels. *Speech Communication, 38*(1), 141-160. doi:10.1016/S0167-6393(01)00049-8

Venkatakrishnan, A., Banquet, J. P., Burnod, Y., & Contreras-vidal, J. (2011). Parkinson's disease differentially affects adaptation to gradual as compared to sudden visuomotor distortions. *Human Movement Science, 30*(4), 760-769. doi:10.1016/j.humov.2010.08.020

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research : JSLHR, 50*(6), 1510-1545. doi:10.1044/1092-4388(2007/104)

Vorperian, H. K., Kent, R. D., Gentry, L. R., & Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: Preliminary results. *International Journal of Pediatric Otorhinolaryngology, 49,* 197-206. doi:10.1016/s0165-5876(99)00208-6

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America, 117*(1), 338-350. doi:10.1121/1.1835958

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., . . . Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America, 125*(3), 1666-1678. doi:10.1121/1.3075589

Webster, W. G. (1997). Principles of human brain organization related to lateralization of language and speech motor functions in normal speakers and stutterers. In W. Hustijn, H. F. M. Peters & P. van Lieshout (Eds.), *Speech production: Motor control, brain research and fluency disorders.* (pp. 119-139). Amsterdam, The Netherlands: Elsevier.

Webster, W. G., & Ryan, C. R. (1991). Task complexity and manual reaction times in people who stutter. *Journal of Speech and Hearing Research, 34*(4), 708-714. doi:10.1044/jshr.3404.708

Weiner, M. J., Hallett, M., & Funkenstein, H. H. (1983). Adaptation to lateral displacement of vision in patients with lesions of the central nervous system. *Neurology, 33*(6), 766-772.

Werner, S., Bock, O., Gizewski, E. R., Schoch, B., & Timmann, D. (2010). Visuomotor adaptive improvement and aftereffects are impaired differentially following cerebellar lesions in SCA and PICA territory. *Experimental Brain Research.Experimentelle Hirnforschung.Experimentation Cerebrale, 201*(3), 429-439. doi:10.1007/s00221-009-2052-6

Werner, S., Bock, O., & Timmann, D. (2009). The effect of cerebellar cortical degeneration on adaptive plasticity and movement control. *Experimental Brain Research.Experimentelle*

Hirnforschung.Experimentation Cerebrale, 193(2), 189-196. doi:10.1007/s00221-008-1607-2

Williams, K. T., & Williams, K. T. (2007). *EVT-2: Expressive vocabulary test* Pearson Assessments.

Wolpert, D. M., & Miall, R. C. (1996). Forward models for physiological motor control. *Neural Networks : The Official Journal of the International Neural Network Society, 9*(8), 1265-1279.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology: CB, 11*(18), 729. doi:10.1016/s0960-9822(01)00432-8

Wolpert, D. M., & Flanagan, J. R. (2009). Forward models. *The oxford companion to consciousness* (pp. 295-296) OUP Oxford.

Wu, J. C., Maguire, G., Riley, G., Fallon, J., LaCasse, L., Chin, S., . . . Lottenberg, S. (1995). A positron emission tomography [18F]deoxyglucose study of developmental stuttering. *Neuroreport, 6*(3), 501-505.

Wu, J. C., Maguire, G., Riley, G., Lee, A., Keator, D., Tang, C., . . . Najafi, A. (1997). Increased dopamine activity associated with stuttering. *Neuroreport, 8*(3), 767-770. doi:10.1097/00001756-199702100-00037

Wymbs, N. F., Ingham, R. J., Ingham, J. C., Paolini, K. E., & Grafton, S. T. (2013). Individual differences in neural regions functionally related to real and imagined stuttering. *Brain and Language, 124*(2), 153-164. doi:10.1016/j.bandl.2012.11.013

Zelaznik, H. N., Smith, A., Franz, E. A., & Ho, M. (1997). Differences in bimanual coordination associated with stuttering. *Acta Psychologica, 96*(3), 229-243. doi:10.1016/s0001-6918(97)00014-0

Chapter II

It's About Time: Minimizing Hardware and Software Latencies in Speech Research with Real-Time Auditory Feedback

Kwang Seob Kim
Department of Speech and Hearing Sciences
University of Washington, Seattle, WA
Hantao Wang
Department of Speech and Hearing Sciences
University of Washington, Seattle, WA
Ludo Max
Department of Speech and Hearing Sciences
University of Washington, Seattle, WA
Haskins Laboratories, New Haven, CT

Conflict of Interest Statement: The authors declare that there is no conflict of interest.

Funding Statement: This work was supported in part by grants from the National Institute on Deafness and Other Communication Disorders Grants R01DC014510 and R01DC17444 (principal investigator: L. Max) and the Canadian Institutes of Health Research Grant MOP-137001 (principal investigator: V. L. Gracco).

Correspondence concerning this work should be addressed to Ludo Max, Ph.D., Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd St., Seattle, WA 98105-6246. E-mail: LudoMax@uw.edu. Phone: 206 543 2674.

Abstract

Purpose: Various aspects of speech production related to auditory-motor integration and learning have been examined through auditory feedback perturbation paradigms in which participants' acoustic speech output is experimentally altered and played back via earphones/headphones "in real-time." Scientific rigor requires high precision in determining and reporting the involved hardware and software latencies. Many reports in the literature, however, are not consistent with the minimum achievable latency for a given experimental set-up. Here, we focus specifically on this methodological issue associated with implementing real-time auditory feedback perturbations, and we offer concrete suggestions for increased reproducibility in this particular line of work.

Method: Hardware- and software latencies as well as total feedback loop latency were measured for formant perturbation studies with the Audapter software. Measurements were conducted for various audio interfaces, desktop and laptop computers, and audio drivers. An approach for lowering Audapter's software latency through non-default parameter specification was also tested.

Results: Oft-overlooked hardware-specific latencies were not negligible for some of the tested audio interfaces (adding up to 15 ms). Total feedback loop latencies (including both hardware and software latency) were also generally larger than claimed in the literature. Non-default parameter values can improve Audapter's own processing latency without negative impact on formant tracking.

Conclusions: Audio interface selection and software parameter optimization substantially affect total feedback loop latency. Thus, the actual total latency (hardware plus software) needs to be correctly measured and described in all published reports. Future speech research with "real-time" auditory feedback perturbations should increase scientific rigor by minimizing this latency.

Key words: Speech production; Instrumentation; Auditory feedback; Perturbation; Signal processing

It's About Time: Minimizing Hardware and Software Latencies in Speech Research with Real-Time Auditory Feedback

During the last two decades, there has been a strong interest in various aspects of speech production related to auditory-motor interactions and auditory-motor learning (e.g., Cai et al., 2012; Cai et al., 2014; Daliri et al., 2018; Daliri & Max, 2018; Feng et al., 2018; Houde & Jordan, 1998; Keough et al., 2013; Max et al., 2003; Mollaei et al., 2013; Mollaei et al., 2016; see also Caudrelier & Rochet-Capellan, 2019, for an extensive review). Many experiments in this general area of work involve auditory feedback perturbation paradigms in which participants' acoustic speech output is experimentally altered and played back via earphones/headphones "in real-time." In an ideal set-up, the perturbed auditory feedback would be provided at the same time as when it is produced. In reality, however, there will always be an inevitable delay inherent in the audio and/or computer hardware and the signal processing algorithms implemented by those hardware components. For example, in order to track the speaker's fundamental frequency or formant frequencies – and then manipulate those frequency values prior to outputting the feedback signal – the involved signal processing algorithms require a minimum window length, causing some amount of delay in the feedback output. In addition, analog-to-digital (A/D) and digital-to-analog (D/A) converters built into the audio or computer hardware add to the overall input-output latency. Although the auditory feedback in these experiments is often referred to as a "real-time" signal, certain combinations of hardware and software components might result in latencies that are sufficiently long to potentially have an impact on the experimental results.

It has been shown, for example, that auditory feedback delays disrupt aspects of ongoing speech production when the delay reaches 50 ms or more (Stuart et al., 2002). Additionally, studies on speech auditory-motor learning across trials have demonstrated that adaptation to frequency-shifted auditory feedback is reduced or eliminated when the feedback signal is delayed by 100 ms (Max & Maffett, 2015; Mitsuya et al., 2017; work currently in progress in our lab is testing the effect on adaptation of delays shorter than 100 ms). In other experiments, including those focused on measuring participants' immediate phonatory or articulatory compensation for unpredictable within-trial perturbations in the auditory feedback, participant response latency often is one of the dependent variables, and the validity of such measurements depends on determining and reporting the feedback loop's input-output latency. Thus, in speech and voice studies that provide participants with auditory feedback there are numerous experimental situations where scientific rigor demands an accurate determination and reporting of the involved hardware and software latencies. Here, we focus specifically on methodological issues involved in studies that implement real-time auditory feedback perturbations, and we will offer concrete suggestions for increased rigor and reproducibility in this particular line of work.

Generally speaking, there are two options for an experimental set-up that implements auditory feedback perturbations. The first option is to use a commercially available vocal processor (e.g., Bauer & Larson, 2003; Behroozmand et al., 2009; Feng et al., 2011; Heller Murray et al., 2019; Loucks et al., 2012; Max et al., 2003; Rochet-Capellan & Ostry, 2011). One device that has been used in speech research, although no longer manufactured at the present time, is TC Helicon's VoiceOne, a rack-mount digital processor that is computer controllable by means of Musical Instrument Digital Interface (MIDI) signals. The VoiceOne processor is able to implement independent pitch and formant shifting with a confirmed total latency of 10-11 ms when using the "live" setting (e.g., Feng et al., 2011; Feng et al., 2018; Jones & Keough, 2008; Keough et al., 2013; Lametti et al., 2014; Max & Maffett, 2015; Mollaei et al., 2013; Mollaei et al., 2016; Rochet-Capellan & Ostry, 2011). The second option consists of using custom-designed signal processing software with a digital signal processing board (e.g., Houde & Jordan, 2002; Purcell & Munhall, 2006; Villacorta et al., 2007) or a commercially available consumer-grade audio interface (Cai et al., 2008; Tourville et al., 2013). In the latter category—custom signal processing software used with an audio interface—the MATLAB-based application Audapter (a MEX interface built from C++ source code) has gained great popularity due to its capability to implement many different real-time perturbations (e.g., Abur et al., 2018; Ballard et al., 2018; Cai et al., 2008; Cai et al., 2010; Cai et al., 2012; Cai et al., 2014; Daliri et al., 2018; Daliri & Dittman, 2019; Franken, Eisner et al., 2018; Franken, Acheson et al., 2018; Klein et al., 2018; Lametti et al., 2018; Reilly & Pettibone, 2017; Sares et al., 2018; Sato & Shiller, 2018; Stepp et al., 2017).

Interestingly, for auditory feedback perturbations with Audapter, the overall system input-out latency has been reported to be as short as 10 ms and as long as 45 ms, but in most studies it has been listed in the range 10-15 ms. Unfortunately, measurement methodology issues have caused misleading information to be reported. As we demonstrate below, when the Audapter latency is claimed to be as short as 10-15 ms, the measurement refers to *only* the software processing time, and it does *not* take into account the equally important and additive latency associated with the Audio Stream Input/Output (ASIO) hardware audio interface. The relevant steps that *should* be taken into account for such a latency report include all three of (1) audio signal input to the interface (Analog/Digital conversion), (2) software signal processing, and (3) audio signal output from the interface (Digital/Analog conversion). Indeed, Audapter's own software processing latency (with default settings) already is 10-14 ms, and the input-output temporal asynchrony data that are available within the Audapter program refer exclusively to this software processing time (i.e., it reports the time delay between the digital input signal after it has already been A/D converted by the audio interface and the processed digital output signal before it is D/A converted and output by the audio interface). Thus, it is necessary to bring attention to the fact that—contrary to common assumptions—the additional hardware delays associated with the audio interface are not at all negligible (as we show below, they can be even longer than the software processing latencies) and need to be accurately taken into account and reported.

In order to truly know the total system latency that accounts for both the hardware latencies and the Audapter (or other software) latency, the only option is to simultaneously record—and then measure the temporal asynchrony between—the (a) original microphone signal *before* it is digitized by the audio interface and (b) the processed signal to the headphones *after* it has been output by the audio interface. To date, there is no published documentation of such correctly measured latencies for set-ups where Audapter is linked with any of the commonly used audio interfaces. In the present study, we therefore investigated for two different computers and five different ASIO compatible audio interfaces the devices' own intrinsic round-trip latency (RTL), the Audapter processing time, and the total system input-output latency when implementing a typical formant-shift perturbation. We also tested non-default Audapter parameter values as a strategy to optimize feedback latency. The data provide realistic measures of the overall feedback delay in typical Audapter set-ups, and demonstrate that some audio interfaces and software options perform significantly better in terms of avoiding unwanted feedback delays.

Method

Instrumentation

Computers.

We ran the Audapter software on two different computers (referred to as operating computers to distinguish them from a simultaneously used recording computer) in conjunction with the different audio interfaces in order to verify whether processing and overall latency differ depending on computing power. Given that Audapter is often used with laptop-based systems for portability, we first tested a Dell laptop (Latitude E5570) with Intel core i7-6820HQ processor, 16 GB RAM memory, and Windows 7 Professional operating system. The laptop ran on external power and in high performance mode. The second operating computer was a Dell workstation (Precision Tower 7510) with Intel Xeon E5-2640 v4 processor, 32 GB RAM, and Windows 7 Professional operating system. During the tests, no other computer programs were open except for the Sophos anti-virus program that ran in the background as required by the University of Washington.

We used an additional Dell laptop (Dell Inspiron 14R-N4010) as the recording computer. This recording computer's stereo line input was used to record on separate channels (a) the direct microphone signal (i.e., prior to reaching the audio interface and Audapter) and (b) the final output signal (i.e., after being input to the audio interface, processed in Audapter, and output from the audio interface). This recording laptop was also connected to external power and also operated in high performance mode. Offline, the latency between the recorded stereo channels was measured using Praat, a free software for analyzing speech sounds (Boersma & Weenink, 2019).

Audio interfaces.

For real-time processing, Audapter requires an ASIO compatible audio interface. The interface most extensively tested by the developers of Audapter, the MOTU MicroBook (Cai, 2014) is now outdated. We therefore included in our tests a more recent model of the MicroBook series, the MOTU MicroBook IIc. Another interface mentioned in the Audapter manual is the MOTU Ultralite, but this model is also outdated now. Thus, again, we included in our testing a more recent model of the Ultralite series, the MOTU Ultralite AVB. In addition, we also tested our own lab's preferred interface, the RME Babyface Pro, as well as two more popular interfaces, the Presonus Studio 192 Mobile and Tascam US 2x2. All tested interfaces are portable (as opposed to rack-mounted).

All audio interfaces were set up to use a sampling rate of 48kHz and a buffer size (or frame size) of 96 or 128 samples, depending on which options were supported by the device. The MOTU MicroBook IIc and RME Babyface Pro support a buffer size of 96 samples (which is used in examples in the Audapter manual), thus the smaller buffer size was used for these two interfaces. The other interfaces did not support a buffer size of 96 samples; only 64 and 128 samples. According to our tests, a buffer size of 64 samples is too short for the present purposes as the 1.3 ms frame length (64 buffer size/48 kHz before downsampling) resulted in unreliable formant tracking. Therefore, for those interfaces that did not support buffer size 96, we tested the next shortest buffer size of 128 samples. Given that the RME Babyface Pro supports both buffer sizes 96 and 128, an additional test was completed for this device with the buffer size set to this greater value of 128 samples.

Audio drivers and software setting.

In addition to each manufacturer's default ASIO driver for the operating system, we also installed ASIO4ALL, a hardware-independent low latency ASIO driver. This allowed us to test whether there is any advantage to using the manufacturers' own default drivers.

Unless specified otherwise[1], the Audapter software itself was used with its default parameters from getAudapterDefaultParams.m as included in the software package (Cai et al., 2008; Tourville et al., 2013). We implemented an auditory perturbation that consisted of upward shifts of Formant 1 (F1) and Formant 2 (F2). To achieve this perturbation, we set *bRatioShift* = 1 for a ratio shift, all 257 values of *pertPhi* = π/4 radians for a 45 degree shift in the F1-by-F2 vowel space, and all 257 values of *pertAmp* = 0.2 for a 20% shift in that direction (thus, a 20% formant frequency upshift with a 45 degree direction in the vowel space). Consequently, both F1 and F2 received a 14% upshift in Hz (sin(π/4 radians) × 20%) which corresponds to 229 cents ($1200 \times \log_2(1.14/1)$).

The downsampling rate parameter, *downFact*, was set to 3 (downsampling the original 48 kHz signal to 16 kHz) for most of the tests with interfaces that used a buffer size of 96 samples , as suggested in the Audapter manual (Cai, 2014). An exception was a separate test conducted to compare the effect on processing latency of setting this parameter to 3 vs. 4 (i.e., increased downsampling). For interfaces with a buffer size of 128 samples, we always used *downFact* 4 (downsampling 48 kHz to 12 kHz). Audapter was operated with Audapter('start') and Audapter('stop') functions to start and stop the auditory perturbations. It was operated for 5 seconds before it was stopped. Audapter recordings were retrieved via the AudapterIO('getdata') command and saved into wav files, resampling to 44.1 kHz.

Users have the option of shortening Audapter's software processing latency by lowering the *nDelay* value. Given that only odd positive integers are accepted, we tested Audapter's processing latency with *nDelay* values of 1, 3, and 5 in addition to the default value of 7. With

---

[1] In a later section, we describe how the *nDelay* parameter was also deliberately manipulated in order to test Audapter's formant shift capability with a shorter processing delay.

nDelay of 1, reliable formant tracking was not possible, so only the results from *nDelay* = 3 and *nDelay* = 5 are reported and compared with *nDelay* = 7. According to the manual (Cai, 2014), the *nDelay* parameter determines the length of the internal processing window used by Audapter ($\text{internal processing window} = 2\ \times nDelay - 1$) for formant tracking. Audapter then filters the $(nDelay)th$ frame to perturb the formant frequencies by the desired amount (chosen by the user), and sends this altered audio signal to the output channel. Hence, in the software, the resulting temporal asynchrony between the input and output channel becomes: $2\ \times nDelay - 1 - nDelay =\ nDelay - 1$ frames. For the default *nDelay* value of 7, there would be a latency of 6 frames or 12 ms ($2\text{ ms }[96\text{ buffer}/48\text{ kHz}]\ \times\ 6\text{ frames}$) (Cai, 2014).[2] We emphasize again that this is software latency[3] only, without taking account of the additional hardware latencies, and thus not an accurate representation of the total feedback loop delay.

It should be noted that Audapter's processing latency is determined by different parameters depending on which type of auditory perturbation is being applied. This report tested the latencies during formant-shift perturbations, and thus, the results are only directly applicable to this type of experimental manipulation. Nevertheless, the procedures that we describe for measuring the latencies accurately, and most of the resulting recommendations for minimizing those latencies, are generalizable to other types of perturbations (such as pitch shifts).

Testing Procedure

Hardware latency

We used the RTL Utility software tool (Oblique Audio, www.oblique-audio.com) to

---

[2] This assumes an original sampling rate of 48 kHz and a buffer size of 96 samples, making each frame size 2 ms.

[3] A distinction could be made between latency that is due to software design (e.g., placing samples in a buffer) vs. actual processing (e.g., doing spectral calculations and manipulations, which could depend on computer processor speed and the specific type of signal manipulation). However, results presented later in this paper demonstrate that, in Audapter, the latency determined by *nDelay* encompasses any separate processing latency. Hence, we use the terms total software latency and software processing latency interchangeably.

measure round trip latency (input latency plus output latency) of the audio interfaces. This is accomplished by looping the audio interface's output back to its own input (Figure 1). When the RTL Utility program causes the interface to generate an audio signal, this signal goes through the output component of the interface and then comes back through the input component of interface to reach the computer. The RTL Utility measures the latency between the two events (sending and receiving the signal) to determine the round trip latency as this measurement reflects the sum of the input and output latencies. We refer to this round trip latency also as the hardware latency. For the present tests, we ran the RTL Utility five times for each individual interface.

Total feedback loop latency

In order to measure the *total feedback loop latency* that occurs in a speech experiment, we needed to simultaneously record the unperturbed subject ouput (i.e., the original microphone signal) and the perturbed feedback signal (i.e., the formant-shifted signal routed to the subject's earphones). To record the first of these two signals, we split the signal from the microphone (AKG C544, Shure) with an XLR splitter such that both the tested USB audio interface and one audio channel of the recording laptop received the exact same microphone signal at the same time. To record the formant-shifted output, the signal that had been processed by Audapter and that was output by the audio interface was routed to a mixer (MMX-24, Monacor) in order to be amplified and sent to the second channel of the recording laptop's stereo line-in input (Figure 2). The two channels of data were saved on the recording computer using the Praat software. Recordings were made while a male speaker's production of the word *"tuck"* was played back from a loudspeaker at least 5 times for each test. We then measured, from the Praat recordings, the temporal asynchrony between the original, unperturbed microphone signal and the perturbed feedback signal for five trials.

Audapter software processing latency

A first, “indirect” method of determining the *Audapter software processing latency* was based on the fact that additional latencies due to mixer and audio cables were negligible (less than 0.05 ms). Thus, any part of the total feedback loop latency not caused by the interface hardware can be attributed to Audapter processing. Consequently, it was possible to subtract the measured hardware latencies from the overall feedback loop latencies to obtain an estimate of Audapter’s signal processing time. As a second, “direct” method of determining *Audapter’s software processing latency*, we also compared the temporal asynchrony between the two channels (input and output corresponding to unperturbed and perturbed signals) stored within Audapter’s own recording on the operating computer. Using Praat to analyze the recordings from five trials, we measured the time difference between the two Audapter channels. These measures of the time offset between the unperturbed and perturbed signals stored by Audapter (i.e., this second, direct method) can then be compared with the estimates obtained by subtracting hardware latency from total feedback loop latency (i.e., the first, indirect method).

Effect of Audapter’s *downFact* parameter

In Audapter, the default value of *downFact* 3 downsamples the original audio input from sampling rate 48 kHz and buffer size 96 samples to sampling rate 16 kHz and buffer size 32 samples. However, some researchers have used *downFact* 4 to downsample the original signal to a rate of 12 kHz and buffer size of 24 samples (e.g., Daliri et al., 2018). Given that Audapter does allow users to specify this amount of downsampling in order to lower computational load, we tested whether or not the amount of downsampling affects processing latency. For this purpose, we used the RME Babyface Pro audio interface and both the laptop and desktop operating computers with *downFact* was set to either 3 or 4.

Effect of Audapter's *nDelay* parameter

As mentioned above, we also aimed to examine whether Audapter's processing time can be shortened by lowering the *nDelay* parameter from its default value of 7 to 5 or 3. Caution is warranted, however, given that a shorter processing window could degrade the accuracy of the formant tracking and, thus, also the intended formant perturbation. We therefore completed a more extensive test to verify formant tracking accuracy for each of the three *nDelay* values 7, 5, and 3. For this purpose, we used a data set consisting of 30 repetitions of the words *"tech," "tuck,"* and *"talk"* spoken by an adult male speaker. The set of 90 trials was played back three times at identical volume from a computer loudspeaker (AX210, Dell) into a microphone (SM58, Shure) connected to the above described RME Babyface Pro audio interface. For each run of 90 trials, Audapter processed the incoming signal and tracked the formants at a given *nDelay* value. Offline, F1 and F2 were calculated for each trial as the average value across the middle 40-60% of the corresponding formant track.

Effect of Windows operating system version installed on the operating computer

As mentioned above, testing was generally completed with both the laptop and desktop operating computers running Audapter software within the Windows 7 Professional operating system. However, given that Microsoft will no longer provide support for Windows 7 starting in 2020, we decided to also do a test of Windows 7 vs. Windows 10 on one of these operating computers. Therefore, after completing all the above described tests, we upgraded the Dell Latitude E5570 machine's operating system to Windows 10 Enterprise LTSC (Long Term Service Channel), and we repeated a set of total latency measurements with the RME Babyface Pro interface and the Audapter parameter *downFact* set to 3. As we had done in Windows 7, we ran this set of tests again with each of the three *nDelay* values 7, 5, and 3.

Results

Hardware Latency

Hardware-related round trip latencies for all tested combinations of audio interfaces and operating computers are listed in Table 1. First, it is clear that use of the ASIO4ALL driver should be avoided. In all cases, use of the audio interface manufacturer's own driver resulted in much shorter hardware latencies than the ASIO4ALL driver. For example, the two fastest devices (RME Babyface Pro and MOTU Ultralite AVB with the HSO (Host Safety Offset) setting at 16) showed hardware latencies that were ~8 ms shorter when operated with the default driver. In addition, hardware latency for the fastest device (RME Babyface Pro) was much more consistent with the default driver (the standard deviation across repeated runs with a buffer size of 96 samples was 0.001 ms) than with the ASIO4ALL driver (standard deviation > 1.5 ms). Both these results suggest significant compatibility issues between the ASIO4ALL driver and some audio interfaces. We will therefore focus our remaining description of the test results on those obtained with the manufacturers' own drivers.

It is also clear from Table 1 that the RME Babyface Pro outperformed all other tested interfaces. The Babyface Pro had the shortest latency regardless of whether it was tested with buffer size 96 or 128 samples and regardless of which operating computer was used. The advantage of this RME Babyface Pro over all other interfaces is substantial and not only due to its ability to use a smaller buffer size (96 vs. 128 samples). Even when tested with the same 128 buffer size used by most other interfaces, the RME Babyface Pro was still ~30% faster than the second best performer, the MOTU Ultralite AVB (6.65 ms vs. 9.54 ms). When leveraging the RME Babyface Pro's ability to operate with a 96 buffer size, its latency advantage over the second best performer was as much as ~45% (5.30 ms vs. 9.54 ms). For all devices, hardware latency was essentially the same when tested with the laptop operating computer and the more powerful workstation desktop computer. Lastly, it is noteworthy that the MOTU Ultralite AVB dropped from second fastest to slowest when the HSO setting was at 128 (14.2 ms).

Total Feedback Loop Latency

The middle column in Table 2 lists the total feedback loop latencies (i.e., hardware latency plus Audapter signal processing latency) as measured with the procedure displayed in Figure 2. Given the above reported finding that the manufacturers' drivers performed better than ASIO4ALL, Table 2 includes only total feedback loop latencies obtained with the manufacturers' own drivers (except for the Tascam US 2x2 interface whose driver was not compatible with Audapter). As expected based on our finding that hardware latencies (Table 1) were shorter for devices with a buffer size of 96 samples (and, in fact, the same was true also for software latencies discussed below and listed in Table 3), total feedback loop latencies were better for devices with a buffer size of 96 samples.

The most important observation in these results is that the total feedback loop latencies observed in this study (from ~19 ms with the fastest device to ~39 ms with the slowest device; ~27 ms for the popular MOTU MicroBook IIc device) are substantially larger than claimed in most published work in which the Audapter software was used (as mentioned in the Introduction, the vast majority of published studies have claimed a feedback delay of only 10-15 ms). It seems clear that the previously reported latencies took into account only the software-based signal processing latency (which is the only latency accounted for when comparing the unperturbed and perturbed signals stored by the Audapter software) while ignoring the sometimes substantial additional delays associated with the audio interface's signal input and output components.

With the appropriate measurement protocol, the overall best total feedback loop latency (19.3 ms) was observed when using the RME Babyface Pro with a buffer size of 96 samples. The second best performance (25.5 ms) was observed when the same device was used with a buffer size of 128 samples.

Audapter Software Processing Latency (indirect method)

The right-side column of Table 2 summarizes the estimates of *Audapter software processing latency* that were obtained by subtracting hardware latencies from overall latencies. Results suggest that the software processing latency was ~14 ms for a buffer size of 96 samples and ~18.9 ms for a buffer size of 128 samples. Interestingly, both these values are slightly larger (2 ms in the case of the 96 buffer size and 2.7 ms for buffer size 128, values corresponding to the respective frame sizes) than what was expected based on the Audapter manual (Cai, 2014) and the results of the direct method discussed below.

Audapter Software Processing Latency (direct method)

Table 3 includes results for *Audapter software processing latency* when measured directly as the temporal offset between the unperturbed and perturbed channels stored by the Audapter software itself. Our measurements of this offset are very consistent: 12 ms for the tests with a buffer size of 96 samples (for which the downsampling factor was set to 3) and 16 ms for the tests with a buffer size of 128 samples (for which the downsampling factor was set to 4).

These estimates closely match the processing latencies that one would predict based on the algorithm information provided in the Audapter manual (Cai, 2014). For buffer size 96, we predicted a processing latency of 12 ms (2 ms [96 buffer/48 kHz] × 6 input frames per input frame = 12 ms). For buffer size 128, we predicted a processing latency of 16.2 ms (2.7 ms [128 buffer/48 kHz] × 6 input frames = 16.2 ms). Importantly, the shorter processing latency for buffer size 96 vs. 128 again gives an advantage to the RME Babyface Pro device. Although processing latency was essentially the same for the MOTU Microbook IIc, the latter interface's hardware in/out latency (Table 1) was more than twice as long as that of the Babyface Pro.

As one would expect, Audapter's software processing latency was not affected by selecting the manufacturer vs. ASIO4ALL driver. Similarly, it was not affected by the use of any specific audio interface (when compared across interfaces operating with the same buffer size). And lastly, based on the two operating computers used in our testing, running Audapter on a more powerful computer (with faster processor as well as more RAM) also did not affect software processing latency.

Effect of Audapter's parameter ***downFact***

Audapter offers the option of downsampling the input signal before processing in order to lower computational load. Results for our comparison of latencies obtained with *downFact* 3 vs. 4 are listed in Table 4. The total feedback loop latency did not improve with more downsampling, and this was true for both the laptop and desktop workstation operating computers.

Effect of Audapter's parameter ***nDelay***

We tested whether reducing Audapter's *nDelay* parameter from its default value of 7 to either 5 or 3 can be used as a way to further minimize software processing latency (and thus also total feedback loop latency) without sacrificing accuracy of the formant shift perturbation. Table 5 shows that, in comparison with the default setting, an additional 4 ms processing time can be save by switching to *nDelay* = 5, and 8 ms processing time can be saved with *nDelay* = 3. With the latter value, Audapter processing latency was reduced from 14 or 12 ms (for the direct and indirect measurement methods, respectively) to only 6 or 4 ms. Used in combination with the

RME Babyface Pro hardware, this strategy is capable of reducing the total feedback loop latency from 19.4 to 11.4 ms—thereby achieving a total feedback delay as short as that provided by the commercially produced VoiceOne processor (see above).

In theory, the short processing window resulting from setting *nDelay* as low as 3 could degrade the accuracy of Audapter's formant tracking (thus affecting any applied formant perturbation). We therefore verified formant tracking accuracy for each of the three *nDelay* values 7, 5, and 3. Figure 3 shows example spectrograms with overlaid formant tracks for each *nDelay* setting (top row), each individual trial's first (F1) and second (F2) formant values determined with each *nDelay* setting (middle row), and trial-by-trial differences in these formant values across the different *nDelay* settings (bottom row). Small differences in formant tracking can be observed, most notably when comparing the different measures on a per-trial basis (bottom row of Figure 3). For example, *nDelay 5* was associated with slightly higher F1 values, especially in comparison with *nDelay 7*. However, formant tracking for the entire data set of 90 trials was generally very similar across the different *nDelay* parameters, and there is no indication that results for *nDelay* 3 are inferior to those for *nDelay* 7.

Effect of Windows operating system version installed on the operating computer

Lastly, we examined whether different latencies are obtained when running Audapter within the newer Windows 10 operating system as opposed to the Windows 7 operating system used for all tests reported above. A direct comparison between latencies obtained with the two different operating systems installed on the same Dell Latitude E5570 operating computer is presented in Table 6. The total feedback loop latencies obtained when running Audapter with each of the three different *nDelay* values did not differ between the two operating systems.

Discussion

An increasing body of work on laryngeal and supralaryngeal aspects of speech production makes use of "real-time" auditory feedback perturbations. To date, little attention has been paid to quantifying the hardware and software latencies involved in the instrumentation set-up for such experiments. Quantifying these latencies is critical, however, as it is known that delays in the auditory feedback loop may affect both online control and auditory-motor learning of speech (Max & Maffett, 2015; Mitsuya et al., 2017; Stuart et al., 2002). Here, we investigated hardware latency, software processing latency, total feedback loop latency, and software parameter optimization strategies for a widely used set-up combining the Audapter software package (Cai, 2014) with different brands and models of commercially available audio interfaces.

First, it is clear from our tests that the choice of software driver is important. In all cases, use of the audio interface manufacturer's own proprietary driver resulted in shorter and more consistent latencies as compared with the generic ASIO4ALL driver. For our fastest audio interface, the RME Babyface Pro, using its default drivers as opposed to the ASIO4ALL driver eliminated more than 5 ms of latency. Consequently, at least for the audio interfaces tested in this study, each interface's proprietary driver should be used for all speech experiments.

Second, despite this being ignored in most published work to date, there are substantial delays associated with the input and output components of the audio interfaces themselves, even in the absence of any software processing. In our tests with the manufacturers' faster proprietary drivers, these hardware-induced latencies ranged between ~5 ms and ~15 ms across the various audio interfaces included here. Our results very closely match the latencies reported in online forums for and by audio gear enthusiasts. For example, our 5.299 ms latency for the RME

Babyface Pro interface operated with a buffer size of 96 samples at 48 kHz is very similar to individual user reports available online (ModMeister, 2018). It needs to be kept in mind that, in speech experiments, these hardware latencies additively combine with software processing latencies. Thus, the additional hardware-specific latency of 5-15 ms cannot be considered negligible. To the contrary, it should become standard practice for all auditory feedback perturbation experiments to report the *total feedback loop latency* that was in effect during data collection. Only with the latter information available can readers make informed decisions about the potential impact of the auditory feedback delay experienced by the participants.

Third, to promote widespread efforts toward this goal, we have described and illustrated above how the *total feedback loop latency* can be determined by comparing the time difference between a processed signal output from the audio interface (i.e., the perturbed signal) and the original unprocessed microphone signal (i.e., the unperturbed signal). With default Audapter software settings (options for further optimization of these software settings are discussed below), our measurements of total feedback loop latency ranged from 19 ms to 39 ms across the different interfaces. These delays are considerably longer than most of the latencies previously reported in publications based on speech experiments with Audapter. Indeed, it seems clear that many prior reports of feedback latency in speech experiments with Audapter failed to capture the additional, hardware-specific latency. This may happen if researchers compare the perturbed channel with an unperturbed channel that is also taken from the output of the audio interface (Figure 4) rather than directly from the microphone. In that case, both channels include the audio interface’s hardware latency (because even the unperturbed channel was first A/D-converted by the interface, passed to the computer, and then received back from the computer), and the temporal difference between these channels does *not* reflect the true latency between a human subject's speech output and the altered auditory feedback signal. Instead, it reflects only the software latency, and, thus, underestimates the total latency. Hence, we emphasize here once again the importance of measuring *total feedback loop latency* as shown in Figure 2.[4]

Fourth, comparing the various audio interfaces tested here, we found that the RME Babyface Pro has the shortest latency. In addition, this RME Babyface Pro supports a buffer size of 96 samples which shortens Audapter's software processing latency as compared with situations in which a buffer size of 128 samples needs to be used. Thanks to these advantages, our formant-shift tests showed a total feedback loop latency of 19.4 ms when using the RME Babyface Pro with Audapter (assuming default settings; down to 11.4 ms with parameter optimization as described below). In contrast, when using Audapter with some of the other interfaces, latencies approached 30 ms or, when the ASIO4ALL driver was used, even 40 ms. It has been reported that a delay greater than 25 ms can already slow down the speech rate and a delay greater than 50 ms may induce disfluencies (Stuart et al., 2002). In addition, our ongoing work suggests that auditory-motor learning in clinical populations may be differentially affected by feedback latency. It is therefore critical to run Audapter with an audio interface that has a short intrinsic latency (i.e., round trip latency).

MOTU Ultralite AVB with the HSO setting of 16 had the second shortest latency after RME Babyface Pro. However, whether or not this HSOs setting affects Audapter's performance is not known and must be examined. The manufacturer states that host software may experience performance issues when the parameter is set too low. Hence, it is possible that the low HSO

---

[4] Of course, if a given auditory feedback perturbation system has *additional* delay before the perturbation is actually implemented after onset of the feedback signal, then researchers should report both the signal latency as discussed in the present paper and the additional latency until the perturbation effectively appears in that signal. In Audapter, even the first sample of the feedback signal has gone through all processing steps. Thus, there is no additional perturbation latency beyond the feedback signal latency.

setting may compromise the quality of auditory feedback perturbations due to formant mistracking. We recommend that researchers interested in using this particular interface test it in depth before using it with the Audapter software. Further, although it is theoretically possible to shorten the hardware latency by using rack-mounted systems (e.g., RME Fireface UFX) or internal PCI-e cards (e.g., RME HDSPe AES or Lynx AES 16e/Aurora 16), various databases informing on the round trip latency of different interfaces indicate that any additional improvement would be marginal (i.e., < 1 ms) as compared with the RME Babyface Pro tested here (Tafkat, 2018).

Fifth, our results suggest caution when estimating Audapter's own signal processing latency. In the absence of additional latencies beyond the audio interface and Audapter software, this signal processing latency can be estimated by subtracting the hardware latency from the total latency. In our tests, this indirect estimate always slightly exceeded the latency measure obtained by measuring the temporal offset between the unperturbed and perturbed channels saved by Audapter as well as the prediction based on the software's *nDelay* parameter. Although the reasons for this discrepancy remain unknown, we strongly advise against using Audapter's own recordings for latency measures. They certainly do not take account of the audio interface's intrinsic latency, and may not even capture the software processing latency accurately.

Sixth, in terms of trying to further optimize performance, we found that operating the interfaces and software with a more powerful computer did not yield an actual benefit. There was no difference in the total feedback loop latency across the tested operating machines (Dell Latitude laptop vs. Dell Precision desktop). Similarly, upgrading the Dell Latitude laptop's operating system from Windows 7 Professional to Windows 10 Enterprise LTSC did not affect total feedback loop latency. We also did not see any improvement in total latency when using a higher downsampling rate (e.g., Audapter parameter *downFact*) as a strategy to lower computational load as suggested in the Audapter manual (Cai, 2014). It is possible, of course, that limitations associated with the older computer technology that was available when Audapter was initially developed caused higher downsampling rates to offer significant improvements. However, most computers in use today should be adequately powerful for this aspect to be a non-issue. In contrast, our tests showed that lowering Audapter's *nDelay* parameter from the default value of 7 to 3 did substantially shorten the software processing latency. Specifically, this latency improved by 8 ms as compared with the default setting. As a result, this test situation where we implemented our formant-shift perturbation by using the REM Babyface Pro interface in combination with Audapter *nDelay* 3 was the only instance in which we were able to achieve a total feedback loop latency of only 11.4 ms. Thus, shortening Audapter's processing latency by lowering the *nDelay* parameter is recommended.

Concluding Remarks

Audapter is a software package that is considered to be the most comprehensive and flexible tool for perturbing auditory feedback. Most likely, it will continue to be heavily used in future studies of voice and speech motor control (Tourville et al., 2013). It is crucial, however, for researchers to be cognizant of the fact that the reproducibility of results from such studies depends on accurately measuring and reporting the overall feedback loop latency that was in effect during the experiment. As described above, this reported latency should take account of both the audio interface hardware latency and the software signal processing latency. In addition, researchers should be aware of currently available options to minimize overall feedback delays through the selection of short-latency hardware (e.g., RME Babyface Pro audio interface) and software parameters (e.g., Audapter's *nDelay*).

Undoubtedly, additional means of optimizing hardware and software components need to be investigated. For example, new technologies are available for the communication between audio interfaces and the operating computer (i.e., USB-C, Thunderbolt 3, etc.). In fact, in our tests, the Presonus 192 Studio Mobile interface used a USB 3.0 connection, but it was considerably slower than the RME Babyface Pro interface which used a USB 2.0 connection. Nevertheless, as manufacturers introduce interfaces with more advanced connections, the effect on device latency should be examined. Another potential influence may relate to the use of Windows vs. Mac vs. Linux as the computer operating system (depending on availability of native drivers for the audio interface), and it should be recalled that all tests reported here were conducted only with Windows computers.

Lastly, different types of perturbation in Audapter depend on different parameters, and each of these parameters may have its own specific effect on processing delay. For example, analogous to our above manipulation of the *nDelay* parameter to shorten latency in formant-shift experiments, the *delayFrames* parameter could be used to shorten latency for pitch-shift perturbations (Cai, 2014). Investigating all additional perturbations that are possible with Audapter (e.g., pitch-shift, delay, time-warping) was well beyond the scope of the present work, but researchers using those feedback manipulations are strongly encouraged to apply the measurements described in this work in order to report accurate feedback loop latencies.

Acknowledgment

The authors thank Shanqing Cai for assistance with back-end information on Audapter and its various parameters. This research was supported in part by grants from the National Institute on Deafness and Other Communication Disorders (R01DC014510 and R01DC17444) and the Canadian Institutes of Health Research (MOP-137001). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, or the Canadian Institutes of Health Research.

References

Abur, D., Lester-Smith, R. A., Daliri, A., Lupiani, A. A., Guenther, F. H., & Stepp, C. E. (2018). Sensorimotor Adaptation of Voice Fundamental Frequency in Parkinson's Disease. *PloS One, 13*(1), e0191839. https://doi.org/10.1371/journal.pone.0191839

Ballard, K. J., Halaki, M., Sowman, P., Kha, A., Daliri, A., Robin, D. A., Tourville, J. A., & Guenther, F. H. (2018). An Investigation of Compensation and Adaptation to Auditory Perturbations in Individuals with Acquired Apraxia of Speech. *Frontiers in Human Neuroscience, 12*, 510. https://doi.org/10.3389/fnhum.2018.00510

Bauer, J. J., & Larson, C. R. (2003). Audio-Vocal Responses to Repetitive Pitch-Shift Stimulation during a Sustained Vocalization: Improvements in Methodology for the Pitch-Shifting Technique. *The Journal of the Acoustical Society of America, 114*(2), 1048-1054. https://doi.org/10.1121/1.1592161

Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. R. (2009). Vocalization-Induced Enhancement of the Auditory Cortex Responsiveness during Voice F0 Feedback Perturbation. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology, 120*(7), 1303-1312. https://doi.org/10.1016/j.clinph.2009.04.022

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [computer software]. http://www.praat.org/

Cai, S. (2014). *A Manual of Audapter.* http://sites.bu.edu/guentherlab/files/2016/11/AudapterManual.pdf

Cai, S., Beal, D. S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2014). Impaired Timing Adjustments in Response to Time-Varying Auditory Perturbation during Connected Speech Production in Persons Who Stutter. *Brain and Language, 129*, 24-29. https://doi.org/10.1016/j.bandl.2014.01.002

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak Responses to Auditory Feedback Perturbation during Articulation in Persons Who Stutter: Evidence for Abnormal Auditory-Motor Transformation. *PloS One, 7*(7), e41830. https://doi.org/10.1371/journal.pone.0041830

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S.A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/. Paper presented at the *The 8th Intl. Seminar on Speech Production,* 65-68.

Cai, S., Ghosh, S., Guenther, F., & Perkell, J. (2010). Adaptive Auditory Feedback Control of the Production of Formant Trajectories in the Mandarin Triphthong /Iau/ and its Pattern of Generalization. *The Journal of the Acoustical Society of America, 128*(4), 2033-2048. https://doi.org/10.1121/1.3479539

Caudrelier, T., & Rochet-Capellan, A. (2019). Changes in speech production in response to formant perturbations: An overview of two decades of research. In S. Fuchs, J. Cleland & A. Rochet-Capellan (Eds.), *Speech Production and Perception: Learning and Memory* (pp. 11-75). Peter Lang.

Daliri, A., & Dittman, J. (2019). Successful Auditory Motor Adaptation Requires Task-Relevant Auditory Errors. *Journal of Neurophysiology, 122*(2), 552-562. https://doi.org/10.1152/jn.00662.2018

Daliri, A., & Max, L. (2018). Stuttering Adults' Lack of Pre-Speech Auditory Modulation Normalizes when Speaking with Delayed Auditory Feedback. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 99*, 55-68. https://doi.org/10.1016/j.cortex.2017.10.019

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. (2018). Auditory-Motor Adaptation is Reduced in Adults Who Stutter but Not in Children Who Stutter. *Developmental Science, 21*(2)https://doi.org/10.1111/desc.12521

Feng, Y., Gracco, V., & Max, L. (2011). Integration of Auditory and Somatosensory Error Signals in the Neural Control of Speech Movements. *Journal of Neurophysiology, 106*(2), 667-679. https://doi.org/10.1152/jn.00638.2010

Feng, Y., Xiao, Y., Yan, Y., Max, L., & Max, L. (2018). Adaptation in Mandarin Tone Production with Pitch-Shifted Auditory Feedback: Influence of Tonal Contrast Requirements. *Language, Cognition and Neuroscience, 33*(6), 734-749. https://doi.org/10.1080/23273798.2017.1421317

Franken, M., Acheson, D., McQueen, J., Hagoort, P., & Eisner, F. (2018). Opposing and Following Responses in Sensorimotor Speech Control: Why Responses Go both Ways. *Psychonomic Bulletin & Review, 25*(4), 1458-1467. https://doi.org/10.3758/s13423-018-1494-x

Franken, M., Eisner, F., Acheson, D. J., McQueen, J. M., Hagoort, P., & Schoffelen, J. (2018). Self-Monitoring in the Cerebral Cortex: Neural Responses to Small Pitch Shifts in Auditory Feedback during Speech Production. *NeuroImage, 179*, 326-336. https://doi.org/10.1016/j.neuroimage.2018.06.061

Heller Murray, E. S., Lupiani, A. A., Kolin, K. R., Segina, R. K., & Stepp, C. E. (2019). Pitch Shifting with the Commercially Available Eventide Eclipse: Intended and Unintended Changes to the Speech Signal. *Journal of Speech, Language, and Hearing Research: JSLHR, 62*(7), 2270-2279. https://doi.org/10.1044/2019_JSLHR-S-18-0408

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor Adaptation in Speech Production. *Science (New York, N.Y.), 279*(5354), 1213-1216. https://doi.org/10.1126/science.279.5354.1213

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech I: Compensation and Adaptation. *Journal of Speech, Language, and Hearing Research : JSLHR, 45*(2), 295-310. https://doi.org/10.1044/1092-4388(2002/023)

Jones, J. A., & Keough, D. (2008). Auditory-Motor Mapping for Pitch Control in Singers and Nonsingers. *Experimental Brain Research, 190*(3), 279-287. https://doi.org/10.1007/s00221-008-1473-y

Keough, D., Hawco, C., & Jones, J. A. (2013). Auditory-Motor Adaptation to Frequency-Altered Auditory Feedback Occurs when Participants Ignore Feedback. *BMC Neuroscience, 14*, 25. https://doi.org/10.1186/1471-2202-14-25

Klein, E., Brunner, J., & Hoole, P. (2018). Which Factors Can Explain Individual Outcome Differences When Learning a New Articulatory-to-Acoustic Mapping? In Q. Fang, J. Dang, P. Perrier, J. Wei, L. Wang & N. Yan (Eds.), *Studies on Speech Production* (pp. 158-172). Springer International Publishing.

Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014). Brief Periods of Auditory Perceptual Training can Determine the Sensory Targets of Speech Motor Learning. *Psychological Science, 25*(7), 1325-1336. https://doi.org/10.1177/0956797614529978

Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust Sensorimotor Learning during Variable Sentence-Level Speech. *Current Biology : CB, 28*(19), 3106-3113.e2. https://doi.org/10.1016/j.cub.2018.07.030

Loucks, T., Chon, H., & Han, W. (2012). Audiovocal Integration in Adults Who Stutter. *International Journal of Language & Communication Disorders, 47*(4), 451-456. https://doi.org/10.1111/j.1460-6984.2011.00111.x

Max, L., & Maffett, D. G. (2015). Feedback Delays Eliminate Auditory-Motor Learning in Speech Production. *Neuroscience Letters, 591*, 25-29. https://doi.org/10.1016/j.neulet.2015.02.012

Max, L., Wallace, M. E., & Vincent, I.Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. Paper presented at the *International Congress of Phonetic Sciences,* 1053-1056.

Mitsuya, T., Munhall, K. G., & Purcell, D. W. (2017). Modulation of Auditory-Motor Learning in Response to Formant Perturbation as a Function of Delayed Auditory Feedback. *The Journal of the Acoustical Society of America, 141*(4), 2758. https://doi.org/10.1121/1.4981139

ModMeister. (2018). *What Latency are You Getting with Your USB-Capable Modeller Or Audio Interface? (+ My Own Results)* . https://www.thegearpage.net/board/index.php?posts/26980846/

Mollaei, F., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2016). Sensorimotor Control of Vocal Pitch and Formant Frequencies in Parkinson's Disease. *Brain Research, 1646*, 269-277. https://doi.org/10.1016/j.brainres.2016.06.013

Mollaei, F., Shiller, D. M., & Gracco, V. L. (2013). Sensorimotor Adaptation of Speech in Parkinson's Disease. *Movement Disorders : Official Journal of the Movement Disorder Society, 28*(12), 1668-1674. https://doi.org/10.1002/mds.25588

Purcell, D. W., & Munhall, K. G. (2006). Compensation Following Real-Time Manipulation of Formants in Isolated Vowels. *The Journal of the Acoustical Society of America, 119*(4), 2288-2297.

Reilly, K. J., & Pettibone, C. (2017). Vowel Generalization and its Relation to Adaptation during Perturbations of Auditory Feedback. *Journal of Neurophysiology, 118*(5), 2925-2934. https://doi.org/10.1152/jn.00702.2016

Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous Acquisition of Multiple Auditory-Motor Transformations in Speech. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 31*(7), 2657-2662. https://doi.org/10.1523/JNEUROSCI.6020-10.2011

Sares, A. G., Deroche, M. L. D., Shiller, D. M., & Gracco, V. L. (2018). Timing Variability of Sensorimotor Integration during Vocalization in Individuals Who Stutter. *Scientific Reports, 8*(1), 16340-1. https://doi.org/10.1038/s41598-018-34517-1

Sato, M., & Shiller, D. M. (2018). Auditory Prediction during Speaking and Listening. *Brain and Language, 187*, 92-103. https://doi.org/10.1016/j.bandl.2018.01.008

Stepp, C. E., Lester-Smith, R. A., Abur, D., Daliri, A., Pieter Noordzij, J., & Lupiani, A. A. (2017). Evidence for Auditory-Motor Impairment in Individuals with Hyperfunctional Voice Disorders. *Journal of Speech, Language, and Hearing Research : JSLHR, 60*(6), 1545-1550. https://doi.org/10.1044/2017_JSLHR-S-16-0282

Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of Delayed Auditory Feedback on Normal Speakers at Two Speech Rates. *The Journal of the Acoustical Society of America, 111*(5 Pt 1), 2237-2241. https://doi.org/10.1121/1.1466868

Tafkat. (2018). *LLP Database Update: June 2018.* https://www.gearslutz.com/board/showpost.php?p=13368351&postcount=2186

Tourville, J. A., Cai, S., & Guenther, F. (2013). Exploring Auditory-Motor Interactions in Normal and Disordered Speech. *Proceedings of Meetings on Acoustics, 19*(1)https://doi.org/10.1121/1.4800684

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor Adaptation to Feedback Perturbations of Vowel Acoustics and its Relation to Perception. *The Journal of the Acoustical Society of America, 122*(4), 2306-2319. https://doi.org/10.1121/1.2773966

Figure 1. Set-up for measuring hardware-only round trip latencies. The audio interface’s output was looped back to its input and the latencies were measured with the *RTL Utility* software. RME Babyface Pro depicted here as an example audio interface.

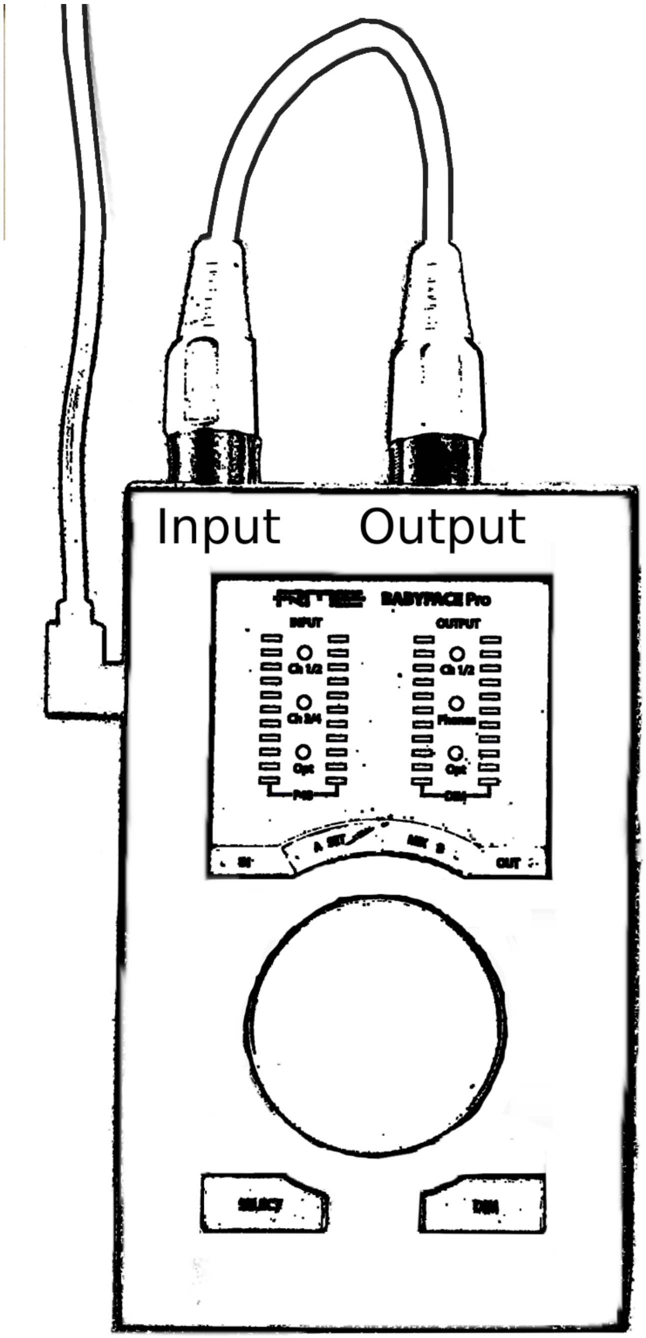


Figure 2. Correct set-up for measuring total feedback loop latency. The unperturbed signal arrives at the recording computer without travelling to the audio interface, while the perturbed signal travels through the audio interface (input), operating computer (software processing), audio interface again (ouput), and a mixer or amplifier before arriving at the recording computer. The total latency, accounting for both hardware and software latencies, is measured.

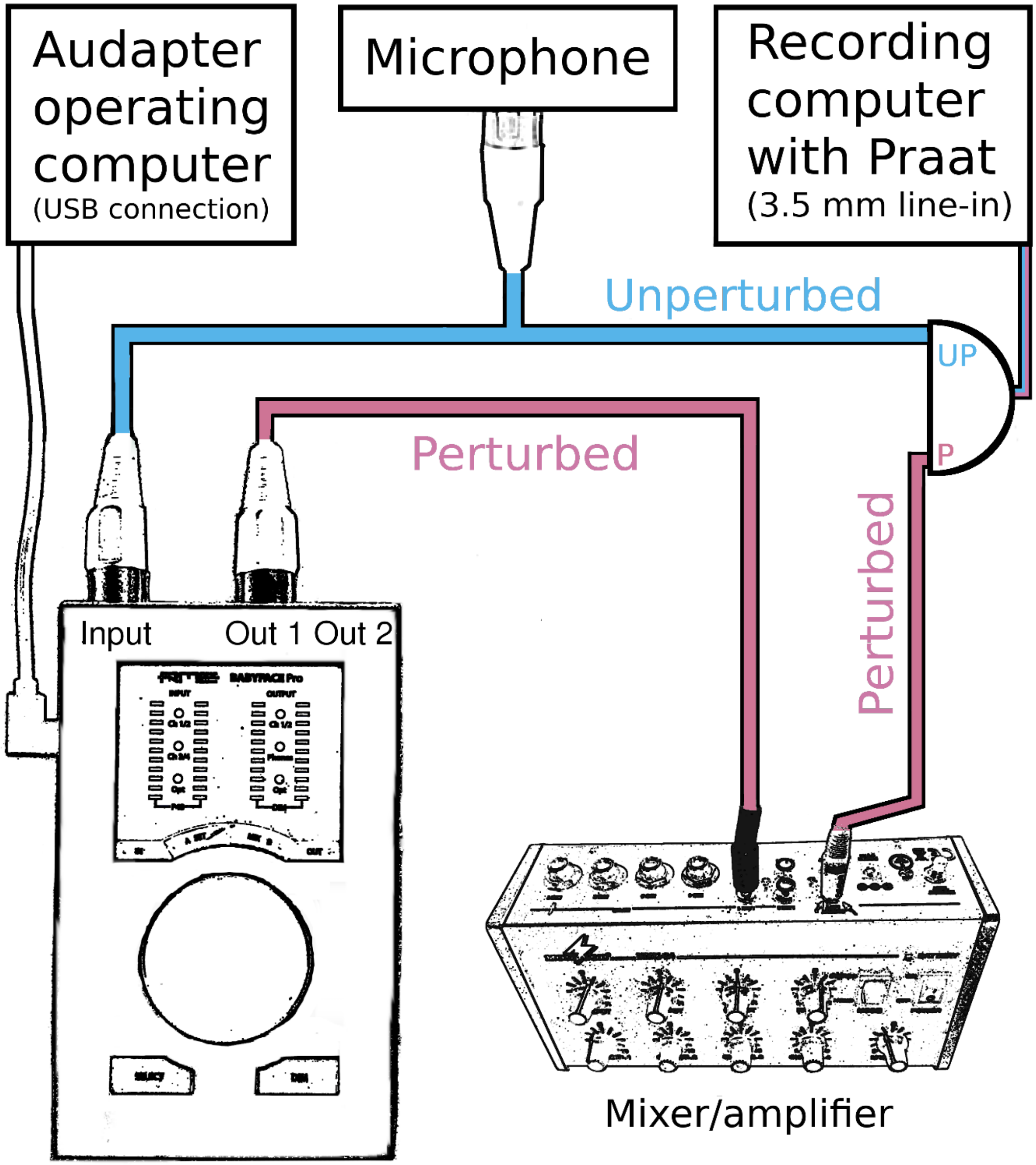


Figure 3. Effect on formant tracking of different values for the Audapter parameter *nDelay*. Top row: Spectrograms of a male speaker's productions of the test words *"tech"*, *"tuck,"* and *"talk"* with overlaid F1 and F2 tracks as calculated by Audapter with the three *nDelay* values. Middle row: Audapter's F1 (left) and F2 (right) values for all 90 individual trials analyzed with the three *nDelay* values (within each trial, F1 and F2 were calculated as the average value across the middle 40-60% of the corresponding formant track). Bottom row: Individual trial differences in formant estimates for pairs of the three analysis conditions.

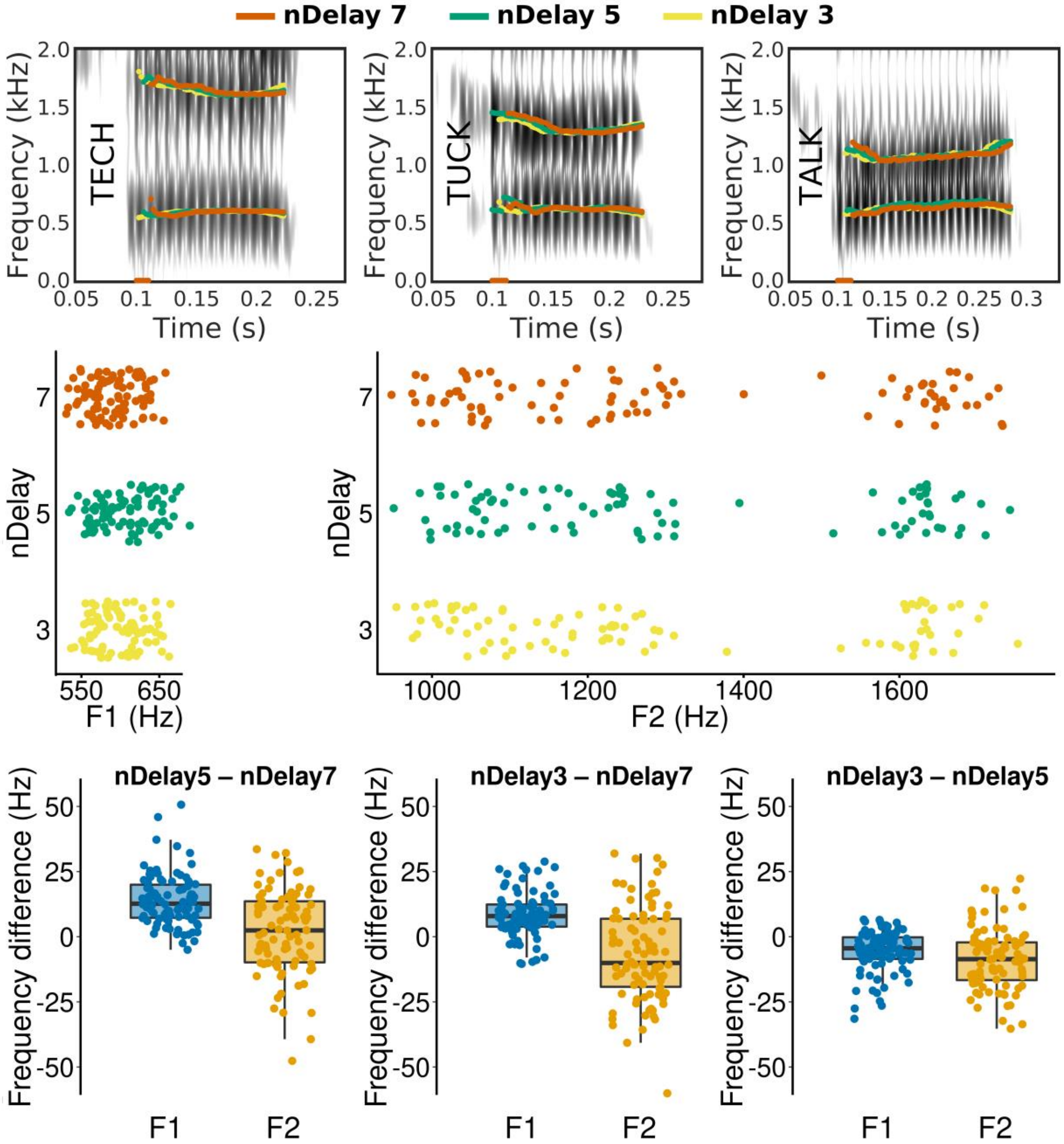


Figure 4. Incorrect set-up for measuring total feedback loop latency. Both the unperturbed and perturbed signals arrive at the recording computer after travelling to the audio interface. As a result, the latency measured by the recording computer ignores the hardware-specific latency.

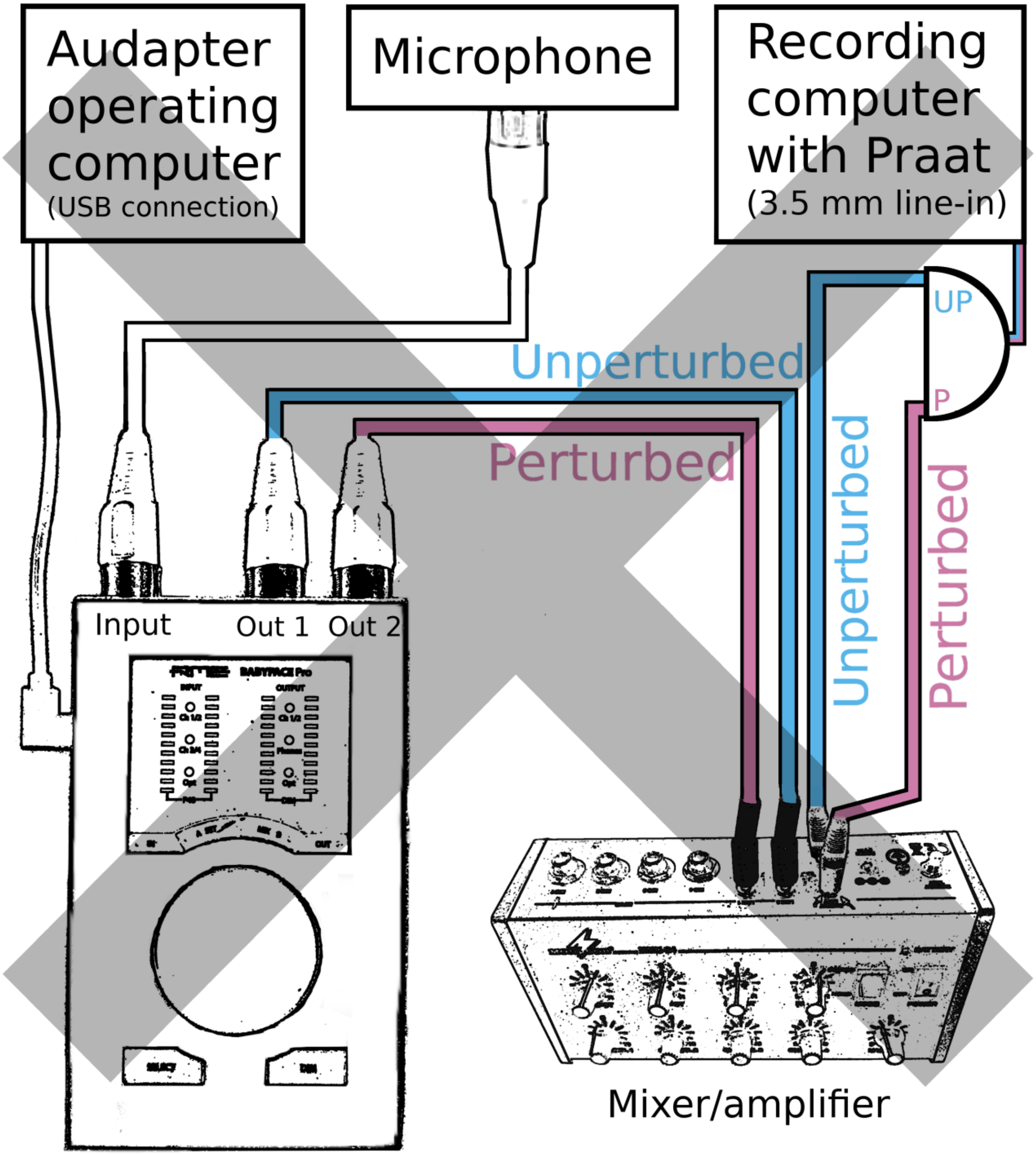


Table 1. Hardware-only round trip latencies (RTL; in milliseconds), measured with the *RTL Utility* (setup shown in Figure 1), for all combinations of audio interface and operating computer. All tests completed with original sampling rate 48 kHz and separate runs for the manufacturer's ASIO driver and ASIO4ALL driver. Note that the MOTU Ultralite AVB has a Host Safety Offset (HSO) setting that influences the latency; hence, we tested this device at both 16 and 128 HSO. Means ± 1 SD.

Audio interface and computer	Hardware RTL with manufacturer ASIO driver	Hardware RTL with ASIO4ALL driver
Devices with buffer size 96 samples		
MOTU MicroBook IIc Latitude E5570	13.081 ± 0.002	14.082 ± 0.000
RME Babyface Pro Latitude E5570	5.299 ± 0.001	11.287 ± 1.528
RME Babyface Pro Precision 7510	5.307 ± 0.001	11.070 ± 1.552
Devices with buffer size 128 samples		
MOTU Ultralite AVB (HSO 16) Latitude E5570	9.546 ± 0.065	17.517 ± 0.056
MOTU Ultralite AVB (HSO 16) Precision 7510	9.544 ± 0.066	17.492 ± 0.010
MOTU Ultralite AVB (HSO 128) Latitude E5570	14.196 ± 0.056	22.138 ± 0.001
MOTU Ultralite AVB (HSO 128) Precision 7510	14.177 ± 0.060	22.169 ± 0.006
Presonus Studio 192 Mobile Latitude E5570	12.550 ± 0.065	16.130 ± 0.000
Presonus Studio 192 Mobile Precision 7510	12.472 ± 0.011	16.126 ± 0.065
RME Babyface Pro Latitude E5570	6.648 ± 0.001	30.570 ± 11.738
RME Babyface Pro Precision 7510	6.645 ± 0.005	13.748 ± 0.895
Tascam US 2x2 Latitude E5570	13.018 ± 0.002	19.488 ± 0.000
Tascam US 2x2 Precision 7510	12.890 ± 0.213	19.042 ± 0.373

Table 2. Total feedback loop latencies (TFLL) comprising both hardware round-trip latency (RTL) and Audapter signal processing latency (setup shown in Figure 2) together with estimated Audapter software latencies (TFLL minus hardware RTL) (in milliseconds). All tests completed with an original sampling rate of 48 kHz and with each manufacturer's default driver (except for one device where, as indicated in the table, the ASIO4ALL driver had to be used due to an incompatibility between the manufacturer driver and Audapter). Means ± 1 SD.

Audio interface and computer	Total feedback loop latency (TFLL)	TFLL minus hardware RTL (estimated software latency)
Devices with buffer size 96 samples, applied downsampling factor = 3		
MOTU MicroBook IIc Latitude E5570	27.214 ± 0.099	14.133
RME Babyface Pro Latitude E5570	19.382 ± 0.062	14.083
RME Babyface Pro Precision 7510	19.302 ± 0.121	13.995
Devices with buffer size 128 samples, applied downsampling factor = 4		
MOTU Ultralite AVB (HSO 16) Latiude E5570	28.496 ± 0.095	18.950
MOTU Ultralite AVB (HSO 16) Precision 7510	28.473 ± 0.105	18.929
MOTU Ultralite AVB (HSO 128) Latiude E5570	33.093 ± 0.080	18.897
MOTU Ultralite AVB (HSO 128) Precision 7510	33.040 ± 0.141	18.863
Presonus Studio 192 Mobile Latitude E5570	31.353 ± 0.125	18.803
Presonus Studio 192 Mobile Precision 7510	31.389 ± 0.143	18.917
RME Babyface Pro Latitude E5570	25.509 ± 0.098	18.861
RME Babyface Pro Precision 7510	25.484 ± 0.048	18.839
Tascam US 2x2 (ASIO4ALL) Latitude E5570	38.972 ± 0.204	19.484
Tascam US 2x2 (ASIO4ALL) Precision 7510	38.567 ± 0.079	19.525

Table 3. Audapter processing latencies (in milliseconds) quantified by using Praat software to measure the temporal difference between the unperturbed and perturbed channels recorded within Audapter. Thus, measures reflect software processing latency only. All tests completed with an original sampling rate of 48 kHz and separate runs for the manufacturer's ASIO driver and ASIO4ALL driver. N/A indicates that the particular interface did not function correctly with Audapter when using the manufacturer's driver. Means ± 1 SD.

Audio interface and computer	Audapter latency with manufacturer ASIO driver	Audapter latency with ASIO4ALL driver
Devices with buffer size 96 samples, applied downsampling factor = 3		
MOTU MicroBook IIc Latitude E5570	12.017 ± 0.073	11.990 ± 0.043
RME Babyface Pro Latitude E5570	11.923 ± 0.123	11.918 ± 0.080
RME Babyface Pro Precision 7510	11.977 ± 0.031	12.024 ± 0.037
Devices with buffer size 128 samples, applied downsampling factor = 4		
MOTU Ultralite AVB (HSO 16) Latiude E5570	16.000 ± 0.046	15.974 ± 0.052
MOTU Ultralite AVB (HSO 16) Precision 7510	16.040 ± 0.037	16.025 ± 0.056
MOTU Ultralite AVB (HSO 128) Latiude E5570	15.984 ± 0.040	15.991 ± 0.074
MOTU Ultralite AVB (HSO 128) Precision 7510	16.020 ± 0.051	16.022 ± 0.054
Presonus Studio 192 Mobile Latitude E5570	15.964 ± 0.030	15.993 ± 0.050
Presonus Studio 192 Mobile Precision 7510	15.988 ± 0.056	16.022 ± 0.066
RME Babyface Pro Latitude E5570	15.987 ± 0.033	15.986 ± 0.041
RME Babyface Pro Precision 7510	15.974 ± 0.057	16.013 ± 0.030
Tascam US 2x2 Latitude E5570	N/A	16.018 ± 0.029
Tascam US 2x2 Precision 7510	N/A	16.011 ± 0.051

Table 4. Total feedback loop latencies (comprising hardware latency and Audapter signal processing latency; in milliseconds) for the RME Babyface Pro audio interface with two different downsampling factors in Audapter. All tests completed with an original sampling rate of 48 kHz and with the manufacturer's default driver. Means ± 1 SD.

Audio interface and computer	Total feedback loop latency with *downFact* = 3	Total feedback loop latency with *downFact* = 4
RME Babyface Pro Latitude E5570	19.382 ± 0.062	19.418 ± 0.181
RME Babyface Pro Precision Tower 7510	19.302 ± 0.121	19.462 ± 0.182

Table 5. Effect of different values for the Audapter parameter *nDelay* on total feedback loop latency and Audapter's own software processing latency (in milliseconds). All measurements were completed with the RME Babyface Pro audio interface, its own manufacturer-issued software driver, an original sampling rate of 48 kHz, and Audapter *downFact* 3. Means ± 1 SD.

Measurements	*nDelay 7*	*nDelay 5*	*nDelay 3*
Total feedback loop latency (TFLL)	19.384 ± 0.092	15.432 ± 0.010	11.373 ± 0.107
Audapter processing latency (Indirect method)	14.084	10.133	6.074
Audapter processing latency (Direct method)	12.003 ± 0.041	7.995 ± 0.025	4.006 ± 0.007

Table 6. Effect of different operating systems (Windows 7 Professional vs. Windows 10 Enterprise LTSC) on total feedback loop latency. All measurements were completed with the Dell Latitude E5570 computer, the RME Babyface Pro audio interface, its own manufacturer-issued software driver, an original sampling rate of 48 kHz, and Audapter *downFact* 3. Means ± 1 SD.

Measurements	*nDelay 7*	*nDelay 5*	*nDelay 3*
Windows 7	19.384 ± 0.092	15.432 ± 0.010	11.373 ± 0.107
Windows 10	19.381 ± 0.058	15.446 ± 0.062	11.431 ± 0.049

Chapter III

Auditory-motor learning differences between stuttering and nonstuttering individuals reflect implicit learning processes.

Kwang S. Kim
University of Washington, Seattle, WA

Ludo Max
University of Washington, Seattle, WA
Haskins Laboratories, New Haven, CT

Correspondence concerning this work should be addressed to Ludo Max, Ph.D., Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd St., Seattle, WA 98105-6246. E-mail: LudoMax@uw.edu. Phone: 206 543 2674.

Abstract

The neural mechanisms underlying stuttering continue to remain poorly understood. Several recent studies have focused on sensorimotor integration in people who stutter, including the capacity for sensorimotor *learning*. Typically, sensorimotor learning is assessed with adaptation paradigms in which one or more sensory feedback channels is experimentally perturbed in real-time. Our lab's previous work on speech with perturbed auditory feedback found strong evidence of auditory-motor learning limitations in both children and adults who stutter (AWS). It remains unknown, however, which sub-processes of sensorimotor learning may be implicated. Indeed, new insights from research on upper-limb motor control indicate that sensorimotor learning involves at least two distinct component: (a) an explicit component that includes intentional strategy use and presumably is driven by target error, and (b) an implicit component that updates an internal model without awareness of the learner and presumably is driven by sensory prediction error. We report an initial attempt at dissociating these components for speech auditory-motor learning in AWS vs. adults who do not stutter (AWNS). First, results replicated our previous finding that such learning is limited in AWS. Second, in neither group did participants report any awareness at all of changing their productions in response to the perceived auditory feedback. These findings suggest that speech auditory-motor adaptation relies exclusively on implicit learning, and, therefore, that the limited adaptation found in AWS is likely due to poor implicit learning mechanisms. This conclusion is in direct agreement with other recent lines of evidence implying sensory prediction deficits in stuttering.

Keywords: stuttering, adaptation, auditory feedback, sensorimotor learning, implicit learning

Auditory-motor learning differences between stuttering and nonstuttering individuals reflect implicit learning processes.

Introduction

The neural mechanisms underlying stuttering remain largely unclear, but it has been hypothesized that this disorder of speech fluency may be due to unstable, noisy, or incorrect neural representations (i.e., internal models) of the multiple motor-to-sensory transformations involved in generating the acoustic speech output (Cai et al., 2012; Daliri, Prokopenko, & Max, 2013; Max, 2004; Max et al., 2004; Neilson & Neilson, 1987, 1991; Hickok et al., 2011). In one of the earliest formulations of this perspective, our research group postulated that difficulties with the appropriate updating of such internal models would manifest as limited sensorimotor learning when exposed to experimentally altered sensory feedback (Max 2004; Max et al., 2004). Since then, a small number of studies have examined sensorimotor learning in individuals who stutter by means of adaptation paradigms in which real-time spectral changes in participants' auditory feedback lead to compensatory adjustments in their articulation. Despite using different spectral manipulations (e.g., formant shifts that did or did not change the phonological identity of produced vowels), these studies consistently found that the extent of auditory-motor adaptation is limited in adults who stutter (AWS) as compared with adults who do not stutter (AWNS) (Daliri, Wieland, Cai, Guenther, & Chang, 2018; Daliri & Max, 2018; Kim, Flanagan, & Max, in preparation; Sengupta, Shah, Gore, Loucks, & Nasir, 2016). Although Daliri et al. (2018) failed to find adaptation problems in *children* who stutter (CWS) with a paradigm where the phonological identity of vowels was changed (e.g., hearing "bad" when saying "bed"), our recent study with a more subtle auditory perturbation (that did not change phonological identity of the vowels) demonstrated that the between-group difference was even larger in children than in adults. In fact, CWS in the range 3-6 years of age (the group closest to the onset of their stuttering) exhibited a complete absence of auditory-motor adaptation. Hence, stuttering individuals' speech auditory-motor learning difficulties are already present at a very young age, and, therefore, have the potential to be fundamentally related to the onset of the disorder (Kim et al., submitted; Max 2004).

The accumulating evidence for auditory-motor learning difficulties in both CWS and AWS raises the question specifically which sub-processes are affected in stuttering. In studies of upper-limb visuomotor learning, for example, it has been demonstrated that at least two different sub-processes or components contribute to the learning. One is an implicit component (i.e., occurring without awareness) that involves the updating of an internal model (i.e., a neural representation of the motor-to-sensory mapping for the effector system in a given environment). This implicit component is believed to be driven by sensory prediction errors; that is, a mismatch between the actual sensory consequences and those predicted based on the generated motor command The second component is explicit (i.e., with awareness) and involves intentional strategy selection. This explicit component is believed to be driven mostly by target error – a discrepancy between movement target and achieved performance. Some of the strongest support for the independence of explicit and implicit components in sensorimotor learning is found in work by Mazzoni & Krakauer (2006). Those investigators asked participants to make reaching movements while the location of a cursor representing hand position was rotated 45° counter-clockwise around the center of the workspace. When participants were informed about the visual perturbation and instructed to use an explicit strategy (i.e., aiming 45° clockwise relative to the target), target error was minimized. Over subsequent trials, however, participants started to reach even further clockwise relative to the target – thus, target error gradually increased. In other words, even when target error was eliminated by explicitly adopting an aiming strategy that countered the visuomotor rotation, sensorimotor learning continued in an implicit manner and such that it minimized sensory prediction error.

Based on those insights, other investigators studying the adaptation of reaching movements have developed experimental methods to estimate the relative contributions of the explicit and implicit learning components. One approach has been to experimentally restrict movement reaction time so that there is insufficient time for the completion of explicit cognitive processes prior to movement onset (Fernandez-Ruiz, Wong, Armstrong, & Flanagan, 2011; Huberdeau, Krakauer, & Haith, 2015). A different approach has been to ask participants to verbally report their intended aiming direction before each trial by naming a number along the circumference of the circular workspace (McDougle, Bond, & Taylor, 2015; McDougle, Ivry, & Taylor, 2016; Taylor, Krakauer, & Ivry, 2014). The reported aiming direction was interpreted as reflecting the participant's explicit strategy for that trial. The contribution of implicit learning for the trial was computed as the difference between intended aiming direction and actual movement direction. The time course of both learning components across trials revealed that explicit learning was dominant immediately after introduction of the perturbation but decreases over time whereas implicit learning increased over time (McDougle et al., 2015; McDougle et al., 2016; Taylor et al., 2014).

Converting this limb paradigm to the domain of speech production is not straightforward. For example, it is not possible to ask naïve subjects where in the formant space (i.e., the two-dimensional acoustic space defined by the first two resonance frequencies of the vocal tract) they will "aim" their acoustic output for the next trial. Naïve speakers know none of the three critical aspects that would be necessary to answer this question: (a) the relative position of different vowels in this acoustic formant space; (b) the intricate relationships between vocal tract postures and acoustic output; and (c) the specific nature of how the experimental perturbation is manipulating their acoustic output in the formant space. In fact, due to these complexities, speech auditory-motor learning – unlike reach visuomotor learning – may be an entirely implicit process. This would be consistent with observations on typical, nonstuttering speakers showing no difference in the amount of adaptation to formant-shifted feedback when instructed to compensate, to ignore the feedback, or to avoid compensating (Munhall, MacDonald, Byrne, & Johnsrude, 2009). Similarly, in a pitch-shift study with trained singers, the amount of adaptation was not affected by instructions to either compensate or ignore the feedback (Keough, Hawco, & Jones, 2013).

It is therefore important to start developing speech-appropriate paradigms that can be used for distinguishing between explicit and implicit aspect of speech auditory-motor learning in individuals who stutter. Determining whether stuttering individuals’ aforementioned auditory-motor learning impairment relates to explicit strategy selection, implicit internal model updating, or both, would provide additional insights into both the sensorimotor mechanisms and neural substrates underlying the disorder. Here, we test AWS and AWNS with a novel paradigm developed to obtain information on participants’ awareness and intent related to adaptive articulatory behavior when speaking with formant-shifted feedback. After each trial, participants were asked to report on a visual analog scale to what extent they intentionally changed their speech, and these ratings were used to quantify the explicit component of learning. As a second innovation, the paradigm also asked participants at regular intervals to select the acoustic stimulus that best represented their auditory target for the test word. The latter addition sought to examine whether AWS and AWNS differ in the extent to which they experience drift of the perceptual target when repeatedly producing speech with spectrally manipulated auditory feedback (Shiller, Sato, Gracco, & Baum, 2009).

Materials and Methods

Participants

We recruited 15 stuttering adults and 15 nonstuttering adults who were individually matched based on age (± 3 years), sex, and handedness. All participants were (a) between 18 and 50 years old, (b) native speakers of American English, (c) without history of speech, language, and hearing disorders (other than stuttering in case of people who stutter), (d) naive to the purpose of the study, and (e) not taking medications with a possible effect on sensorimotor functioning at the time of the study. Importantly, all participants had no prior experience of participating in speech experiments involving auditory perturbation. All participants had normal binaural hearing thresholds (i.e., ≤20 dB HL for the octave frequencies 250–4000 Hz) except for the three participants who stutter. Of the three participants, one participant’s hearing thresholds remain unknown as the audiogram measurement could not be collected due to an equipment malfunction. The other two participants had a slightly higher hearing thresholds at one or two frequencies: one had 25 dB HL at 1 kHz and 4 kHz for both ears, and another had a threshold of 25 dB HL at 4 kHz for his right ear. All participants signed the informed consent form approved by the University of Washington Institutional Review Board.

All stuttering participants reported that they stuttered since early childhood (i.e., before the age of 8 years) and the presence of stuttering was confirmed by the experimenter on the day of the experiment. Most data for stuttering participants were collected at the 2018 National Stuttering Association Annual Conference in Chicago, IL, but their speech samples for Stuttering

Severity Index (SSI-4, Riley, 2009) were collected after the conference via video call due to the limited availabilities at the conference. Unfortunately, two participants were not reachable after the study so their SSI-4 samples were not collected. An American Speech-Language-Hearing Association-certified speech-language pathologist scored the rest of the SSI-4 samples. One participant was excluded from the analyses because the speech samples failed to reach the minimum required overall SSI score of 10 or more (i.e., SSI-4 score did not reach the ”very mild” category). Two additional stuttering participants were excluded because they stuttered for most of the words in the adaptation task, preventing reliable formant extractions.

Consequently, the final data set included 12 individuals who stutter (age *M* = 27.75 years, *SD* = 7.65, range 18-47 years) and 12 individuals who do not stutter (age *M* = 28.08 years, *SD* = 7.45, range 20-48 years). In each group there were 9 male and 3 female participants. The participants were right-handed except one left-handed individual in each group (per self-report). Stuttering severity for the participants (other than the two individuals with unknown SSI-4) ranged from very mild to very severe (2 very mild, 3 mild, 3 moderate, 1 severe, and 1 very severe).

Experimental setup

In a quiet room, participants wore a headset microphone (AKG C544, Shure) connected to a USB audio interface (Babyface Pro, RME) operated by a laptop (Latitude E5570, Dell). The laptop executed a custom MATLAB user-interactive program (MATLAB, The MathWorks, Inc.) implemented with the Audapter software (Cai, Boucek, Ghosh, Guenther, & Perkell, 2008; Cai, Ghosh, Guenther, & Perkell, 2010; Cai, Ghosh, Guenther, & Perkell, 2011; Tourville, Cai, & Guenther, 2013) for formant perturbation. The perturbed speech signal was then routed to a mixer (MMX-24, Monacor) via the output of the interface. The auditory feedback was provided to the participants via insert-earphones (ER-3A, Etymotic Research) connected to the mixer. The Audapter software settings had the *downFact* of 3, the sampling rate of 48 kHz, and the *nDelay* parameter of 5. The total latency between the microphone input and headphone output measured following the recommended procedure (Kim, Wang, & Max., submitted) was 15.3 ms. As recommended in the Audapter manual, we used different online formant tracking parameters for male participants and female participants by running a MATLAB function included in the Audapter package (getAudapterDefaultParams.m) with an input "male" or "female." Prior to each recording session, the amplification in the audio interface and the mixer were calibrated in such a manner that a speech signal with an intensity of 90.4 dB SPL at the headset microphone placed 2.54 cm from a loudspeaker[1] resulted in an output of 73 dB SPL in the earphones (measured in a 2 cc coupler Type 4946 connected to a sound level meter Type 2250A with Type 4947 1⁄2" pressure field microphone, Bruel & Kjaer).

Auditory-motor adaptation task

The custom MATLAB program asked the participants to read monosyllabic words ("bed" and "pet") that appeared on a touchscreen monitor (P2314T, Dell or P2418HT, Dell). Each word appeared at a time in an alternating order, starting with the first trial "bed." There were 10 utterances of *baseline*, 60 of *full perturbation,* and 10 of *after-effects* for each word (80 utterances × 2 words = 160 trials). In the full perturbation phase, only F1 was shifted up by 400 cents which perturbed "bed" and "'pet" to sound more like "bad" and "pat" (Figure 1). The 400 cents perturbation in F1 was applied by setting *bRatio* to 1, all 257 elements of *pertAmp* to 1.26

---

[1] This is equivalent of 75 dB SPL when a microphone is placed 15 cm from a speaker.

(i.e., 26% increase), and all 257 elements of *pertPhi* = 0 radian (i.e., setting a 0 radian in the F1-by-F2 vowel space would shift along the horizontal F1 axis) in Audapter.

After each trial the MATLAB program asked the participants whether they employed an explicit strategy by asking the question “did you TRY to change your speech? In which direction?” (Figure 2). Subjects were instructed to move a slider on the touchscreen to indicate to what extent, if any, they changed their production in terms of raising F1 (e.g., spoke more like ‘bat’) or lowering F1 (i.e., spoke more like ‘bit’). Note that this is different from Taylor et al. (2014) in that we asked about the explicit strategy after, rather than before, each movement. In our initial attempts of designing a task, asking participants about their strategy plan ahead of a trial (e.g., “how are you going to say bed?”) was found to be too confusing even for speech pathology students with explicit knowledge of formant frequencies. Therefore, we decided to obtain information on participants’ awareness and intent related to adaptive articulatory behavior after each word.

Importantly, before the beginning of the experiment, the program recorded productions of the two words for each participant (i.e., pre-task recording). The program then perturbed the pre-task recording to generate different versions (i.e., different F1 shifts) used as options for participants to express their auditory targets in response to the target questions (e.g., “what SHOULD bed sound like?”). Participants responded by dragging and rotating a “wheel” on the touchscreen monitor (Figure 3). Depending on which location of the wheel the participants selected by double tapping, the program played different formant-shifted productions of pre-task recording. The amount of the possible shift in F1 was between -746 cents (35% decrease in Hz) to 520 cents (35% increase in Hz) arranged in a continuous spectrum. The participants were encouraged to try several different locations of the wheel so that they can pick the best auditory target. We prevented the participants from choosing the same location of the wheel across multiple trials without relying on auditory feedback, by randomizing formant shifts that corresponds to the locations of the wheel. In other words, the participants had to rely solely on auditory feedback rather than visual or cognitive information when choosing the best targets. The final decisions were recorded in cents normalized to F1 in the pre-task recording. In order to prevent influencing auditory-motor adaptation, the targets were acquired every 10th trial (i.e., before 4th trial, 14th trial, 24th trial, 34th trial and so on) with an exception of the first targets that were obtained before the very first trial.

Data extraction and statistical analyses

Formant frequencies were extracted using a custom MATLAB program with Praat (Boersma & Weenink, 2019) formant tracking algorithm. The analysis program rejected trials for various reasons that interfered with formant tracking (e.g., participants missing the time window to speak, speech signal containing yawning, stuttering moments, or other noise). After those trials were rejected (8.7% of “bed” trials and 11.8% of “pet”), the missing data points were interpolated using 4 neighboring data points for each word. The F1 was then normalized to the average of baseline trials (block 6 to 10).

Repeated measures analyses of variance (ANOVA) were conducted with the *ezANOVA* function in the *ez* package (Lawrence, 2016) in R (R Core Team, 2018) for early and late adaptation. The analyses included a between-groups (i.e., AWS vs. AWNS) test, and a within-groups test for different words (e.g., bed vs. pet). The analyses also tested for the interactions between the two variables (Group × Word). We also reported effect sizes with partial omega-squared (ω_p^2) for all tests.

The explicit component was extracted from the location of the slide bar that the participant placed as his/her response to the explicit strategy question (i.e., “did you TRY to change your speech? In which direction?”). We reported this intent rating in percentage, -100% would be when the participant moves the slide bar completely to “more like bid”, indicating that he was trying his best to sound more like “bid.” 100% would mean that the participant moves the bar to “more like bad” instead. If a participant left the slide bar in the middle to report no change in their production, the response would correspond to 0% intent rating.

For perceptual target shifts, the F1 (in cents) of the sounds that the participant chose as his/her final answer to questions “what SHOULD bed sound like?” and “what SHOULD pet sound like?” were included in the statistical analyses. We analyzed both the explicit component and target shifts with independent one-sample t-tests to compare whether the values are different from 0. For the t-tests done in a given post hoc analysis or the same family of data, we adjusted the p-values with Holm−Bonferroni method using *p.adjust* function in the R stats package (R Core Team, 2018).

Lastly, we tested for any correlations between adaptation and stuttering frequencies using Pearson’s correlation in R. As previously described, we were able to include only 10 AWS’ data in this analysis given that SSI speech samples could not be collected for two stuttering individuals. We included the stuttering frequency in both reading and speaking samples for the correlation tests.

Results

Auditory-motor adaptation

Both groups showed adaptation, lowering F1 relative to their baseline in response to the perturbation (Figure 4). A repeated measures ANOVA for early adaptation (i.e., adaptation in the first five perturbation blocks) revealed a significant main effect of Group, $F(1, 22) = 8.756$, $p = 0.007$, $\omega_p^2 = 0.144$ (see Figure 4, Early), suggesting that AWS adapted with significantly slower initial rates compared to AWNS. According to the ANOVA, there was no significant main effect of Word (i.e., bed vs. pet), $F(1, 22) = 0.094$, $p = 0.761$, $\omega_p^2 < 0.001$. There was also no significant interaction, Group × Word, $F(1, 22) = 0.755$, $p = 0.394$, $\omega_p^2 < 0.001$.

The group difference continued throughout the perturbation phase, and AWNS adapted ~50% of the perturbation while AWS adapted ~25% by the end of the phase. As expected from this observation, we found a significant main effect of Group on the amount of late adaptation (i.e., last five perturbation blocks), $F(1, 22) = 11.095$, $p = 0.003$, $\omega_p^2 = 0.180$ (see Figure 4, Late). This result revealed that the stuttering group adapted less than the nonstuttering group by the end of the perturbation phase. The ANOVA for the amount of late adaptation did not significantly differ across different words, Word, $F(1, 22) = 1.145$, $p = 0.296$, $\omega_p^2 = 0.001$. There was also no significant interaction between Group and Word, $F(1, 22) = 0.036$, $p = 0.852$, $\omega_p^2 < 0.001$.

Explicit component

Most of the participants continuously placed the slide bar in the middle to report that they were not changing their productions even after the perturbation started (Figure 5). During the baseline phase (i.e., from block 6 to 10), both groups' response (i.e., intent rating) did not differ from 0 for both AWNS, $t(11) = 0.506$, $p = 1.000$, $d = 0.146$ and AWS, $t(11) = -0.641$, $p = 1.000$, $d = 0.185$. The intent ratings, again, continued to stay around 0 in the early perturbation phase (i.e., from block 11 to 15) for AWNS, $t(11) = 1.399$, $p = 1.000$, $d = 0.404$, and AWS, $t(11) = 1.053$, $p = 1.000$, $d = 0.304$. The values at the end of the perturbation phase (i.e., from block 66 to 70) also did not differ from 0 for AWNS, $t(11) = -0.665$, $p = 1.000$, $d = 0.192$ and for AWS, *t*(11) = 0.524, *p* = 1.000, *d* = 0.151. Taken together, both AWS and AWNS consistently indicated that there was no intent to change their speech production throughout the adaptation task.

Interestingly, even though on average the responses were mostly around 0, there were two AWNS who consistently reported that they were explicitly trying to change their speech (Figure 6). One participant consistently expressed that he was trying to change his speech more like “bid” or more like “pit” throughout the experiment, congruent with his adaptation data. When asked about his responses after the experiment, he explained that he changed his speech to sound more like “bid” or “pit” because he thought his speech sounded too much like “bad” or “pat.” Another participant reported responses in an opposite direction, expressing that he tried to speak more like “bad,” and “pat.” In this case, it is likely that this person mistakenly thought that the auditory perturbation was caused by himself given that the F1 in the recorded speech actually decreased during the perturbation phase (i.e., spoke more like “bid” and “pit” compared to the baseline level). Taken together, other than the one participant who correctly indicated his explicit strategy, the rest of the participants clearly did not use any explicit strategy.

Perceptual target shifts

F1 extracted from the perceptual targets selected by the participants were measured in cents, relative to the pre-task recording baseline. Even though the F1 in perceptual targets seem to slightly larger than 0 cents (Figure 7), they did not significantly differ from 0 cents for both AWNS, *t*(11) = 2.310, *p* = 0.207, *d* = 0.667 and AWS, *t*(11) = 2.796, *p*= 0.613, *d* = 0.807. The targets also did not change after the perturbation started; the target F1 was not different from that of the original production in AWNS, *t*(11) = 1.534, *p* = 0.619, *d* = 0.443 and AWS, *t*(11) = -0.412, *p* = 0.104, *d* = 0.119. The pattern continued towards the end of the perturbation phase: the responses stayed around 0 for AWNS, *t*(11) = 1.343, *p* = 1.000, *d* = 0.388 as well as for AWS,

$t(11) = 0.008$, $p = 1.000$, $d = 0.002$. In sum, both groups consistently reported the target sounds that do not differ from their original productions throughout the experiment, suggesting that there was no perceptual target shift.

Stuttering frequency and auditory-motor adaptation

The amount of auditory-motor adaptation in the early perturbation phase was not significantly correlated with stuttering frequency (averaged across the reading and speaking samples from the SSI evaluations), $r(8) = -0.149$, $p = 0.680$. There was also no significant correlation between late auditory-motor adaptation and stuttering frequency, $r(8) = -0.246$, $p = 0.493$.

Discussion

In this study, we attempted to dissociate explicit and implicit components of auditory-motor learning in stuttering and nonstuttering individuals. In this novel paradigm, the participants were asked to speak monosyllabic words ("bed" and "pet") while auditory perturbation (i.e., 400 cents up-shift in F1) was introduced in near real-time. In order to quantify the explicit component, the participants were asked to report how they changed their speech production, if any, after every trial. In addition, we also asked the participants to report their auditory targets (i.e., how the words should sound like) in order to determine any internal auditory target shifts. Several important findings and implications are discussed in this section.

The first important finding is that stuttering individuals adapted significantly less in both early and late adaptation compared to nonstuttering individuals. This finding replicates the previous findings on auditory-motor learning difficulties in adults who stutter (Daliri et al., 2018; Daliri & Max, 2018; Kim et al., in preparation; Sengupta et al., 2016) as well as children who stutter (Kim et al., in preparation). Importantly, it should be noted that this current study had more perturbation trials both per each word and in total number of utterances compared to previous studies (e.g., 80 trials for each word in this current study vs. 30 in Kim et al., submitted, 18 in Daliri et al., 2018, 35 in Daliri & Max, 2018). Yet, stuttering individuals exhibited a clear limitation, not mitigated by the additional exposure to more trials and duration of the perturbation. Taken together, the results add further support to the idea that stuttering individuals have difficulties in updating of the internal models (Max et al., 2004)

The second important finding is that there was no explicit learning component found in all participants but one[2]. The vast majority of the participants consistently responded 0 (i.e., no changes intended) to the explicit strategy question (i.e., "did you TRY to change your speech? In which direction?") during the adaptation task. This is surprising because the first formant frequency (i.e., F1) was shifted up 400 cents in their auditory feedback, enough for the vowel perception to change. In fact, many participants reported after the study that they had noticed vowel perception changes (i.e., verbally reporting that their productions of "bed" and "'pet" sounded more like "bad" and "pat"), but had reported no explicit strategy. Similarly, one participant expressed after the study that he was very certain that he produced "bed" and "pet," and there was no change in his speech. However, he was surprised when an experimenter played back his speech that was recorded during the adaptation that sounded more like "bid" and "pit." Along with these observations, participants' intent ratings suggest that there was nearly no explicit learning component, and thus, auditory-motor learning may be comprised of mostly implicit learning component.

---

[2] One individual who reported an explicit strategy that is congruent with the direction of the adaptation (Figure 5, green),

One implication of this finding is that the auditory-motor system in nonstuttering individuals demonstrated a remarkable ability to update the internal models and generate adaptive motor commands based on the internal models without any explicit strategy. At the same time, auditory-motor learning being mostly implicit may be also the reason that amount of auditory-motor adaptation tends to be limited (~40% of the perturbation across different studies) compared to a typical amount of visuo-motor adaptation (e.g,. ~80%). According to Taylor et al. (2014), an explicit component accounted for about one-third of the total amount of visuo-motor learning. Therefore, if somehow auditory-motor learning could be assisted with an explicit component, we may be able to achieve a comparable amount of adaptation to visuo-motor learning. In fact, the only individual who indicated a clear explicit strategy that corresponded to the direction of adaptation (Figure 5, green) also happened to be the largest adapter (-296 cents). However, the implication of this observation remains inconclusive given that another individual who adapted nearly the same amount (-292 cents) consistently indicated that he did not employ any explicit strategy.

The clear dominance of implicit component in auditory-motor learning may also explain why auditory-motor adaptation is more susceptible to external time delay in sensory feedback compared to visuo-motor adaptation. It has been demonstrated that when auditory feedback is delayed by 100 ms, auditory-motor adaptation was either completely eliminated (Max & Maffett, 2015) or significantly reduced (Mitsuya, Munhall, & Purcell, 2017). This amount of delay is much shorter than the amount required to affect visuo-motor learning (e.g., 200 ms or longer). This notion make sense because implicit component, unlike explicit component, is attenuated in the presence of delays in the sensory feedback (Brudner, Kethidi, Graeupner, Ivry, & Taylor, 2016; Schween & Hegele, 2017), because it is driven by sensory prediction error, a difference between predicted sensory feedback and actual sensory feedback. When predicted sensory feedback and actual sensory feedback are not aligned in the temporal domain due to delay, it would result in inaccurate sensory prediction error, leading to deficient implicit component. Explicit component, on the other hand, may be more resilient when faced with an artificial delay in sensory feedback, because it relies on performance error that is not as dependent on time-sensitive processes. Auditory-motor learning, being mostly implicit, may be more susceptible to the delays as opposed to visuo-motor learning in which explicit component plays a significant role.

It is important to note that our finding of the absence of explicit component in auditory-motor adaptation does not imply that explicit component cannot play a role in auditory-motor learning with different paradigms. For example, inducing explicit component of learning by instructing participants to rely on target error remains entirely plausible. In the upper-limb visuo-motor learning that is known to have explicit component, participants can easily perceive target error (i.e., a cursor is not hitting the target), and thus can effectively employ an explicit strategy. Likewise, one could explicitly instruct participants to speak in a way that their perturbed auditory feedback to sound “normal” again.

The third important finding is that there was no perceptual target shift in both stuttering and nonstuttering groups, suggesting that the perceptual knowledge of auditory targets did not change due to adaptation. Although this finding may seem odd with Shiller et al. (2009) study, there are several major methodological differences between the studies. The previous study had measured perceptual boundaries by asking participants to do a discrimination task for a psychometric function. In contrast, the current study asked the participants to indicate the best perceptual targets throughout the experiment. Given the difference between the two measures, adaptation may affect only perceptual boundaries but not perceptual targets. Another difference between the studies is that Shiller et al. (2009) examined consonant perception rather than vowel perception, and to our knowledge it remains unknown whether vowel perceptual boundaries change due to adaptation.

The fourth important finding is that, given no explicit component and perceptual auditory target shifts found, auditory-motor learning differences between stuttering and nonstuttering individuals are very likely due to implicit learning processes. One implication of this finding is that sensory prediction error that drives implicit learning processes may be deficient in stuttering individuals. This idea is directly in line with hypotheses that the sensorimotor system in people who stutter may not accurately predict the sensory consequences of planned motor commands (Hickok, Houde, & Rong, 2011; Max, 2004; Max, Guenther, Gracco, Ghosh, & Wallace, 2004). Importantly, Max et al. (2004) suggested that such erroneous sensory prediction may result in consequent mismatches between predicted and desired sensory feedback that lead to corrections and interruptions in speech, causing stuttering moments.

In addition, there are several lines of evidence that support the idea of sensory prediction deficits in individuals who stutter. One line of evidence comes from electrophysiological studies on abnormal preparation of the auditory system in prediction of upcoming auditory feedback in stuttering individuals (Daliri & Max, 2015a; Daliri & Max, 2015b; Daliri & Max, 2018; Max & Daliri, 2019). In these studies, nonstuttering individuals consistently showed modulated N100 response in response to auditory stimuli (probe tones) before speaking compared before silently reading (i.e., pre-speech auditory modulation or PSAM). However, compared to nonstuttering individuals, stuttering individuals showed less PSAM, which may suggest deficient processes involved in priming the auditory cortex in preparation of monitoring upcoming auditory feedback. This evidence corroborates our finding in that prediction of auditory feedback is deficient and thus would lead to mal-priming of the auditor system prior to speech onset.

Our overall finding warrants further research on deficient sensory prediction in the context of speech sensorimotor *control*. It has been hypothesized that such deficient sensory prediction would lead to inaccuracies in the movement planning (i.e., feedforward control), and consequently the CNS has to rely more on online feedback control (Max et al., 2004; Max & Daliri, 2019). In order to test this idea, ongoing work from our laboratory is investigating contributions from online feedback control in stuttering individuals using kinematic analyses (see Kim & Max, 2014 for estimating feedforward vs. feedback control contributions in nonstuttering individuals).

Some limitations of the study include the possibility that the adaptation limitations found in stuttering individuals may be due to their poor hearing abilities. Even though all nonstuttering individuals had hearing thresholds less than 20 dB HL, two stuttering individuals had a slightly higher hearing threshold (i.e., 25 dB) at a few frequencies (see Methods). Additionally, the hearing thresholds were not measured for one participant who stutter. However, the possibility that the limited adaptation is a result of poor hearing thresholds is very unlikely given that the amount of adaptation among these individuals do not seem to be related to their hearing abilities. The stuttering individual with the worst hearing thresholds in the group (25 dB HL at both 1 kHz and 2 kHz in both ears) adapted the third most in his group (-183 cents from baseline), an amount that is close to the group average for nonstuttering individuals. In addition, the stuttering individual whose hearing thresholds were not measured also adapted more than the group average for stuttering individual. Another participant who had a very mild hearing difficulty (i.e., 25 dB at only 4 kHz in only right ear) did not show adaptation at all (46 cents from baseline).

The lack of relationship between hearing thresholds and adaptation can also be found in three other individuals (one nonstuttering individual and two stuttering individuals) who also showed nearly no adaptation (around -10 cents from the baseline) despite having less than 20dB HL thresholds at all frequencies tested for both ears. Taken together, the group difference in adaptation is very likely from sensorimotor deficits rather than differences in the hearing abilities.

Another limitation is that we did not find any significant correlation between the stuttering frequency and adaptation, replicating the previous study's finding (Kim et al., submitted). As mentioned by Daliri et al. (2014), however, the current measures of stuttering severity do not adequately capture the severity of stuttering experienced by the speakers. First, stuttering frequency tends to be variable over time and it remains difficult to capture a measure of stuttering severity that is experienced in a longer period from a short speech sample of several minutes. Second, stuttering frequency does not include measures of severity on different aspects of the stuttering problem (e.g., avoidance/substitution of feared words, physical struggle). Therefore, without a comprehensive measure of severity of stuttering, the relationship between severity of stuttering and adaptation difficulties, or any other sensorimotor d, remain unclear (see Daliri et al. 2014 for more details).

In summary, we found that both stuttering and nonstuttering individuals, in most cases, did not use any explicit strategy during typical auditory-motor adaptation, suggesting that both groups relied solely on the implicit component. Taken together, our findings suggest that auditory-motor learning difficulties found in stuttering individuals are likely due to deficient implicit learning processes. Consequently, these data add support to the hypothesis that poor sensory prediction may be associated with stuttering. Future studies are warranted to further examine sensory prediction and its role in speech sensorimotor control and stuttering.

Acknowledgements

The authors thank National Stuttering Association (NSA) Research Committee for accommodating and assisting this research. We are also grateful for Nathalie Farage, Elizabeth Rylance, Elise LeBovidge, and Hantao Wang for help with data extraction. This research was supported in part by grants from the National Institute on Deafness and Other Communication Disorders (R01DC014510 and R01DC17444) and the Canadian Institutes of Health Research (MOP-137001). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, or the Canadian Institutes of Health Research.

References

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [computer software]. http://www.praat.org/:

Brudner, S. N., Kethidi, N., Graeupner, D., Ivry, R. B., & Taylor, J. A. (2016). Delayed feedback during sensorimotor learning selectively disrupts adaptation but not strategy use. *Journal of Neurophysiology, 115*(3), 1499-1511. doi:10.1152/jn.00066.2015

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S.A system for online dynamic perturbation of formant frequencies and results from perturbation of the mandarin triphthong /iau/. Paper presented at the *The 8th Intl. Seminar on Speech Production,* 65-68.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 31*(45), 16483-16490. doi:10.1523/JNEUROSCI.3653-11.2011

Cai, S., Ghosh, S., Guenther, F., & Perkell, J. (2010). Adaptive auditory feedback control of the production of formant trajectories in the mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America, 128*(4), 2033-2048. doi:10.1121/1.3479539

Daliri, A., & Max, L. (2015a). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain and Language, 150*, 37-44. doi:10.1016/j.bandl.2015.08.008

Daliri, A., & Max, L. (2015b). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language, 143*, 59-68. doi:10.1016/j.bandl.2015.03.002

Daliri, A., & Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 99*, 55-68. doi:10.1016/j.cortex.2017.10.019

Daliri, A., Prokopenko, R. A., Flanagan, J. R., & Max, L. (2014). Control and prediction components of movement planning in stuttering versus nonstuttering adults. *Journal of Speech, Language, and Hearing Research: JSLHR, 57*(6), 2131-2141. doi:10.1044/2014_JSLHR-S-13-0333

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science, 21*(2) doi:10.1111/desc.12521

Fernandez-Ruiz, J., Wong, W., Armstrong, I. T., & Flanagan, J. R. (2011). Relation between reaction time and reach errors during visuomotor adaptation. *Behavioural Brain Research, 219*(1), 8-14. doi:10.1016/j.bbr.2010.11.060

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69*(3), 407-422. doi:10.1016/j.neuron.2011.01.019

Huberdeau, D. M., Krakauer, J. W., & Haith, A. M. (2015). Dual-process decomposition in human sensorimotor adaptation. *Current Opinion in Neurobiology, 33*, 71-77. doi:10.1016/j.conb.2015.03.003

Keough, D., Hawco, C., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience, 14*, 25. doi:10.1186/1471-2202-14-25

Kim, K. S., Flanagan, R. J., & Max, L. (in preparation). Dissociated speech and limb sensorimotor learning limitations in stuttering: Speech auditory-motor learning is impaired in both children and adults who stutter
. *Doi,*

Kim, K. S., & Max, L. (2014). Estimating feedforward vs. feedback control of speech production through kinematic analyses of unperturbed articulatory movements. *Frontiers in Human Neuroscience, 8*, 911. doi:10.3389/fnhum.2014.00911

Lawrence, M. A. (2016). *Ez: Easy analysis and visualization of factorial experiments*

Max, L. (2004). Stuttering and internal models for sensorimotor control: A theoretical perspective to generate testable hypotheses. *Speech motor control in normal and disordered speech* (pp. 357–388) Oxford University Press.

Max, L., & Daliri, A. (2019). Limited pre-speech auditory modulation in individuals who stutter: Data and hypotheses. *Journal of Speech, Language, and Hearing Research: JSLHR, 62*(8S), 3071-3084. doi:10.1044/2019_JSLHR-S-CSMC7-18-0358

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (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 Sciences and Disorders, 31*, 105-122. doi:10.1044/cicsd_31_S_105

Max, L., & Maffett, D. G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience Letters, 591*, 25-29. doi:10.1016/j.neulet.2015.02.012

Mazzoni, P., & Krakauer, J. W. (2006). An implicit plan overrides an explicit strategy during visuomotor adaptation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 26*(14), 3642-3645. doi:10.1523/JNEUROSCI.5317-05.2006

McDougle, S. D., Bond, K. M., & Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *The Journal of Neuroscience, 35*(26), 9568-9579. doi:10.1523/JNEUROSCI.5061-14.2015

McDougle, S. D., Ivry, R. B., & Taylor, J. A. (2016). Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends in Cognitive Sciences, 20*(7), 535-544. doi:10.1016/j.tics.2016.05.002

Mitsuya, T., Munhall, K. G., & Purcell, D. W. (2017). Modulation of auditory-motor learning in response to formant perturbation as a function of delayed auditory feedback. *The Journal of the Acoustical Society of America, 141*(4), 2758. doi:10.1121/1.4981139

Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America, 125*(1), 384-390. doi:10.1121/1.3035829

R Core Team. (2018). R: A language and environment for statistical computing [computer software]. Vienna, Austria:

Riley, G. (2009). *SSI-4: Stuttering severity instrument*

Schween, R., & Hegele, M. (2017). Feedback delay attenuates implicit but facilitates explicit adjustments to a visuomotor rotation. *Neurobiology of Learning and Memory, 140*, 124-133. doi:10.1016/j.nlm.2017.02.015

Sengupta, R., Shah, S., Gore, K., Loucks, T. M., & Nasir, S. M. (2016). Anomaly in neural phase coherence accompanies reduced sensorimotor integration in adults who stutter. *Neuropsychologia, 93*(Pt A), 242-250. doi:10.1016/j.neuropsychologia.2016.11.004

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America, 125*(2), 1103-1113. doi:10.1121/1.3058638

Taylor, J. A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 34*(8), 3023-3032. doi:10.1523/JNEUROSCI.3619-13.2014

Tourville, J. A., Cai, S., & Guenther, F. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics, 19*(1) doi:10.1121/1.4800684

Figure 1. The vowel portion of the speech production (top spectrogram) and perturbed auditory feedback (bottom spectrogram) in an example trial ("bed") during perturbation phase. The red markers on the spectrograms represent the first and second formant frequencies. In this example, the first formant frequency was shifted up 400 cents, resulting in the received auditory feedback that sounded more like "bad." The phonetic vowel symbols corresponding to the vowels in the spectrograms are shown on a vowel space on the left.

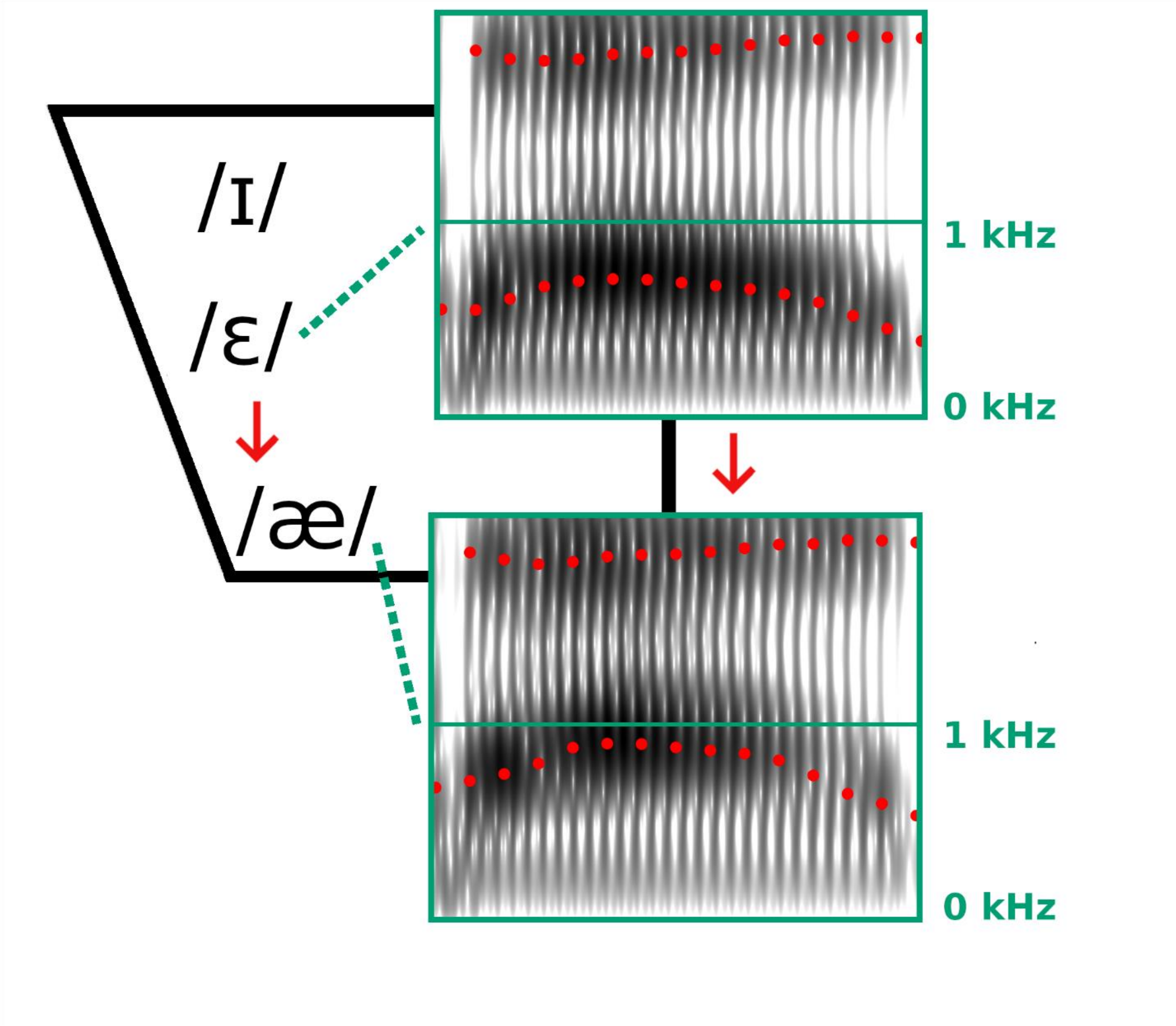
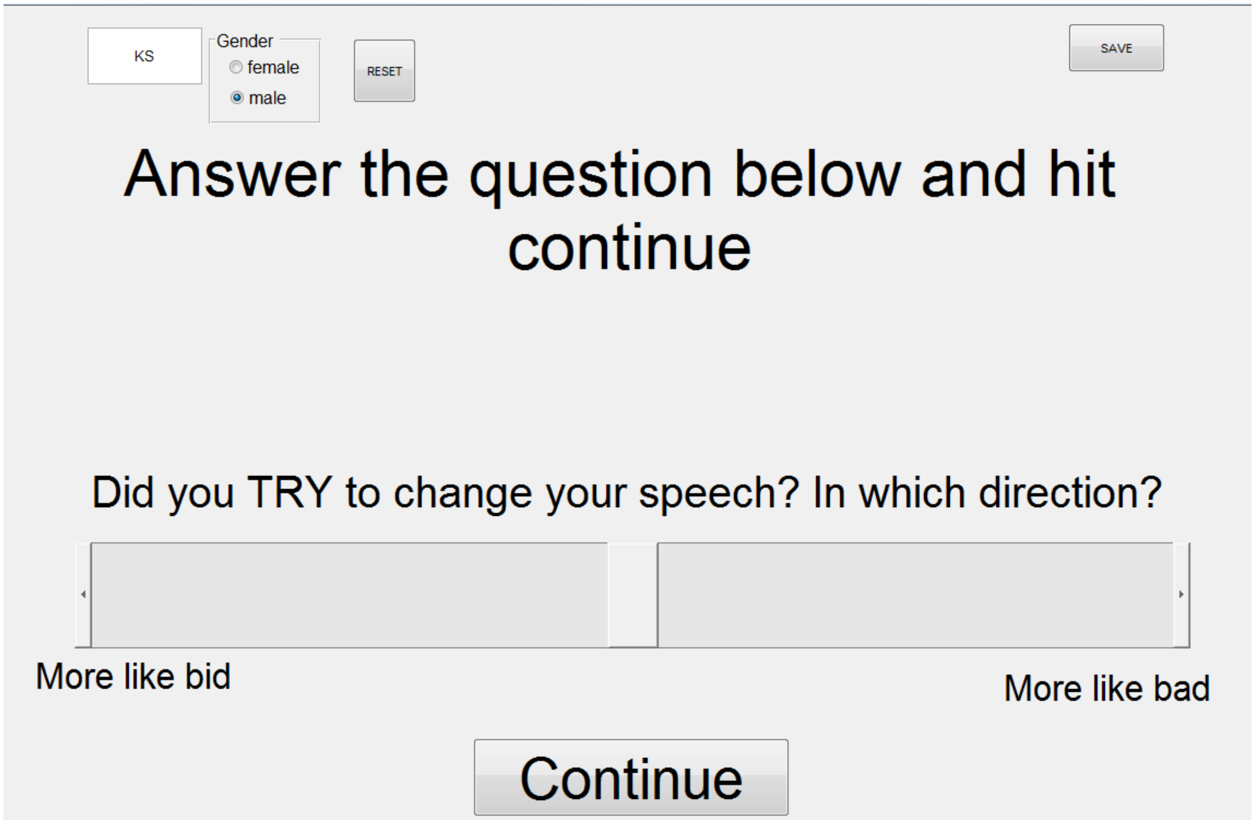


Figure 2. A screen shot of the graphic user interface (GUI) of the custom MATLAB program recording explicit component for “bed.” After every trial, the custom program asked the participants to report whether they intended to change their speech (i.e., more like bid or more like bad) by moving the slide bar on the touchscreen. They were instructed to leave the bar in the middle if they did not try to change their speech.



KS

Gender

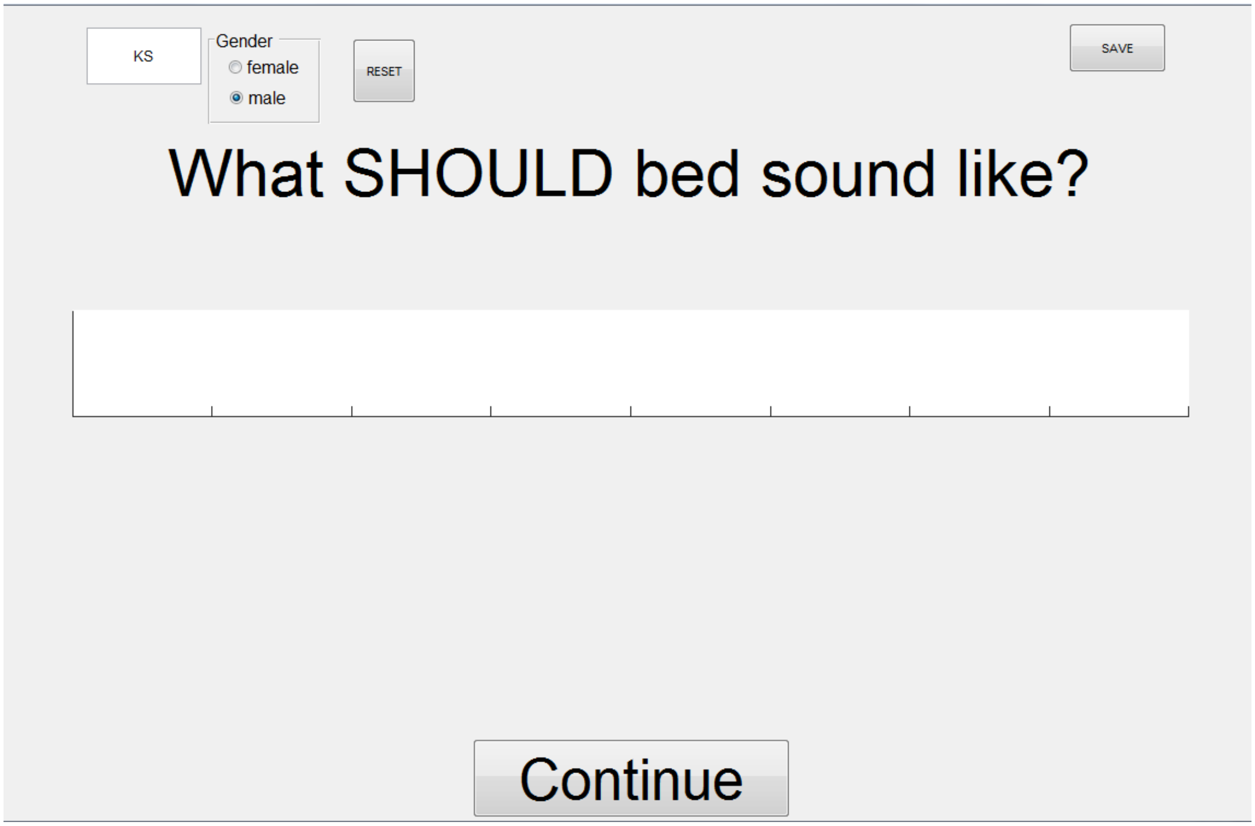
○ female

● male

Answer the question below and hit continue

Did you TRY to change your speech? In which direction?

Figure 3. A screen shot of the GUI recording auditory target for “bed.” After every 10 trials (with an exception of the first target obtained before the very first trial), the custom MATLAB program asked participants “what SHOULD bed sound like?” and “what SHOULD pet sound like?” The participants were instructed to drag/move the “wheel” shown in the middle of the screen (the white box in the middle) and try several different locations of the wheel, which generated their own pre-task recorded speech with different amount of F1 shift, before they made the final decision.



KS

Gender

○ female

● male

What SHOULD bed sound like?

Figure 4. Auditory-motor adaptation for AWS (Pink) and AWNS (Blue). All F1 are shown in cents, normalized to the baseline blocks (i.e., 6-10). Top: group mean F1 for each block across the baseline, perturbation, and aftereffect phases. Shaded areas correspond to standard errors. The gray lines represent perfect adaptation data (i.e., full adaptation in the perturbation phase, no adaptation whereas there is no perturbation). Bottom: the amount of early and late adaptation for each participant. Early adaptation (i.e., average of the first 5 blocks of the perturbation phase) shown in left, and late adaptation (i.e., average of last 5 blocks of the perturbation phase) shown in right.

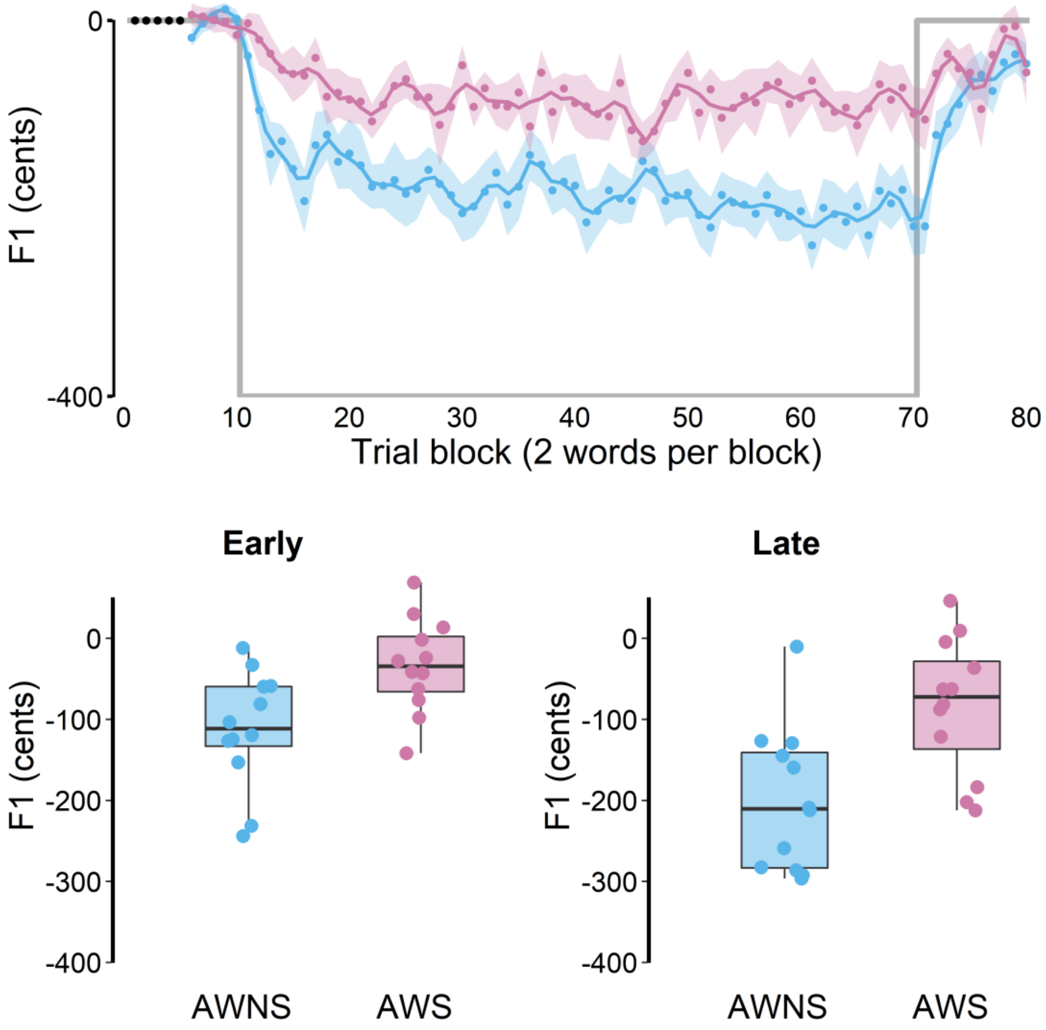


Figure 5. Group mean (AWS in pink and AWNS in blue) explicit component extracted from the responses to the question “Did you TRY to change your speech? In which direction?” throughout the experiment. The intent rating corresponds to the location of the slide bar that participants placed. -100% intent rating corresponds to slide bar placed in the most left (e.g., more like bid) and 100% intent rating corresponds to slide bar placed in the most right (e.g., more like bad). 0% intent rating represents that the bar was placed in the middle to indicate no change in their speech productions. The vertical gray lines represent when the perturbation started (left) and ended (right). The shaded areas correspond to standard errors.

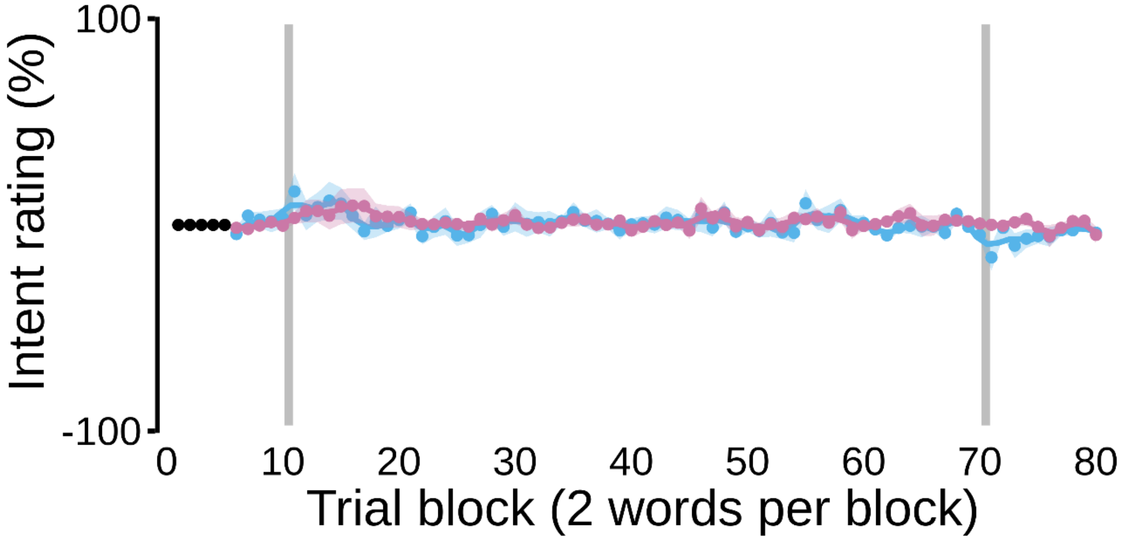


Figure 6. Explicit component responses from two participants (both AWNS) who consistently and clearly expressed some explicit components. One participant (shown yellow) indicated that he was trying to say more like “bad” and “pat” in the perturbation phase. Another participant indicated the opposite: he expressed that he was trying to say more like mount of F1 shift, moving the slide bar to the left. Each dot represents a block average. The vertical gray lines represent where the perturbation started (left) and ended (right).

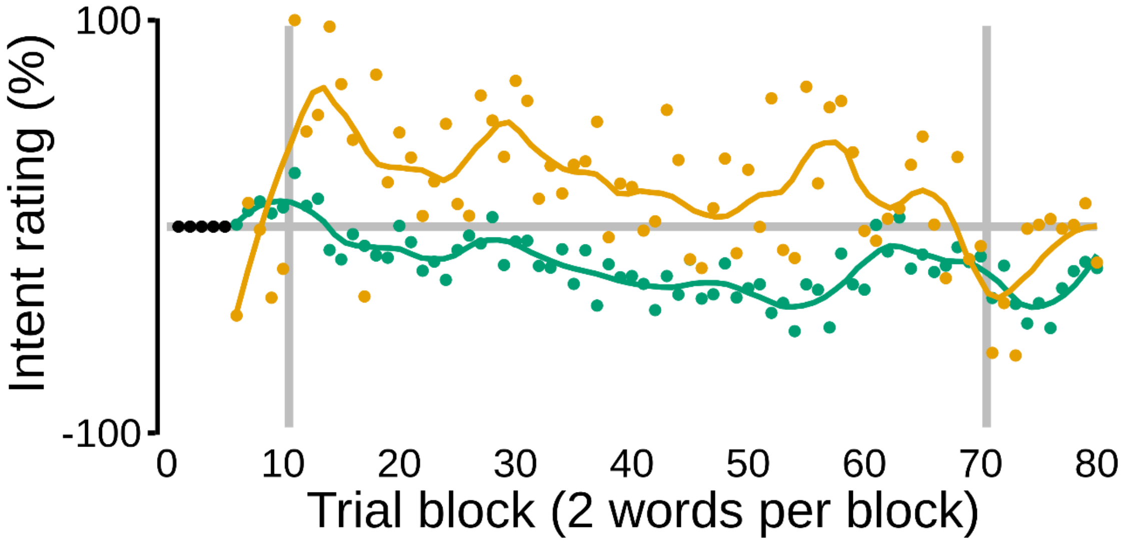
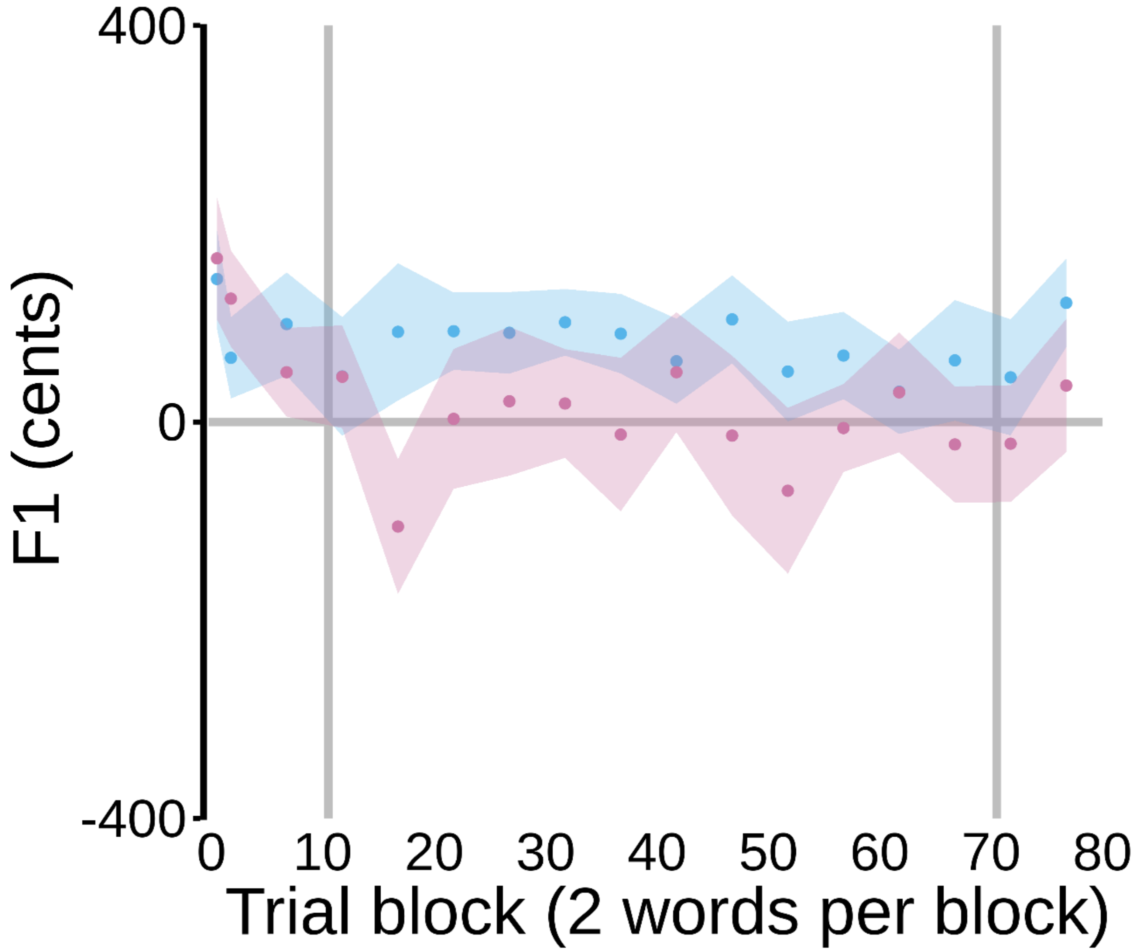


Figure 7. Group averages of F1 of auditory targets (AWS in pink and AWNS in blue), extracted from the participants' responses to the question "what should bed sound like?" and "what should pet sound like?" Each dot represents an average of the response for the two words. The shaded areas correspond to standard errors.



Chapter IV

Feedforward vs. feedback control of unperturbed speech movements in stuttering and nonstuttering adults

Kwang S. Kim[1], Vincent L. Gracco[2,3], Ludo Max[1,2,*]

[1] University of Washington, Dept. of Speech and Hearing Sciences, 1417 N.E. 42nd St., Seattle, WA 98105-6246 (USA)

[2] Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511 (USA)

[3] McGill University, School of Communication Sciences and Disorders, 2001 McGill College, 8th Floor, Montreal, QC H3A 1G1 (CAN)

[*] Corresponding author. E-mail: LudoMax@uw.edu. Phone: (206) 543-2674. Fax: (206) 543-1093

Abstract

Stuttering is associated with deficiencies in speech sensorimotor control. One suggestion has been that stuttering individuals excessively rely on afferent input to control ongoing movements. Feedback-driven control may interfere with speech fluency by triggering maladaptive movement corrections and reflect compensation for an inefficient feedforward control system. Surprisingly, the idea of over-reliance on feedback in stuttering has never been investigated in speech movements performed without experimentally induced mechanical or auditory perturbations. Here, we recorded kinematic data from unperturbed articulatory movements using a statistical approach previously described to estimate the relative contributions of feedforward and feedback mechanisms in reaching (Heath et al., 2004; Messier & Kalaska, 1999), saccades (West et al., 2009) and speech movements (Kim & Max, 2014). Specifically, we compared 13 stuttering adults with age- and sex-matched nonstuttering adults in their tongue and jaw opening and closing movements for different combinations of consonants and vowels. We determined how well initial kinematic landmarks (peak acceleration, peak velocity) and found strong correspondence between those initial and final kinematics, which suggests that both groups are largely under feedforward control. Yet, this relationship was significantly weaker for stuttering speakers. Additionally, estimates of feedback-based adjustments in movement duration were significantly larger for stuttering speakers. Thus, findings are consistent with the notion that sensorimotor control processes in stuttering adults show an increased reliance on online sensory feedback control.

Keywords: stuttering; speech; motor control; sensorimotor integration; feedback; kinematics

Introduction

Stuttering is a disorder of speech fluency characterized by involuntary repetitions of sounds or syllables, prolongations of sounds, and complete blocks during which sounds cannot be initiated. A large body of evidence suggests that stuttering is, in essence, a movement disorder associated with deficiencies in sensorimotor integration (see reviews in, among others, Bloodstein & Ratner, 2008; Max, 2004). Parallel lines of work demonstrate that these speech sensorimotor problems are accompanied by structural and functional brain differences in sensory, pre-motor, and motor regions (e.g., Beal et al., 2013; Chang et al., 2008; Chang & Zhu, 2013; Neef et al., 2015; Watkins et al., 2008) and even sensorimotor differences in the control of orofacial and limb nonspeech movements (e.g., Daliri et al., 2013; Daliri et al., 2014; Max et al., 2003).

The exact neurophysiological mechanisms underlying stuttering remain unknown, but recent electrophysiological work from our own laboratory has revealed that an atypical motor-to-sensory modulation can be detected in adults who stutter (AWS) even during speech planning prior to movement onset (Daliri & Max, 2015a; Daliri & Max, 2015b; Daliri & Max, 2018; Max & Daliri, 2019). In addition, auditory-motor adaptation—driven entirely by discrepancies between auditory prediction and actual auditory outcome (Kim & Max, in preparation)—seems to be reduced in AWS (Daliri et al., 2018; Daliri & Max, 2018; Kim et al., in preparation; Loucks et al., 2012; Sengupta et al., 2016) and absent in children who stutter (Kim et al., in preparation). The clear limitations in auditory-motor learning add support to the idea that at least some aspects of sensory prediction and/or planning accurate motor commands (i.e., feedforward control) that can account for previously experienced sensory errors may be deficient in individuals who stutter (Max et al., 2004; Max, 2004).

In light of the above findings, it may be productive to consider stuttering within a general theoretical framework that suggests combined feedforward control (i.e., preplanning a movement) and feedback control (i.e., adjustments of the ongoing movement) for voluntary movements (e.g., Desmurget & Grafton, 2000; Seidler et al., 2004; Wolpert & Flanagan, 2001). As a result of abnormal sensorimotor prediction and feedforward control, PWS may exhibit an over-reliance on feedback—in particular auditory feedback—for the online control of speech movements (Max, 2004; Max et al., 2004; Max & Daliri, 2019). According to Max et al. (2004), such over-reliance on movement corrections would result in instability in the sensorimotor system, leading to stuttering moments.

A convincing evidence for over-reliance on feedback control in PWS comes from a computational simulation by the Directions Into Velocities of Articulators (DIVA) model (Civier et al., 2010; Max et al., 2004). When a bias toward feedback control was introduced in the DIVA model, a “reset” signal was triggered, and this led to stuttering-like behavior (i.e., sound repetition). In other words, an over-reliance on feedback caused the system to interrupt sounds and attempt repairs. When the model’s outputs were compared to acoustic data from AWS’ fluent speech, and to combined acoustic and articulatory movement data from AWS’ dysfluent speech, the model resembled important aspects of the human data (Civier et al., 2010).

In addition, both in laboratory experiments and in the clinical management of stuttering, many individuals who stutter experience partial alleviation of their symptoms when auditory feedback is made unavailable (e.g., masked auditory feedback, MAF) or unreliable (e.g., delayed auditory feedback, DAF, or frequency altered auditory feedback, FAF) (see Lincoln et al., 2006; Ritto et al., 2016). Thus, consistent with the hypothesis of an over-reliance on auditory feedback under typical circumstances, a fluency-enhancing effect occurs when auditory feedback is distorted. Furthermore, a recent study found that DAF influenced the perception of simultaneity in AWS to a greater extent compared to AWNS (Iimura et al., 2019).

Prior empirical studies have investigated the reliance on feedback mechanisms for speech by comparing AWS and control subjects in terms of their responses to perturbations in the somatosensory (e.g., tendon vibration) or auditory (e.g., formant-shifted feedback) domain (Cai et al., 2012; Loucks & De Nil, 2006; Namasivayam et al., 2009). Apparently contradicting the hypothesis that suggests an over-reliance on feedback in stuttering, these studies typically have found that in the presence of perturbed afferent input, AWS show *weaker* responses in both online movement control (Cai et al., 2012; Cai, Beal et al., 2014; Loucks & De Nil, 2006). Nevertheless, there are several potential explanations for these findings. First, the perturbations applied in these studies are rather large, which are not often experienced in everyday life. Thus, the sensorimotor system may perceive, especially with poor sensory prediction, the huge and unexpected perturbation in sensory feedback as irrelevant sensory errors that cannot be controlled. As a result, the system may not attempt to correct for the error even if it over-relies on the online feedback control. Second, it is also possible that even if the sensorimotor system attempts to compensate for the perturbation, the resulting compensatory response generated by poor sensory prediction may be suboptimal in PWS.

Therefore, it is highly desired to examine feedforward vs. feedback contributions in unperturbed speech movements. In nonspeech studies, individual subjects' movement control strategies in terms of feedforward vs. feedback contributions have been studied using at least one method in which movements remain entirely unperturbed (Heath et al., 2004; Kim & Max, 2014; Messier & Kalaska, 1999; West et al., 2009). If movement parameters are completely determined prior to onset (i.e., feedforward control), the initial kinematics (peak acceleration, peak velocity, and effector position at the time of peak acceleration and velocity) can be expected to show a high degree of correspondence with the final kinematics at movement endpoint. If, during execution, movements are modified based on incoming sensory feedback, the initial kinematics can be expected to be less closely related to movement endpoint (Heath et al., 2004; Messier & Kalaska, 1999). Our laboratory has previously demonstrated the feasibility of using this same approach to examine feedforward vs. feedback mechanisms in the control of speech movements in nonstuttering adults (Kim & Max, 2014). Instead of limb reaching movements to targets at different distances, participants perform articulatory movements of different amplitude for the production of consonant-vowel (opening movements) or vowel-consonant sequences (closing movements). The targets contain vowels of different heights (high, mid, low) to elicit a wide variation of movement amplitudes. Tongue, lip, and jaw movements are transduced with an electromagnetic motion tracking system. The dependent variables, correlational analyses, and mathematically-derived estimate of feedback contributions are as described by Messier & Kalaska (1999). Our correlational results for typical speakers (Kim & Max, 2014) were highly similar to those previously reported for the upper limb during reaching: articulatory speech movements are predominantly, but not entirely, pre-determined by moment onset (i.e., feedforward control).

Here, we report (a) an application of the same mathematical approach to a data set obtained from thirteen AWS who individually match in age, sex, and handedness and the thirteen AWNS (five of which reported in Kim & Max, 2014), and (b) a statistical between-group comparison between stuttering versus nonstuttering speakers.

Methods

Subjects

Participants were thirteen stuttering males (*M* = 29, *SD* = 9.87, min = 18, max = 49) with no other diagnosed communication disorders other than stuttering and their age-matched (± 3 years) nonstuttering males (M=30.46, SD= 10.10, min = 18, max = 48, the data from five participants from the group previously reported in Kim and Max, 2014). All participants were native speakers of American English. Most participants passed a pure tone behavioral hearing screening (20 dB HL or better for the octave frequencies from 250 to 4000 Hz) except one pair of participants who each had a higher threshold at 4000 Hz: one AWS had 30 dB in the left ear whereas his control match had 30 dB in the right ear (AWNS). Most participants were matched according to handedness as well (10 right-handed and 1 ambidextrous), however, one right handed AWS was matched with an ambidextrous AWNS and one left-handed AWS was matched with a right-handed AWNS. Stuttering Severity Index (SSI-4, Riley) was administered on the day of the experiment. The severity ranged from very severe to very mild (1 very severe, 1 severe, 3 moderate, 5 mild, and 3 very mild).

Procedure

After giving written informed consent, each participant produced 7-20 trials for each of 36 different utterances containing target syllables of consonant-vowel (CV) or vowel-consonant (VC) form. The latter type of syllable corresponds to articulatory closing movements (e.g., the tongue tip moving up and obstructing the vocal tract against the alveolar ridge during the production of “eat”) whereas the former type corresponds to articulatory opening movements without mechanically constrained endpoint (e.g., the tongue tip moving down for “tea”). We attempted to record a full set of 540 (15×36) utterances per participant, but some recording sessions were slightly shorter due to dissolving of the intra-oral adhesive that secured motion tracking sensors to the tongue and gums (see below).

Target syllables combined one of two voiceless stop consonants (/t, k/) or the voiceless fricative (/s/) with a high, mid, or low front (/i, ɛ, æ/) or back (u, ɔ, ɑ) vowel in order to elicit movements of different amplitude (e.g., small amplitude opening movement for a high vowel, large amplitude opening movement for a low vowel). These target syllables were embedded in utterances constrained by the following rules to allow unambiguous movement segmentation in the offline data analyses: (a) the consonants immediately preceding and following the target vowel always shared the same place and manner of articulation as the target consonant, and (b) the vowel preceding the target consonant of a CV syllable or following the target consonant of a VC target was always schwa. As a result, the recorded movement trajectories contained clear reversals in effector movement direction before and after the targeted opening or closing movements. Within these constraints, the following utterances were created:

He said a t*V* to me.	He said *V*t again.
He spoke a k*V* quietly.	He spoke *V*k again.
He says a s*V* so well.	He says *V*s again.

The use of non-words as target syllables can lead to pronunciation ambiguity if these syllables are displayed in orthographic spelling. Therefore, subjects were first familiarized (by means of example words) with the six International Phonetic Alphabet symbols for the target vowels. Then, during the actual data recording, the utterance for each trial was shown on a computer monitor with the target vowel displayed as the corresponding phonetic (rather than orthographic) character and with the example word also shown at the bottom of the monitor at the same time. In addition, when the utterance appeared on the monitor, a computer loudspeaker played back a pre-recorded sound file of that same utterance spoken by a young adult woman. After this combined visual-auditory presentation, the participant produced the utterance.

Data recording

Movements from were recorded with electromagnetic articulographs (EMA; Carstens AG200 for five participants in each group, AG500 for the rest). Sensors were attached to the tongue and jaw with a cyanoacrylate adhesive (Cyanodent, Ellman International). Three sensors were placed on the tongue: the first sensor (T1) approximately 1 cm from the tongue tip, the second and third sensors (T2, T3) approximately 1.5 and 3 cm posterior to T1. One sensor for the jaw (J) was attached to the mandibular gums below the lower central incisors. For AG200, two reference sensors were attached to the bridge of the nose and the maxillary gums above the upper central incisors (with double-sided adhesive tape and Cyanodent, respectively). For AG500, instead of the upper incisors, two sensors on each mastoid were attached using double-sided adhesive tape. All data were then corrected relative to the head movements using the reference sensors (i.e., by calculating the coordinates of moving sensors T1, T2, T3, and J relative to the stationary reference sensors). At each recording session, the participant was asked to briefly clench a bite plate between their upper and lower teeth (a custom-made for AG200 and a manufacturer made for AG500). The bite recording made it possible to transform all data from the original coordinate system into an anatomically-defined coordinate system that is consistent across participants. The AG200 was transformed into the anatomical coordinate system that places the x-axis in the individual participant's occlusal plane and the origin at the tip of the upper incisors (Max et al., 2003; Okadome & Honda, 2001; Perkell & Zandipour, 2002). For the AG500 data, we used the rotation-preset provided by the manufacturer which places the x-axis also in the occlusal plane and the origin at the bottom tip of the upper incisors.

Signal processing and data extraction

First, kinematic data were zero-phase filtered with 100th order FIR low-pass filters designed with a hamming window. The cut-off frequency was 15 Hz for sensors on moving articulators (T1, T2, T3, J) and 5 Hz for stationary reference sensors (Green et al., 2000; Green et al., 2002; Lucero et al., 1997). Second, signals from the moving articulators were corrected for head movement, and then transformed into the anatomically-defined coordinate system. Third, for each sensor, all movement paths for a given combination of subject, consonant, and syllable type were spatially shifted such that they started from a common point with coordinates corresponding to the average x and y coordinates of all starting points within the high, mid, or low vowels produced in that particular condition (Kim & Max, 2014). For the tongue tip movements associated with the alveolar consonants /t/ and /s/, data from sensor T1 were used for all analyses. For the tongue dorsum movements associated with the velar consonant /k/, data from sensor T3 were used.

Directly based on the limb motor control work by Messier & Kalaska (1999) and then adapted for speech motor control in Kim & Max (2014), we measured peak tangential acceleration, peak tangential velocity, and movement extent. Our Matlab analysis code first detected a peak in the tangential velocity profile within the time window of interest. The local minima immediately preceding and following this tangential velocity peak were then automatically identified as the movement’s start and end times. In a few cases, one or both of these local minima in the tangential velocity signal could not be identified automatically. For those trials, visual inspection of the acceleration profile was used to determine movement start and end times (given that a tangential acceleration peak near the expected time region indicates that tangential velocity increased/decreases relatively suddenly). Peak tangential acceleration was defined as the maximum acceleration value within the time window from movement onset to peak tangential velocity. Movement extent was defined as the straight-line distance between the positions at movement start and end time.

Data Selection

In the previous study (Kim & Max, 2014), we included high, mid, and low vowels in the target syllables to elicit movements of different amplitude and found that it was generally successful, but not in all cases. Thus, given that the underlying rationale assumed the inclusion of movements with different target distances, we previously selected all productions with the consonants (/t, s/) and the front vowels (/i, ɛ, æ/). Before analyzing the data, all utterances with perceptually noticeable stuttering or speech errors had to be excluded. We also excluded the utterances if any other syllables in the given sentence had stuttering moments due to their potential influences on the target syllable. These judgements were made by a certified speech-language pathologist and a graduate student with appropriate training.

In addition, utterances that were spoken incorrectly spoken or contained noise in the kinematic data had to be rejected. In fact, for one AWNS consistently spoke all *V*t utterances without the actual consonant (e.g., "eh"), so the data for those utterances had to be rejected. Importantly, as in the previous study, utterances with multiple tangential velocity peaks were found and discarded due to their ambiguous kinematic parameters (i.e., multiple velocity and acceleration peaks). We reported the proportion of these movements that had to be rejected and examined any group differences with Welch's t-test along with Holm-Bonferroni adjusted p-values for multiple comparisons (Holm, 1979). The resulting final data set were similar between the two groups, containing 1,964 (jaw) and 1,790 (tongue) utterances for the stuttering participants (81-226 per participant for tongue, 88-226 per participant for jaw) and 1,779 (jaw) and 1,725 (tongue) utterances for the nonstuttering participants (69-206 per participant for tongue, 66-211 per participant for jaw).

Data analysis

For feedforward control estimation, we calculated—for each combination of individual subject, articulator, syllable type, and consonant—the correlation coefficients between peak acceleration and movement extent and between peak velocity and movement extent. These correlations were calculated across the three front vowels paired with a given consonant (/t/ or /s/) and syllable type (CV or VC). Hence, for each participant, there were a total of four correlation coefficients between peak acceleration and movement extent, and another four coefficients between peak velocity and movement extent.

For the estimation of online feedback control contributions, we first checked if the peak kinematic parameters also correlate with movement duration. Interestingly, the correlations are mostly positive across the vowels, but mostly negative *within* each vowel. This assured that for a given target, the movement duration was shorter for larger peak acceleration/velocity and longer for smaller peak acceleration/velocity. It should be noted that such relationship is different from (a) peak acceleration and peak velocity scaling with movement extent for movements across and within different targets (Kim & Max, 2014) and (b) the positive correlation found between movement duration and initial kinematics *across vowels*. These differences suggest that the movement duration was not planned as peak velocity and acceleration variables were, rather, it was adjusted especially towards the latter part of the movement in order for the movement to reach the given target. Therefore, we added movement duration into a multiple correlation model in order to examine whether it significantly increases the explained variance for movement extent, compared to the amount explained by the bivariate correlation between peak acceleration or velocity and movement extent alone (Gordon & Ghez, 1987; Kim & Max, 2014). If movement duration was planned with acceleration/velocity and varied in the same way with distance, then adding duration to a multiple regression model should not significantly increase the explained variance. If adding movement duration to such a model significantly increases explained variance in movement distance compared with the bivariate model, it would suggest that duration plays a compensatory role, with adjustments occurring after peak acceleration/velocity have already been reached. For mathematical details, please refer to Kim & Max, (2014).

To examine any group differences in feedforward and feedback estimations, linear mixed effects models were constructed using lmer function from R package lme4 (Bates et al., 2015). We computed maximum likelihood estimates for fixed effects of group (AWNS vs. AWS), other experimental variables (consonants: /t/ vs. /s/, syllable types: CV vs. VC, articulators: Jaw vs. Tongue), as well as the interactions between the variables while keeping subjects as a random variable. Type III analysis of variance tests using Satterwaite approximation from lmerTest package were conducted to calculate the degrees of freedom and p-values (Kuznetsova et al., 2017).

Results

General observations

Representative tongue movements from a single control subject's CV trials for the consonant /t/ are illustrated in Figure 1(A). Figure 1(C) shows that of a stuttering subject for the same utterance. Overall, the tangential velocity profiles are generally bell-shaped, scale with movement extent and contain typical acceleration and deceleration phases. Both the individual subject data in Figure 1(A) and Figure 1(C) and the group mean data for jaw extent in Figure 1(B) confirm that, for these front vowels, the speech task achieved its goal of eliciting movements that differed in extent across the three categories of vowel height. Although not included in Figure 1, jaw movements generally followed the same pattern as described above for tongue movements.

When we tested for difference in kinematic measures between the two groups using linear mixed effects model, there was no group difference in peak acceleration, peak velocity, and movement extent ($p > 0.05$). For each model constructed, there was a significant group × vowel interaction, however, post-hoc pairwise comparisons revealed that there was no group difference in those kinematic measures for each vowel ($p > 0.05$). We also compared the temporal measures in our data, namely time of peak acceleration, time of peak velocity, and movement duration. Again, there was no significant fixed effect for group, but there was a significant interaction between group and consonant for time of peak acceleration and a three-way significant interaction between group, articulators, and syllables. For both of these cases, post-hoc analyses pairwise comparisons revealed that there was again no group difference in those temporal measures given each consonant or articulator/syllable type. Combined together, the result suggests that there is essentially no difference in kinematic measures between the two groups and the movement kinematic profiles between the two populations. Given that some trials were removed before this analysis and only non-stuttering movements that followed the standard bell-shaped velocity profiles (see Methods) remained for both populations, this finding is not surprising.

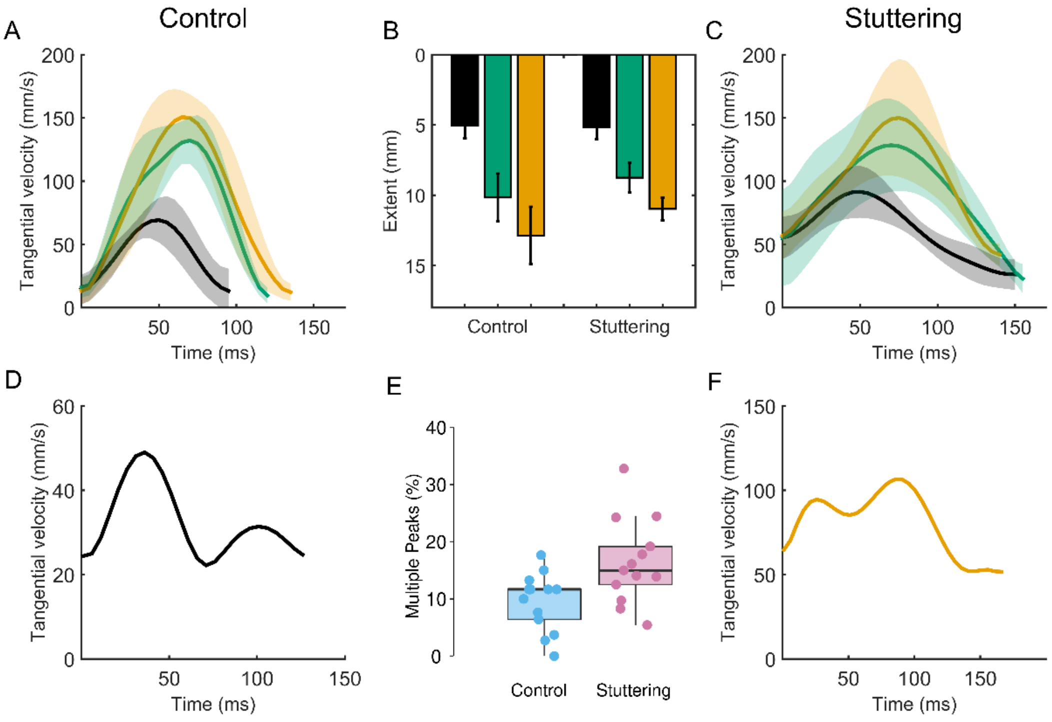


Figure 1. (A) Example of a single nonstuttering subject's tongue movement data for CV targets with consonant /t/. Each color shows the time series of tangential velocity averaged for high (black), mid (green), and low vowels (yellow) (shaded areas indicate SE).

(B) Group mean (with SE) movement extent for high, mid, and low vowels in the same target syllables for the tongue movements.

(C) Example of a single stuttering subject's tongue movement data for CV targets with consonant /t/ to compare with panel A.

(D) Example of a nonstuttering subject's trial with multiple peaks in tangential velocity (high vowel).

(E) The proportion of the tongue movements that were removed due to having multiple peaks in the tangential velocity profiles.

(F) Example of a stuttering subject's trial with multiple peaks in tangential velocity (low vowel)

AWS have more multiple peaks in tongue movements

As previously mentioned, some trials were removed before further analyses because the tangential velocity profiles contained double or multiple peaks (examples shown in Figure 1D and 1F). The proportion of jaw movements that were removed due to having multiple velocity peaks did not differ for the groups ($p > 0.05$). However, for tongue movements, stuttering individuals had nearly significantly more movements with multiple peaks than nonstuttering individuals, t(21.243) = -2.778, p = 0.022, d = -1.089 (see Figure 1E). Note that utterances that had to be rejected for other reasons (e.g., incorrect production, stuttering, among others) were counted as non-multiple peak movements.

AWS have lower correlation coefficients between peak acceleration and extent

The correlation coefficient was computed between peak acceleration and movement extent across all three vowels for all possible combinations of the syllable and consonant types (see examples in Figure 2). In most cases, peak acceleration was highly correlated with movement extent. For AWNS, correlation coefficients (between peak acceleration and extent) averaged across the four different syllable combinations (/t/ + V, /s/ + V, V + /t/, V + /s/) ranged .66 -.91 for the tongue and .79-.96 for the jaw. For AWS, the ranges were slightly lower, .65-.86 for the tongue and .73-.91 for the jaw movements (Figure 3). When tested with linear mixed effects model for the correlation coefficients, there was a significant group difference, F(1, 25.806) = 7.564, p =0.011. There was also a significant fixed effect of the articulator variable (the tongue vs. jaw movements). There were also two significant interactions: syllable × consonant, F(1, 180.43) = 4.071, p = 0.045, and articulator × consonant, F(1, 179.83) = 4.630, p = 0.033. However, there was no significant interaction with the group effect, suggesting that the group difference was consistent across different utterances and articulators.

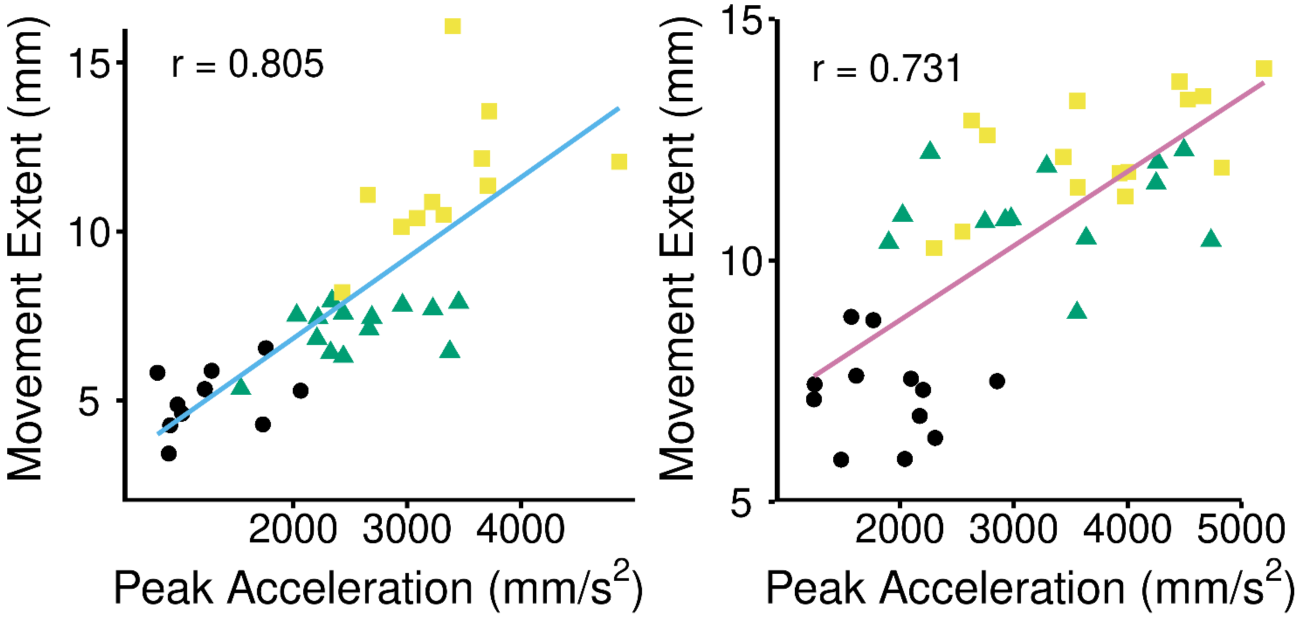


Figure 2. An example relationship between peak acceleration and movement extent from AWNS (left) and for AWS (right). In both cases, the utterances illustrated were t+*V*. High vowels (/ti/) shown in black circle, mid vowels (/tɛ/) shown in green triangles, and low vowels (/æ/) shown in yellow squares.

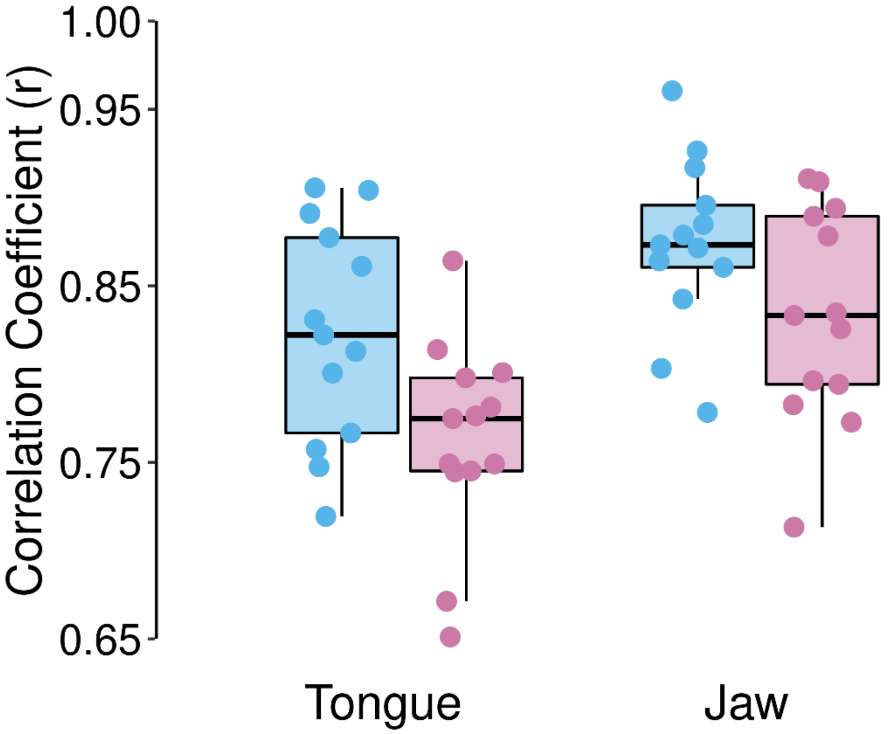


Figure 3. The correlation coefficient between peak acceleration and movement extent for each participant, averaged across all possible combinations of the syllable and consonant types. AWNS (blue) and AWS (pink).

AWS have lower correlation coefficients between peak velocity and extent

For both AWS and AWNS, the correlation coefficients between peak velocity and movement extent were larger than those between peak acceleration and movement extent. This is expected given that peak velocity arises much after peak acceleration, peak velocity can predict the movement extent much better than peak acceleration. For AWNS, the correlation coefficients (between peak velocity and movement extent) averaged across the four different syllable combinations (/t/ + V, /s/ + V, V + /t/, V + /s/) ranged from .92-.98 for tongue movements and .93-.99 for jaw movements. Again, these values were mostly slightly lower in AWS, .86-96 for tongue movements and .94-.98 for jaw movements (Figure 4). Analyzing the correlation coefficients with a linear mixed effects model, we found a significant fixed effect for group, $F(1, 24.793) = 14.061$, $p < 0.001$. The correlation coefficients also differed across different articulators, $F(1, 178.805) = 19.127$, $p < 0.001$, and syllables, $F(1, 179.634) = 25.184$, $p < 0.001$. There were also three significant interactions, each between syllable and consonant types, $F(1, 179.634) = 6.008$, $p = 0.015$, articulators and consonant types, $F(1, 178.805) = 6.039$, $p = 0.015$, and syllable types and articulators, $F(1, 178.805) = 5.855$, $p = 0.017$. Again, there was no significant interaction with group, suggesting that the group difference found is independent of the utterances.

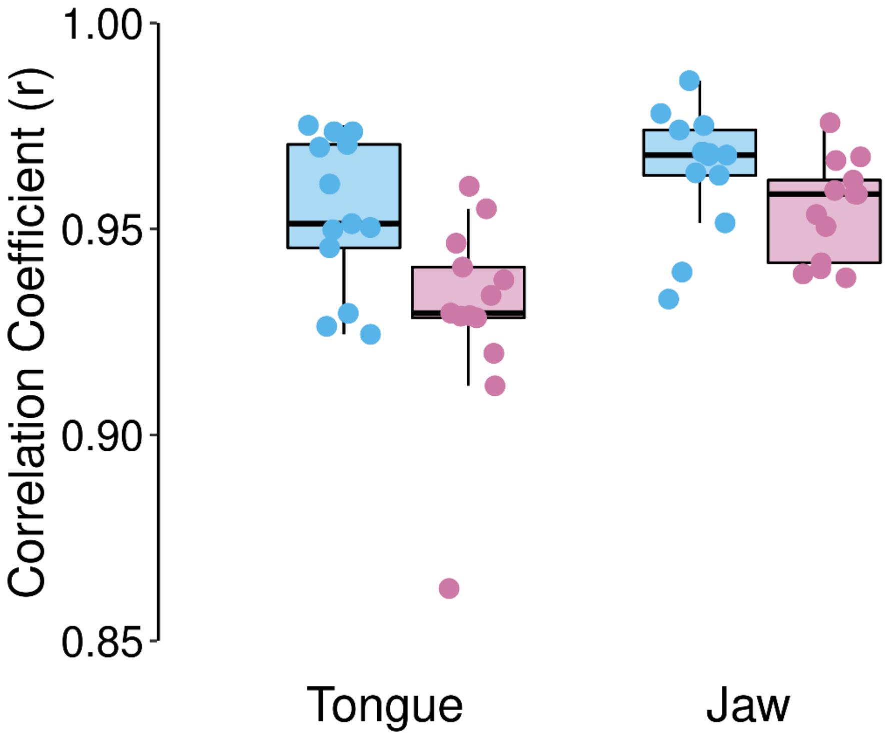


Figure 4. The correlation coefficient between peak velocity and movement extent for each participant, averaged across all possible combinations of the syllable and consonant types. AWNS (blue) and AWS (pink).

Both groups show significant increases in the multiple regression model when duration is added

A previous study reported that movement duration may be adjusted online to generate feedback control based on the afferent information from the time of peak acceleration (see Kim & Max, 2014). The finding was supported by the apparent negative relationships between the duration and peak acceleration across the movements of the same target (within vowels) among nonstuttering individuals. Here, we also found a similar result for both groups. Both groups had mostly positive correlation between peak acceleration and duration across vowels (75.5% for nonstuttering individuals and 67.65% for stuttering individuals) but mostly negative correlation across the movements within each vowel (76.39% for nonstuttering individuals and 88.71% for stuttering individuals).

In this study, we also found that peak velocity (when scaled with movement duration) follows the same pattern (see an example in Figure 5). Across vowels, the correlation coefficients were mostly positive (87.25% for nonstuttering individuals and 85.29% for stuttering individuals) whereas they were mostly negative within each vowel (53.53% for nonstuttering individuals and 68.12% for stuttering individuals). If movement duration was mostly planned along with kinematic parameters, it is very unlikely that the CNS plans the variable to be proportional to movement extent across vowels but inversely proportional to movement extent within each vowel at the same time. The discrepancy between correlation coefficients across vowels and within each vowel shows that movement duration is not planned at the movement onset but rather changing online along the movements.

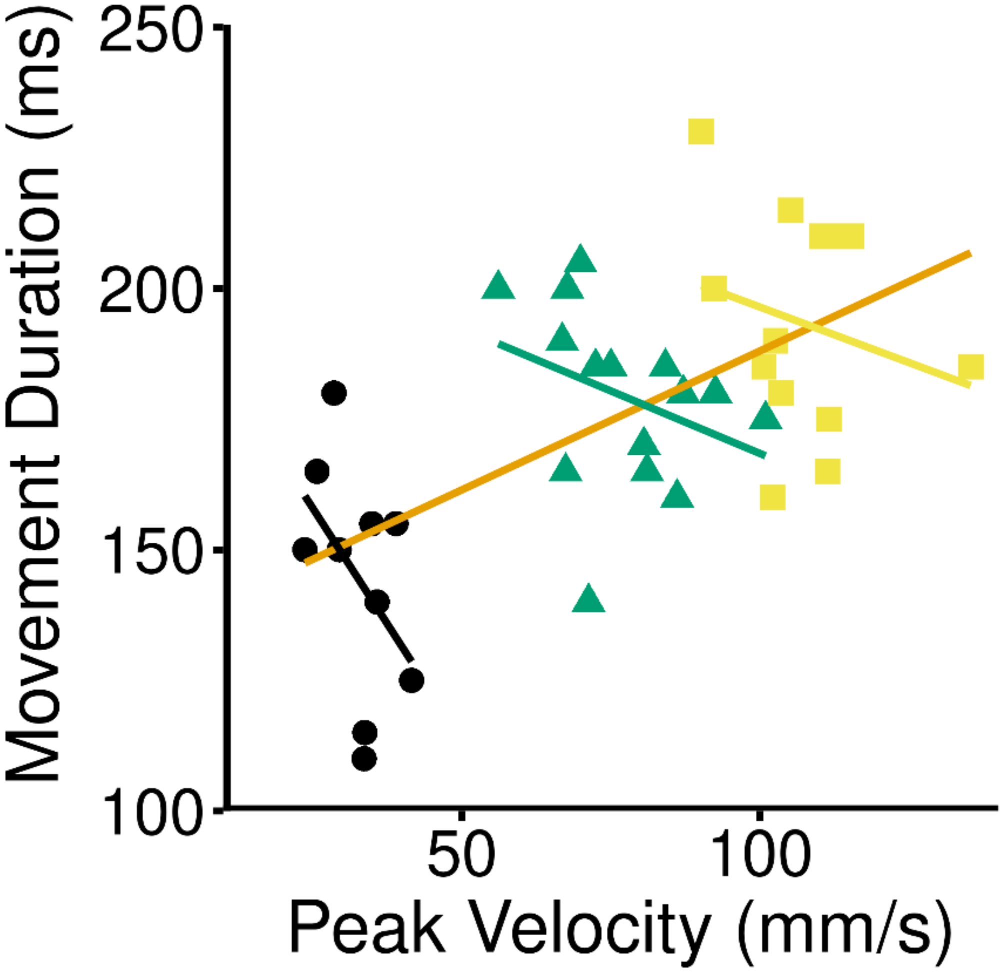


Figure 5. The positive correlation between peak velocity and movement duration *across vowels*, but negative correlation *within each vowel*. These suggest that movement duration was not planned, but rather adjusted for the movement to reach each target vowel. This is an example data from an AWS for *V* + /s/ utterances. Black circles = /is/, green triangles = /ɛs/, and yellow squares = /æs/.

Therefore, in order to estimate online compensation (i.e., feedback control), we calculated the *increase* in explained movement extent variance that resulted from taking movement duration into account in addition to peak acceleration ranged across subjects and target syllables. In case of peak acceleration, this *increase* in explained movement extent variance ranged from 8 to 28% for tongue movements and 3 to 24% for jaw movements for AWNS whereas they ranged 13 to 32% for tongue and 5 to 25% for jaw in AWS. Importantly, the *F* statistic calculated to test the increase in explained variance for each individual combination of subject and target syllable was statistically significant (uncorrected p values < 0.05) in 102/103 cases for nonstuttering individuals and 98/104 cases for stuttering individuals (i.e., 13 subjects × 2 syllable types × 2 consonants × 2 articulators = 104 cases, but notice that it is 103 due to an AWNS' data being rejected for all *V* + /t/, see Methods for more details).

The movement variance could also be explained more by taking movement duration into account in addition to peak velocity (see an example in Figure 6). For nonstuttering individuals, the increases ranged from 2 to 9% for tongue movements and 1 to 6% for jaw movements, whereas stuttering individuals had 3 to 13% increases for tongue movements and 3 to 7% increases in jaw movements. Despite of the generally small increases, the vast majority of added variances were mostly statistically significant according to the *F* statistic (101/103 for nonstuttering individuals and 102/104 cases for stuttering individuals.).

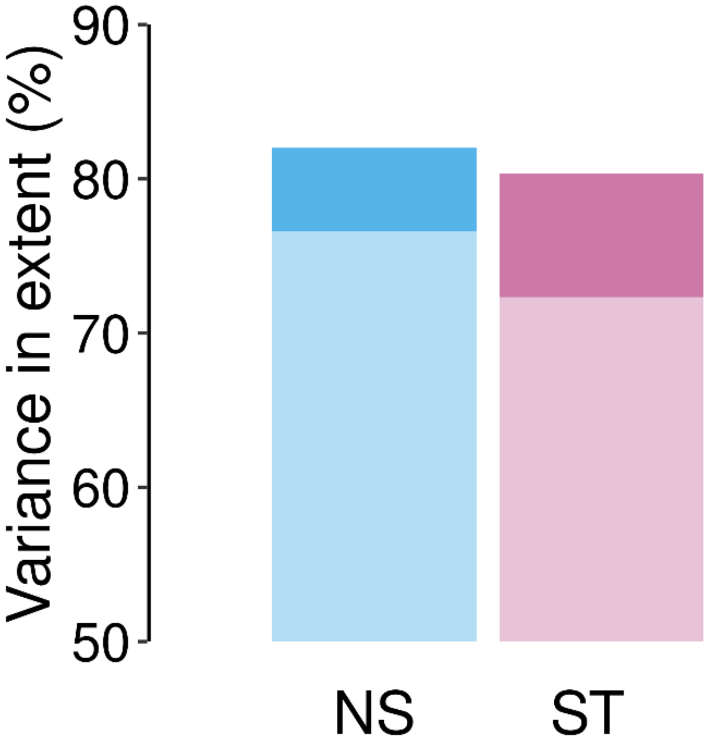


Figure 6. An example variance for movement extent in an AWS and an AWNS. The top portions of each bar illustrate by darker colors indicate the variance in movement extent that was explained by adding the duration. The bottom portions of the bars illustrated by lighter colors show the variance explained by the bivariate correlation coefficients with peak velocity. These increases in multiple regression models were mostly statistically significant in both groups. NS = nonstuttering individuals, ST = stuttering individuals.

AWS show more increases in the multiple regression model when duration is added together with peak velocity, but not with peak acceleration

The linear mixed effects model revealed that the increases in variance were not different between the groups when movement duration was added to peak acceleration, F(1, 24.865) = 1.675, p = 0.207 (Figure 7). There was a significant fixed effects for syllable types, F(1, 179.504), = 6.613, p = 0.011, and articulators, F(1, 178.848) = 9.089, p = 0.003. There were no significant interactions. Interestingly, the increases in variance when movement duration was added to *peak velocity*, the increases were significantly different between the groups, F(1, 23.247) = 5.563, p = 0.027 (Figure 8). There was no significant interaction with the group but one fixed effects by the syllable types, There was also a significant fixed effect for syllable types were found, F(1, 177.761) = 8.850, p = 0.003.

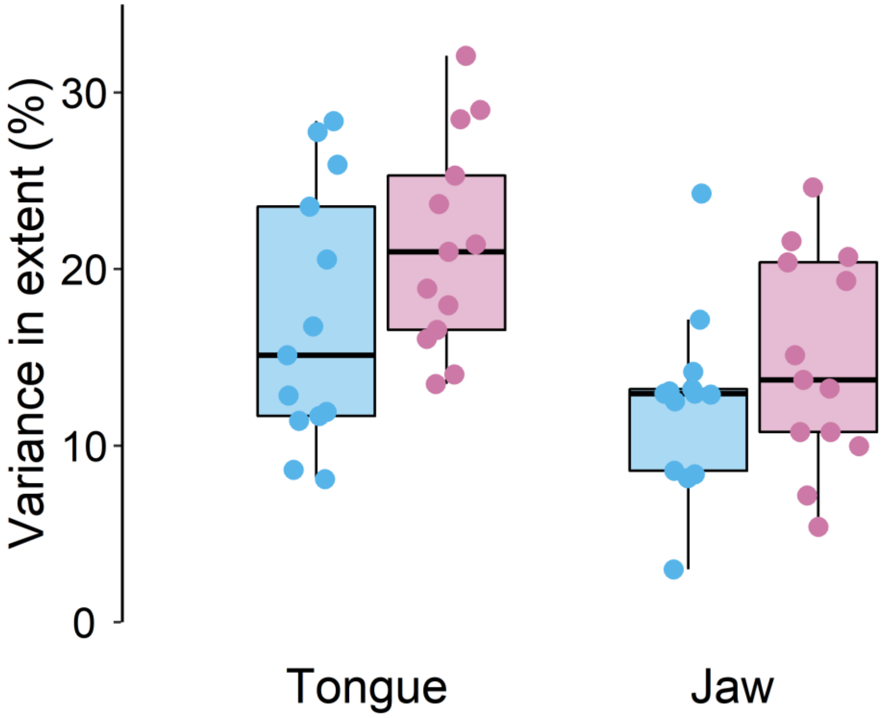


Figure 7. The variance in movement extent increased by adding duration in the multiple regression models with peak acceleration. Each dot represents average variance across the four different utterances for each participant. AWNS (blue) did not significantly differ from AWS (pink).

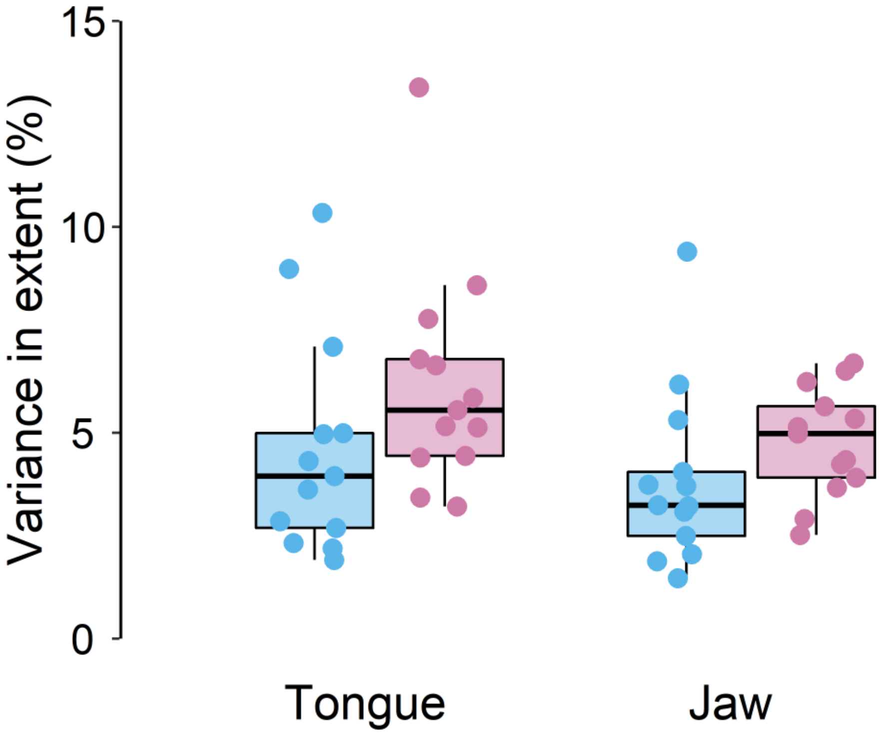


Figure 8. The variance in movement extent increased by adding duration in the multiple regression models with peak velocity. Each dot represents average variance across the four different utterances for each participant. AWS (pink) had significantly larger variance than AWNS (blue).

Discussion

In this study, we estimated contributions from feedforward and online feedback control mechanism in speech movements of AWS and AWNS. After recording speech articulatory movements using electromagnetic articulographs (AG200, AG500, Carstens), several kinematic landmarks (peak acceleration, peak velocity, movement endpoint) were extracted. The correlations between initial kinematic parameters (i.e., peak acceleration, peak velocity) and movement extent were then computed to estimate the feedforward control contributions. In addition to determining the correlation coefficients between the pairs of early and endpoint measures, we also used a statistical model to directly estimate feedback-based compensatory adjustments (Kim & Max, 2014; Messier & Kalaska, 1999). Specifically, the difference between (a) the multiple correlation coefficient for acceleration (or velocity), movement extent, and one dependent variable that may reflect compensatory influences (i.e., movement duration) and (b) the univariate correlation between acceleration (or velocity) and movement extent was used to estimate the contribution from online feedback control.

Double or multiple velocity peaks

Prior to estimating feedforward vs. feedback control, we found that stuttering individuals had more tongue movements with double or multiple tangential velocity peaks than nonstuttering individuals. One possible interpretation may be that the double/multiple velocity peaks in AWS represent online corrections. In the upper-limb motor literature, the secondary submovements in movements with double peaks have been suggested to be associated with sensory adjustments (Crossman & Goodeve, 1983; Meyer et al., 1990) and corrective responses (e.g., Desmurget et al., 2004; Milner, 1992) to successfully reach targets. Therefore, the larger proportion of multiple velocity peaks may suggest that AWS was correcting their movements online more compared to AWNS (i.e., larger online feedback control) in order to reach the target vowel.

Some studies have suggested that there are several different types of submovements and at least some of them might not be corrective in nature (e.g., Fradet et al., 2008). In fact, double peaks in the tangential velocity profile can also be found in smaller movements (Fradet et al., 2008) or movements with curvilinear trajectories (Abend et al., 1982; Quinn et al., 1997). However, it is unlikely that the tongue movements had double peaks due to being small given that some of the trials for larger movements in both AWS and AWNS were found to have double peaks. In addition, the movements had trajectories that were not necessarily more curvilinear than movements did not have multiple peaks.

Deficient/weaker feedforward control in AWS

Both AWS and AWNS showed generally large correlation coefficients between the initial kinematics and extents, suggesting that speech movements are mostly pre-planned (i.e., mostly feedforward control). Yet, comparing the two groups, we found overall lower correlations between initial kinematics (peak acceleration and peak velocity) and movement extent in AWS compared to AWNS. It should be noted that these significant group differences were found despite of the lack of any between-group difference in general movement kinematics (i.e., movement duration, peak velocity, peak acceleration, movement extent, peak acceleration latency, and peak velocity latency). In addition, we also removed more atypical tongue movements multiple velocity profiles from AWS data compared to AWNS, which may have made the two data sets more similar. Nevertheless, these group differences were found, showing that even in unperturbed and fluent speech with very similar kinematic profiles, initial kinematic parameters in stuttering individuals cannot predict movement outcome as well as in nonstuttering individuals. Therefore, this finding suggests that the speech movements in AWS are not as pre-planned (i.e., less feedforward) compared to AWNS.

Consistent with our results, a number of recent studies have presented evidence for potential impaired feedforward control in AWS. One prominent evidence is a structural abnormality in white matter fibers (WMF) beneath the left ventral sensorimotor cortex in adults and children who stutter (e.g., Cai, Tourville et al., 2014; Chang et al., 2008; Cykowski et al., 2010; Sommer et al., 2002; Watkins et al., 2008), which connect brain regions that contribute to planning and executing movements and are thought to be responsible for feedforward control for speech production. In addition, movement inaccuracies have been reported even in nonspeech sensorimotor control (i.e., ballistic arm-reaching movements) of AWS, suggesting that there may be general weak or deficient feedforward control signals associated with stuttering (Daliri et al., 2014).

More online feedback control in AWS in the second half of the movement

As in the previous study (Kim & Max, 2014), we found significant increases in the movement variance from adding the movement duration to the multiple regression model. The significant increases in variance were found in both groups, suggesting that speech movements in both AWS and AWNS are adjusted/corrected online. Interestingly, comparing the amount of increases in the variance, we found that AWS have significantly larger increases than AWNS in the model with peak velocity. This finding clearly suggests that stuttering individuals are adjusting their movements more compared to nonstuttering individuals, an evidence for over-reliance on online feedback control.

For smooth and efficient motor control, it is important for movements to be adjusted early during the movement (or even prior to the onset) with sensory prediction of the forward model (e.g., Desmurget & Grafton, 2000; Desmurget et al., 2004). However, if feedforward control motor plan or the sensory prediction of the feedforward motor plan is inaccurate/deficient, one must rely considerably more on unstable and delayed online feedback control, resulting in late and suboptimal error corrections (Tseng et al., 2007; Wagner & Smith, 2008). More recently, a series of electrophysiological work from our own laboratory has revealed that an abnormal sensorimotor prediction processes are found in adults who stutter

(AWS) during speech planning prior to the movement onset (Daliri & Max, 2015a; Daliri & Max, 2015b; Daliri & Max, 2018; Max & Daliri, 2019). In addition, auditory-motor adaptation—which may be driven entirely by sensory prediction errors (Kim & Max, in preparation)—seems to be reduced in AWS (Daliri & Max, 2018; Daliri et al., 2018; Kim & Max, in preparation; Kim et al., in preparation; Loucks et al., 2012; Sengupta et al., 2016) and absent in children who stutter (Kim et al., in preparation). The clear limitations in auditory-motor learning add support to the idea that at least some aspects of auditory prediction and/or generating accurate motor commands (feedforward control) that account for previously experienced auditory errors may be deficient in individuals who stutter (Max et al., 2004). In turn, the deficiencies in both generating appropriate feedforward signals (Daliri et al., 2014) and early correction based on sensory prediction may leave no option other than having to rely less on feedforward control and more on inefficient online feedback-driven control.

The current study's results may also explain the previous findings that found *smaller* extents of responses to different kinds of perturbation than nonstuttering individuals (Cai et al., 2012; Loucks & De Nil, 2006). Given our results where stuttering individuals already rely more on online feedback control even in unperturbed movements, their system may not have capacity to adequately respond to external perturbation. Hence, such weak response to external perturbation may be consistent with our finding.

In summary, our findings add further support to the hypothesis that stuttering individuals may over-rely on online feedback control and less on feedforward control (Max et al., 2004). Especially our finding on over-reliance on online feedback control in the second half of the movement suggests that the online feedback control mechanism that PWS over-rely on may be delayed/inefficient in nature. The idea is also highly consistent with recent reports on weak/deficient feedforward control (Daliri et al., 2014) as well as abnormal sensory prediction (see Max & Daliri, 2019). Future studies should further examine the relationship between over-reliance on online feedback control and sensory prediction deficits. In addition, examining the hypothesized causal relationship between the over-reliance on online feedback control and stuttering moments is highly desired as well.

Acknowledgement

This work was funded by grants R01DC007603, R01DC014510, and R01DC017444 from the National Institute on Deafness and Other Communication Disorders (NIDCD) and grant MOP-137001 from the Canadian Institutes of Health Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, or the Canadian Institutes of Health Research.

References

Abend, W., Bizzi, E., & Morasso, P. (1982). Human arm trajectory formation. *Brain: A Journal of Neurology, 105*(Pt 2), 331-348. 10.1093/brain/105.2.331

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package ‘lme4’. *Convergence, 12*(1), 2.

Beal, D. S., Gracco, V. L., Brettschneider, J., Kroll, R. M., & De Nil, L. F. (2013). A voxel-based morphometry (VBM) analysis of regional grey and white matter volume abnormalities within the speech production network of children who stutter. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 49*(8), 2151-2161. 10.1016/j.cortex.2012.08.013

Bloodstein, O., & Ratner, N. B. (2008). *A Handbook on Stuttering*. Thomson Delmar Learning.

Cai, S., Beal, D. S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2014). Impaired timing adjustments in response to time-varying auditory perturbation during connected speech production in persons who stutter. *Brain and Language, 129*, 24-29. 10.1016/j.bandl.2014.01.002

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: evidence for abnormal auditory-motor transformation. *PloS One, 7*(7), e41830. 10.1371/journal.pone.0041830

Cai, S., Tourville, J. A., Beal, D. S., Perkell, J. S., Guenther, F. H., & Ghosh, S. S. (2014). Diffusion imaging of cerebral white matter in persons who stutter: evidence for network-level anomalies. *Frontiers in Human Neuroscience, 8*10.3389/fnhum.2014.00054

Chang, S., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. L. (2008). Brain anatomy differences in childhood stuttering. *NeuroImage, 39*(3), 1333-1344. 10.1016/j.neuroimage.2007.09.067

Chang, S., & Zhu, D. C. (2013). Neural network connectivity differences in children who stutter. *Brain : A Journal of Neurology, 136*(Pt 12), 3709-3726. 10.1093/brain/awt275

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders, 35*(3), 246-279. 10.1016/j.jfludis.2010.05.002

Crossman, E. R., & Goodeve, P. J. (1983). Feedback control of hand-movement and Fitts' Law. *The Quarterly Journal of Experimental Psychology.A, Human Experimental Psychology, 35*, 251-278. http://www.ncbi.nlm.nih.gov/pubmed/6571310

Cykowski, M. D., Fox, P. T., Ingham, R. J., Ingham, J. C., & Robin, D. A. (2010). A study of the reproducibility and etiology of diffusion anisotropy differences in developmental stuttering: a potential role for impaired myelination. *NeuroImage, 52*(4), 1495-1504. 10.1016/j.neuroimage.2010.05.011

Daliri, A., & Max, L. (2015a). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain and Language, 150*, 37-44. 10.1016/j.bandl.2015.08.008

Daliri, A., & Max, L. (2015b). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language, 143*, 59-68. 10.1016/j.bandl.2015.03.002

Daliri, A., & Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 99*, 55-68. 10.1016/j.cortex.2017.10.019

Daliri, A., Prokopenko, R. A., Flanagan, J. R., & Max, L. (2014). Control and prediction components of movement planning in stuttering versus nonstuttering adults. *Journal of Speech, Language, and Hearing Research: JSLHR, 57*(6), 2131-2141. 10.1044/2014_JSLHR-S-13-0333

Daliri, A., Prokopenko, R. A., & Max, L. (2013). Afferent and efferent aspects of mandibular sensorimotor control in adults who stutter. *Journal of Speech, Language, and Hearing Research: JSLHR, 56*(6), 1774-1788. 10.1044/1092-4388(2013/12-0134)

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science, 21*(2)10.1111/desc.12521

Desmurget, M., Gaveau, V., Vindras, P., Turner, R. S., Broussolle, E., & Thobois, S. (2004). On-line motor control in patients with Parkinson's disease. *Brain: A Journal of Neurology, 127*(Pt 8), 1755-1773. 10.1093/brain/awh206

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences, 4*(11), 423-431. 10.1016/s1364-6613(00)01537-0

Fradet, L., Lee, G., & Dounskaia, N. (2008). Origins of submovements in movements of elderly adults. *Journal of Neuroengineering and Rehabilitation, 5*, 28. 10.1186/1743-0003-5-28

Gordon, J., & Ghez, C. (1987). Trajectory control in targeted force impulses. III. Compensatory adjustments for initial errors. *Experimental Brain Research*, *67*(2), 253-269. 10.1007/bf00248547

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: lip and jaw coordination. *Journal of Speech, Language, and Hearing Research: JSLHR, 43*(1), 239-255. 10.1044/jslhr.4301.239

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research: JSLHR*, *45*(1), 66-79. 10.1044/1092-4388(2002/005)

Heath, M., Westwood, D. A., & Binsted, G. (2004). The control of memory-guided reaching movements in peripersonal space. *Motor Control-Champaign-*, *8*(1), 76-106.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70. http://www.jstor.org/stable/4615733

Iimura, D., Asakura, N., Sasaoka, T., & Inui, T. (2019). Abnormal Sensorimotor Integration in Adults Who Stutter: A Behavioral Study by Adaptation of Delayed Auditory Feedback. *Frontiers in Psychology, 10*, 2440. 10.3389/fpsyg.2019.02440

Kim, K. S., Flanagan, R. J., & Max, L. (in preparation). Dissociated speech and limb sensorimotor learning limitations in stuttering: speech auditory-motor learning is impaired in both children and adults who stutter
. *Doi,*

Kim, K. S., & Max, L. (2014). Estimating feedforward vs. feedback control of speech production through kinematic analyses of unperturbed articulatory movements. *Frontiers in Human Neuroscience, 8*, 911. 10.3389/fnhum.2014.00911

Kim, K. S., & Max, L. (in preparation). Auditory-motor learning differences between stuttering and nonstuttering individuals reflect implicit learning processes. *Doi,*

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13)

Lincoln, M., Packman, A., & Onslow, M. (2006). Altered auditory feedback and the treatment of stuttering: A review. *Journal of Fluency Disorders, 31*(2), 71-89. 10.1016/j.jfludis.2006.04.001

Loucks, T. M., & De Nil, L. F. (2006). Anomalous sensorimotor integration in adults who stutter: A tendon vibration study. *Neuroscience Letters, 402*(1-2), 195-200. 10.1016/j.neulet.2006.04.002

Loucks, T. M., Chon, H., & Han, W. (2012). Audiovocal integration in adults who stutter. *International Journal of Language & Communication Disorders, 47*(4), 451-456. 10.1111/j.1460-6984.2011.00111.x

Lucero, J. C., Munhall, K. G., Gracco, V. L., & Ramsay, J. O. (1997). On the registration of time and the patterning of speech movements. *Journal of Speech, Language, and Hearing Research: JSLHR, 40*(5), 1111-1117. 10.1044/jslhr.4005.1111

Max, L. (2004). Stuttering and internal models for sensorimotor control: A theoretical perspective to generate testable hypotheses. *Speech motor control in normal and disordered speech* (pp. 357–388). Oxford University Press.

Max, L., Caruso, A. J., & Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech, Language, and Hearing Research: JSLHR, 46*(1), 215-232. 10.1044/1092-4388(2003/017)

Max, L., & Daliri, A. (2019). Limited Pre-Speech Auditory Modulation in Individuals Who Stutter: Data and Hypotheses. *Journal of Speech, Language, and Hearing Research: JSLHR, 62*(8S), 3071-3084. 10.1044/2019_JSLHR-S-CSMC7-18-0358

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (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 Sciences and Disorders, 31*, 105-122. 10.1044/cicsd_31_S_105

Messier, J., & Kalaska, J. F. (1999). Comparison of variability of initial kinematics and endpoints of reaching movements. *Experimental Brain Research, 125*(2), 139-152. 10.1007/s002210050669

Meyer, D. E., Smith, J. E., Kornblum, S., & Abrams, R. A. (1990). Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In M. Jeannerod (Ed.), *Attention and Performance Xiii: Motor Representation and Control* (). Psychology Press.

Milner, T. E. (1992). A model for the generation of movements requiring endpoint precision. *Neuroscience, 49*(2), 487-496. http://www.ncbi.nlm.nih.gov/pubmed/1436478

Namasivayam, A. K., van Lieshout, P., McIlroy, W. E., & De Nil, L. (2009). Sensory feedback dependence hypothesis in persons who stutter. *Human Movement Science, 28*(6), 688-707. 10.1016/j.humov.2009.04.004

Neef, N. E., Hoang, T. N. L., Neef, A., Paulus, W., & Sommer, M. (2015). Speech dynamics are coded in the left motor cortex in fluent speakers but not in adults who stutter. *Brain: A Journal of Neurology, 138*, 712-725. 10.1093/brain/awu390

Okadome, T., & Honda, M. (2001). Generation of articulatory movements by using a kinematic triphone model. *The Journal of the Acoustical Society of America, 110*(1), 453-463. 10.1121/1.1377633

Perkell, J. S., & Zandipour, M. (2002). Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements. *The Journal of the Acoustical Society of America, 112*(4), 1642-1651. 10.1121/1.1506368

Quinn, L., Hamel, V., Flanagan, J. R., Kaminski, T., & Rubin, A. (1997). Control of multijoint arm movements in Huntington's disease. *Journal of Neurologic Rehabilitation, 11*(1), 47-60.

Ritto, A. P., Juste, F. S., Stuart, A., Kalinowski, J., & de Andrade, Claudia Regina Furquim. (2016). Randomized clinical trial: the use of SpeechEasy® in stuttering treatment. *International Journal of Language & Communication Disorders, 51*(6), 769-774. 10.1111/1460-6984.12237

Seidler, R. D., Noll, D. C., & Thiers, G. (2004). Feedforward and feedback processes in motor control. *NeuroImage, 22*(4), 1775-1783. 10.1016/j.neuroimage.2004.05.003

Sengupta, R., Shah, S., Gore, K., Loucks, T. M., & Nasir, S. M. (2016). Anomaly in neural phase coherence accompanies reduced sensorimotor integration in adults who stutter. *Neuropsychologia, 93*(Pt A), 242-250. 10.1016/j.neuropsychologia.2016.11.004

Sommer, M., Koch, M. A., Paulus, W., Weiller, C., & Büchel, C. (2002). Disconnection of speech-relevant brain areas in persistent developmental stuttering. *Lancet (London, England), 360*(9330), 380-383. 10.1016/S0140-6736(02)09610-1

Tseng, Y., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of Neurophysiology, 98*(1), 54-62. 10.1152/jn.00266.2007

Wagner, M. J., & Smith, M. A. (2008). Shared Internal Models for Feedforward and Feedback Control. *Journal of Neuroscience, 28*(42), 10663-10673. 10.1523/JNEUROSCI.5479-07.2008

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain: A Journal of Neurology, 131*(Pt 1), 50-59. 10.1093/brain/awm241

West, G. L., Welsh, T. N., & Pratt, J. (2009). Saccadic trajectories receive online correction: evidence for a feedback-based system of oculomotor control. *Journal of Motor Behavior, 41*(2), 117-127. 10.3200/JMBR.41.2.117-127

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology: CB, 11*(18), 729. 10.1016/s0960-9822(01)00432-8