Impaired timing adjustments in response to time-varying auditory perturbation during connected speech production in persons who stutter

Shanqing Cai [a,*], Deryk S. Beal [b], Satrajit S. Ghosh [c], Frank H. Guenther [a,d], Joseph S. Perkell [a]

[a] Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, USA
[b] Institute for Stuttering Treatment and Research and Department of Speech Pathology and Audiology, Faculty of Rehabilitative Medicine, University of Alberta, Edmonton, Alberta, Canada
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[d] Department of Biomedical Engineering, Boston University, Boston, MA, USA

* Corresponding author: Address: 677 Beacon St, Room 102, Boston, MA 02215, USA. Fax: 617-353-7755.
Email: scai@bu.edu (S. Cai)

Abstract

Auditory feedback (AF), the speech signal received by a speaker's own auditory system, contributes to the online control of speech movements. Recent studies based on AF perturbation provided evidence for abnormalities in the integration of auditory error with ongoing articulation and phonation in persons who stutter (PWS), but stopped short of examining connected speech. This is a crucial limitation considering the importance of sequencing and timing in stuttering. In the current study, we imposed time-varying perturbations on AF while PWS and fluent participants uttered a multisyllabic sentence. Two distinct types of perturbations were used to separately probe the control of the spatial and temporal parameters of articulation. While PWS exhibited only subtle anomalies in the AF-based spatial control, their AF-based fine-tuning of articulatory timing was substantially weaker than normal, especially in early parts of the responses, indicating slowness in the auditory-motor integration for temporal control.

1. Introduction

Auditory feedback (AF) refers to the speech sounds received by the speaker's own auditory system during speech production. AF is an important component of the mechanisms underlying the online control of speech movements. There is evidence (Kalinowski et al., 1993; Foundas et al., 2004) for abnormalities in the utilization of AF by the speech motor system in stuttering, a developmental disorder of speech fluency in which the production of speech is interrupted by sound or syllable repetitions, prolongations, and silent blocks.

When sudden-onset perturbations are introduced to specific acoustic parameters of AF, normal speakers show online corrections in their production, in directions opposite to the perturbations. Such short-latency (~150 ms) compensatory responses have been demonstrated for fundamental frequency (F0; e.g., Chen, Liu, Xu, & Larson, 2007) and formant frequencies (e.g., Purcell & Munhall, 2006; Tourville, Reilly, & Guenther, 2008), highlighting the active role AF plays in assisting feedforward mechanisms (Guenther, Ghosh, & Tourville, 2006) during online control of phonation and articulation. Recent studies have shown weaker-than-normal compensatory responses to these types of AF perturbation in PWS (for F0, see Loucks, Chon, & Han, 2012; for formant, see Cai et al., 2012). These results indicate that the speech motor system of a PWS cannot compare the desired and actual auditory outcome of speech movements and/or transform the difference (i.e., termed *auditory error*) to proper corrective movements as effectively as non-stutterers can.

How may this subnormal auditory-motor interaction in online speech motor control be manifested during multisyllabic, connected speech? In stuttering, dysfluencies are more likely to occur during multiword utterances than during single words; the frequency of stuttering is positively related to utterance length and complexity (e.g., Soderberg, 1966). Thus examining connect speech production appears to be important for understanding the role of abnormal AF-based speech motor control in this disorder. However, the aforementioned AF perturbation studies (Loucks et al., 2012; Cai et al. 2012) used sustained phonation and isolated monosyllabic words, which were not suitable for probing auditory-motor interaction in stutterers' connected speech.

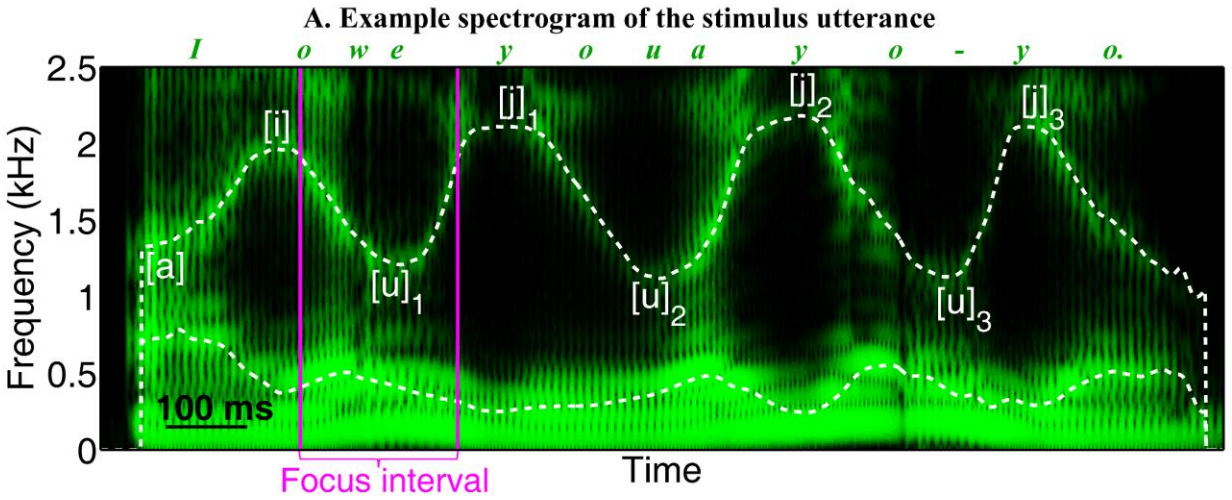
We have used the technique of time-varying formant perturbation to demonstrate the role of AF in the online control of multisyllabic articulation in normal speakers (Cai, Ghosh, Guenther, & Perkell, 2011). By introducing different types of manipulations of the formant trajectories during the utterance "I owe you a yo-yo", this technique separately examined the spatial and temporal aspects of the control. First, the spatial perturbation altered the formant frequencies at specific local extrema in the AF, while preserving the timing of the extrema. In articulatory terms, this perturbation corresponded to perceived changes in the positions of the articulators (tongue and lips). Under the spatial perturbation, typically fluent participants were shown to compensate by altering formant frequencies produced in the ensuing part of the utterance. Second, the temporal perturbation altered the timing of the formant-frequency extrema in the AF, while preserving the values at those extrema, which corresponded to changes in the timing of the phonemes. Healthy speakers showed timing adjustments in their articulation after the onset of the temporal perturbation and throughout the rest of the utterance.

The goal of the current study was to examine whether PWS show deficits in the online AF-based control of multisyllabic articulation using the same technique as Cai et al. (2011). The compensatory responses by a group of PWS to the spatial and temporal perturbations were compared with the responses from fluent controls. Differences in the magnitude and timing of the compensation were analyzed to reveal anomalies in the auditory-motor interaction during multisyllabic articulation in PWS.

2. Results

PWS and matched controls produced the utterance “I owe you a yo-yo”. The choice of this utterance was based on the consideration that it consisted of only vowels and semivowels and hence elicited continuous phonation. This allowed us to indirectly measure the spatial positions and timing of the articulation using formant trajectories throughout the utterance.

As Figure 1 illustrates, there is a set of well-defined local minima and maxima in the second-formant (F2) trajectory, due to lip rounding and the alternation between front and back tongue positions. These extrema were used as landmarks for defining the onsets and offsets of syllables in this utterance, so that we could extract articulatory timing, as well as measure the formant values at the landmarks, which reflect the underlying articulatory positions. Both the spatial and temporal types of AF perturbation occurred during the word “owe” and the transition from “owe” to the following word “you”, as indicated by the focus interval in Fig. 1A. As an initial part of each experiment, the participant was trained to produce the sentence within medium ranges of speech intensity (74–84 dB SPL at 10 cm from mouth) and speaking rate (sentence duration: 1.0–1.4 s).


A. Example spectrogram of the stimulus utterance

B. List of *ad hoc* phonetic symbols used in the current paper

[a]	Onset of "I"	**[u]**$_2$	F2 minimum at the end of "you"
[i]	F2 maximum at the end of "I"	**[j]**$_2$	F2 maximum at the onset of the 1st "yo"
[u]$_1$	F2 minimum at the end of "owe"	**[u]**$_3$	F2 minimum at the end of the 1st "yo"
[j]$_1$	F2 maximum at the onset of "you"	**[j]**$_3$	F2 maximum at the onset of the 2nd "yo"

[Figure 1]

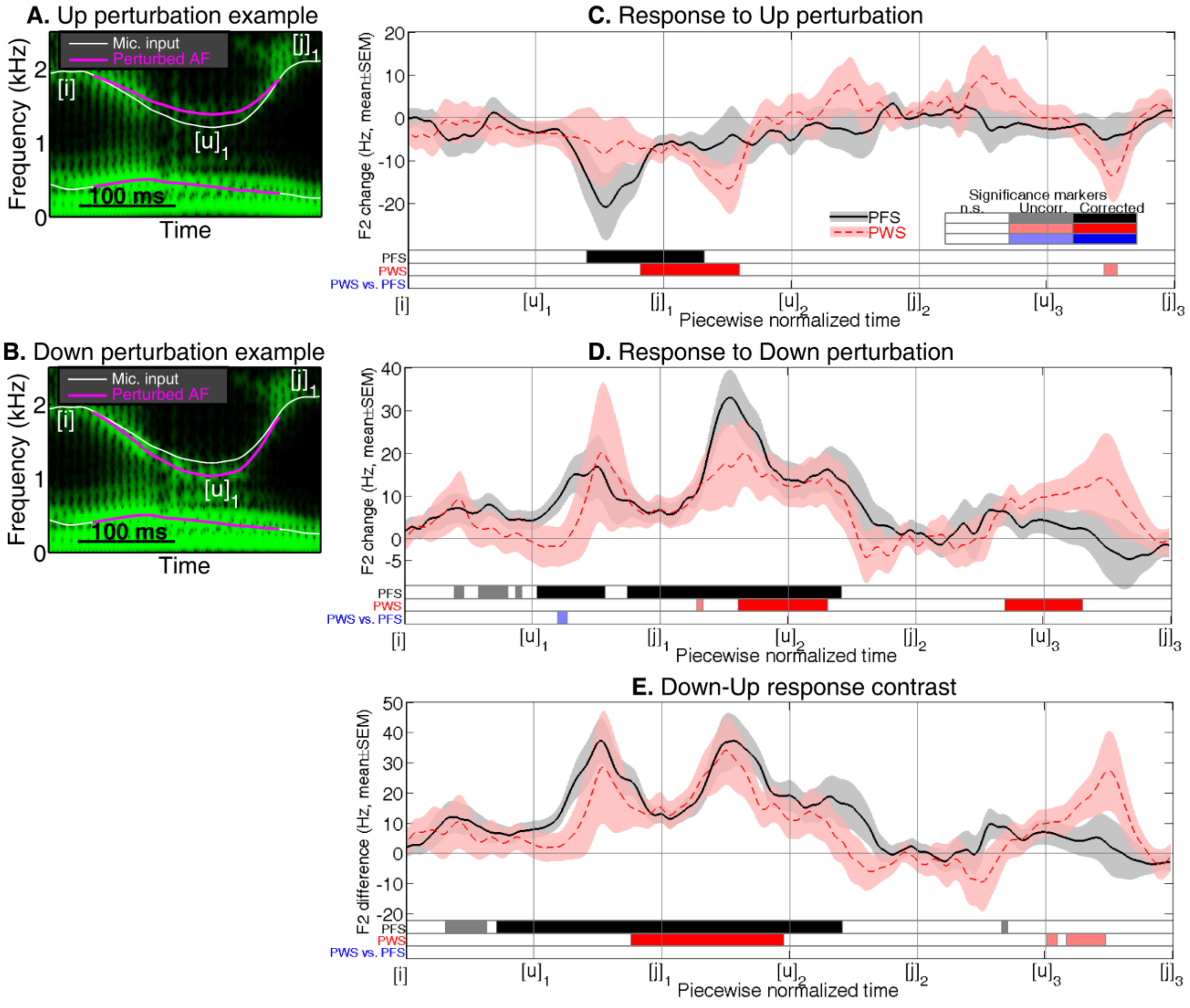
We conducted two experiments on a group of adults with persistent developmental stuttering confirmed by a certified speech-language pathologist (D.S.B.), in addition to two different groups of persons with fluent speech (PFS) as matched controls. Experiment 1 focused on perturbations of spatial parameters in the F2 trajectory; Experiment 2 used perturbations of temporal parameters. Each PWS undertook both Experiments 1 and 2, in randomized order. Two different but partially overlapping groups of controls participated in Experiments 1 and 2. In the following, we visit the results from the spatial perturbations in Experiment 1, then we examine the results from the temporal perturbations in Experiment 2.

2.1. Experiment 1: Spatial perturbation

Twenty PWS and 37 PFS participated in Experiment 1, which focused on the AF-based control of the spatial parameters of multisyllabic articulation, as reflected in formant values. The age distributions of the two groups were similar (mean±1 SD: PWS 27.0±7.7; PFS: 24.9±5.6; t-test: p=0.24); so were the gender distributions (PWS: 4F16M; PFS: 6F31M; χ^2-test: p=0.94). The stuttering severity of the PWS participants, as measured with Stuttering Severity Instrument version 4 (Riley, 2008), ranged from 13 to 43 (median: 25.4; interquartile range: 11.5).

As the examples in Fig. 2A illustrates, the Up perturbation increased the value of F2 at the local minimum corresponding to the end of the word "owe", in a way that preserved the smoothness of the F2 trajectory. The Down perturbation decreased the F2 at the local minimum (Fig. 2B). Such changes in the F2 value would result naturally from changes in the front-back position of the tongue and/or the degree of lip rounding during the [u] sound in "owe". Both the Up and Down perturbations preserved the timing of the local F2 minimum. Therefore they focused on altering the acoustic correlates of the spatial parameters of articulation.

To analyze the compensatory changes in the F2 values produced by the participants under the perturbations, we manually extracted the seven local extrema ([i] to $[j]_3$ as listed in Fig. 1B) as landmark points from each trial. We manipulated the time axis in each trial in a piece-wise linear fashion, so as to aligned all trials at these landmarks. Specifically, the time between each pair of adjacent landmarks were linearly interpolated at 250 evenly spaced points, giving rise to a single piecewise-normalized time axis (e.g., Fig. 2C-E) on which the F2 values were analyzed. This time normalization followed the approach of Cai et al. (2011). Comparisons between the perturbation conditions and between the groups were performed on this piecewise-normalized time axis using Monte Carlo permutation tests (see Methods).



[Figure 2]

The PFS responded to the Down and Up perturbations by altering the F2 values in their production in directions opposite to the perturbations (Fig. 2C and D: black curves). Under the Up perturbation, the earliest significant compensation could be seen between $[u]_1$ and $[j]_1$ (i.e., during the transition from “owe” to “you”). Under the Down perturbation, a significant response (corrected) started shortly after $[u]_1$ (the end of “owe”) and exhibited multiple local maxima between $[u]_1$ and $[j]_1$, between $[j]_1$ and $[u]_2$, and between $[u]_2$ and $[j]_2$. The contrast between the Down and Up F2 trajectories (black curve in Fig. 2E) showed a similar pattern, with the significant compensation seen as early as between [i] and $[u]_1$ (i.e., during “owe”) and as late as between $[u]_2$ and $[j]_2$ (i.e., after "you"). In general, the compensatory responses were longer and slightly greater in magnitude under the Down perturbation than under the Up one. This counteracting and slightly asymmetric pattern of response is highly similar to the results in Cai et al. (2011).

As shown by the red curves in Fig. 2C-D, the mean responses to the spatial perturbations in PWS group were similar to those from the PFS group in that they opposed the directions of perturbation. However, compared to the PFS, trends of later response onset and slower ramping to peak response can be seen PWS group (Fig. 2C-D). Under the Up perturbation, the peak compensation from the PWS was seen during the word "you", instead of before the word "you" as in the PFS group. Under the Down perturbation, between-group comparison revealed a period between $[u]_1$ and $[j]_1$ in which the magnitude of the F2 change was significantly lower in the PWS than in the PFS, although this difference was not significant under the permutation-based correction for multiple comparisons (see Methods). The Down-Up contrast from the PWS group showed a pattern qualitatively similar to that from the PFS group (Fig. 2E). However, the interval of significant difference was substantially later in onset and shorter compared to PFS, although the between-group comparison revealed no significant differences.

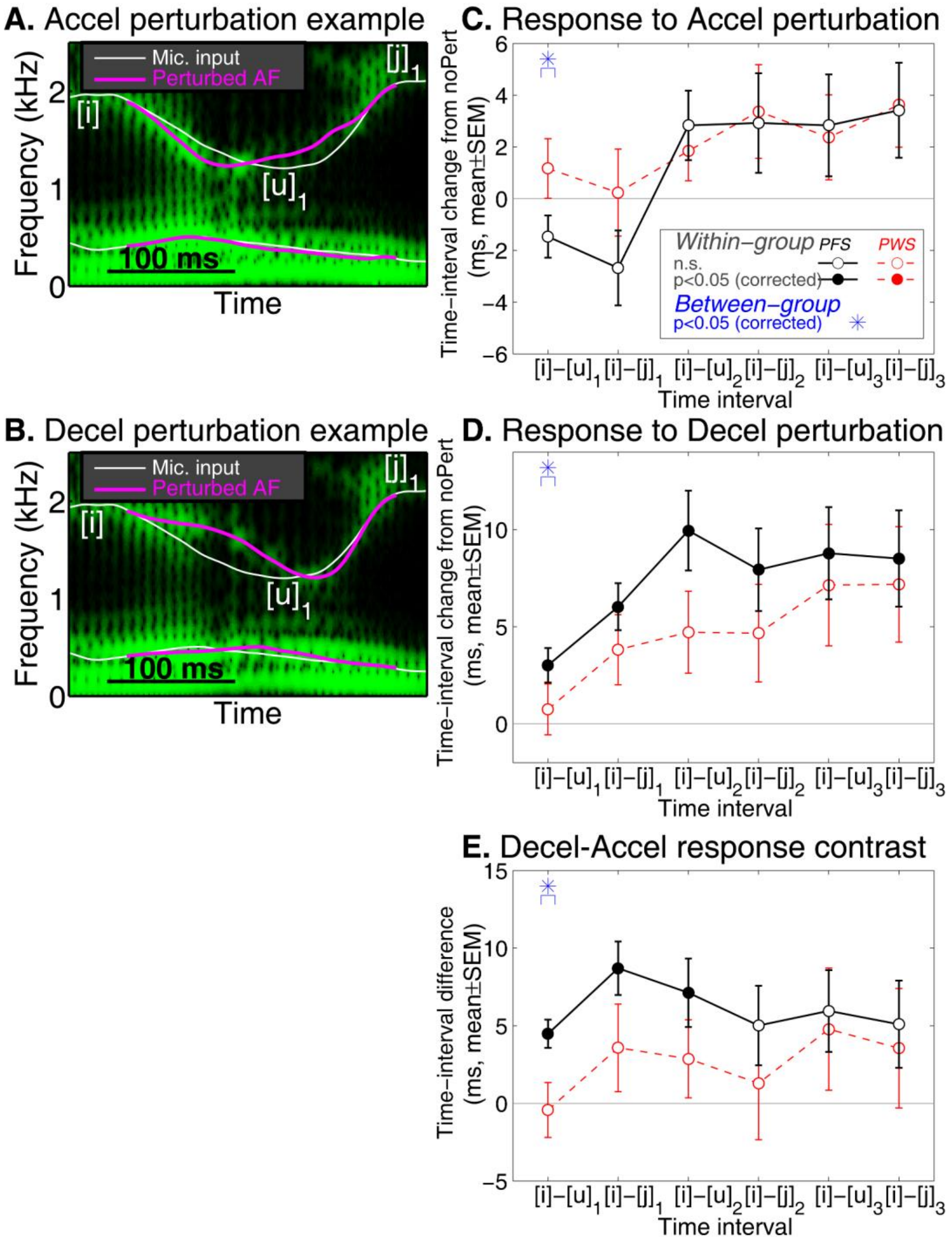
The average magnitude of the first formant (F1) changes in response to the auditory perturbation of F2 was only approximately 10% of that of the F2 changes. The F1 changes along the normalized time axis did not reveal any intervals with significant between-group differences under uncorrected $p<0.05$.

Summarizing the results from Experiment 1, PWS showed qualitatively normal compensatory articulatory adjustments (as reflected by the formant changes) under unexpected perturbations to the spatial parameters of AF, indicating that the spatial component of auditory-motor articulatory control is largely functional in PWS, at least under the form of AF perturbation in this experiment. However, some marginally significant differences hinted at possibilities of slower compensation onset in PWS compared to fluent speakers.

2.1. Experiment 2: Temporal perturbation

In Experiment 2, we applied temporal AF perturbations, including the Accelerating (*Accel*) and Decelerating (*Decel*) types. The temporal perturbations were distinguished from the spatial perturbations in Experiment 1, in that they altered the timing of a landmark acoustic event (F2 minimum at $[u]_1$) in the AF while preserving the F2 value at the landmark. Figure 3A shows an example of the Accel perturbation, in which the F2 minimum in the AF was advanced in time by 47 ms; Fig. 3B shows an example of the Decel perturbation, which delayed the F2 minimum by 25 ms in the AF.

Twenty PWS (same individuals as in Experiment 1) and 29 PFS participated in Experiment 2. The age distributions were similar between the groups (PWS: 27.0±7.7; PFS: 26.0±6.7; t-test: p=0.63). So were the gender distributions (PWS: 4F16M; PFS: 4F25M; χ^2-test: p=0.85).



[Figure 3]

The adjustments in production timing were extracted from the six F2-based acoustic landmarks. The F1 trajectory was not analyzed because it did not contain salient local extrema for timing measurement and because the AF perturbation was focused on the landmarks ($[u]_1$) in

F2. The black curves in Fig. 3C and D show the average timing adjustment in response to the Accel and Decel perturbations in the PFS group. An asymmetric pattern of timing compensation can be seen: while the timing corrections were small and statistically non-significant under the Accel perturbation, the timing adjustments in responses to the Decel perturbation were larger in magnitude and reached statistical significance at all six landmarks ($p<0.05$, permutation-corrected). In contrast to the opposing response to the spatial perturbations (Experiment 1), the responses to the Decel temporal perturbation were in the same direction as (i.e., "followed") the perturbation, highlighting a fundamental difference in the way the spatial and temporal parameters are controlled by the speech motor system. The timing adjustments under Decel showed an increasing trend from the early landmarks to the later ones. The amount of lengthening increased gradually from the first ([i]-$[u]_1$) time interval to the third ([i]-$[u]_2$) and leveled off thereon.

As the red curves in Fig. 3C-D illustrate, the average timing adjustments exhibited by the PWS generally had a smaller magnitude as compared to the controls' responses and failed to reach significance at any of the acoustic landmarks. In addition, the Decel-Accel timing difference was not significant at any of the six landmarks. This was in contrast to the PFS pattern, which showed significant contrast in the first three time intervals (Fig. 3E). Permutation-based comparison revealed significant between-group differences in the adjustment of the [i]-$[u]_1$ interval for both Accel and Decel (asterisks in Fig. 3C and D). Most noticeably, under the Decel perturbation, the [i]-$[u]_1$ interval change in the PFS had an average ratio of 12.2% in relation to the $[u]_1$ time shift introduced to the AF by the perturbation, but this ratio was merely 3.3% in the PWS group (Fig. 3D). A significant between-group difference was seen for the [i]-$[u]_1$ interval in the Decel-Accel contrast as well (Fig. 3E: asterisk). No significant between-group differences were seen at later landmarks. These findings indicate that PWS show deficits in the online fine-tuning of articulatory timing based on AF and these deficits are more pronounced at early moments after the onset of the temporal perturbation than at later ones, indicating a limit in the speed (i.e., “loop duration”) in the AF-based timing control.

3. Discussion

In the current study, perturbations were employed to separately examine the spatial and temporal components of the AF-based multisyllabic articulatory control in PWS and fluent speakers. It was observed that while PWS showed largely normal, but marginally slower, compensation to the spatial perturbations, they showed significantly smaller-than-normal articulatory timing adjustments under the temporal perturbations, especially in the early moments of the response. These findings highlight deficits in feedback-based timing control in PWS, and in particular, indicate a pronounced deficit in the rapid (short-latency) integration of auditory state information with ongoing motor planning and control. Stuttering tends to occur during rapid production of speech sound sequences that involves high demand on precision and timing. The perturbations used in this study examined the interaction between AF and such multisyllabic articulation. As such, our findings bring us a step closer to the relations between auditory-motor interaction and the core motor behavior of stuttering than previous studies have (Loucks et al., 2012; Cai et al., 2012).

The lack of unambiguously weaker-than-normal responses to the spatial perturbation in Experiment 1 appears to contradict our previous finding based on perturbation of the vowel [ɛ] (Cai et al., 2012). In the previous study, PWS showed online compensations about 50% smaller than the average PFS response, under perturbations of the quasi-static first formant (F1) during the vowel, which can be considered as a type of spatial perturbation. There are a number of possible explanations for this apparent contradiction. First, the perturbation used in Cai et al. (2012) had a sudden, step-like onset (see also Loucks et al., 2012), whereas the spatial perturbation in the present study ramped gradually from zero to maximum (e.g., Fig. 2A and B). It is possible that this smooth perturbation profile was less taxing on the AF-based control mechanism than the sudden-onset one, hence partially obscuring the deficits in PWS. Second, in the current study, the period of response to the spatial perturbation involved a semivowel consonant [j] (in “you”), which, due to the contact between the tongue blade and the palate, entailed more somatosensory information than the vowel [ɛ]. It is possible that the heightened involvement of somatosensory feedback, which was unperturbed and therefore conflicted with the perturbed AF, masked deficits in the auditory-motor interaction.

The finding of slower response onset under the temporal perturbation may be related to the repeated findings of longer-than-normal simple motor reaction times under auditory cues in PWS (see Bloodstein & Ratner, 2008, pp. 166-174 for a review). It is also interesting to note the consistency of the results with a previous study (Nudelman et al., 1992). Nudelman et al. (1992) reported slower initiation of pitch correction in a humming pitch tracking task in PWS compared to fluent controls, a finding similar to our observation of weaker temporal and spatial adjustment in early parts of the response. However, to our knowledge, our findings constitute the first demonstration of slower responses during ongoing speech production in PWS. The slowness in auditory-motor integration may form the basis for the well-known speaking rate effect, which refers to decreases in the frequency of stuttering under slower speaking rate (e.g., Adams, Lewis, & Besozzi, 1973). Longer syllable durations under slower speaking rate may give a PWS more time to react to timing information from AF and to implement appropriate adjustments in the articulation, ensuring more accurate production.

Possible neural correlates of the deficit in timing control based on AF in PWS can be found in previous neuroimaging studies on stuttering. For example, the abnormal latencies of the M50 and M100 magnetoencephalography response to auditorily presented and self-produced vowels may be a related neural anomaly (Beal et al., 2010; 2011). In addition, the SMA has been shown to be involved in the initiation and sequencing of speech units (e.g., Bohland & Guenther, 2006). Presumably, its interaction with the auditory cortical area and the cortico-basal-ganglia loop forms the neural basis of the AF-based online temporal control. Previous MRI studies have reported abnormal functional (Lu et al., 2009) connectivity involving the SMA that are possible correlates of the auditory-motor deficit observed in the PWS by current study, which can be tested in future functional neuroimaging studies that use AF perturbation during connected speech.

An important question raised by our findings is whether and how the auditory-motor under-compensation in PWS may lead to breakdowns in fluency. As present, it cannot be ruled out that instead of being involved in the cause of disfluencies, this under-compensation reflects a general lack of flexibility in online responses to unexpected changes in PWS, which would be consistent with the limited speech motor skill hypothesis (van Lieshout, Hulstijn, and Peters, 2004). It is also possible that this under-compensation reflects a defensive compensatory strategy that adult PWS developed to cope with intrinsically unstable speech movements or with an intrinsically defective auditory-motor mechanism for online speech sequencing and timing control, which if engaged to a full extent, would lead to fluency breakdown. The latter possibility is potentially consistent with the fluency enhancing effects of noise masking and global AF delay (e.g., Kalinowski et al. 1993), if it can be assumed that such conditions force the speech motor system to temporarily abandon all (defective) dependency on AF for sequencing and timing. However, under this hypothesis, the nature of the intrinsic deficits remains to be elucidated.

This study has other limitations. First, since we did not measure the participants' capacity to perceive the time-varying formant-trajectory manipulations, future studies are needed to rule out the possibility of perceptual deficits forming the basis of the under-compensation observed in Experiment 2. Second, we used an utterance consisting of only vowels and semivowels. Hence our results cannot provide information about the AF-based control of articulation during broader categories of consonants (e.g., stops and fricatives). We are currently using new AF manipulation techniques (e.g., Tourville et al., 2013) to examine the AF-based control of articulation during more general types of utterances.

4. Methods

Perturbations and experiment design

The methodology of the formant-trajectory manipulation (Fig. 2A-B and 3A-B) has been described previously (Cai et al., 2011) and will not be elaborated here. The design of Experiments 1 and 2 was identical to Cai et al. (2011), in which the noPert and perturbation trials were intermingled and randomized in order. Experiment 1 consisted of 120 noPert, 20 Down and 20 Up trials; Experiment 2 consisted of 120 noPert, 20 Accel and 20 Decel trials. In addition, sentences different from the main stimulus utterances ("I owe you a yo-yo") were inserted to reduce the repetitiveness of the task.

Partly due to the simplicity of the stimulus utterance and the large number of repetition, very few productions of the sentence "I owe you a yo-yo" contained audible dysfluencies. In Experiment 1, only one trial from the PWS group and five from the PFS group were excluded from further analysis due to dysfluency or speech error. In Experiment 2, two trials from the PWS group and 10 trials from the PFS were discarded due to dysfluency or speech error.

Because responses to perturbation may depend on the magnitude of the perturbation, it was important to make sure that the perturbation magnitudes were approximately equal in the two groups. This was indeed the case. In Experiment 1, the average peak magnitudes of the Up perturbation were 172.4±48.4 and 160.4±55.3 Hz (±1 SD) in the PFS and PWS groups, respectively, and did not differ significantly (t-test: p=0.42). The same was true for the Down perturbation (PFS: 169.1±51.0 Hz; PWS: 160.0±54.3 Hz; p=0.53).

In Experiment 2, the amount of timing shift in $[u]_1$ introduced to the AF by the Accel perturbation was not significantly different between the two participant groups (PFS: -44.7±14.7 ms; PWS: -49.33±17.4 ms; t-test: p=0.32); neither was the timing shift introduced by the Decel perturbation (PFS: 22.7±8.7 ms; PWS: 24.7±5.6 ms; p=0.33).

Permutation correction for multiple comparisons

When analyzing the F2 compensation profiles from Experiment 1, a large number of within- or between-group comparisons were performed along the piecewise-normalized time axis (Fig. 2C-E). To correct for multiple comparisons, we used Monte Carlo permutation tests (Westfall & Young, 1993). Briefly, during each permutation, if a between-group comparison is being performed, the group labels (PWS, PFS) are randomly shuffled among the participants. If a within-group test of significance is concerned, the signs of the values are randomly reassigned. Then the statistical test in question (e.g., between-group t-test) is performed at all points along the time axis, giving rise to a number of contiguous intervals of significant between-group differences. The durations of these significant (uncorrected) intervals are calculated and the maximum duration recorded. A number of permutations lead to an approximated null distribution of the maximum interval durations, with which the actual duration of each significant (uncorrected) interval from the un-permuted data are compared to generate the corrected p-value.

Similarly, the analysis of the time-interval change data from Experiment 2 involved statistical comparisons on the six different landmarks (Fig. 3C-D). Similar permutation tests were used for multiple-comparison corrections. In this study, we used 10,000 iterations for each permutation test.

Acknowledgments

This work was supported by National Institutes of Health grants R56-DC0010849 (PI: J.S.P.) and R01-DC007683 (PI: F.H.G.). Adriana DiGrande and Diane Parris assisted with PWS participant recruitment.

References

Adams, M. R., Lewis, J. I., & Besozzi, T. E. (1973). The effect of reduced reading rate on stuttering frequency. J Speech Hear Res, 16(4), 671-675.

Beal, D. S., Cheyne, D. O., Gracco, V. L., Quraan, M. A., Taylor, M. J., & De Nil, L. F. (2010). Auditory evoked fields to vocalization during passive listening and active generation in adults who stutter. Neuroimage, 52(4), 1645-1653.

Beal, D. S., Quraan, M. A., Cheyne, D. O., Taylor, M. J., Gracco, V. L., & De Nil, L. F. (2011). Speech-induced suppression of evoked auditory fields in children who stutter. Neuroimage, 54(4), 2994-3003.

Bloodstein, O., & Ratner, N. B. (2008). A handbook on stuttering (6th ed.). Clifton Park, NY: Thomson/Delmar Learning.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. Neuroimage, 32(2), 821-841.

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

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. J Neurosci, 31(45), 16483-16490.

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. J. Acoust. Soc. Am., 121(2), 1157-1163.

Foundas, A. L., Bollich, A. M., Feldman, J., Corey, D. M., Hurley, M., Lemen, L. C., et al. (2004). Aberrant auditory processing and atypical planum temporale in developmental stuttering. Neurology, 63(9), 1640-1646.

Guenther, F. H., Ghosh, S. S., Tourville J. A. (2006). “Neural modeling and imaging of the cortical interactions underlying syllable production”, Brain Lang., 96(3), 280-301.

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

Loucks, T., Chon, H., & Han, W. (2012). Audiovocal integration in adults who stutter. Int J Lang Commun Disord, 47(4), 451-456.

Lu, C., Ning, N., Peng, D., Ding, G., Li, K., Yang, Y., et al. (2009). The role of large-scale neural interactions for developmental stuttering. Neuroscience, 161(4), 1008-1026.

Nudelman, H. B., Herbrich, K. E., Hess, K. R., Hoyt, B. D., & Rosenfield, D. B. (1992). A Model of the Phonatory Response-Time of Stutterers and Fluent Speakers to Frequency-Modulated Tones. Journal of the Acoustical Society of America, 92(4), 1882-1888.

Purcell DW, Munhall KG. 2006. Compensation following real-time manipulation of formants in isolated vowels. J Acoust Soc Am. 119: 2288-2297.

Riley, G. D. (2008). SSI-4: Stuttering Severity Instrument: ProEd.

Stager, S. V., Denman, D. W., & Ludlow, C. L. (1997). Modifications in aerodynamic variables by persons who stutter under fluency-evoking conditions. J Speech Lang Hear Res, 40(4), 832-847.

Soderberg, G. (1966). The relations of stuttering to word length and word frequency. J. Speech Hear. Res., 9, 584-589.

Tourville, J. A., Cai, S., & Guenther, F. H. (2013). Exploring auditory-motor interactions in normal and disordered speech. Proceedings of Meetings in Acoustics, 19, 060180.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. Neuroimage, 39(3), 1429-1443.

Van Lieshout, P. H. H. M., Hulstijn, W., & Peters, H. F. M (2004). Searching for the weak link in the speech production chain of people who stutter: A motor skill approach. In B. Maassen, R. Kent, H. F. M. Peters, P. Van Lieshout, & W. Hulstijn (Eds.), Speech motor control in normal and disordered speech (pp. 313–355). Oxford, England: Oxford University Press.

Westfall, P. H., & Young, S. S. (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment (Vol. 279): Wiley-Interscience.

Figure Captions

Figure 1. An example spectrogram of the stimulus utterance “I owe you a yo-yo”, with the F1 and F2 trajectories overlaid (Panel A, dashed curves). The set of local minima and maxima in F2 (landmarks) are labeled by the phonetic symbols (Panel B). The focus interval is the time period containing the AF perturbation.

Figure 2. Perturbations of the spatial parameters of AF: Up and Down. Panels **A** and **B** show **e**xamples of the Up and Down perturbations. **C:** Average responses to the Up perturbation in the PFS and PWS groups, shown as group-mean differences between the F2 trajectories produced under the Up and no-perturbation (noPert) conditions. The shading show ±1 standard error of mean (SEM). **D:** Group-mean responses to the Down perturbation (same format as Panel C). **E:** Group-mean Down-Up contrast (same format as Panel C). In panels C-E, the three bars at the bottom of each panel indicate the time intervals in which significant differences (corrected and uncorrected) were reached. The top two bars show significance of the F2 changes (from zero) in the PFS and PWS groups, respectively; the bottom bars show the significance of the between-group difference in the F2 change curves. The color coding scheme for statistical significance is illustrated in the “Significance Marker” inset. White: non-significant (n.s.) differences; lighter colors: significance at uncorrected (uncorr.) p<0.05; deeper colors: significance at permutation-corrected p<0.05.

Figure 3. Perturbations of the temporal parameters of AF, *Accel* and *Decel*. Panels **A** and **B** show examples of the Accel and Decel perturbation. **C.** Responses to the Accel perturbation in the PFS and PWS groups, shown as group-average change in the timing of the six acoustic landmarks (see Fig. 1). **D:** Responses to the Decel perturbation (same format as Panel C). **E:** Contrast between the time-interval changes between the Decel and Accel conditions. In Panels C-E, filled symbols represent time-interval changes or contrasts that are significant at permutation-corrected $p<0.05$. Asterisks indicate significant between-group difference ($p<0.05$, permutation-corrected). Note the different y-axis scales in Panels C-E