Published in final edited form as:
J Fluency Disord. 2023 March ; 75: 105943. doi:10.1016/j.jfludis.2022.105943.

Auditory feedback control in adults who stutter during metronome-paced speech I. Timing Perturbation

Saul A. Frankford[a], **Shanqing Cai**[a], **Alfonso Nieto-Castañón**[a], **Frank H. Guenther**[a,b]

[a]Department of Speech, Language, & Hearing Sciences, Boston University, Boston, MA 02215

[b]Department of Biomedical Engineering, Boston University, Boston, MA 02215

Abstract

Purpose: This study determined whether adults who stutter (AWS) exhibit deficits in responding to an auditory feedback timing perturbation, and whether external timing cues, which increase fluency, attenuate any disruptions due to altered temporal auditory feedback.

Methods: Fifteen AWS and sixteen adults who do not stutter (ANS) read aloud a multisyllabic sentence either with normal pacing or with each syllable paced at the rate of a metronome. On random trials, an auditory feedback timing perturbation was applied, and timing responses were compared between groups and pacing conditions.

Results: Both groups responded to the timing perturbation by delaying subsequent syllable boundaries, and there were no significant differences between groups in either pacing condition. Furthermore, no response differences were found between normally paced and metronome-paced conditions.

Conclusion: These findings are interpreted as showing that 1) AWS respond normally to pure timing perturbations, and 2) metronome-paced speech has no effect on online speech timing control as assessed in the present experiment.

Keywords

stuttering; auditory feedback; metronome; speech timing control

1. Introduction

Persistent developmental stuttering is characterized by speech disfluencies such as sound repetitions, prolongations and blocks. It affects up to 8% of preschool-age children and persists into adulthood for 1% of the population (Yairi & Ambrose, 2013). Despite its prevalence and the expansive body of behavioral and neural stuttering research, the

Send Correspondence to: Saul A. Frankford, Department of Speech, Language, & Hearing Sciences, Boston University, 677 Beacon St., Boston, MA 02115, sauf@bu.edu.

Conflicts of Interest: All authors declare no competing interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

mechanisms underlying stuttering remain poorly understood. One set of theories put forward in the literature posits that individuals who stutter have difficulty with properly timing the initiation and/or termination of speech segments (Alm, 2004; Etchell et al., 2014; Guenther, 2016; Howell, 2004; MacKay & MacDonald, 1984; Wingate, 2002). Following Alm's influential 2004 review of evidence for basal ganglia involvement in stuttering, a number of other researchers have suggested that this difficulty relates to neurally integrating sensory and motor information to initiate speech motor programs (Chang & Guenther, 2020; Guenther, 2016). In this hypothesis, the three main types of stuttering events can be explained as follows: prolongations result from difficulty cuing the termination of a motor program; blocks result from difficulty cuing the initiation of a motor program; and repetitions result from repeated dropouts in initiation signals (Chang & Guenther, 2020). This is supported by numerous behavioral and neural studies. Reducing speech rate, which decreases stuttering in people who stutter (PWS; Andrews et al., 1982), may allow for more time to resolve discrepencies between neural cues from among sensory and motor regions. Furthermore, "external" timing cues from a metronome or second speaker reduce stuttering (e.g., Andrews et al., 1982; Brady, 1969; Toyomura et al., 2011), potentially helping to resolve discrepencies between sensorimotor timing cues. In addition, PWS exhibit delayed reaction times and abnormal variability in motor coordination measures during both speech and nonspeech motor tasks (Falk et al., 2015; Howell et al., 1997; Kleinow & Smith, 2000; Max et al., 2003; McClean & Runyan, 2000; Starkweather et al., 1984), suggesting difficulties integrating sensory and motor signals for precise motor timing.

In the neural domain, PWS show differences in the cortico-basal ganglia motor network (Chang & Zhu, 2013; Giraud, 2008; Lu et al., 2010) and auditory sensory areas (Chang & Zhu, 2013; De Nil et al., 2000; Foundas et al., 2001, 2004; Fox et al., 2000; Yang et al., 2016) as compared to individuals who do not stutter. This network is implicated in selectively releasing motor programs for action (Alm, 2004; Mink, 1996), and may therefore be involved in timing the onsets and offsets of speech segments. Specifically, the striatum in the basal ganglia receives input from large portions of sensory cortex, potentially aggregating sensory and motor information to guide speech timing online. Damage to this pathway has been associated with neurogenic stuttering (Ludlow et al., 1987; Theys et al., 2013), and there is evidence that modulation of dopamine receptors in the basal ganglia can lead to reduced disfluencies in adults who stutter (AWS; Alm, 2004). In summary, difficulty combining sensory and motor information to precisely time speech is a potential underlying mechanism of stuttering.

A common experimental paradigm for testing interactions between sensory and motor processes examines online responses to perturbed sensory feedback. In studies using auditory feedback perturbations, one or more components of a participant's speech signal, such as voice fundamental frequency (f0; Burnett et al., 1998; Chen et al., 2007) or vowel formants (Niziolek & Guenther, 2013; Purcell & Munhall, 2006), are altered and fed back to them and the ensuing responses are measured. Another type of auditory feedback perturbation more directly tests the role of sensorimotor integration for *timing* of ongoing speech by modifying the perceived timing of a self-produced speech gesture (henceforth, "timing perturbation"). A timing perturbation temporally stretches or compresses a short segment of the speech signal (either online or using pre-recorded samples), so the duration of a phoneme sounds altered to the speaker (e.g., prolonging the /s/ sound in the word “steady”). Comparing speech timing during perturbed trials with that of non-perturbed trials yields a measure of the effect of the timing perturbation on speech. This measure can potentially provide information on the extent to which auditory timing cues are used to sequence speech.

Previous work shows that in response to timing perturbations, typically fluent speakers will delay the production of subsequent speech gestures (Floegel et al., 2020; Mitsuya et al., 2014; Ogane & Honda, 2014; Oschkinat & Hoole, 2020). In adults who stutter (AWS), these delays are reduced suggesting altered sensorimotor integration for timing control in stuttering (Cai et al., 2014). However, responding to the perturbation in this study required tracking formant changes in order to infer timing, and other work has demonstrated that AWS show reduced compensation to auditory feedback formant perturbations (Cai et al., 2012; Daliri et al., 2018). As a result, it is unclear whether a pure timing perturbation that alters timing but not formant trajectories would show the same response reduction in AWS.

As mentioned previously, it is well-documented that speaking with an external timing cue like a metronome reduces disfluencies in people who stutter (e.g., Andrews et al., 1982; Brady, 1969; Braun et al., 1997; Davidow, 2014; Stager et al., 2003; Toyomura et al., 2011). It has been suggested that external cues allow people who stutter to rely less on inefficient or impaired “internal” timing mechanisms to sequence speech utterances (Alm, 2004; Etchell et al., 2014; Guenther, 2016). As the effect is also present when speech timing cues are imagined or retained in memory (Barber, 1940; Stager et al., 2003), “external” refers to any source outside of habitual or automatic speech sequencing mechanisms, even if they are generated by the speaker. Prior neuroimaging work has indicated that external pacing may “normalize” the level of speech activation in brain regions supporting speech timing such as the basal ganglia and supplementary motor area (Toyomura et al., 2011) and/or recruit alternative brain networks involving the cerebellum (Frankford et al., 2021) to restore speech timing function and increase fluency. Behaviorally, then, speaking in a manner that references an external stimulus may lead to normalized auditory motor integration for speech timing.

Therefore, the present study had two primary aims. The first aim was to test whether a pure timing perturbation was sufficient to elicit reduced timing delays in AWS compared to adults who do not stutter (ANS). To achieve a pure timing perturbation, the speech spectrum was stretched in time without altering spectral information such that the boundary between a fricative and stop consonant was delayed in auditory feedback. The second aim was to test whether externally paced speech leads to fluency by helping to resolve temporal sensorimotor integration disruptions. This was carried out by applying the timing perturbation to the same speech segment during normal and metronome-paced speech. It was hypothesized that AWS would show reduced speech timing delays in response to the timing perturbation during normal speech (as in Cai et al., 2014), but that during metronome-paced speech no group differences would be observed.

2. Methods and Materials

2.1. Participants

Fifteen adults who stutter (AWS; 12 Males/3 Females, aged 18–44 years, mean age = 25.73 years, SD = 8.37) and 16 adults who do not stutter (ANS; 12 Males/4 Females, aged 18–44 years, mean age = 26.69 years, SD = 6.79) participated in this study. This unbalanced male-to-female ratio mirrors the prevalence of persistent developmental stuttering in the population (Bloodstein & Ratner, 2008). All participants were native speakers of American English with no prior history of speech, language, or hearing disorders (other than stuttering for the AWS group), and all participants passed audiometric screenings with binaural pure-tone hearing thresholds of less than 25 dB HL at 500, 1000, 2000, and 4000 Hz. All AWS self-identified as stuttering during a screening interview. AWS were video-recorded to collect a sample of speech during three tasks: 1) an in-person conversation, 2) a phone conversation, and 3) reading the Grandfather passage (Van Riper, 1963) aloud. Based on these samples a speech-language pathologist board certified in fluency and fluency disorders with extensive experience diagnosing stuttering evaluated stuttering severity using the Stuttering Severity Instrument – Fourth Edition (SSI-4; Riley, 2008). Individual scores ranged from 9 (subclinical) to 42 (very severe) with a mean score of 23. One individual had a score below the lowest percentile on the SSI-4, but self-reported that they currently stutter and that certain sounds and social contexts impact their overt symptoms. As SSI-4 scores can vary significantly from day to day (Constantino et al., 2016), they were included in the AWS group. Participants provided informed written consent and the study was approved by the Boston University Institutional Review Board.

2.2. Experimental Setup

Figure 1 provides a schematic diagram of the experimental setup. Participants were seated upright in front of a computer monitor in a sound-attenuating booth. They were fitted with ER1 earphones (Etymotic Research, Inc.) and an AT803 microphone (Audio-Technica) mounted to a headband via an adjustable metal arm. This arm was positioned in front of the mouth at approximately 45° below the horizontal plane such that the mouth to microphone distance was 10 cm for all participants. The microphone signal was amplified and digitized using a MOTU Microbook external sound card and sent to a computer running a custom experimental pipeline in MATLAB (Mathworks; version 2013b), including Audapter (Tourville et al., 2013) and Julius (Lee & Kawahara, 2009). Auditory feedback was sent back through the Microbook and amplified using a Xenyx 802 (Behringer) analog mixer such that the signal played through the earphones sounded 4.5 dB louder than the microphone input. This amplification helped reduce participants' ability to hear their non-perturbed feedback through air or bone conduction, and is consistent with feedback levels that enable reliable perturbation responses elsewhere in the literature (Abur et al., 2021; Weerathunge et al., 2020).

2.3. Stimuli

Stimuli for this study mainly consisted of one “target” sentence (“The steady bat gave birth to pups”) on which all experimental manipulations were applied. This sentence was constructed such that the target of perturbation, /s/, occurred early in the sentence allowing for several timing measures following the perturbation. In addition, subsequent syllables all began with stop consonants that could be used to clearly identify syllable boundaries. Fifteen “filler” sentences, selected from the Harvard sentence pool (i.e., the Revised List of Phonetically Balanced Sentences; IEEE Recommended Practice for Speech Quality Measurements, 1969), were also included to reduce boredom and keep participants attending to the task. All sentences contained eight syllables.

2.4. Procedure

Participants were instructed to read aloud sentences displayed on the computer monitor under two speaking conditions: either with each syllable evenly spaced (*metronome-paced speech* condition) or with a normal (unmodified) timing pattern (*normal speech* condition). At the beginning of every trial, participants viewed white crosshairs on a grey screen while eight isochronous tones (1000Hz pure tone, 25ms, 5ms ramped onset) were played with an inter-onset interval of 270ms. This resulting rate of approximately 222 beats/min was chosen so that participants’ speech would approximate the rate of the *normal* condition (based on estimates of mean speaking rate in English; Davidow, 2014; Pellegrino et al., 2004). The trial type then appeared on the screen (“Normal” or “Rhythm”) followed by a stimulus sentence. On *metronome-paced speech* trials (corresponding to the “Rhythm” cue), participants were cued to read the sentence with evenly paced syllables at the rate of the tone stimuli aligning each syllable to a beat, while on *normal speech* trials, participants ignored the pacing tones, and spoke with normal rate and rhythm. The font color was either blue for *metronome-paced speech* trials or green for *normal speech* or vice versa, and colors were counterbalanced across participants. Tones were presented at the beginning of both trial types to provide consistency with a related functional neuroimaging study where this feature was necessary (Frankford et al., 2021).

The experiment comprised three brief training runs followed by three experimental runs. During the first training run, participants received visual feedback on their loudness (a horizontal bar in relation to an upper and lower boundary) – males were trained to speak between 65 and 75 dB SPL, and females were trained to speak between 62.5 and 72.5 dB SPL. This difference was meant to account for natural differences in the speech sound intensity of males and females during conversational speech (Gelfer & Young, 1997). On the second training run, participants also were trained to speak with mean inter-syllable duration (ISD) between 220ms and 320ms (centered around the inter-onset interval of the tones). ISD was calculated as the time between the midpoints of successive vowels in the sentence. On the third training run, in addition to visual feedback on loudness and speaking rate, participants received feedback on the isochronicity of their speech. Isochronicity was measured using the coefficient of variation (standard deviation/mean) of inter-syllable duration (CV-ISD), such that lower coefficients of variance indicate greater isochronicity. Using this feedback, participants were trained to speak with a CV-ISD less than 0.25 for the *metronome-paced speech* trials. This training procedure was carried out to ensure approximate consistency in intensity, speech rate, and isochronicity across participants and conditions.

Each experimental run contained 80 speech trials, half *metronome-paced* and half *normal.* The target sentence (see "2.3. Stimuli") appeared in 80% of trials in each condition, while filler sentences comprised the remaining 20%. Normal feedback was provided on half of these target trials, while one quarter included a timing perturbation - a brief delay in the transition between the /s/ and /t/ phonemes in the word "steady" (see Section 2.5. for details). The other quarter included a vowel formant perturbation but this condition was not analyzed for the present paper and will not be discussed further. The order of these trials in each run was pseudo-randomized such that every set of 10 trials contained two timing perturbations, one *metronome-paced* and one *normal*. In total, participants completed 240 trials, 192 of which contained target sentences. Of these, 96 were *metronome-paced* trials and 96 were *normal* trials, each containing 24 trials with a timing perturbation, and 48 unperturbed trials (the remaining 24 trials contained the vowel formant perturbation and are not discussed here). One participant only completed 200 trials (160 target trials, 80 *normal*, 80 *metronome-paced*, each containing 40 unperturbed trials and 20 trials with a timing perturbation) due to a technical error.

2.5. Timing Perturbation

This study used fine-scale temporal processing previously described in Tourville et al. (2013) to employ temporal dilation (slowing down and speeding up) of auditory feedback using a phase vocoder (Bernsee, 1999). This focal timing perturbation was applied to the /s/ in "steady" and feedback was returned to normal by the end of the word (Figure 2). To apply this perturbation at the desired time, Audapter relied on the online detection of the /s/, carried out in two steps. First, voicing onset (/ə/ in the preceding word, "the") was detected when the amplitude of the speech signal surpassed a short-time root-mean square (RMS) adaptive threshold (Equation A1) for at least 20 ms. Following this, the onset of the /s/ was detected when the ratio of the pre-emphasized (i.e. high-pass filtered) RMS and the unfiltered RMS exceeded another adaptive threshold (Equation A2) for at least 20 ms. See the Appendix for details on how these adaptive thresholds were calculated.

On a given timing-perturbed trial, once the /s/ was detected, Audapter downsampled the digitized 48000 Hz microphone signal to 16000 Hz and applied a short-time Fourier transform (STFT) on frames of 16ms (sliding by 4ms), saving the Fourier spectrum in memory. Through linear interpolation and inverse STFT re-synthesis, Audapter slowed down the auditory feedback to half speed for a participant-specific interval. This interval was equal to the average duration of the /s/ across non-perturbed target trials in previous runs as determined by the forced-alignment speech recognition software Julius (Lee & Kawahara, 2009; see section 2.6.1. Data Processing), carried out separately for the *metronome-paced* and *normal* conditions. The delayed feedback was maintained at normal speed for the same interval, then feedback was accelerated to double speed until it realigned with the incoming microphone signal. This delayed the boundary between the /s/ and /t/ in the auditory signal by ~50ms and returned feedback to normal by the end of the following syllable. The gradual delay onset and offset assured feedback timing continuity that made the perturbed utterances sound qualitatively natural to the participant. Auditory feedback remained unperturbed for the rest of the trial.

The total feedback latency was approximately 32 ms during non-perturbed speech. This was determined by turning on (unaltered) feedback in Audapter and using a second computer with a microphone to record both a brief stimulus (snap) produced near the experimental microphone and the "echo" of it coming over the headphones. The latency in Audapter between the microphone signal and the processed headphone signal was 16 ms for all trials. Therefore, an additional 16 ms of delay was incurred by the hardware setup. While this additional latency could affect the results, because it was consistent across groups and speaking conditions (our main comparisons of interest), it should have minimal impact on the interpretation of these effects (see section 4.3 for more discussion on this feedback latency).

2.6. Analyses

2.6.1. Data Preparation—An automatic speech recognition (ASR) engine, Julius (Lee & Kawahara, 2009), was used in conjunction with the free *VoxForge* American English acoustic models (voxforge.org) to determine phoneme boundary timing information for every trial. All trials (presented in random order and blinded to condition) were manually inspected by a research assistant trained to identify errors and stutters (initial rater). Any cases where the initial rater was not confident (applied liberally) were noted and resolved by the first author. Trials where there were gross ASR errors were removed. For each trial that included a perturbation, the speech spectrogram from the microphone and the headphone signal were compared to determine whether the perturbation occurred at the proper time within the utterance (i.e. during the /s/-/t/ boundary for the timing perturbation). Trials where this was not the case were discarded from further analysis. Any trials in which the participant made a reading error or a condition error (i.e. spoke isochronously when they were cued to speak normally or vice versa) were eliminated from further analysis. Trials that were identified by the initial rater as potentially containing a stutter (applied liberally to capture as many trials as possible) were reviewed by the first author and a speech-language therapist with experience in treating stuttering and were either confirmed or dismissed by consensus. Trials judged to contain a stutter were not included when calculating timing responses to the perturbation. The proportion of removed trials that contained stutters were, however, used to calculate the experimental stuttering rate for each participant. Because the timing perturbation can lead to a stutter-like prolongation, only unambiguous stutters were eliminated. Finally, trials where participants spoke outside of the trained mean ISD (220 ms – 320 ms) were also eliminated. In total, these procedures excluded an average of 13.8% of trials for ANS (SD: 6.5%) and 15.8% of trials for AWS (SD: 10.8%). This was not significantly different between groups ($t = 0.63$, $p = .53$).

Measures of the total sentence duration and intersyllable timing from each trial were also extracted to determine the rate and isochronicity of each production. Within a sentence, the average time between the centers of the eight successive vowels was calculated to determine the ISD. The reciprocal of the ISD from each sentence (1/ISD) was then calculated, resulting in a measure of speaking rate in units of syllables per second.

Perturbation magnitudes for each timing-perturbed trial were defined as the maximum difference in ASR phoneme boundaries between the microphone input signal and auditory feedback during the word “steady.” For example, if the /s/-/t/ boundary was delayed by 50 ms in auditory feedback and the /t/-/ɛ/ boundary was delayed by 60 ms, the perturbation magnitude for that trial would be 60 ms. One trial from each of three participants was removed due to ASR errors in the auditory feedback that led to erroneous perturbation magnitudes.

To assess speech timing changes in response to the perturbation, the durations from the onset of /s/ in “steady” to the s-t boundary in “steady” (/t1/) and to each of six subsequent syllable boundaries (see Table 2 for a description of these boundaries) in the ASR segmentation were calculated, similar to previous literature (Cai et al., 2011, 2014; Howell & Sackin, 2000). These values were then averaged within each of four condition combinations (perturbed/normal, non-perturbed/normal, perturbed/metronome-paced, non-perturbed/metronome-paced) in each participant. Then, the non-perturbed average was subtracted from the perturbed average (in both normal and metronome-paced conditions) to yield cumulative timing response curves in normal and metronome-paced conditions.

2.6.2. Statistical Analysis—To evaluate whether there was a fluency-enhancing effect of isochronous pacing, the percentage of trials eliminated due to stuttering in the AWS group was compared between the two speaking conditions using a non-parametric Wilcoxon signed-rank test.

Rate and CV-ISD were compared between groups and conditions using linear mixed effects models with group, condition, and group × condition interaction as fixed effects and participants as random effects. For these analyses, type III Wald F tests with Kenward-Roger degrees of freedom were used to assess significance.

Because the exact magnitude of the perturbation depended on a) the parameters that defined the time dilation, which were derived from previous productions, and b) the duration of the /s/ and /t/ phonemes in a given trial, it was difficult to ensure complete consistency across participants and trials. To determine if any differences in perturbation magnitude existed between groups and conditions, a linear mixed effects model with group, condition, and group x condition interaction as fixed effects and participants as random effects was carried out.

A multivariate general linear model (GLM) framework was used to determine whether the magnitude of the timing response curves were dependent on group, condition, or stuttering severity. In order to avoid having too many outcome variables and to account for the inherent correlation between the timing of adjacent syllable boundaries, a principal components analysis (PCA) was performed on the timing response curves from all participants to extract the components that characterize at least 95% of the variance in the responses. The responses of each participant for each condition were projected onto these principal components and used as the set of dependent variables. F-tests (implemented using the *conn_glm* function from the CONN toolbox; Nieto-Castañón, 2020) assessed whether each independent variable had a significant effect on any of the dependent variables (capturing *overall* timing responses). For stuttering severity, two separate measures were used. The first was a modification of the SSI-4 score, heretofore termed “SSI-Mod.” SSI-Mod removes the secondary concomitants subscore from each participant's SSI-4 score, thus focusing the measure on speech-related function. This procedure is empirically supported by Mirawdeli & Howell (2016). The second measure was the percentage of trials removed due to stuttering during the *normal* conditions (stuttering rate). Therefore, two separate models were evaluated; one that included the SSI-Mod scores and a second that included the stuttering rates[1].

To test whether timing responses were correlated with isochronicity in the *metronome-paced* condition, CV-ISD was added as a covariate to the original model.

3. Results

3.1. Stuttering Rate

For most participants, stuttering occurred infrequently over the course of the experiment, with 6 out of 15 AWS producing no stuttering disfluencies. However, AWS produced significantly fewer stuttering disfluencies during the *metronome-paced* condition (1.2%) than in the *normal* condition (7.4%; $W = 52$, $p = .012$; Figure 3; this result is also reported in Frankford et al., 2022).

3.2. Speaking Rate and Isochronicity

For rate (Table 3), there was no significant effect of group, $F(1, 51.9) = 3.17$, $p = .08$, but there was a significant effect of condition, $F(1, 29) = 65.27$, $p < .001$, that was modulated by a significant group × condition interaction, $F(1, 29) = 6.17$, $p = .02$. In this case, participants in both groups spoke at a slower rate in the *metronome-paced* condition, but this difference was larger for ANS (ANS *normal*: 4.0 ± 0.2 syl/sec, ANS *metronome-paced*: 3.6 ± 0.1 syl/sec, AWS *normal*: 3.9 ± 0.2 syl/sec, AWS *metronome-paced*: 3.7 ± 0.1 syl/sec). Because of these significant effects, rate was included as a covariate in the timing perturbation analysis. To examine whether this reduction in rate led to increased fluency rather than the isochronous pacing, we tested for a correlation between the change in speech rate and the reduction in stutters. These two measures were not significantly correlated, Spearman's $\rho = .10$, $p = .73$. As expected, there was a significant effect of condition on CV-ISD, $F(1, 29) = 163.80$, $p < .001$, but no effect of group, $F(1, 57.8) = 0.94$, $p = .34$, or interaction, $F(1, 29) = 0.41$, $p = .53$. These results are also reported in Frankford et al. (2022).

3.3. Timing Perturbation

Figure 4 shows the average cumulative speech timing response between the perturbed and non-perturbed conditions across all time points in each group and condition. On average, both groups show a significant timing response to the perturbation in both conditions. In the *normal* condition, both groups first show this significant timing response at landmark /b1/ while in the *metronome-paced* condition, the groups exhibit significant timing responses after landmark /d1/ (the onset of the /d/ in "steady"). This difference between conditions makes sense since in the *normal* condition, the /tɛ/ in "steady" is produced with a shorter

[1] Note that to perform the GLM using *conn_glm*, each severity measure was mean-centered within the AWS group and zero values were entered for ANS, allowing for separate estimation of group and severity.

duration (mean: 172 ms, SD: 22 ms) than in the *metronome-paced* condition (mean: 206 ms, SD: 41 ms). Assuming that the neural mechanism for delaying speech in response to a perturbation has some specific latency due to auditory processing and integration time, by the time the response begins the sentence may have already progressed to the next syllable in the *normal* condition but not in the *metronome-paced* condition. A similar effect for a vocal pitch perturbation in typically fluent adults can be found elsewhere (Natke & Kalveram, 2001).

Since perturbation magnitude was not precisely controlled and may have had an effect on timing responses, a linear mixed effects model was carried out to determine whether perturbation magnitudes varied across groups and conditions. While there was not a significant effect of group, $F(1, 29) = 0.93$, $p = .34$, or interaction between group and condition, $F(1, 29) = 0.31$, $p = .58$, there was a main effect of condition, $F(1, 29) = 22.15$, $p < .001$, such that there were smaller perturbations in the *metronome-paced* condition than the *normal* condition (ANS *normal*: 59.4 ± 11.1 ms, ANS *metronome-paced*: 53.8 ± 11.0 ms, AWS *normal*: 56.5 ± 13.4 ms, AWS *metronome-paced*: 49.4 ± 8.9 ms). Because of this, perturbation magnitude was entered as a covariate in the main timing perturbation analysis.

To determine the effects of group, condition, and severity on timing responses and control for factors like perturbation magnitude and rate, a multivariate GLM was performed (see Table 4 for complete results). A PCA on the timing responses across seven time-points found that the first three principal components accounted for >95% of the variance in the data set (PC1: 84.8%, PC2: 9.5%, PC3: 2.2%; see Supplementary Figure 1). Therefore, timing responses for each condition were projected onto each of these principal components and used as the set of dependent variables. For estimating effects that were agnostic to condition (i.e., group, severity), covariate values for speaking rate and perturbation magnitude were averaged between normal and metronome-timed conditions. For estimating effects related to changes across conditions (i.e., condition, group x condition, and severity x condition), the difference in the covariate values for rate and magnitude between conditions was used in the model.

Among the variables of interest, there was no significant effect of group or group x condition interaction on timing response magnitude, but there was a significant effect of condition, $F(3, 24) = 3.15$, $p = .04$. There was also no significant effect of either SSI-mod or SSI-mod x condition (estimated for AWS). Among control covariates, perturbation magnitude had a significant effect on timing response, $F(3, 24) = 3.42$, $p = .03$ (larger perturbation magnitudes led to larger responses), while mean rate did not. To further quantify the null group effect, confidence intervals were computed for the difference in response between AWS and ANS. To generate interpretable confidence intervals, the same model was run with the original data (i.e. all seven time-points with no PCA) and mean response differences and confidence intervals were averaged across the last three timepoints (when the group means of the response curves appear to reach a plateau). Using these criteria, the mean effect size was 6.4 ms (AWS delayed their speech 6.4 ms more than the ANS response; 90% CI, −9.9, 22.6).To follow up on the significant effect of condition, the effect sizes were projected back into syllable-boundary time. This analysis showed that the difference was mainly due to the earlier onset responses in the metronome-paced condition (with respect to syllable boundary in the sentence) rather than the total cumulative timing responses as measured at the end of the sentence. To confirm this, a GLM was performed where the independent variables were the same as above, but the dependent variable was the perturbation timing response at syllable boundary 7 (/s/-/p/). This analysis found no significant effect of condition, *F*(1, 26) = 0.001, *p* = .98.

We then re-ran the model substituting in stuttering rate from the experimental session for SSI-Mod, and found a main effect of stuttering rate, *F*(3, 24) = 5.22, *p* = .006, that was significantly modulated by condition, *F*(3, 24) = 3.48, *p* = .03 (see Figure 5; for complete results of this new model, see Supplementary Table 1). Accounting for all other variables, participants with more stutters during the task had larger speech timing delays as a result of the perturbation. Finally, to see if CV-ISD score was associated with the response, we added it into the original model and found that it was not a significant predictor of response, *F*(3, 23) = 1.35, *p* = .28.

4. Discussion

The present experiment examined the effects of perturbations in the timing of auditory feedback on speech timing when adults who do and do not stutter read sentences aloud. In contrast with a previous study that used dynamic vowel formant cues to evoke such feedback timing perturbations (Cai et al., 2014), the present study employed an algorithm that stretched the entire speech signal. While Cai et al. (2014) found that AWS altered their speech timing to a lesser extent than ANS, the present study found no such group effects. In addition, this study examined whether pacing speech to an external stimulus (which is known to increase fluency in AWS) alters auditory feedback timing control processes in AWS. While AWS did stutter less frequently in the *metronome-paced* condition, changes in the overall timing responses in either group were not observed. These results are discussed in further detail below with respect to prior literature.

4.1. Auditory feedback timing control in AWS

It has been suggested that stuttering results from difficulty integrating self-generated sensory and motor cues to properly control the timing of speech segments (Chang & Guenther, 2020). As such, altering auditory feedback timing cues and measuring subsequent speech timing can be used to probe these sensorimotor integration processes. The first paper to investigate auditory feedback-based speech timing control in AWS was Cai et al. (2014). Taking advantage of the continuous formant trajectories in the carrier phrase "I owe you a yo-yo," the authors applied either an advancement (~45 ms) or delay (~24 ms) to these trajectories at the local minimum of the second formant (F2) during the first /o/ ("owe"), and measured changes in the timing of subsequent F2 landmarks compared to a non-perturbed condition. They found that while neither group responded to the advanced feedback, only ANS significantly responded to delayed feedback with a delay in subsequent landmarks. Furthermore, AWS' reduced responses were most pronounced earlier in the phrase. This study supported the theory that AWS exhibit an impairment in utilizing sensory cues for timing ongoing speech (although it is possible that reduced responses reflected a compensatory strategy to minimize the influence of unpredictable or unreliable feedback).

Given these findings, the lack of a group difference in responses to the timing perturbation in the present study might seem surprising. However, Cai et al. (2014) created a timing perturbation by applying an F1 and F2 perturbation that remapped the formants on a time lag to delay the F2 local minimum. Thus, the delayed timing signals were embedded in a task that required precise spectral tracking of the acoustic signal. Previous studies have demonstrated that AWS show decreased spectral acuity when tracking pitch changes (Nudelman et al., 1992), and decreased compensation to formant perturbations (Cai et al., 2012; Daliri et al., 2018; Daliri & Max, 2018), indicating a potential deficit in spectral tracking abilities. In contrast, the timing perturbation in the present study was applied to the middle of the /s/ and the occlusion of the /t/ in “steady” and involved a temporal prolongation of the entire speech signal with no modification of the spectral content. The present results, in combination with those of Cai et al. (2014), suggest that AWS only have a deficit in their ability to use auditory feedback timing cues to sequence speech when those cues require tracking spectral features like formant frequencies.

This dichotomy between spectro-temporal and pure temporal perturbations can be thought of in terms of two motor timing theories described in the speech motor control literature: intrinsic (state) timing vs. extrinsic (clock) timing (Fowler, 1980; Kelso & Tuller, 1987). Extrinsic (clock) timing refers to the idea that the timing of subsequent speech segments in an utterance is planned in relation to an absolute timekeeper (e.g., in millisecond time). For intrinsic (state) timing, the planned temporal relations between adjacent speech segments are determined based on the relative progression of the speech system through a series of states (articulator positions and velocities, evident as formant trajectories in the acoustic signal). In the present study, extrinsic time (clock time) was perturbed, whereas Cai et al. (2014) applied more of an intrinsic timing perturbation, i.e. changing the state of the system (formants) to change the perception of time. Put together, the results of these two studies suggest that AWS have difficulty responding to intrinsic timing manipulations, but not external “clock time” manipulations. It is proposed here that intrinsic timing is more closely related to the sensorimotor integration role of the cortico-basal ganglia motor loops, the disruption of which is thought to subserve stuttering (Chang & Guenther, 2020).

It has been suggested that speaking isochronously, as in the metronome-paced condition, may bias syllable timing away from an intrinsic timing mechanism and toward an extrinsic timing mechanism, circumventing the impaired intrinsic timing mechanism in AWS and leading to greater fluency (Etchell et al., 2014). Because the perturbation in the present study more likely recruited an extrinsic timining control system (therefore yielding similar responses in both groups), it was not possible to examine the effects of switching from intrinsic to extrinsic timing control in AWS. This hypothesis could be tested with an additional study investigating the effect of metronome-timed speech using the intrinsic timing perturbation from Cai, et al. (2014).

An alternative possibility regarding the different responses found in the present study versus Cai et al. (2014) involves the magnitude of the perturbation. While the present study delayed the auditory feedback signal by 50 – 60 ms, the perturbation in Cai et al. (2014) only introduced a delay of 20 – 25 ms. This could indicate that AWS have a more difficult time detecting and/or responding to more fine-grained temporal perturbations.

Indeed, recent work indicates that AWS have less sensitivity to judging time intervals of various types (Devaraju et al., 2020; Schwartze & Kotz, 2020), although no studies have yet examined perceptual acuity for speech segment timing in AWS. Future studies would need to directly compare spectro-temporal and pure temporal perturbations of the same magnitude to confirm that magnitude differences did not lead to the differences found between Cai et al. (2014) and the present study.

4.2. Additional considerations

Previous studies of formant perturbations show that response magnitudes are not correlated with stuttering severity (Cai et al., 2012; Daliri et al., 2018) but response timing variability is (Sares et al., 2018), such that more severe AWS are more variable. In the previous timing perturbation study (Cai et al., 2014), correlation of response to severity was not reported. Therefore, correlations between responses and stuttering severity were examined in the present study. Despite there being no group differences between AWS and ANS for the timing perturbation, there was a significant positive correlation between experimental stuttering (during the *normal* condition) and the size of compensatory responses to the timing perturbation. This correlation indicates that those with a propensity to stutter during the normal speech task had a more sensitive response to extrinsic auditory feedback timing cues than those who stuttered less. The fact that, unlike stuttering rate during the experiment, stuttering severity as measured by SSI-Mod did not correlate with response magnitude may indicate that sensitivity to auditory feedback timing cues varies across time, so only the most local measure (i.e., within the same experimental session) of severity has a significant relationship. However, because there was not a significant group difference in response to the timing perturbation, it is difficult to determine how this within-group effect relates to responses of typically fluent speakers.

One potential limitation in this study pertains to the use of a stimulus phrase that is repeated almost 200 times throughout the experiment. Prior work has shown than under repeated readings of the same utterance, stuttering tends to decrease (the so-called “adaptation effect”; (Max & Baldwin, 2010). This could have an impact on the data in the present study given that there was a between-participant positive correlation between stuttering rate and timing delay in responses to the perturbation. Specifically, if stuttering rate is related to timing perturbation responses at a within-participant level in AWS, adaptation could have muted overall enhanced responses in this group. However, from informally examining the occurance of trials containing stuttering across the duration of the experiment, we did not notice a trend supporting adaptation in either the *normal* or *metronome-timed* conditions (see Supplementary Figure 3). In fact, of the four AWS who stuttered on greater than 10% of *normal* trials, only one (AWS12) had a stuttering pattern resembling adaptation; the other three showed an increase in stuttering over the course the study, possibly due to fatigue. Therefore, it is unlikely the that adaptation effect had a significant impact on the data.

A final consideration concerns the hardware and software delays that led to a global auditory feedback delay throughout the experiment. It is well-known that delayed auditory feedback (applied to an entire utterance rather than a short interval as in the timing perturbation in the present study) has a fluency-enhacing effect in people who stutter, even at delays as short as 25ms (Kalinowski et al., 1996). This may have contributed to the relative infrequency of stuttering throughout the experiment. In addition, if this global delay led to changes in sensorimotor processing in AWS, it may have masked perturbation responses that are sensitive to this type of manipulation. Unfortunately, delays of this magnitude are intrinsic to the software and hardware processing needed to accomplish online auditory feedback perturbations and exist throughout the literature to varying degrees (Kim et al., 2020). These feedback delays may therefore have an outsized effect on populations like AWS that demonstrate speech changes in response to delayed auditory feedback. However, despite such delays, several prior studies have found significant auditory feedback perturbation response differences between AWS and ANS (Bauer et al., 2007; Cai et al., 2012, 2014; Daliri et al., 2018; Daliri & Max, 2018; Loucks et al., 2012; Sares et al., 2018), suggesting that modest global delays do not prevent detection of such differences. Given this consideration, global auditory feedback delays should, to the extent possible, be minimized in future perturbation studies involving people who stutter.

5. Conclusion

The present study demonstrated that a pure temporal auditory feedback perturbation does not elicit the same response deficit in AWS that was previously found in a spectro-temporal perturbation. In addition, metronome-timed speaking may not impact the auditory feedback control processes related to speech timing, at least as implemented herein. Finally, stuttering rate during the task was a significant predictor of responses to timing perturbations in AWS such that AWS who stuttered more had larger responses than those who stuttered less. These results help clarify the nature of online auditory feedback control of speech timing in AWS and the influence of external pacing on this process.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We are grateful to Diane Constantino, Barbara Holland, and Erin Archibald for assistance with subject recruitment and data collection and Ina Jessen, Lauren Jiang, Mona Tong and Brittany Steinfeld for their help with behavioral data analysis. This work benefited from helpful discussions with, or comments from, Ayoub Daliri and other members of the Boston University Speech Lab.

Funding Statement:

The research reported here was supported by grants from the National Institutes of Health under award numbers R01DC007683 and T32DC013017.

Appendix

Equations for reference:

The onset of the /ə/ in the word “the” was detected using an adaptive short-time signal intensity root-mean-square (RMS) threshold determined by

$$thresh_{/ə/} = \frac{q_{90}(minRMS_{/ðə/}) + \ q_{10}(maxRMS_{/ðə/})}{2}, \qquad \text{(Equation A1)}$$

where $minRMS_{/ðə/}$ is the set of lowest RMS values from previous trials during production of “the”, $maxRMS_{/ðə/}$ is the set of the highest RMS values from previous trials during production of “the”, and q_x is the x^{th} quantile of the distribution of values. Note that RMS was computed in successive 32-sample (2ms) frames.

The onset and offset of the /s/ in the word “steady” were detected using an adaptive RMS threshold determined by

$$thresh_{/s/} = \frac{q_{90}(minRAT_{/ə/}) + \ q_{10}(minRAT_{/st/})}{2}, \qquad \text{(Equation A2)}$$

where $minRAT_{/ə/}$ is the set of lowest values from previous trials of the ratio between pre-emphasized RMS and non-filtered RMS during production of the /ə/ (in the word “the”), $minRAT_{/st/}$ is the set of lowest values from previous trials of the ratio between pre-emphasized RMS and non-filtered RMS during production of /st/ (in the word “steady”), and q_x is the x^{th} quantile of the distribution of values.

Vitae

Saul Frankford is a postdoctoral research fellow in the Department of Speech, Language, & Hearing Sciences at Boston University His research interests include using neural and behavioral techniques to study the mechanisms of speech production and speech disorders.

Shanqing Cai is a Staff Software Engineer at Google, focusing on tool development for deep learning and artificial intelligence.

Alfonso Nieto-Castañón is a Senior Research Scientist in the Department of Speech, Language, and Hearing Sciences at Boston University. He specializes in computational modeling, statistical analysis, and software development for neuroimaging.

Frank Guenther is a faculty member in the Departments of Speech, Language, and Hearing Sciences and Biomedical Engineering at Boston University and in Speech and Hearing Bioscience and Technology at Harvard/MIT. Guenther’s research combines computational modeling and neuroimaging to investigate the neural bases of speech in normal and disordered populations. He also develops brain-machine interfaces aimed at restoring speech capabilities to profoundly paralyzed patients.

References

Abur D, Subaciute A, Daliri A, Lester-Smith RA, Lupiani AA, Cilento D, Enos NM, Weerathunge HR, Tardif MC, & Stepp CE (2021). Feedback and feedforward auditory-motor processes for voice and articulation in parkinson’s disease. Journal of Speech, Language, and Hearing Research, 64(12), 4682–4694. 10.1044/2021_JSLHR-21-00153

Alm PA (2004). Stuttering and the basal ganglia circuits: A critical review of possible relations. Journal of Communication Disorders, 37(4), 325–369. 10.1016/j.jcomdis.2004.03.001 [PubMed: 15159193]

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

Barber V (1940). Studies in the psychology of stuttering, XVI: Rhythm as a distraction in stuttering. Journal of Speech Disorders, 5(1), 29–42. 10.1044/jshd.0501.29

Bauer JJ, Hubbard Seery C, LaBonte R, & Ruhnke L (2007). Voice F0 responses elicited by perturbations in pitch of auditory feedback in individuals that stutter and controls. The Journal of the Acoustical Society of America, 121(5), 3201–3201. 10.1121/1.4782465

Bernsee S (1999, August 18). Time stretching and pitch shifting of audio signals – an overview. Stephan Bernsee's Blog. http://blogs.zynaptiq.com/bernsee/time-pitch-overview/

Bloodstein O, & Ratner NB (2008). A handbook on stuttering (6th ed). Thomson/Delmar Learning.

Brady JP (1969). Studies on the metronome effect on stuttering. Behaviour Research and Therapy, 7(2), 197–204. 10.1016/0005-7967(69)90033-3 [PubMed: 5808691]

Braun AR, Varga M, Stager S, Schulz G, Selbie S, Maisog JM, Carson RE, & Ludlow CL (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering. An H2 (15) O positron emission tomography study. Brain, 120(5), 761–784. [PubMed: 9183248]

Burnett TA, Freedland MB, Larson CR, & Hain TC (1998). Voice F0 responses to manipulations in pitch feedback. The Journal of the Acoustical Society of America, 103(6), 3153–3161. 10.1121/1.423073 [PubMed: 9637026]

Cai S, Beal DS, Ghosh SS, Guenther FH, & Perkell JS (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 [PubMed: 24486601]

Cai S, Beal DS, Ghosh SS, Tiede MK, Guenther FH, & Perkell JS (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 [PubMed: 22911857]

Cai S, Ghosh SS, Guenther FH, & Perkell JS (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. Journal of Neuroscience, 31(45), 16483–16490. 10.1523/JNEUROSCI.3653-11.2011 [PubMed: 22072698]

Chang S-E, & Guenther FH (2020). Involvement of the cortico-basal ganglia-thalamocortical loop in developmental stuttering. Frontiers in Psychology, 10. 10.3389/fpsyg.2019.03088

Chang S-E, & Zhu DC (2013). Neural network connectivity differences in children who stutter. Brain, 136(12), 3709–3726. 10.1093/brain/awt275 [PubMed: 24131593]

Chen SH, Liu H, Xu Y, & Larson CR (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. The Journal of the Acoustical Society of America, 121(2), 1157–1163. [PubMed: 17348536]

Constantino CD, Leslie P, Quesal RW, & Yaruss JS (2016). A preliminary investigation of daily variability of stuttering in adults. Journal of Communication Disorders, 60, 39–50. 10.1016/j.jcomdis.2016.02.001 [PubMed: 26945438]

Daliri A, & Max L (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. Cortex, 99, 55–68. 10.1016/j.cortex.2017.10.019 [PubMed: 29169049]

Daliri A, Wieland EA, Cai S, Guenther FH, & Chang S-E (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. Developmental Science, 21(2), e12521. 10.1111/desc.12521

Davidow JH (2014). Systematic studies of modified vocalization: The effect of speech rate on speech production measures during metronome-paced speech in persons who stutter: Speech rate and speech production measures during metronome-paced speech in PWS. International Journal of Language & Communication Disorders, 49(1), 100–112. 10.1111/1460-6984.12050 [PubMed: 24372888]

De Nil LF, Kroll RM, Kapur S, & Houle S (2000). A positron emission tomography study of silent and oral single word reading in stuttering and nonstuttering adults. Journal of Speech, Language, and Hearing Research, 43(4), 1038–1053. 10.1044/jslhr.4304.1038

Devaraju DS, Maruthy S, & Kumar AU (2020). Detection of gap and modulations: auditory temporal resolution deficits in adults who stutter. Folia Phoniatrica et Logopaedica, 72(1), 13–21. 10.1159/000499565 [PubMed: 31132766]

Etchell AC, Johnson BW, & Sowman PF (2014). Behavioral and multimodal neuroimaging evidence for a deficit in brain timing networks in stuttering: A hypothesis and theory. Frontiers in Human Neuroscience, 8. 10.3389/fnhum.2014.00467

Falk S, Müller T, & Dalla Bella S (2015). Non-verbal sensorimotor timing deficits in children and adolescents who stutter. Frontiers in Psychology, 6. 10.3389/fpsyg.2015.00847

Floegel M, Fuchs S, & Kell CA (2020). Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. Nature Communications, 11(1), 2839. 10.1038/s41467-020-16743-2

Foundas AL, Bollich AM, Corey DM, Hurley M, & Heilman KM (2001). Anomalous anatomy of speech-language areas in adults with persistent developmental stuttering. Neurology, 57(2), 207–215. 10.1212/WNL.57.2.207 [PubMed: 11468304]

Foundas AL, Bollich AM, Feldman J, Corey DM, Hurley M, Lemen LC, & Heilman KM (2004). Aberrant auditory processing and atypical planum temporale in developmental stuttering. Neurology, 63(9), 1640–1646. 10.1212/01.WNL.0000142993.33158.2A [PubMed: 15534249]

Fowler CA (1980). Coarticulation and theories of extrinsic timing. Journal of Phonetics, 8(1), 113–133. 10.1016/S0095-4470(19)31446-9

Fox PT, Ingham RJ, Ingham JC, Zamarripa F, Xiong JH, & Lancaster JL (2000). Brain correlates of stuttering and syllable production. A PET performance-correlation analysis. Brain: A Journal of Neurology, 123 (Pt 10), 1985–2004. [PubMed: 11004117]

Frankford SA, Cai S, Nieto-Castañón A, & Guenther FH (2022). Auditory feedback control in adults who stutter during metronome-paced speech II. Formant Perturbation. Journal of Fluency Disorders, 74, 105928. 10.1016/j.jfludis.2022.105928 [PubMed: 36063640]

Frankford SA, Heller Murray ES, Masapollo M, Cai S, Tourville JA, Nieto-Castañón A, & Guenther FH (2021). The neural circuitry underlying the “rhythm effect” in stuttering. Journal of Speech, Language, and Hearing Research, 64(6S), 2325–2346. 10.1044/2021_JSLHR-20-00328

Gelfer MP, & Young SR (1997). Comparisons of intensity measures and their stability in male and female speakers. Journal of Voice, 11(2), 178–186. 10.1016/S0892-1997(97)80076-8 [PubMed: 9181541]

Giraud A (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. Brain and Language, 104(2), 190–199. 10.1016/j.bandl.2007.04.005 [PubMed: 17531310]

Guenther FH (2016). Neural control of speech. MIT Press.

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

Howell P, Au-Yeung J, & Rustin L (1997). Clock and motor variances in lip-tracking: A comparison between children who stutter and those who do not. In Hulstijn W, Peters HFM, & Van Lieshout P, Speech production: Motor control, brain research and fluency disorders (pp. 573–578). Elsevier Scientific.

Howell P, & Sackin S (2000). Speech rate modification and its effects on fluency reversal in fluent speakers and people who stutter. Journal of Developmental and Physical Disabilities, 12(4), 291–315. 10.1023/A:1009428029167 [PubMed: 18259598]

IEEE Recommended Practice for Speech Quality Measurements (No. 17; pp. 227–246). (1969). IEEE Transactions on Audio and Electroacoustics. 10.1109/IEEESTD.1969.7405210

Kalinowski J, Stuart A, Sark S, & Armson J (1996). Stuttering amelioration at various auditory feedback delays and speech rates. International Journal of Language & Communication Disorders, 31(3), 259–269. 10.3109/13682829609033157

Kelso JAS, & Tuller B (1987). Intrinsic time in speech production: Theory, methodology, and preliminary observations. In Keller E & Gopnik M (Eds.), Motor and sensory processes of language (pp. 203–222). Erlbaum.

Kim KS, Wang H, & Max L (2020). It's about time: Minimizing hardware and software latencies in speech research with real-time auditory feedback. Journal of Speech, Language, and Hearing Research, 63(8), 2522–2534. 10.1044/2020_JSLHR-19-00419

Kleinow J, & Smith A (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. Journal of Speech Language and Hearing Research, 43(2), 548. 10.1044/jslhr.4302.548

Lee A, & Kawahara T (2009, January). Recent development of open-source speech recognition engine julius. Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference.

Loucks T, Chon H, & Han W (2012). Audiovocal integration in adults who stutter: Audiovocal integration in stuttering. International Journal of Language & Communication Disorders, 47(4), 451–456. 10.1111/j.1460-6984.2011.00111.x [PubMed: 22788230]

Lu C, Peng D, Chen C, Ning N, Ding G, Li K, Yang Y, & Lin C (2010). Altered effective connectivity and anomalous anatomy in the basal ganglia-thalamocortical circuit of stuttering speakers. Cortex, 46(1), 49–67. 10.1016/j.cortex.2009.02.017 [PubMed: 19375076]

Ludlow CL, Rosenberg J, Salazar A, Grafman J, & Smutok M (1987). Site of penetrating brain lesions causing chronic acquired stuttering. Annals of Neurology, 22(1), 60–66. 10.1002/ana.410220114 [PubMed: 3631921]

MacKay DG, & MacDonald MC (1984). Stuttering as a sequencing and timing disorder. In Curlee RF & Perkins WH (Eds.), Nature and treatment of stuttering: New directions (pp. 261–282). College-Hill Press.

Max L, & Baldwin CJ (2010). The role of motor learning in stuttering adaptation: Repeated versus novel utterances in a practice–retention paradigm. Journal of Fluency Disorders, 35(1), 33–43. 10.1016/j.jfludis.2009.12.003 [PubMed: 20412981]

Max L, Caruso AJ, & Gracco VL (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. [PubMed: 12647900]

McClean MD, & Runyan CM (2000). Variations in the relative speeds of orofacial structures with stuttering severity. Journal of Speech Language and Hearing Research, 43(6), 1524. 10.1044/jslhr.4306.1524

Mink JW (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. Progress in Neurobiology, 50(4), 381–425. [PubMed: 9004351]

Mirawdeli A, & Howell P (2016). Is it necessary to assess fluent symptoms, duration of dysfluent events, and physical concomitants when identifying children who have speech difficulties? Clinical Linguistics & Phonetics, 30(9), 696–719. 10.1080/02699206.2016.1179345 [PubMed: 27315282]

Mitsuya T, MacDonald EN, & Munhall KG (2014). Temporal control and compensation for perturbed voicing feedback. The Journal of the Acoustical Society of America, 135(5), 2986–2994. 10.1121/1.4871359 [PubMed: 24815278]

Natke U, & Kalveram KT (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. Journal of Speech, Language, and Hearing Research, 44(3), 577–584. 10.1044/1092-4388(2001/045)

Nieto-Castañón A (2020). Handbook of functional connecitivity Magnetic resonance Imaging methods in CONN. Hilbert Press.

Niziolek CA, & Guenther FH (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. Journal of Neuroscience, 33(29), 12090–12098. 10.1523/JNEUROSCI.1008-13.2013 [PubMed: 23864694]

Nudelman HB, Herbrich KE, Hess KR, Hoyt BD, & Rosenfield DB (1992). A model of the phonatory response time of stutterers and fluent speakers to frequency-modulated tones. The Journal of the Acoustical Society of America, 92(4), 1882–1888. 10.1121/1.405263 [PubMed: 1401532]

Ogane R, & Honda M (2014). Speech compensation for time-scale-modified auditory feedback. Journal of Speech, Language, and Hearing Research, 57(2). 10.1044/2014_JSLHR-S-12-0214

Oschkinat M, & Hoole P (2020). Compensation to real-time temporal auditory feedback perturbation depends on syllable position. The Journal of the Acoustical Society of America, 148(3), 1478–1495. 10.1121/10.0001765 [PubMed: 33003874]

Pellegrino F, Farinas J, & Rouas J-L (2004). Automatic estimation of speaking rate in multilingual spontaneous speech. Speech Prosody 2004, 517–520.

Purcell DW, & Munhall KG (2006). Compensation following real-time manipulation of formants in isolated vowels. The Journal of the Acoustical Society of America, 119(4), 2288. 10.1121/1.2173514 [PubMed: 16642842]

Riley GD (2008). SSI-4, Stuttering severity intrument for children and adults (4th ed.). Pro Ed.

Sares AG, Deroche MLD, Shiller DM, & Gracco VL (2018). Timing variability of sensorimotor integration during vocalization in individuals who stutter. Scientific Reports, 8(1), 16340. 10.1038/s41598-018-34517-1 [PubMed: 30397215]

Schwartze M, & Kotz SA (2020). Decreased sensitivity to changing durational parameters of syllable sequences in people who stutter. Language, Cognition and Neuroscience, 35(2), 179–187. 10.1080/23273798.2019.1642499

Stager SV, Jeffries KJ, & Braun AR (2003). Common features of fluency-evoking conditions studied in stuttering subjects and controls: An PET study. Journal of Fluency Disorders, 28(4), 319–336. 10.1016/j.jfludis.2003.08.004 [PubMed: 14643068]

Starkweather CW, Franklin S, & Smigo TM (1984). Vocal and finger reaction times in stutterers and nonstutterers: differences and correlations. Journal of Speech, Language, and Hearing Research, 27(2), 193–196. 10.1044/jshr.2702.193

Theys C, De Nil L, Thijs V, van Wieringen A, & Sunaert S (2013). A crucial role for the cortico-striato-cortical loop in the pathogenesis of stroke-related neurogenic stuttering: Neural Network of Neurogenic Stuttering. Human Brain Mapping, 34(9), 2103–2112. 10.1002/hbm.22052 [PubMed: 22451328]

Tourville JA, Cai S, & Guenther F (2013). Exploring auditory-motor interactions in normal and disordered speech. 060180–060180. 10.1121/1.4800684

Toyomura A, Fujii T, & Kuriki S (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. NeuroImage, 57(4), 1507–1516. 10.1016/j.neuroimage.2011.05.039 [PubMed: 21624474]

Van Riper C (1963). Speech correction (4th ed.). Prentice-Hall.

Weerathunge HR, Abur D, Enos NM, Brown KM, & Stepp CE (2020). Auditory-motor perturbations of voice fundamental frequency: feedback delay and amplification. Journal of Speech, Language, and Hearing Research, 63(9), 2846–2860. 10.1044/2020_JSLHR-19-00407

Wingate ME (2002). Foundations of stuttering. Academic Press.

Yairi E, & Ambrose N (2013). Epidemiology of stuttering: 21st century advances. Journal of Fluency Disorders, 38(2), 66–87. 10.1016/j.jfludis.2012.11.002 [PubMed: 23773662]

Yang Y, Jia F, Siok WT, & Tan LH (2016). Altered functional connectivity in persistent developmental stuttering. Scientific Reports, 6(1). 10.1038/srep19128

Highlights

- Adults who stutter show typical responses to explicit speech timing perturbations
- Metronome-timed speech does not alter responses to timing perturbations
- Auditory feedback is used to control speech timing during naturalistic utterances

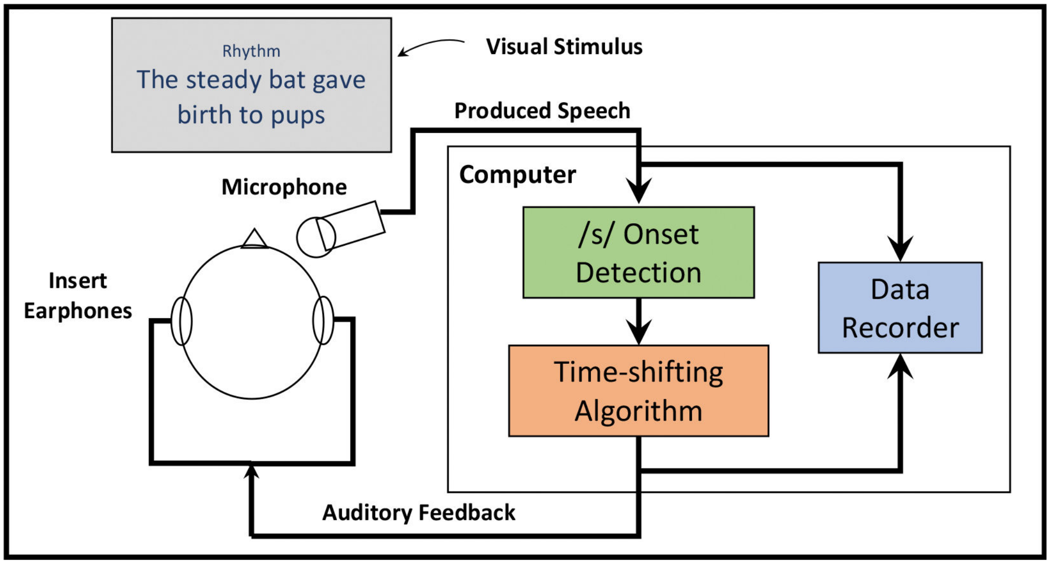

Figure 1.

A schematic diagram showing the setup for the experiment. Following the presentation of an orthographic stimulus sentence and a condition cue (“Normal” or “Rhythm”), participants read the sentence according to the cue. Participants’ speech signal was recorded and fed to an experimental computer running Audapter. On perturbed trials, detection of the onset of /s/ in “steady” was used to initiate the pre-programmed auditory perturbation which was fed back to the participant via insert earphones. Both the perturbed and unperturbed signals were recorded for further analysis.

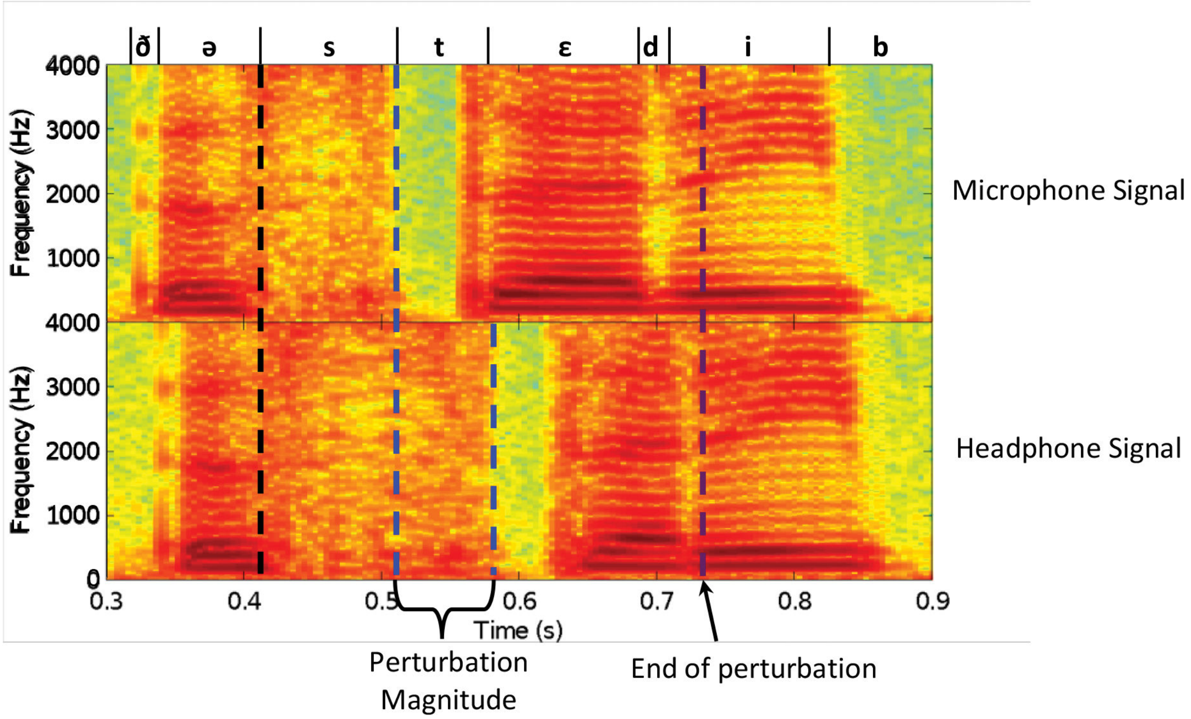

Figure 2.

: Example spectrograms of “The steady bat” from a timing perturbation trial generated from the recorded microphone signal (top) and headphone signal (bottom). The dashed black line indicates the onset of the /s/ in “steady” in the microphone signal. The first dashed blue line indicates the offset of the /s/ in “steady” in the microphone signal, and the second dashed blue line indicates the offset of the /s/ in “steady” in the headphone signal. The dashed purple line indicates when auditory feedback is returned to normal. Phoneme boundaries and international phonetic alphabet symbols are indicated above the microphone signals. Hz = hertz.

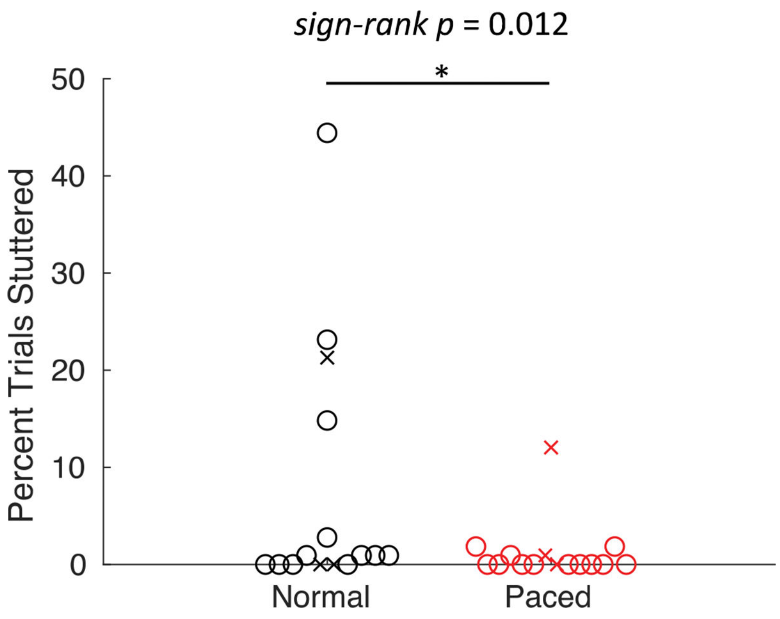

Figure 3.

Comparison of stuttering between the *normal* and *metronome-paced* conditions for AWS. Circles and ‘x’s represent individual male and female participants, respectively.

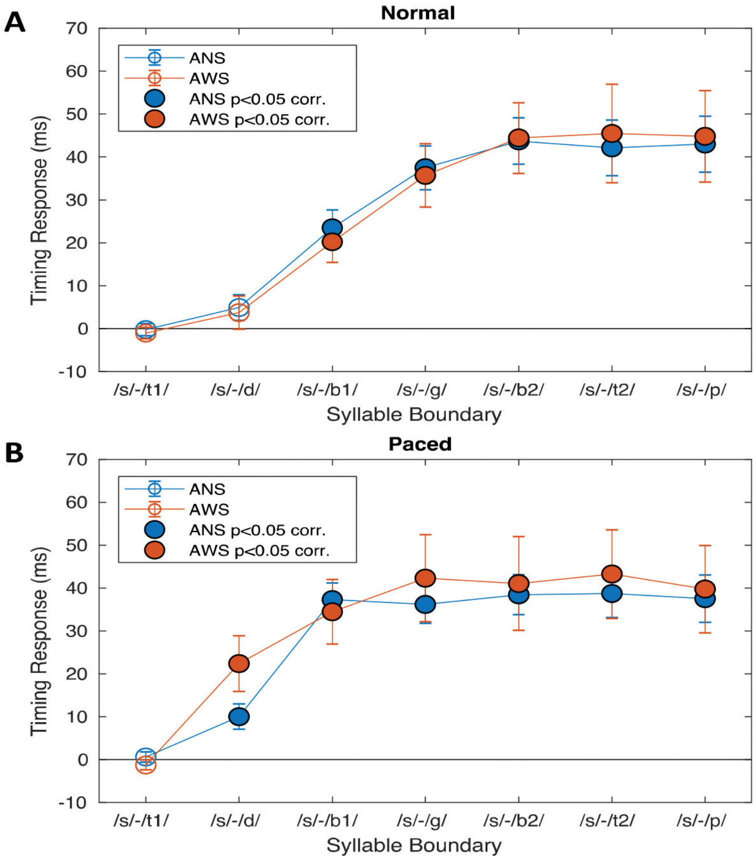

Figure 4.

Cumulative speech timing response between the perturbed and non-perturbed conditions at each of seven sound/syllable boundaries (see Table 2). A. Responses during the *normal* speaking condition. The blue and orange curves correspond to the ANS and AWS groups, respectively. Filled circles indicate responses that differ significantly from 0 (one-sample t-test, p <.05, Bonferroni-corrected for 7 time-points). B. Responses during the *metronome-paced* speaking condition. Colors represent the same as in A. Error bars indicate the standard error of the mean of each group.

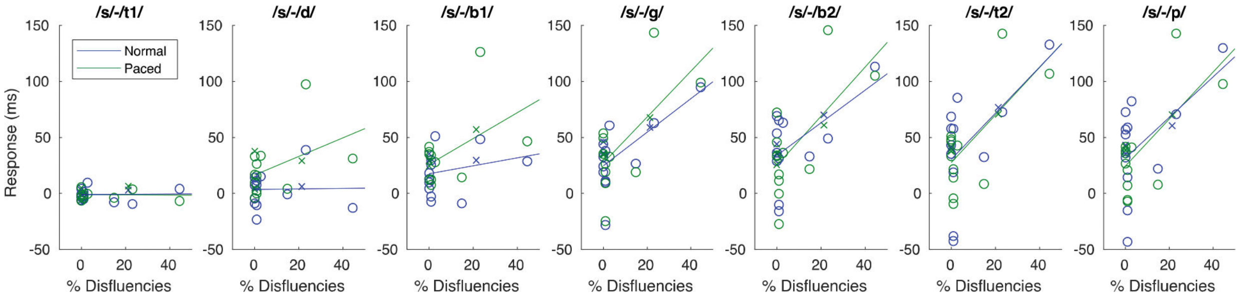

Figure 5.

Scatterplots comparing stuttering rate during the *normal* condition with cumulative timing perturbation responses in AWS at each of seven sound/syllable boundaries (see Table 2). Circles and ‘x’s indicate individual male and female AWS, respectively, for either the *normal* (blue) or *metronome-paced* (green) conditions. Least squares lines across all participants for each condition are superimposed on the data.

Table 1.

Demographic and stuttering severity data from adults who stutter (AWS).

Participant ID	Age	Gender	SSI-4 Composite	SSI-Mod	Stuttering Rate - Normal	Stuttering Rate - Paced
AWS01	19	F	28	19	0%	0%
AWS02	22	F	31	26	21.30%	12.04%
AWS03	18	F	14	11	0%	0.93%
AWS04	23	M	20	15	0%	0%
AWS05	20	M	9	7	0.93%	0%
AWS06	23	M	42	29	0.93%	0%
AWS07	44	M	20	16	2.78%	0%
AWS08	24	M	21	15	0%	0%
AWS09	24	M	20	15	14.81%	0.93%
AWS10	29	M	14	12	0.93%	0%
AWS11	20	M	18	15	0%	0%
AWS12	43	M	27	25	44.44%	1.85%
AWS13	35	M	30	19	0.93%	0%
AWS14	22	M	27	18	0%	0%
AWS15	20	M	24	14	23.15%	1.85%

SSI-4 = Stuttering Severity Index–Fourth Edition; SSI-Mod = a modified version of the SSI-4 that does not include a subscore related to concomitant movements; Stuttering rate = the percentage of trials containing stutters during the normal or metronome-paced conditions; F = female; M = male.

Table 2.

Symbols used to denote sound/syllable boundaries in the present study.

Symbol	Landmark
/s/	Onset of “steady”
/t1/	s-t boundary in “steady”
/d/	Onset of /d/ in “steady”
/b1/	Onset of “bat”
/g/	Onset of “gave”
/b2/	Onset of “birth”
/t2/	Onset of “to”
/p/	Onset of “pups”

Table 3.

Descriptive and inferential statistics for speaking rate and CV-ISD.

Measure	*ANS*		*AWS*		Main effect of Group:	Main effect of Condition:	Interaction:
	Normal	***Paced***	***Normal***	***Paced***			
Speaking rate (ISD/sec)	4.0 ± 0.2	3.6 ± 0.1	3.9 ± 0.2	3.7 ± 0.1	*F*(1, 51.9) = 3.17, *p* = .08	***F*(1, 29) = 65.27, *p* < .001**	***F*(1, 29) = 6.17, *p* = .02**
CV-ISD	0.27 ± 0.06	0.10 ± 0.02	0.26 ± 0.05	0.10± 0.02	*F*(1, 57.8) = 0.94, *p* = .34	***F*(1, 29) = 163.80, *p* < .001**	*F*(1, 29) = 0.41, *p* = .53

Error estimates indicate standard deviations. Significant effects are highlighted in bold.

ANS = adults who do not stutter, AWS = adults who stutter, ISD = intersyllable duration, CV-ISD = coefficient of variation of the ISD within a trial.

Table 4.

Results from a multivariate general linear model predicting perturbation responses. The dependent variables were the cumulative perturbation timing delays from each participant in each condition projected onto the first three principal components of a principal components analysis.

Predictor	***df***	***F***	***p***
Group	3, 24	0.45	.72
Perturbation Magnitude	3, 24	3.42	.03 *
Mean Speaking Rate	3, 24	0.96	.43
SSI-Mod	3, 24	2.39	.09
Condition	3, 24	3.15	.04 *
Group × Condition	3, 24	0.91	.45
SSI-Mod × Condition	3, 24	1.39	.27

SSI-Mod = a modified versionof the SSI-4 that does not include a subscore related to concomitant movements, df = degrees of freedom

* = $p < .05$.