DIFFERENT SENSITIVITY TO TEMPORAL AUDITORY FEEDBACK PERTURBATION IN ADULTS WHO STUTTER: SYLLABLE STRUCTURE EFFECTS

Miriam Oschkinat, Philip Hoole

Institute for Phonetics and Speech Processing (IPS), LMU Munich
miriamo@phonetik.uni-muenchen.de

ABSTRACT

Ten adults who stutter (AWS) and ten fluent speakers were tested in three conditions with a real-time temporal auditory feedback perturbation paradigm (Onset, Vowel, and Coda Condition). In each condition, the perturbation target was stretched (either the onset consonant, the vowel, or the coda consonant of the German mono-syllabic word "Schaf", /ʃa:f/, *sheep*). No compensatory shortening for either of the stretched segments was observed, but both groups lengthened the segment(s) following the perturbed segment in the Onset and the Vowel Condition (reactive feedback control). In comparing both groups, the clearest tendency was that in the Onset Condition, AWS showed earlier and stronger responses by lengthening the following vowel. No group differences were observed in the Vowel or the Coda Condition. The data support the idea that persons who stutter use auditory feedback to a greater extent in onsets than fluent speakers and rely less on their internal representation.

Keywords: Temporal auditory feedback perturbation, stuttering, reactive feedback control, adaptation

1. INTRODUCTION

The integration of auditory feedback into the speech production process has proven crucial in the acquisition, planning, and control of fluent speech. In the online control of ongoing speech movements, the auditory feedback is used to compare the produced outcome against a stored representation. If the acoustic outcome does not match the prediction, corrections can be made in the online control with a latency of 120 to 200 ms (online compensation), or in future productions (adaptation). In adults who stutter (AWS), it is assumed that prolongations, blocks or repetitions occur because of a malfunctioning integration of sensory feedback information into the speech production process [1-3]. This assumption is supported by studies that found enhanced fluency in AWS under conditions of delayed auditory feedback [4]. Further, auditory feedback perturbation studies showed that online responses [5] or adaptation [6-8] to shifted vowel formants are weaker in AWS than in fluent speakers, supporting the idea that AWS rely less on auditory feedback and show deficits in integrating the auditory feedback into the speech production process. However, these studies investigated spectral parameters in syllable nuclei (vowels). Since stuttering symptoms occur in onsets rather than in syllable nuclei/codas, it is assumable that the integration of auditory feedback is differently weighted in AWS. [9] suggested that AWS rely too heavily on auditory feedback in onsets as compared to fluent speakers, who rely more on their internal representations. The latter was also suggested by [10-12] since fluent speakers showed no significant reaction to *temporally* perturbed syllable onsets, but to vowels and codas. To date, two studies investigated responses to temporal perturbations in adults who stutter. With unexpected random perturbations (online compensation paradigms), [13] found weaker responses to a time-shifted spectral vowel target (F2 minimum of /u/ in "owe" either accelerated or decelerated) in adults who stutter, [14] did not find a group difference between AWS and fluent speakers when the onset /s/ in the word "steady" was stretched followed by a compression phase later in the syllable. However, since online responses to a perturbation of speech timing are hardly feasible, the current study uses an adaptation paradigm to investigate responses to temporally perturbed auditory feedback in different parts of a syllable in AWS and fluent speakers. We expect different sensitivity to temporal real-time manipulations in AWS compared to fluent speakers with respect to syllable structure.

2. METHODS AND HYPOTHESES

Ten individuals who stutter (mean age 28.8 years, 3 females, mean age at onset of stuttering 5.5) and 10 fluent speakers (mean age 29.6 years, 3 females) were tested with a real-time temporal auditory feedback adaptation paradigm in Matlab using the Audapter software for real-time time-warping [15, 16]. Speakers spoke into a Sennheiser headset microphone and received auditory feedback via E-A-RTone 3A in-ear earphones with foam eartips. In each of three experimental conditions, speakers uttered the German phrase "mein Schaf" (*my sheep*) 95 times. The first 25 trials served as a Baseline with no perturbation, followed by a Ramp phase with increasing perturbation over 20 trials up to maximum perturbation which was held for another 30 trials (Hold phase). In the last 20 trials, normal feedback was restored (After-effect phase). The perturbation target was an 80 ms sequence of the onset consonant /ʃ/ in "Schaf" in the Onset Condition, the vowel /a:/ in the Vowel Condition, or the coda consonant /f/ in the Coda Condition. The respective perturbation target was triggered by Audapter's online status tracking with heuristic rules searching for certain landmarks in the signal such as the onset of a fricative or a vowel with predefined intensity or pre-emphasized intensity thresholds. The online status tracking was implemented for the purposes of the current study and programmed to trigger late in the respective segment of interest to avoid manipulation of formant transitions. In each condition, the perturbed sequence was stretched by a factor of 1.9 during maximum perturbation in the Hold phase (perturbed sequence stretched by 72 ms to approx. 152 ms). The signal following the perturbed sequence was therefore delayed by the amount of stretching, and the "catch-up" or compression phase was implemented into the silence after the end of the utterance, eventually reverting the signal back to real-time. The implementation of only stretching the signal without compressing an immediately following part had two main goals: Firstly, the expected compensatory response to a stretched signal would be a shortening in production of the respective segment. This shortening would necessarily be adaptive, since an online shortening as response to a stretched signal is physically not possible. Secondly, examining the duration of segments after the perturbation target would indicate adjustments due to reactive feedback control. Based on previous findings, we do not expect adaptive shortening of the stretched onset target, but for the stretched vowel and the stretched coda. Reactive feedback control effects should be observable in lengthened segments after the perturbation target in the Onset and Vowel Condition. If AWS rely more on auditory feedback in onsets, we expect them to show greater reactive feedback control responses in the Onset condition than in the Vowel or the Coda Condition. Further, we expect greater responses of AWS to the Onset condition compared to fluent speakers. Secondly, if AWS show deficits in auditory-motor mapping, we expect AWS to be less efficient in adapting with shortening responses to the stretched perturbed segments than fluent speakers.

3. ANALYSES

Dysfluent and erroneous trials were removed before calculations. However, due to the speech material, almost none of the AWS produced dysfluent trials. All following analyses were performed in RStudio using mostly packages from the *tidyverse* [17]. The data were grouped by participant, condition, segment, and phase; outliers were removed. To estimate the difference between Baseline productions and responses to maximum perturbation during the Hold phase, linear-mixed models [18, 19] were calculated per group (fluent speakers vs. AWS) and condition (Onset, Vowel, and Coda Condition). The last 20 trials of the Baseline and the last 20 trials of the Hold phase were included into calculations. Models were fitted with absolute durations in ms as dependent variable, with phase (Baseline or Hold phase) and segment (onset consonant, vowel, coda consonant) as predictors as well as the interaction between phase and segment. Random effects included by-subject intercepts and random slopes for phase. Per model, the predictor *segment* included the perturbed and following segments. For the Coda Condition the inclusion of segment was dropped since there was only one segment to be analysed.
Conditional R-squared values of the full models were provided by MuMin's *r.squaredGLMM* function [20] and *emmeans'* [21] pairwise comparison revealed the difference between Baseline and Hold phase for the respective segment.

4. RESULTS

The following section reports the conditional R-squared values for the linear mixed models and the estimates along with the standard error (SE), degrees of freedom (df), t-ratio, and p-value as provided by *emmeans*' pairwise comparison for the contrast between Hold phase and Baseline. Positive estimates indicate longer, and negative estimates shorter productions in Hold phase than Baseline (see Fig.1).

4.1. Onset Condition

The linear mixed models revealed that in the Onset Condition, the Hold-Baseline contrast for the perturbed onset consonant was not significant for the group of fluent speakers (R-squared = 0.75, estimate = 9.17 ms, SE = 5.56, df = 22.2, t.ratio = 1.651, p-value = 0.113) but significant for the group of AWS (R-squared = 0.73, estimate = 22.8 ms, SE = 8.15, df = 16.6, t-ratio = 2.80, p-value = 0.0125*). Both groups significantly lengthened the following vowel in production (fluent speakers: estimate = 33.91 ms, SE = 5.52, df = 21.5, t-ratio = 6.15, p-value < .0001***; AWS: estimate = 57 ms, SE = 8.11, df =

16.3, t-ratio = 7.03, p < .0001***). The coda consonant was significantly lengthened by the group of fluent speakers (estimate = 20.58 ms, SE = 5.58, df = 22.5, t-ratio = 3.688, p-value = 0.0012**), but non-significantly lengthened by the group of AWS (estimate = 16 ms, SE = 8.12, df = 16.4, t-ratio = 1.998, p-value = 0.0626).

4.2. Vowel Condition

In the Vowel Condition, the Hold-Baseline contrast for the perturbed vowel was non-significant in both groups (fluent speakers: R-squared = 0.85, estimate = 3.98 ms, SE = 5.02, df = 15.5, t-ratio = 0.791, p-value = 0.44; AWS: R-squared = 0.72, estimate = 15.5 ms, SE = 7.79, df = 11.7, t-ratio = 1.987, p-value = 0.07). This result is rather unexpected (at least for the group of fluent speakers) since previous studies found shortening responses in reaction to a stretched vowel and will be further discussed in section 5. Both groups significantly lengthened the following coda consonant (fluent speakers: estimate = 22.63 ms, SE = 5.06, df = 16, t-ratio = 4.47, p-value = 0.0004***; AWS: estimate = 21 ms, SE = 7.81, df = 11.8, t-ratio = 2.692, p-value = 0.0199*).

4.3. Coda Condition

The contrast between Hold phase and Baseline for the perturbed coda consonant in the Coda Condition was non-significant for the group of fluent speakers (R-squared = 0.92, estimate = -6.88 ms, SE = 4.91, df = 9, t-ratio = -1.402, p-value = 0.194), as well as for the group of AWS (R-squared = 0.82, estimate = 2.35 ms, SE = 4.79, df = 9, t-ratio = 0.491, p-value = 0.635).

4.4. Group comparison of Hold phase productions

For the group comparison, t-tests were calculated between the group of AWS and fluent speakers with mean Hold phase durations relative to the mean of the last 20 Baseline trials per participant, segment, and condition, after checking the data for a normal distribution. The t-tests indicated no significant group differences for any of the segments per condition (see Fig. 2). Since the means are already provided in the preceding section, the statistical details of the t-tests will not be reported.

4.5. Qualitative analyses of Ramp phase and After-effects

The sensitivity to an auditory shift can be determined by examining the magnitude of perturbation that evokes a change in production. In the Ramp phase, the perturbation is gradually increased by 3.6 ms per trial up to a 72 ms stretch in the Hold phase. In Fig. 1, a sensitivity difference between the two groups is observable in the Onset Condition for the vowel and the onset consonant (left panels, magenta stars and yellow crosses): Both the onset consonant and vowel productions in the group of AWS start to lengthen earlier in the Ramp phase and increase linearly, while the productions for the group of fluent speakers change more slowly over trials, showing less steep curves. The coda consonant seems not to differ visually between groups. In the Vowel Condition, however, the group of fluent speakers seems to be more sensitive to the shift since they lengthen the following coda consonant sooner (right at the beginning at the Ramp phase) and to a greater extent

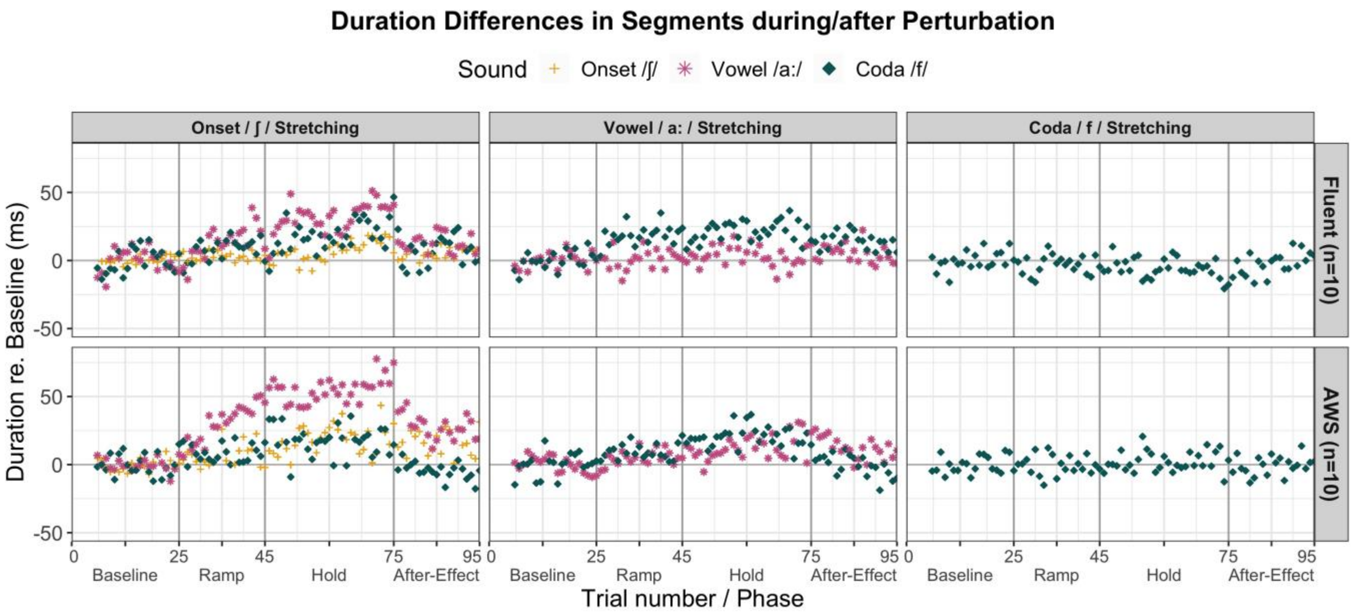

Figure 1: Mean durations (in ms) relative to the Baseline mean per group (upper panels: fluent speakers, lower panels: Adults who Stutter) and condition over the course of the experiment. Onset consonant productions in yellow crosses, vowel productions in magenta stars, and coda consonant productions in green rhombuses.

(middle panels, green rhombuses). Vowel productions themselves seem not to differ much between groups despite a lower general variability in vowel production in the group of AWS (magenta stars). In the Coda Condition, no group differences are observable in the Ramp phase.

The After-effect phase allows for the examination of adaptive behaviour when changes in production from the Hold phase remain. In the Onset Condition, vowel productions show a large drop from the Hold phase to the After-effect phase towards baseline productions in the group of fluent speakers, but remain longer throughout the After-effect phase, at least for the onset consonant and the vowel (upper left panel, magenta stars, yellow crosses). For the group of AWS, the remaining lengthening in the After-effect phase is stronger for vowel and onset consonant compared to the group of fluent speakers, indicating stronger adaptive behaviour. In the Vowel Condition, the group of fluent speakers seems to adapt the lengthened coda consonant (upper middle panel, green rhombuses), while for the group of AWS coda consonant productions revert to Baseline in the After-effect phase (lower middle panel, green rhombuses). Vowel productions are longer than Baseline productions in the After-effect phase for both groups. No differences are observable in the Coda Condition.

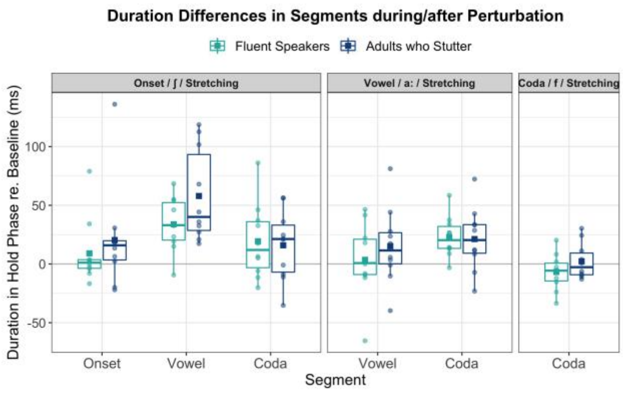

Figure 2: Mean difference in Hold phase relative to the Baseline mean per condition and segment. Left (turquoise) boxplots represent the group of fluent speakers, the right (dark blue) boxplots the group of AWS. Single participants are shown as dots. Boxes correspond to the first and third quartiles, bars represent the median and squares the mean. Whiskers extend from the hinge to the highest/smallest value no further than 1.5*IQR. Data beyond whiskers are outliers.

5. DISCUSSION

The current study investigated differences in responses to an auditorily stretched onset consonant, vowel, and coda consonant of a monosyllabic word in fluent speakers and AWS. Visual examination suggested that AWS show a greater sensitivity to temporally stretched onsets than fluent speakers, visible in earlier reactions in the onset and the following vowel during the Ramp phase. Analyses showed stronger lengthening of the following vowel (reactive feedback control) in AWS than in fluent speakers during maximum perturbation (difference between groups: 23 ms, but non-significant). The lengthening responses remained in the After-effect phase for both groups. In the Vowel Condition, no group compensatorily shortened the vowel in the Hold phase, and responses for the coda consonant were also similar with both groups lengthening the following coda consonant (~22 ms for both groups). The group of fluent speakers seemed to adapt (longer durations in the After-effect phase) while the group of AWS did not. In the Coda Condition, no significant effects were found.

Any group differences in the Hold phase turned out non-significant (section 4.4), presumably due to the currently rather small number of participants and fairly high variability in the data which might be reduced when the full dataset of 20 participants per group is tested/analysed. The findings of the current study support the idea that AWS rely more heavily on auditory feedback in syllable onsets than fluent speakers, and also rely more heavily on auditory feedback in onsets than in nucleus and coda of a syllable. These results are in line with the modelling studies in [9], and represent, to our knowledge, the first empirical demonstration by means of focal perturbation of the model predictions. Further, AWS seem to be able to adapt vowel productions in the Onset Condition, although these adaptive changes are rather an effect resulting from learned reactive feedback control than a direct (compensatory) response to a focal perturbation. The lack of adaptive shortening of the vowel in the Vowel Condition is rather unexpected (at least for the group of fluent speakers), since adaptive shortening for a stretched vowel was previously found in [10]. However, in [10], the perturbation caused a shift in phoneme category from /a/ to /a:/ in the auditory feedback, while the /a:/ in "Schaf" is already long and stretching might not cause an auditory error, or tolerances might be greater towards a longer vowel here.

In summary, the findings of the current study support the assumption of a different weighting of auditory feedback in AWS and fluent speakers with respect to syllable structure. The current investigation thus highlights the importance of taking syllable structure into account when studying the planning and control of speech timing, both in general, as well as in speech fluency impairments such as stuttering.

6. ACKNOWLEDGEMENTS

We would like to thank Nicole Benker, Anton Gadringer, Mona Franke and Charlie Wiltshire for their help in running the tests.

7. REFERENCES

[1] Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., and Wallace, M. E. 2004. "Unstable or Insufficiently Activated Internal Models and Feedback-Biased Motor Control as Sources of Dysfluency: A Theoretical Model of Stuttering," *Contemporary Issues in Communication Science and Disorders*, vol. 31, no. Spring, pp. 105-122, doi: doi:10.1044/cicsd_31_S_105.

[2] Hickok, G., Houde, J., and Rong, F. 2011. "Sensorimotor integration in speech processing: computational basis and neural organization," *Neuron*, vol. 69, no. 3, pp. 407-422.

[3] Harrington, J. 1988. "Stuttering, delayed auditory feedback, and linguistic rhythm," *Journal of Speech, Language, and Hearing Research*, vol. 31, no. 1, pp. 36-47.

[4] Andrade, C. R. F. d. and Juste, F. S. 2011. "Systematic review of delayed auditory feedback effectiveness for stuttering reduction," *Jornal da Sociedade Brasileira de Fonoaudiologia*, vol. 23, pp. 187-191.

[5] Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., and Perkell, J. S. 2012. "Weak responses to auditory feedback perturbation during articulation in persons who stutter: evidence for abnormal auditory-motor transformation," *PloS one*, vol. 7, no. 7, p. e41830.

[6] Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., and Chang, S. E. 2018. "Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter," *Developmental science*, vol. 21, no. 2, p. e12521.

[7] Kim, K. S. and Max, L. 2021. "Speech auditory-motor adaptation to formant-shifted feedback lacks an explicit component: Reduced adaptation in adults who stutter reflects limitations in implicit sensorimotor learning," *European Journal of Neuroscience*, vol. 53, no. 9, pp. 3093-3108.

[8] Kim, K. S., Daliri, A., Flanagan, J. R., and 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*, vol. 451, pp. 1-21, doi: https://doi.org/10.1016/j.neuroscience.2020.10.014.

[9] Civier, O., Tasko, S. M., and Guenther, F. H. 2010. "Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production," *Journal of Fluency Disorders*, vol. 35, no. 3, pp. 246-279, doi: https://doi.org/10.1016/j.jfludis.2010.05.002.

[10] Oschkinat, M. and Hoole, P. 2020. "Compensation to real-time temporal auditory feedback perturbation depends on syllable position," *J. Acoust. Soc. Am.*, vol. 148, no. 3, pp. 1478-1495, doi: 10.1121/10.0001765.

[11] Oschkinat, M. and Hoole, P. 2022. "Reactive feedback control and adaptation to perturbed speech timing in stressed and unstressed syllables," *Journal of Phonetics*, vol. 91, p. 101133, doi: https://doi.org/10.1016/j.wocn.2022.101133.

[12] Karlin, R., Naber, C., and Parrell, B. 2021. "Auditory Feedback Is Used for Adaptation and Compensation in Speech Timing," *Journal of Speech, Language, and Hearing Research*, doi: doi:10.1044/2021_JSLHR-21-00021.

[13] Cai, S., Beal, D. S., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. 2014. "Impaired timing adjustments in response to time-varying auditory perturbation during connected speech production in persons who stutter," *Brain and language*, vol. 129, pp. 24-29.

[14] Frankford, S. A., Cai, S., Nieto-Castañón, A., and Guenther, F. H. 2023. "Auditory feedback control in adults who stutter during metronome-paced speech I. Timing Perturbation," *Journal of Fluency Disorders*, vol. 75, p. 105943.

[15] Tourville, J. A., Cai, S., and Guenther, F. 2013. "Exploring auditory-motor interactions in normal and disordered speech," *Proceedings of Meetings on Acoustics*, vol. 19, 060180, pp. 1-8, doi: 10.1121/1.4800684.

[16] Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., and Perkell, J. S., 2008. „A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/," in *Proceedings of the 8th ISSP*, 2008, pp. 65-68.

[17] Wickham, H. *et al.* 2019. "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4(43):1686, pp. 1-6, Art no. 43, doi: 10.21105/joss.01686.

[18] Bates, D., Maechler, M., Biolker, B., and Walker, S. 2015. "Fitting Linear Mixed-Effect Models Using lme4," *Journal of Statistical Software*, vol. 67(1), pp. 1-48, doi: 10.18637/jss.v067.i01.

[19] Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models," *Journal of Statistical Software*, vol. 82(13), pp. 1–26, doi: 10.18637/jss.v082.i13.

[20] Bartoń, K. 2022. "MuMIn: Multi-Model Inference," *R package version 1.47.1 Available at https://cran.r-project.org/web/packages/MuMIn/index.html (last viewed: January 4, 2023).*

[21] Lenth, R. 2022. "emmeans: Estimated marginal means, aka least-squares means," *R package version 1.8.0. Available at: https://cran.r-project.org/package=emmeans (last viewed January 4, 2023).*