Auditory feedback control in adults who stutter during metronome-paced speech II. Formant 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: Prior work has shown that Adults who stutter (AWS) have reduced and delayed responses to auditory feedback perturbations. This study aimed to determine whether external timing cues, which increase fluency, resolve auditory feedback processing disruptions.

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

Results: During normally paced speech, ANS showed a significant compensatory response to the perturbation by the end of the perturbed vowel, while AWS did not. In the metronome-paced condition, which significantly reduced the disfluency rate, the opposite was true: AWS showed a significant response by the end of the vowel, while ANS did not.

Conclusion: These findings indicate a potential link between the reduction in stuttering found during metronome-paced speech and changes in auditory motor integration in AWS.

Keywords

stuttering; auditory feedback; metronome; speech timing control

1. Introduction

The mechanisms underlying persistent developmental stuttering are not well understood, though over 100 years of anecdotal and experimental evidence points to disruptions in sensorimotor processing. Several theories put forward in the literature suggest that stuttering occurs due to a reduced capacity for transforming sensory information to motor information (and vice versa; Hickok et al., 2011; Max et al., 2004; Neilson & Neilson, 1987). According to these theories, individual moments of stuttering occur when either a) insufficient neural resources are available to make this transformation and the whole system stops (Neilson & Neilson, 1987), or b) false errors are detected due to a disruption of the process that

generates a sensory prediction (Hickok et al., 2011; Max et al., 2004), leading either to delays until the mismatch is resolved (Max et al., 2004), or iterative (erroneous) correction signals (Hickok et al., 2011).

Several studies have used online sensory feedback manipulations to examine the integrity of these sensorimotor processes in people who stutter during speech. In general, these studies use signal processing techniques to modify components of sensory feedback (e.g. voice fundamental frequency [f0] or vowel formants) as a person speaks and report the magnitude and latency of speech responses (see Barrett & Howell, 2021, and Bradshaw et al., 2021, for recent detailed reviews). Speakers typically compensate for these perturbations by adjusting their articulatory musculature so that the modified component of their speech output (f0 or vowel formant) changes in the opposite direction of the shift. For example, if a speaker's first formant frequency (F1) is increased in their auditory feedback (making "bet" sound more like "bat"), they will generally respond by lowering the F1 of their actual speech output (and sounding more like "bit"). Results consistently demonstrate that adults who stutter (AWS) exhibit responses that are delayed (f0 and second vowel formant perturbations; Bauer et al., 2007; Cai et al., 2014; Loucks et al., 2012), reduced (f0, first/second vowel formants; Cai et al., 2012; Daliri et al., 2018; Daliri & Max, 2018; Loucks et al., 2012; Sares et al., 2018), and/or more temporally variable (f0; Sares et al., 2018) compared to adults who do not stutter (ANS). In the somatosensory domain, AWS have shown reduced speech compensatory responses to a bite block (Namasivayam & Van Lieshout, 2011) and reduced attenuation of (non-speech) jaw opening in response to masseter muscle vibration (Loucks & De Nil, 2006). These studies suggest that during sensorimotor tasks like speech, AWS exhibit processing delays and/or reduced scaling of corrective movements in line with the sensorimotor theories discussed above.

Speaking along with an external timing cue like a metronome generally has the effect of reducing overt disfluencies in AWS (e.g., Andrews et al., 1982; Brady, 1969; Braun et al., 1997; Davidow, 2014; Stager et al., 2003; Toyomura et al., 2011). It is unclear, however, whether this is related to changes in sensorimotor processing. Some work posits that an external timing source reduces reliance on impaired "internal" timing mechanisms for speech sequencing (Alm, 2004; Etchell et al., 2014; Guenther, 2016) which depend (in part) on sensory feedback. This would suggest that sensorimotor processes related to sequencing might be affected by external timing cues, but those related to error correction might be unchanged. Other work has found that pacing speech to an external rhythmic stimulus leads to increased vowel duration (Brayton & Conture, 1978)(Klich & May, 1982; Stager et al., 1997) and decreased peak intraoral pressure (Hutchinson & Navarre, 1977; Stager et al., 1997), which are also components of slowed speech rate (Boucher & Lamontagne, 2001; Kessinger & Blumstein, 1998). Even when speaking rate is properly controlled for, metronome-timed speech (as well as other fluency-inducing types of speech) results in a decrease in the proportion of phonated intervals in the 30-100ms range (Davidow, 2014; Davidow et al., 2009, 2010). In the context of the previously mentioned theories, reducing the number of short-duration phonated intervals may allow more time for sensorimotor transformations to take place without error (Max et al., 2004).

The aim of the present study was to test whether externally paced speech leads to fluency by helping resolve speech sensorimotor processing disruptions. This was carried out by examining the effects of paced speech on responses to online auditory feedback perturbations of vowel formant frequencies in AWS and ANS.

2. Methods and Materials

2.1. Participants

Study participants included 31 participants: 15 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]). This distribution of males and females in each group is consistent with the gender breakdown of adults who stutter in general (Bloodstein, 1995). Participants had no prior history of speech, language, or hearing disorders (other than stuttering for the AWS group), were native speakers of American English, and passed audiometric screenings with binaural pure-tone hearing thresholds of less than 25 dB HL at 500, 1000, 2000, and 4000 Hz. Stuttering severity was evaluated by a trained speech-language pathologist using the Stuttering Severity Instrument – Fourth Edition (SSI-4; Riley, 2008), based on video-recorded speech samples during three tasks: 1) an in-person conversation, 2) a phone conversation, and 3) reading the Grandfather passage (Van Riper, 1963) aloud. AWS ranged from subclinical (9) to very severe (42) with a mean score of 23. Though one individual had a score just below the lowest percentile listed in score manual, they identified as a person who stutters during the screening interview. As SSI-4 scores can vary significantly from day to day, even to levels below this arbitrary threshold (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

The experimental setup is schematized in Figure 1. Participants were seated upright in front of a computer monitor in a sound-attenuating booth and fitted with 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. Etymotic ER1 insert earphones (Etymotic Research, Inc.) were used to deliver auditory stimuli. 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 earphone signal was played 4.5 dB louder than the microphone input. This amplification helped reduce participants' ability to hear their air- and bone-conducted non-perturbed feedback, consistent with prior auditory feedback perturbation studies (Abur et al., 2021; Weerathunge et al., 2020).

2.3. Stimuli

This study used one "target" sentence ("The steady bat gave birth to pups") on which all experimental manipulations were applied. To reduce boredom and maintain participant attention, fifteen "filler" sentences, each containing eight syllables, were selected from the Harvard sentence pool (i.e., the Revised List of Phonetically Balanced Sentences; IEEE Recommended Practice for Speech Quality Measurements, 1969).

2.4. Procedure

Participants read aloud sentences displayed on the computer monitor under two speaking conditions: the *metronome-paced* speech condition, where each syllable was evenly spaced, and the *normal* speech condition, where speakers used a natural (unmodified) stress pattern. 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 rate (approximately 222 beats/minute) was chosen so that participants' mean speech rate during the *metronome-paced* condition would approximate that of the *normal* condition (using an estimate 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 read the sentence with even stress at the rate of the tone stimuli aligning each syllable to a beat. On *normal speech* trials, participants ignored the pacing tones, and spoke with normal rate and rhythm.

At the beginning of the experiment, three brief training runs were included to acclimate participants to the task and ensure they spoke with a consistent rate, loudness, and isochronicity. During the first training run, participants received visual feedback on their loudness (a horizontal bar in relation to an upper and lower boundary). To account for natural differences in the speech sound intensity of males and females during conversational speech (Gelfer & Young, 1997), 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. A calibration procedure was carried out prior to each experimental session whereby a sustained complex tone was played 10 cm from the microphone, and the signal in Audapter was associated with an A-weighted decibel level measured by a sound level meter at the position of the microphone. 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), so lower coefficients of variance indicate greater isochronicity. Subjects were trained to speak with a CV-ISD less than 0.25 for the *metronome-paced speech* trials.

Three experimental runs contained 80 speech trials each, 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 perturbation of the first vowel formant

(F1 perturbation) on the phoneme /ɛ/ in "steady" (see Section 2.5. for details). The other quarter included a brief timing perturbation, but this condition was not analyzed for the present paper and will not be discussed further. Within each run, trial order was pseudo-randomized such that every set of 10 trials contained two of each type of perturbation, one *metronome-paced* and one *normal*, and perturbed trials were always separated with at least one non-perturbed trial to minimize adaptation or interactions among the perturbations. Thus, 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 an F1 perturbation and 48 unperturbed trials. One subject only completed 200 trials (160 target trials, 80 *normal*, 80 *metronome-paced*, each containing 40 unperturbed trials and 20 trials with a F1 perturbation) due to a technical error.

2.5. Focal Perturbation

The F1 perturbation was applied to the vowel /ɛ/ in the word "steady" from the target sentence using an adaptive short-time root-mean-square (RMS) amplitude threshold (based on previous trials; see Equation A1 in the Appendix for details) to determine the onset and offset. Perturbation onset occurred at the conclusion of the preceding /s/, when the ratio of the pre-emphasized (i.e. high-pass filtered) RMS and the unfiltered RMS went below an adaptive threshold (Equation A2). The perturbation was removed when the offset of the /ɛ/ in "steady" was detected using an adaptive RMS slope threshold (see Equations A3 and A4 in the Appendix for details).

The perturbation was carried out as previously described in Cai et al. (2012). Briefly, the microphone signal was digitized at a sample frequency of 48000 Hz and downsampled by a factor of 3 to 16000 Hz for processing. An autoregressive linear predictive coding algorithm, followed by a dynamic-programming tracking algorithm (Xia & Espy-Wilson, 2000), was used to estimate the formant frequencies in near real time. The tracked formant frequencies were then mapped to new, shifted values. In this experiment, fixed-ratio (+25%) shifting of F1 was used. Once the shifted formant frequencies were determined, a pole-substituting digital filter served to bring the formant resonance peaks from their original values to the new ones.

The latency in Audapter between the microphone signal and the processed headphone signal was 8 ms for the F1 perturbation trials and 16 ms for the unperturbed trials. An additional 16 ms of delay was incurred by the hardware setup across both conditions. Thus, the total latency was approximately 24 ms for the F1 perturbation trials and 32 ms for the nonperturbed trials. This discrepancy occurred due to the need to match the latency of another condition not discussed in the present manuscript. While this 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.

2.6. Analyses

2.6.1. Data Processing—Phoneme boundary timing information for every trial was determined using an automatic speech recognition (ASR) engine, Julius (Lee & Kawahara, 2009), in conjunction with the free *VoxForge* American English acoustic

models (voxforge.org). Each trial was manually inspected (in random order and blinded to condition) to ensure correct automatic detection of phoneme boundaries. Trials where gross ASR errors occurred were removed. For perturbation trials, the speech spectrogram from the microphone and the headphone signal were compared to ensure the perturbation occurred during the /ɛ/ in "steady." Trials where this was not the case were discarded from further analysis. In addition, trials in which the subject did not read the sentence correctly or spoke using the wrong stress pattern (i.e. spoke isochronously when they were cued to speak normally or vice versa) were eliminated from further analysis. Trials in which the subject produced a disfluency categorized as a stutter (determined by the experimenter and a speech-language therapist) were removed from the perturbation analysis (though the proportion of trials removed due to disfluency were used to calculate disfluency rate for the experiment). To ensure that large deviations in speaking rate across trials did not affect comparisons across trials and participants, trials where subjects spoke outside the trained mean ISD (220 ms – 320 ms) were eliminated. In total, an average of 13.8% of trials for ANS (SD: 6.5%) and 15.8% of trials for AWS (SD: 10.8%) were excluded. This group difference was not significant ($t = 0.63$, $p = .53$).

The fluency-enhancing effect of isochronous pacing in the AWS group was evaluated by comparing the percentage of trials eliminated due to stuttering between the two speaking conditions using a non-parametric Wilcoxon signed-rank test. Intersyllable timing measures from each trial were used to compute 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. Speaking rate was determined by calculating the reciprocal of the ISD from each trial (1/ISD), resulting in a measure with units of syllables per second. Linear mixed effects models were used to compare Rate and CV-ISD between groups and speaking conditions, treating group, condition, and group × condition interaction as fixed effects and subjects as random effects.

2.6.2. F1 Perturbation

Processing.: Formant trajectories for the /ɛ/ in "steady" were extracted using a semi-automated process. First, the Audapter-determined formant trajectories were cropped within the ASR boundaries of the /ɛ/ and subsequent /d/. Only continuous segments surpassing an RMS intensity threshold were retained. This threshold was initially set to the value used for detecting the onset of the perturbation described in Section 2.5. Adjustments to this threshold were made for vowel onset and offset in each subject to eliminate highly variable endpoints of the formant traces caused by pauses in voicing. Formant traces that were less than 30 frames (60 ms) were flagged for manual inspection and relabeling by a researcher. Traces were then time-normalized through linear interpolation for direct comparison across trials, and outliers (traces greater than 3 standard deviations away from the mean at any time point) were also flagged for manual inspection and relabeling. Manual relabeling was carried out blind to perturbation condition and consisted of labeling the onset and offset of a smooth, consistent first formant trace, often bounded by sudden changes in formant velocity due to changes in voicing status, and leaving a 2 ms buffer on either end. If a trace did not have a clear boundary, the entire trace was kept. This process of identifying outliers and relabeling trials was repeated until sudden changes in formant velocity were

not identifiable at the beginning and end of each trace by visual inspection. Trials where a clear formant trajectory was not identifiable, where sudden large deviations occurred in an otherwise smooth formant trace, or where a vowel formant trace was shorter than 30ms were removed. During this process, 0.21% of total vowel traces were removed (0.23% from ANS, 0.19% from AWS). Using a Wilcoxon rank-sum test, this group difference was not found to be statistically significantly ($W = 250$, $p = .82$).

Compensation Analysis: Compared to previous formant perturbation studies, where perturbed vowels were produced in the context of lengthened single-syllable productions (e.g., Tourville et al., 2008) or continuous formant trajectories (Cai et al., 2011, 2014), each /ɛ/ vowel in the present study was relatively short in duration (~ 150 ms) and was followed immediately by the plosive /d/. In the context of consonants, vowel formants follow a trajectory with different values at the beginning, middle, and end. As in Cai et al. (2011, 2014), we wanted to make sure that when averaging formant traces across trials, the corresponding parts of the trajectory were directly compared. In order to account for variations in the duration of /ɛ/ across trials, traces were linearly interpolated to 100 evenly spaced points such that there was a single normalized time axis (see Figure 4). Because the actual duration of these traces could have an effect on responses (e.g., longer vowel duration could provide more time to process and respond to auditory feedback changes), vowel duration was compared between groups and conditions using linear mixed effects models with group, condition, and group × condition interaction as fixed effects and subjects as random effects.

Evaluating differences in the shape of the formant response curves between groups and conditions required an appropriate statistical control. Because of the intrinsic temporal autocorrelation of these timeseries and the arbitrary number of timepoints in the normalized time axis, a simple massive univariate approach with one test per sample/timepoint followed by a family-wise error rate correction like Bonferroni would have been unnecessarily conservative. Therefore, a principal components analysis (PCA) was carried out on all /ɛ/ formant traces across subjects and conditions to reduce 100 time points down to a small number of components that characterized 95% of the variance in individual formant traces. This variance was accounted for by the first 3 principal components, and the data were projected into this low-dimensional component space yielding three component scores for every trial. These scores were averaged across trials in the perturbed and non-perturbed conditions for each subject and speaking condition and subtracted to derive deviation scores.

The PCA scores were then used as the dependent variables in a multivariate general linear model (GLM) framework to determine using single tests whether responses were dependent on group, condition, or stuttering severity, while controlling for vowel duration. These tests simultaneously evaluate differences across any of the principal component scores jointly characterizing the formant response shapes, enabling statistically controlled confirmatory inferences regarding these shapes. For group and severity analyses, the vowel duration regressor was averaged between the two conditions. For the condition, group x condition interaction, and severity x condition interaction analyses, the differences in vowel duration between conditions were included as regressors. 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 subject's SSI-4 score, thus focusing the measure on speech-related function. The second measure was the percentage of disfluent trials during the *normal* conditions (disfluency rate). Therefore, two separate models were evaluated; one that included the SSI-Mod scores and a second that included the disfluency rates.

In addition to these confirmatory tests, exploratory tests were performed to evaluate responses and group differences separately for each sample/timepoint. Formant traces were averaged across trials in the perturbed and non-perturbed conditions for each subject, and a deviation trace was created by subtracting the perturbed trace from the non-perturbed trace at each time point. To test for a response within groups in each speaking condition, deviation traces were compared to 0 using a one-sample t-test at each timepoint with an alpha criterion of 0.05. In addition, deviation traces were compared between groups using a two-sample t-test at each timepoint with an alpha criterion of 0.05. These massive univariate exploratory tests used uncorrected statistics, and they are used only to provide some additional level of temporal detail regarding the formant response time-courses analyzed in the previous confirmatory tests.

3. Results

3.1. Disfluency Rate

The rate of stuttering during the experiment was quite low, with 6 out of 16 AWS producing no disfluencies. However, AWS produced significantly fewer 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. [under review]).

3.2. Speaking Rate and Isochronicity

For rate (Table 1), there was no significant effect of group ($F_{(1,29)} = 0.19$, $p = .67$), but there was a significant effect of condition ($F_{(1,29)} = 76.6$, $p < 0.0001$) that was modulated by a significant group × condition interaction ($F_{(1,29)} = 6.15$, $p = .02$). 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). 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 disfluencies. These two measures were not significantly correlated (Spearman's $\rho = 0.10$, $p = .73$). As expected, there was a significant effect of condition on CV-ISD ($F_{(1,29)} = 294.73$, $p < 0.0001$), but no effect of group ($F_{(1,29)} = 0.53$, $p = .47$) or interaction ($F_{(1,29)} = 0.41$, $p = .52$). These results are also reported in Frankford et al. (under review).

3.3. F1 Perturbation

Figure 4 shows the time-normalized formant change responses to the F1 perturbation during the /ɛ/ in "steady" for each condition. There was a significant main effect of condition on the duration of the /ɛ/ in "steady," where the vowel was longer in the metronome-paced condition than the normal condition ($F_{(1,29)} = 105.9$, $p < 0.0001$), but no significant main

effect of group ($F_{(1,29)}$ = 3.3, p = .08) or group by condition interaction ($F_{(1,29)}$ = 1.3, p = .26).

To determine the effects of group and condition on formant change responses while controlling for vowel duration, a multivariate GLM was performed. There was no significant effect of group ($F_{(3, 25)}$ = 0.75, p = .54) or condition ($F_{(3, 25)}$ = 0.03, p = .99), but there was a significant group x condition interaction ($F_{(3, 25)}$ = 3.18, p = .04) on response magnitude. There was also no significant effect of either SSI-mod ($F_{(3, 25)}$ = 0.22, p = .88) or SSI-mod x condition ($F_{(3, 25)}$ = 0.82, p = .50). Finally, there was no significant effect of vowel duration ($F_{(3, 25)}$ = 1.22, p = .32) or a significant vowel duration by condition interaction ($F_{(3,25)}$ = 0.36, p = .78) on response. After substituting in disfluency rate from the experimental session for SSI-Mod, we re-ran the model. There was no significant main effect of disfluency rate ($F_{(3, 25)}$ = 0.81, p = .50) or disfluency rate by condition interaction ($F_{(3, 25)}$ = 1.60, p = .21; for complete results of this new model, see Supplementary Table 1).

Colored bars in Figure 4 indicate the results of exploratory tests showing when each group's response was different from 0 (uncorrected), and when the groups were significantly different (uncorrected). For the *normal* condition (Figure 4A), ANS exhibit an opposing response that begins approximately 80% of the way through the vowel and continues until the transition to the /d/. AWS do not show a similar response prior to the end of the /ɛ/ vowel. There is a brief period during which the difference between the groups is significant using an uncorrected threshold of p < 0.05. Conversely, in the *metronome-paced* condition (Figure 4B) AWS respond to oppose the perturbation beginning at approximately 70% of the way through the vowel until the end, while ANS do not show any response. This difference in response is reflected as a group difference between approximately 85% and 95% through the vowel.

4. Discussion

The present experiment examined whether pacing speech to an external metronome stimulus (which is known to improve fluency in AWS) alters formant auditory feedback control processes in AWS. While AWS did produce significantly fewer disfluencies in the *metronome-paced* condition, changes in perturbation responses were more complex. AWS showed a significant compensatory formant change response to the F1 perturbation in the *metronome-paced* condition (while they did not in the *normal* condition). At the same time, ANS exhibited a compensatory response in the *normal* condition, but no significant response to the perturbation in the *metronome-paced* condition. These results are discussed in further detail below with respect to the stuttering theories mentioned in the Introduction as well as prior literature.

4.1. The effect of metronome-timed speech on auditory feedback control

The presence of a compensatory formant change for AWS in response to the F1 perturbation during the *metronome-paced* condition makes it tempting to suggest that speaking isochronously corrects the disturbance in auditory-motor transformation or makes this process more efficient. Previous work has shown that speaking along with a metronome

reduces neural activation differences between AWS and ANS (Toyomura et al., 2011; although see Frankford et al. [2021] who did not find this). Furthermore, delayed auditory feedback, which also reduces disfluencies, causes pre-speech neural suppression in AWS to become more similar to ANS (Daliri & Max, 2018). Assuming that pre-speech neural suppression corresponds to speech motor control processes (Max & Daliri, 2019), this finding implicates a change in the forward modeling processes needed for error detection. A similar process could be accompanying fluency in the present study.

Alternatively, responding to feedback perturbations during the *metronome-paced* condition may reflect a sudden freeing of neural resources in AWS due to the timing of the sequence being preplanned. If the speech timing system, which is thought to be disrupted in AWS (e.g., Chang & Guenther, 2020), requires more neural resources to generate phoneme initiation cues during normal speech, there may be reduced resources available to make corrections for spectral feedback errors. When suprasyllabic timing is predetermined as in the *metronome-paced* condition, these resources can be redirected to spectral auditory feedback control as well as fluent speech production.

There may also be a simpler explanation for this finding: speaking isochronously led to an increase in the average duration of the /ɛ/ in "steady." If AWS take longer to generate corrective movements to feedback errors due to inefficient sensorimotor transformations as previously hypothesized (Hickok et al., 2011; Max et al., 2004; Neilson & Neilson, 1987) and supported by auditory feedback perturbation studies (Bauer et al., 2007; Cai et al., 2014; Loucks et al., 2012), the correction may only be registered when the vowel is sufficiently long. Differences in vowel duration across conditions were at least partially accounted for by including vowel duration difference scores in the GLM, however this makes the assumption that the relationship between vowel duration and response is linear. In fact, the relationship is likely more complex; we would not expect any response if the vowel is shorter than the minimum auditory feedback processing delay. As such, this change in response between the *normal* and *metronome-paced* conditions may be a secondary effect of isochronous speech rather than directly related to the "rhythm effect."

An additional surprising result of this study was that ANS did not show any response to the formant perturbation during the *metronome-paced* condition. Because this is the first experiment to examine responses to formant perturbation during isochronous speech in ANS, there is no prior literature to aid in interpretation. This may reflect a general change in the balance of auditory feedback versus feedforward control for spectral cues during the *metronome-paced* condition, where ANS are more concerned with precise control of the timing aspects of the sentence. From this, it may be expected that during isochronous speech, ANS do not make spontaneous adjustments to formants (as in Niziolek et al. [2013]). Future research will need to examine this relationship further to determine why ANS respond to formant perturbations during normal (non-paced) speech but not metronome-paced speech, whereas AWS show the opposite pattern.

At the same time, while these interpretations are potentially compelling, care should be taken in relating the present results to the underlying mechanisms of stuttering. The presently discussed findings, along with most studies of auditory feedback control in

stuttering, only include AWS who have had many years to develop adaptive behaviors that could influence these results. Indeed, the only two auditory feedback manipulation studies including both children who stutter (CWS) and AWS found that while AWS consistently demonstrate diminished adaptation to a repeated formant perturbation, CWS responses are less consistent across studies (Daliri et al., 2018; Kim et al., 2020). This may indicate that impaired sensory-motor transformation abilities found in spectral perturbations are secondary rather than core components of stuttering and may develop slowly over time. Future studies of auditory feedback perturbations should examine responses in CWS as well as AWS in order to disentangle these possibilities.

4.2.   Additional considerations

It should be noted that this is only the third study to examine online auditory feedback control of formants during a multisyllabic speech utterance and at syllable rates comparable to conversational English. Formant perturbations are generally applied during prolonged vowels (Purcell & Munhall, 2006) or consonant-vowel-consonant syllables (Niziolek & Guenther, 2013; Purcell & Munhall, 2006; Tourville et al., 2008), which is helpful for measuring responses because of the time it takes to process the speech signal and generate the appropriate corrective motor command (on the order of 100-150 ms; Guenther, 2016). At the same time, these studies are not as ecologically valid regarding the role of auditory feedback during normal speaking situations. Cai et al. (2011, 2014) achieved higher ecological validity by embedding perturbations during the course of a multisyllabic phrase. This is especially important for examining feedback control in AWS, as stuttering is more likely to occur during longer phrases compared to single words (e.g., Coalson et al., 2012; Soderberg, 1966). However, Cai et al. (2011, 2014) used stimuli comprised only of vowels and semivowels. The perturbation results in this study upheld the prior results within a phrase including a mixture of vowels, stop consonants, and fricatives, indicating that auditory feedback does indeed play a role in speech motor control during naturalistic utterances.

A potential limitation of this study was the effect of global hardware and software delays needed to alter feedback online. While these delays were modest (24-32ms), such delayed auditory feedback (DAF) is known to have a fluency-enhancing effect in AWS (Kalinowski et al., 1996), and may have led to the low overall level of stuttering in the present study. Furthermore, there was a discrepancy in global feedback latency between the perturbed trials (24ms delay) and non-perturbed trials (32ms delay). While this discrepancy was quite small, it is possible that the condition with greater delay led to a reduced speaking rate overall (Chon et al., 2021; Kalinowski et al., 1993; Stuart et al., 2002), increased fluency in AWS (Kalinowski et al., 1993, 1996), and decreased fluency in ANS (Stuart & Kalinowski, 2015). Differences in speech rate could have potentially affected which portions of the formant trajectories are compared between the perturbed and non-perturbed conditions. However, all formant trajectories in the present study were time-normalized at the individual trial level which should have mitigated any effect of speech rate. Regarding fluency, no difference in unambiguous stutters was found between feedback conditions (AWS: sign-rank test: W = 17.5, p = 1; ANS did not produce enough stutter-like disfluencies to make a meaningful comparison). One could argue that even if trials with unambiguous stutters were removed

from analysis, some aspect of speech motor control associated with this (dis)fluency was affected. Chon and colleagues found that across both AWS and ANS, 50ms DAF can lead to greater kinematic variability across trials compared to 25ms DAF, while within trial kinematic stability is similar (Chon et al., 2021). If greater trial-to-trial variability were an issue in the present study, averaging the formant trajectories across trials in perturbed and non-perturbed conditions should have removed this potential confound. Finally, because the difference in feedback delays was much smaller than any comparisons studied in the literature, it is difficult to conclude whether this difference would meaningfully change the results. In follow-up work, it will be necessary to carefully match feedback delays between perturbed and non-perturbed conditions to ensure that the discrepancy included here did not have an important effect on the results.

Another potential limitation of the present study was the inclusion of two different auditory feedback perturbations interspersed among trials (the F1 perturbation and the temporal perturbation not described here). In general, with altered feedback paradigms that measure responses to unexpected perturbations, care must be taken to ensure that cross-trial adaptation (i.e. error-based learning) is minimized. When more than one perturbation is applied, it is even more important to make sure there are no interactions across perturbations. In the present study, design choices were made to mitigate these effects. First, each perturbation was applied to a different part of the stimulus sentence (the temporal perturbation was applied to the /s/ in "steady" while the F1 perturbation affected the /ɛ/). Second, each perturbation only occurred on 20% of the total trials, limiting expectancy of either perturbation. Third, at least one trial without a perturbation always separated trials with any perturbation. Each of these design features served to make the perturbations unpredictable so that cross-trial adaptation or interactions between perturbations were minimized. Furthermore, these perturbations were quite subtle - participants were largely unaware that there were any perturbations and even less aware of what the perturbations actually were. Therefore, it is also less likely that participants consciously altered their speech in response to the perturbations. Nonetheless, it is still possible that the presence of one perturbation affected responses to the other. As such, researchers should take these issues into consideration when designing auditory feedback perturbation experiments with more than one type of perturbation.

Finally, the training that participants received prior to the experiment may have altered participants' natural speech. It is possible that including such training may prime participants to attend to their speech more than usual and potentially increase responses to the perturbation. However, this effect would be constant across all participants and should not have affected comparisons between groups and conditions. The primary aim of the training procedures was to better control speech parameters like loudness and speaking rate in order to minimize unintended effects on responses. This tradeoff between experimental control and ecological validity is certainly an important consideration for any behavioral study. It will be beneficial for future sensory feedback perturbation research to explore the effects of natural variability in these parameters during testing. This will help clarify whether and how altering speech variability for the sake of experimental control affects outcomes in this type of research.

5. Conclusion

The present study demonstrated that metronome-timed speaking may alter the processes involved in responding to spectral auditory feedback perturbations. Specifically, while perturbations of F1 led to a compensatory response in ANS when speaking with natural timing, AWS did not show a significant response. However, during the metronome-timed speech condition, AWS did show a significant response while ANS do not. As stuttering is a developmental speech disorder that generally emerges in early childhood and all participants included herein were adults, it is important to consider that the results of this study may reflect either primary characteristics of stuttering or secondary behaviors developed to compensate for or adapt to these primary characteristics. Future research examining the mechanisms of online auditory feedback processes in children who stutter will be necessary to clarify the roles of disrupted sensorimotor transformations in persistent developmental stuttering.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

Funding Statement:

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

$$thresh_{/ə/} = \frac{q_{90}(minRMS_{/ðə/}) + q_{10}(maxRMS_{/ðə/})}{2}$$ (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 xth 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 xth quantile of the distribution of values.

The offset of the /ɛ/ in the word "steady" was detected when both 1) the RMS slope was negative for a duration determined by

$$dur_{/ɛ/off} = \frac{mean(maxDur_{negRMSslp})}{2} \qquad \text{(Equation A3)}$$

and 2) the cumulative sum of the RMS slope exceeded a value determined by

$$int_{/ɛ/off} = \frac{q_{90}(intRMSslp_{/ɛd/}) + q_{10}(intRMSslp_{/di/})}{2} \qquad \text{(Equation A4)}$$

where $maxDur_{negRMSslp}$ is the set of the maximum duration of negative RMS slope intervals from previous trials during production of /di/ in "steady," $intRMSslp_{/ɛd/}$ is the integral of the RMS slopes from the longest negative RMS slope intervals during /ɛd/ in "steady", $intRMSslp_{/di/}$ is the integral of the RMS slopes from the longest negative RMS slope intervals during /di/ in "steady," and q_x is the xth quantile of the distribution of values.

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]

Barrett L, & Howell P (2021). Altered Sensory Feedback in Speech. Manual of Clinical Phonetics, 461–479.

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

Bloodstein O (1995). A handbook on stuttering (5th ed). Singular Pub. Group.

Boucher V, & Lamontagne M (2001). Effects of Speaking Rate on the Control of Vocal Fold Vibration: Clinical Implications of Active and Passive Aspects of Devoicing. Journal of Speech, Language, and Hearing Research, 44(5), 1005–1014. 10.1044/1092-4388(2001/079)

Bradshaw AR, Lametti DR, & McGettigan C (2021). The Role of Sensory Feedback in Developmental Stuttering: A Review. Neurobiology of Language, 2(2), 308–334. 10.1162/nol_a_00036

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]

Brayton ER, & Conture EG (1978). Effects of Noise and Rhythmic Stimulation on the Speech of Stutterers. Journal of Speech and Hearing Research, 21(2), 285–294. 10.1044/jshr.2102.285

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

Chon H, Jackson ES, Kraft SJ, Ambrose NG, & Loucks TM (2021). Deficit or Difference? Effects of Altered Auditory Feedback on Speech Fluency and Kinematic Variability in Adults Who Stutter. Journal of Speech, Language, and Hearing Research, 64(7), 2539–2556. 10.1044/2021_JSLHR-20-00606

Coalson GA, Byrd CT, & Davis BL (2012). The influence of phonetic complexity on stuttered speech. Clinical Linguistics & Phonetics, 26(7), 646–659. 10.3109/02699206.2012.682696 [PubMed: 22690719]

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]

Davidow JH, Bothe AK, Andreatta RD, & Ye J (2009). Measurement of Phonated Intervals During Four Fluency-Inducing Conditions. Journal of Speech, Language, and Hearing Research, 52(1), 188–205. 10.1044/1092-4388(2008/07-0040)

Davidow JH, Bothe AK, Richardson JD, & Andreatta RD (2010). Systematic Studies of Modified Vocalization: Effects of Speech Rate and Instatement Style During Metronome Stimulation. Journal of Speech, Language, and Hearing Research, 53(6), 1579–1594. 10.1044/1092-4388(2010/09-0173)

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

Frankford SA, Cai S, Nieto-Castanon A, & Guenther FH (under review). Auditory feedback control in adults who stutter during metronome-timed speech I. Timing Perturbation. Journal of Fluency Disorders.

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]

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

Hickok G, Houde J, & Rong F (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. Neuron, 69(3), 407–422. 10.1016/j.neuron.2011.01.019 [PubMed: 21315253]

Hutchinson JM, & Navarre BM (1977). The effect of metronome pacing on selected aerodynamic patterns of stuttered speech: Some preliminary observations and interpretations. Journal of Fluency Disorders, 2(3), 189–204. 10.1016/0094-730X(77)90023-7

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, Armson J, Stuart A, & Gracco VL (1993). Effects of Alterations in Auditory Feedback and Speech Rate on Stuttering Frequency. Language and Speech, 36(1), 1–16. 10.1177/002383099303600101 [PubMed: 8345771]

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

Kessinger RH, & Blumstein SE (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. Journal of Phonetics, 26(2), 117–128. 10.1006/jpho.1997.0069

Kim KS, Daliri A, Flanagan JR, & Max L (2020). Dissociated Development of Speech and Limb Sensorimotor Learning in Stuttering: Speech Auditory-motor Learning is Impaired in Both Children and Adults Who Stutter. Neuroscience, 451, 1–21. 10.1016/j.neuroscience.2020.10.014 [PubMed: 33091464]

Klich RJ, & May GM (1982). Spectrographic Study of Vowels in Stutterers' Fluent Speech. Journal of Speech, Language, and Hearing Research, 25(3), 364–370. 10.1044/jshr.2503.364

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]

Loucks T, & De Nil LF (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 [PubMed: 16698179]

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

Max L, Guenther FH, Gracco VL, Ghosh SS, & Wallace ME (2004). Unstable or Insufficiently Activated Internal Models and Feedback-Biased Motor Control as Sources of Dysfluency: A Theoretical Model of Stuttering. Contemporary Issues in Communication Science and Disorders, 31(Spring), 105–122. 10.1044/cicsd_31_S_105

Namasivayam AK, & Van Lieshout P (2011). Speech compensation in persons who stutter: Acoustic and perceptual data. Canadian Acoustics, 39(3), 150–151.

Neilson MD, & Neilson PD (1987). Speech motor control and stuttering: A computational model of adaptive sensory-motor processing. Speech Communication, 6(4), 325–333. 10.1016/0167-6393(87)90007-0

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]

Niziolek CA, Nagarajan SS, & Houde JF (2013). What Does Motor Efference Copy Represent? Evidence from Speech Production. Journal of Neuroscience, 33(41), 16110–16116. 10.1523/JNEUROSCI.2137-13.2013 [PubMed: 24107944]

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]

Soderberg GA (1966). The Relations of Stuttering to Word Length and Word Frequency. Journal of Speech and Hearing Research, 9(4), 584–589. 10.1044/jshr.0904.584

Stager SV, Denman DW, & Ludlow CL (1997). Modifications in Aerodynamic Variables by Persons Who Stutter Under Fluency-Evoking Conditions. Journal of Speech, Language, and Hearing Research, 40(4), 832–847. 10.1044/jslhr.4004.832

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]

Stuart A, & Kalinowski J (2015). Effect of Delayed Auditory Feedback, Speech Rate, and Sex on Speech Production. Perceptual and Motor Skills, 120(3), 747–765. 10.2466/23.25.PMS.120v17x2 [PubMed: 26029968]

Stuart A, Kalinowski J, Rastatter MP, & 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), 2237. 10.1121/1.1466868 [PubMed: 12051443]

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

Tourville JA, Reilly KJ, & Guenther FH (2008). Neural mechanisms underlying auditory feedback control of speech. NeuroImage, 39(3), 1429–1443. 10.1016/j.neuroimage.2007.09.054 [PubMed: 18035557]

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

Xia K, & Espy-Wilson CY (2000). A new strategy of formant tracking based on dynamic programming. Sixth International Conference on Spoken Language Processing.
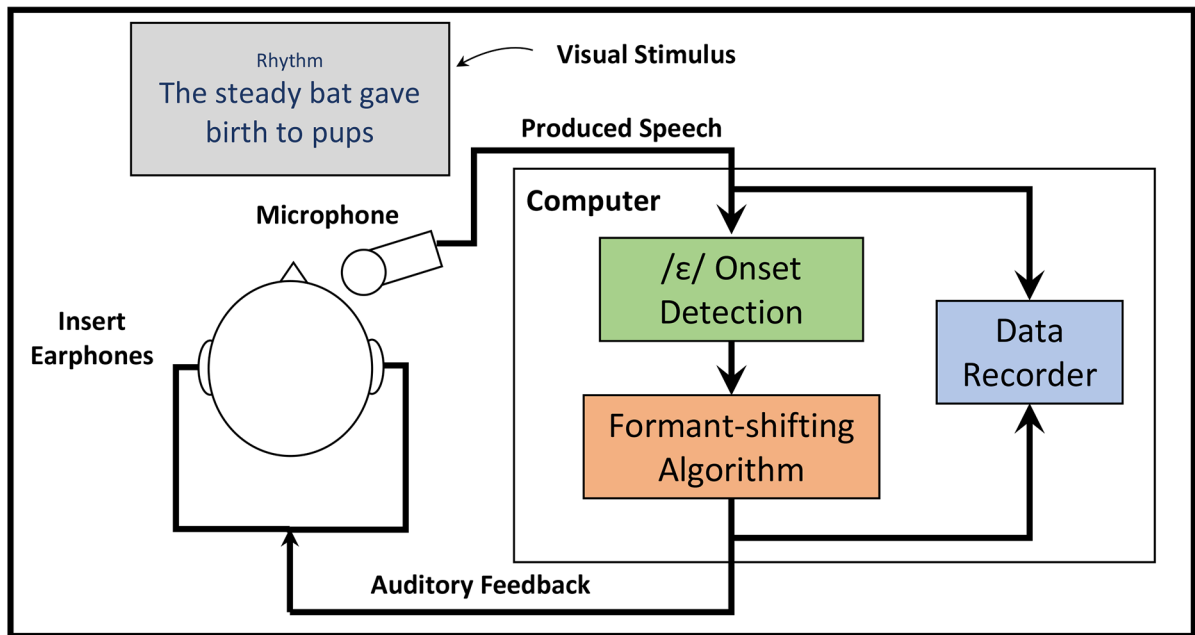
Figure 1.

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 signals were recorded and fed to an experimental computer running Audapter. On perturbed trials, detection of the onset of /ɛ/ 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.
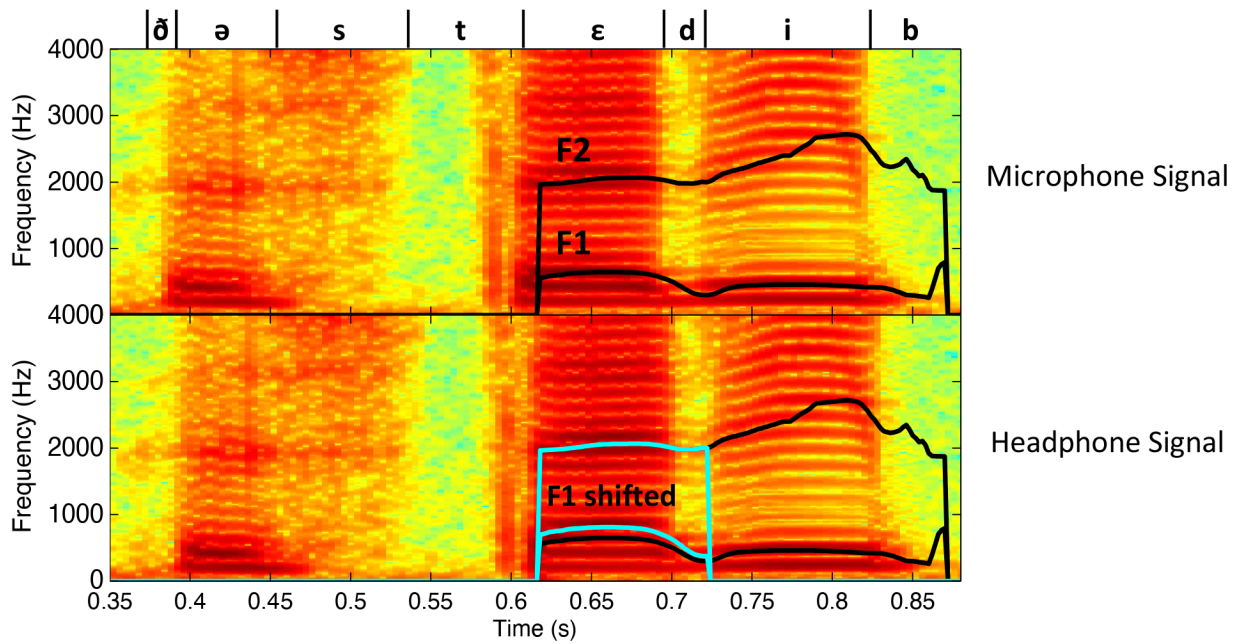
Figure 2:
Example spectrograms of "The steady bat" during a formant perturbation trial generated from the recorded microphone signal (top) and headphone signal (bottom). The first and second formant traces during the word "steady" are displayed. In the headphone signal spectrogram, the original formant traces are displayed in black and the shifted formant traces are overlaid in blue. Note that only the first formant is perturbed. Phoneme boundaries and international phonetic alphabet symbols are indicated above the microphone signals. F1 = first formant, F2 = second formant, Hz = hertz.
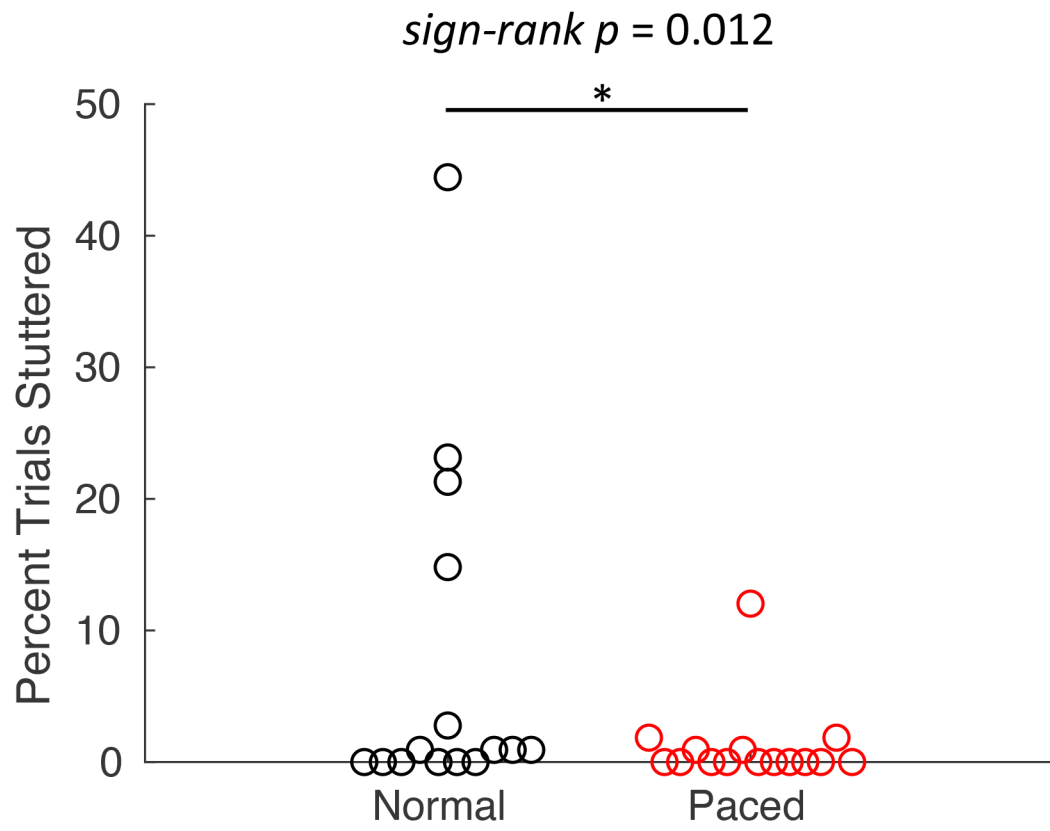
Figure 3.
Disfluency rate in the *normal* and *metronome-paced* conditions for AWS. Circles represent individual participants.
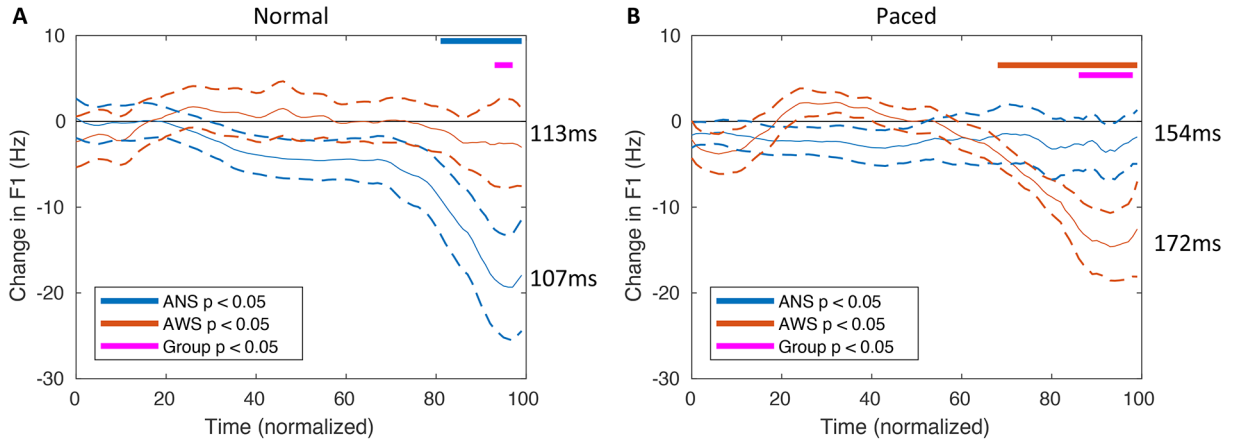
Figure 4.
Time-normalized formant responses to the F1 perturbation during the /ɛ/ in "steady" in the *normal* (A) and *metronome-paced* (B) conditions. Solid curves indicate the average difference in F1 frequency between the perturbed and non-perturbed conditions for ANS (blue) and AWS (orange). Dashed lines indicate mean formant change response +/− standard error of the mean. The blue bars indicate intervals of significant responses in ANS ($p < 0.05$, uncorrected). The orange bars represent intervals of significant responses in AWS ($p < 0.05$, uncorrected). The magenta bars indicate intervals of significant differences between ANS and AWS ($p < 0.05$, uncorrected). Duration of the non-normalized /ɛ/ response curves (averaged between perturbed and non-perturbed trials and across subjects) for each group and speaking condition are included to the right of the response curves for reference.

Table 1.

Descriptive and inferential statistics for speaking rate and CV-ISD. Error estimates indicate 95% confidence intervals. 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.

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	$F(1,29) = 0.19$, $p = .67$	$F(1,29) = 76.60$, $p < .0001$	$F(1,29) = 6.15$ $p = .02$
CV-ISD	0.27 ± 0.06	0.10 ± 0.02	0.26 ± 0.05	0.10 ± 0.02	$F(1,29) = 0.53$, $p = .47$	$F(1,29) = 294.73$, $p < .0001$	$F(1,29) = 0.41$ $p = .52$