SYNCHRONIZATION TYPE MATTERS: ARTICULATORY TIMING IN DIFFERENT RHYTHMIC CONDITIONS IN PERSONS WHO STUTTER

Mona Franke[1,2,3], Nicole Benker[1], Simone Falk[2,3], Philip Hoole[1]

[1]Institute for Phonetics and Speech Processing (IPS), LMU Munich, Germany,
[2]Faculté des arts et des sciences – Départment de linguistique et de traduction, UdeM, Montreal, Canada,
[3]International laboratory for Brain, Music, and Sound Research (BRAMS), Montreal, Canada

mona.franke@phonetik.uni-muenchen.de

ABSTRACT

This study investigates articulatory timing of four persons who stutter (PWS) and four persons who do not stutter (PWNS) in different conditions: Speaking and tapping (self-paced), speaking along with a metronome (externally paced), speaking and tapping to a metronome (Metronome+Tapping). Using electromagnetic articulography, gestures of the articulatory speech onset and the finger taps were recorded and analyzed. Results show that, compared to the metronome beats, finger taps were more closely aligned with the articulatory speech onset supporting the assumption of a close link between articulatory and manual motor systems. Furthermore, our results indicate timing differences between PWS and PWNS, since intervals between metronome beat and articulatory speech onset were shorter in PWS. The Metronome+Tapping condition also led to significantly shorter intervals between articulatory onsets and finger taps in PWS. Our results suggest that PWS time their speech later when synchronizing to a metronome possibly pointing towards difficulties in movement initiation.

Keywords: stuttering, articulatory timing, gestural timing, finger tapping, paced speech.

1. INTRODUCTION

Stuttering is a neurodevelopmental speech fluency disorder that affects approximately 5-8% of children and 1% of the adult population [1]. The most characteristic symptoms of stuttering are involuntary disruptions in the flow of speech, such as pauses before a syllable (blocks), repetitions of sounds, syllables or words (repetitions) and lengthening of sounds (prolongations) [2]. These disfluencies typically occur at the beginning of (stressed) words or syllables, indicating that the speech motor program breaks down at this point. While the cause(s) of stuttering still remain(s) unknown, the breakdown of fluency in persons who stutter (PWS) has been linked to malfunctioning timing mechanisms (see [3], for a review). A recent review by Bradshaw et al. [4] proposed that in PWS the updating and use of internal models in speech motor control are disrupted, affecting both feedback and feedforward control of their speech.

Moreover, malfunctions in feedforward and feedback control in stuttering are linked to disruptions in more general motor networks in the brain, in particular, the basal ganglia-thalamo-cortical loop (e.g., [5,6]). This loop controls, among other processes, the timely initiation and termination of articulatory as well as other movements. There is indeed evidence that PWS also show alterations in non-verbal timing processes, such as finger tapping. Some studies found that PWS were more variable and tapped earlier in relation to the beat compared to persons who do not stutter (PWNS) when synchronizing finger taps to a metronome rhythm [7,8].

Interestingly, speaking with an external rhythm like a metronome reduces the occurrence of stuttered disfluencies in a major way [9]. Potentially, this phenomenon is due to higher reliance on cerebellar-cortical networks for motor control in PWS, circumventing the error-prone basal ganglia motor loop [5,10]. To date, it is unknown how inter-gestural timing (such as joint speaking and tapping) with or without an external rhythm is mastered by PWS.

Studies on joint speech and manual movements indicate that there should be a close coupling between speech and manual motor control systems (e.g., [11]). Hence, tapping and speaking at the same time (in the speech rhythm) could lead to more stable (articulatory) gestures in PWS as it is expected that this inter-gestural timing is closely linked. Moreover, it would be particularly interesting to see how PWS synchronize articulatory gestures and finger-tapping to an external rhythm. This setting can test whether timing information from multiple channels (auditory, manual + articulatory rhythm) is strongly or weakly coupled in PWS and PWNS, which might improve articulatory stability in the former or deteriorate it in the latter case. A study by Hulstijn and colleagues [12] points to weaker integration in PWS. They found that PWS were more variable in coordinating speech

and hand movements to tones than PWNS. However, they did not report how exactly the timing occurred nor whether effects on fluency were found.
Therefore, the aim of the current electromagnetic articulography (EMA) study is to shed further light on timing processes in PWS, by analyzing a) the effect of external pacing (metronome) and self-pacing (tapping) on speech gesture timing (where does the beat occur in relation to articulatory gestures) and b) how inter-gestural timing (non-verbal and verbal gestures) is affected by an external rhythm. It is an open question whether PWS synchronize their speech earlier to the metronome than PWNS, which would mirror results on non-verbal tasks (e.g., [7,8]).
Note that, in general, it is unclear whether metronome and finger-tapping synchronization time-points with respect to articulatory gestures differ or coincide. Therefore, it is another aim of our study to compare these conditions. What also remains to be answered is whether PWS differ from PWNS in the time point of synchronizing non-verbal gestures (finger tapping) and verbal gestures (speech). Following the result of Hulstijn et al. [12], we would expect that speaking along with a combination of motor and auditory pacing would lead to greater timing variability in PWS compared to PWNS.

2. METHODS

EMA data (AG501, Carstens Medizinelektronik) were collected from 10 adults who stutter and 10 adults who do not stutter. For the present paper, data of 4 persons who stutter (mean age = 24.3, 2 female) and 4 persons who do not stutter (mean age= 24.5, 2 female) were analyzed. All participants were native speakers of German and besides stuttering, no other impairments were reported. PWS and PWNS were matched in pairs having similar musical experience, the same age (±1 year), and sex.
Participants produced mono- and disyllabic German target words (cf. Table 1) embedded in the carrier phrase [ˈzeːə ____ ˈan] (Look at ____).

/a/	/o/	/u/
Maß [maːs]	Moos [moːs]	Mus [muːs]
Baden [ˈbaːdn̩]	Boden [ˈboːdn̩]	Buden [ˈbuːdn̩]
Mahl [maːl]	Mohn [moːn]	Buhne [ˈbuːnə]

Table 1: Target words

The experiment comprised 4 conditions (see below for more details) wherein each target word occurred 4 times along with filler words in a quasi-randomized order. The conditions were conducted in the following order:

- Baseline: Reading words embedded in the carrier phrase in a self-chosen speech tempo
- Tapping condition (self-pacing): Baseline + aligning finger tap to each word
- Metronome condition (externally paced): Reading + synchronizing each word to a metronome (90bpm)
- Metronome+Tapping condition: Reading + aligning finger tap to each word while synchronizing speech to a metronome (90bpm)

The metronome tone was presented via an in-ear headphone which participants plugged in their right ear. The onset of the metronome time point closest to the target word was automatically extracted.
For the conditions where tapping was involved, participants were instructed to tap their index finger of the dominant hand on an elevated wooden block that was placed on a table close to the participants. Sensors relevant for the data we report here were placed on the tip of the index finger of the participants' dominant hand and on the upper and lower lip. In addition, we had sensors placed on the tongue and the jaw. Only fluent productions (determined by listening to audio recordings) were analyzed. Therefore, a maximum of 144 target words were analyzed per participant.

2.1. Kinematic measures

Lip activity forming the constriction for the bilabial onset was measured using Lip Aperture (LA), defined as the Euclidian distance between transducers placed on the upper and lower lip. For LA and the finger tap (FT), the gesture onset was semi-automatically detected using a 20% velocity threshold and the onset of the gesture nucleus (start of the plateau) was also semi-automatically detected.
For each target word the relative timing of the consonantal gesture (onset) to the metronome onset and the finger tap onset was calculated as the lag between the onsets of the gesture nuclei of LA and FT and the lag between the onset of the LA gesture nucleus and the acoustic metronome onset.
We will refer to the two resulting intervals as *tap - articulatory onset interval* and *met - articulatory onset interval*. Both intervals are calculated such that positive values result if the articulatory onset is before the tap or the metronome.

2.2. Statistical analyses

For statistical analyses, linear mixed effects models (*lme4* package, [13]) were conducted with R Version 4.0.2 [14]. We are aware that this method was applied to a small group of participants. However, we aim to analyze a larger participant sample size with linear mixed models and we aim to have 10 participants per group ready for presentation at the conference. Therefore, we decided to include this method in the present paper. To determine p-values for the main effects and interactions between factors, a model including the fixed factor/interaction of interest was compared to the same model with no fixed factor/no interaction [15]. Post-hoc Tukey corrected t-tests, using the package *emmeans* [16], were performed to test significant interactions.

Variables that were included in the models as fixed factors were group (PWS and PWNS), as well as condition (Metronome, Tapping, Metronome+Tapping) or synchronization type (finger tapping, metronome) with a two-way interaction term between group and one of the latter two factors. Random intercepts were included for participant, word, and repetition number. Since repetition number did not have an effect on any predicting variables, it was excluded from all final models. Residual plots were visually checked for homoscedasticity of normality before reporting the results.

3. RESULTS

The following figure displays the interval between the articulatory onset and the metronome onset as well as the interval of the articulatory onset and the finger tap for the respective conditions (see Figure 1) and for all participants, separated by group.

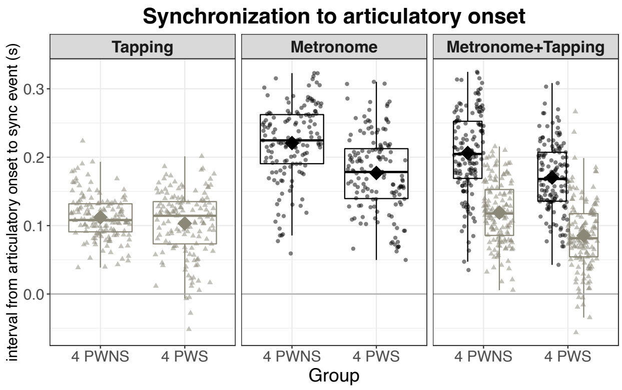

Figure 1: *tap - articulatory onset interval* (light grey, triangles) and *met - articulatory onset interval* (dark grey, dots). 0 seconds indicates the articulatory onset (nucleus onset of the bilabial), positive intervals indicate that the event of synchronization was after articulatory speech onset. Each triangle/dot represents one tap/metronome of one participant. Diamonds display the mean. Groups are displayed on the x-axis, PWNS = persons who do not stutter, PWS = persons who stutter.

As can be seen in Figure 1, finger taps are more closely aligned with the articulatory onset compared to the metronome. Hence, the *tap - articulatory onset interval* is shorter than the *met - articulatory onset interval*. A linear mixed effects model (Conditional $R^2 = 0.50$, Marginal $R^2 = 0.45$) was run in order to test whether the interval duration differs in the Metronome and the Tapping condition in PWS and PWNS. The model revealed that finger taps were aligned significantly earlier to the articulatory onset than the metronome ($p < 0.0001$). Furthermore, the model showed that group also had a significant effect on the interval duration ($p < 0.0001$). The significant interaction between condition (Metronome and Tapping) and group ($p < 0.0001$) revealed that the groups only differed significantly in the Metronome condition ($p = 0.021$) but not in the Tapping condition. Hence, PWS had significantly shorter *met - articulatory onset intervals* than PWNS.

To investigate how the combined condition (Metronome+Tapping) affected synchronization events in PWS and PWNS, we ran three different linear mixed effects models: The first model was run to test whether the synchronization time points of the two different synchronization types also differ even when they occur simultaneously in one task and whether there is a difference in timing between PWS and PWNS. The second and third models were run to test how synchronizing finger taps (second model) and metronome beats (third model) were affected by the combined condition and whether PWS and PWNS changed the timing in the combined condition compared to the single condition.

The first model (Conditional $R^2 = 0.47$, Marginal $R^2 = 0.39$) showed that PWS had significantly shorter intervals, regardless of synchronization type ($p = 0.0438$) and that finger taps were placed closer to the articulatory onset than the metronome ($p < 0.0001$). Thus, in the combined condition, PWS had shorter *met - articulatory onset intervals* as well as *tap - articulatory intervals* than PWNS. Note that this was not the case in the single Tapping condition.

These results led to the question whether synchronization time points with respect to the articulatory onset differed from the single to the combined condition in PWS and PWNS.

For the *tap - articulatory onset interval* the model (Conditional $R^2 = 0.31$, Marginal $R^2 = 0.07$) revealed a significant effect of condition ($p = 0.0003$) and a significant interaction between group and condition ($p = 0.0002$). Pairwise comparisons showed that PWS decreased the *tap - articulatory onset interval* in the combined condition compared to the single Tapping condition ($p = 0.0011$). PWNS on the other hand did not time their finger taps differently in the combined condition.

The third model (Conditional $R^2 = 0.36$, Marginal $R^2 = 0.11$) revealed that the time points of the

metronome beat shifted significantly towards the articulatory onset in the Metronome+Tapping condition compared to the single Metronome condition ($p = 0.0452$). This effect was found independently of group; no interaction was found.
Finally, in order to test if PWS were more variable than PWNS in synchronizing, the standard deviation was calculated for the different intervals per condition.

	SD Met - articulatory onset interval		**SD Tap - articulatory onset interval**	
	Met	Met + Tap	Tap	Met + Tap
PWS	0.069	0.065	0.048	0.051
PWNS	0.056	0.061	0.033	0.046

Table 2: Standard deviations for interval durations

Table 2 shows that PWS were more variable than PWNS in all conditions. However, when comparing intra-group differences between the single conditions (Tapping, Metronome) vs. the complex condition (Metronome+Tapping) it seems that PWNS increase more in variability in the intervals in the combined condition compared to PWS.

4. DISCUSSION

The study revealed both differences in timing when synchronizing speech with an internally generated rhythm (inter-gestural timing) as well as when synchronizing speech with an external rhythm (paced timing). Moreover, the data suggests differences between PWS and PWNS. We will first address differences between conditions and then group differences.
Compared to the metronome beats, finger taps were more closely aligned with the articulatory speech onset. This finding supports the idea of a close relationship between non-verbal and verbal motor systems [11,17]. Thus, joint tapping and speaking could lead to more stable gestures across modalities. Indeed, our results provide initial evidence for this conjecture as the timing of finger taps was more stable (smaller SD) than that of the external pacing with respect to the articulatory speech onset. A future study could therefore focus on the variability of the gestures themselves to test this assumption. The fact that the 8 participants in our study have longer intervals between articulatory speech onset and metronome beats could also indicate that externally paced speech is strongly based on acoustic cues. As previously shown (e.g., [18,19]) in purely perceptual studies, participants place the metronome beat within or close to the acoustic vowel onset of the target word. Hence, it is a possibility that the vowel onset is an anchor for acoustically synchronizing the metronome to one's own speech, while the syllable onset is the reference point for coordinating inter-gestural timing.

In terms of the group effect we found that PWS had shorter intervals between the metronome and the articulatory speech onset. This result indicates that PWS time their speech later to the metronome than PWNS, potentially because of later speech initiation in the group who stutters. This finding would be in line with preliminary results for children who stutter reported by Schreier et al. [20].
However, finger tapping to one's own speech did not differ between PWS and PWNS, indicating similar inter-gestural timing conditions. Interestingly, joint speaking and tapping to an external rhythm (Metronome+Tapping condition) led to a group difference in the interval between articulatory speech onset and finger tap such that PWS have shorter intervals than PWNS. This difference was caused by the fact that compared to the single Tapping condition, PWS decreased the *tap - articulatory onset interval* in the combined condition, whereas PWNS did not change the timing of their finger taps. Therefore, it can be assumed that in PWS inter-gestural timing is more affected by an external rhythmic cue than in PWNS. As previous research showed, PWS engage different timing mechanisms and/or brain circuits to time movements with an external cue [10,21].
Furthermore, it was hypothesized that timing would become more variable in PWS with increasing task complexity, however, this hypothesis was not supported by the results of our study. Despite PWS being more variable than PWNS in general, PWNS have a greater increase in timing variability in the combined condition.
From our study it can be concluded that PWS potentially couple auditory, manual, and articulatory rhythms in a different way, leading to later speech initiation and more temporal variation. This remains to be tested with a greater participant sample, of course. We aim to present data from 10 participants per group at the conference. Finally, our dataset offers the possibility for specific consideration of the vowel gesture, of inter-gestural timing between onset consonants and vowels, as well as on intra-gesture stability in different rhythmic conditions.

5. ACKNOWLEDGEMENTS

This work was supported by the DFG grants [FA 901/4-1 and HO 3271/6-1], as well as the Graduate School “Class of Language” of LMU, the Bavarian Research Alliance, and the German Academic Research Exchange Service. We would also like to thank Charlie Wiltshire for helping with data collection, Henry Derwanz for helping with data segmentation, and all our participants for participating.

6. REFERENCES

[1] Yairi, E., Ambrose, N. G. 2013. Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, 38, 66–87.

[2] Bloodstein, O., Bernstein Ratner, N. 2008. *A handbook on stuttering*, 6th ed. Clifton Park, NY: Thompson/Delmar.

[3] Etchell, A. C., Jonson, B. W., Sowman, P. F. 2014. Behavioral and multimodal neuroimaging evidence for a deficit in brain timing networks in stuttering: a hypothesis and theory. *Frontiers in Human Neuroscience*, 8, 467.

[4] Bradshaw, A. R., Lametti, D. R., McGettigan, C. 2021. The Role of Sensory Feedback in Developmental Stuttering: A Review. *Neurobiology of Language*, 2(2), 308–334. doi: https://doi.org/10.1162/nol_a_00036

[5] Chang, S. E., Guenther, F. H. 2020. Involvement of the Cortico-Basal Ganglia-Thalamocortical Loop in Developmental Stuttering. *Frontiers in psychology*, 10, 3088. https://doi.org/10.3389/fpsyg.2019.03088

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

[7] Sares, A. G., Deroche, M. L. D., Shiller, D. M., Gracco, V. L. 2019. Adults who stutter and metronome synchronization: evidence for a nonspeech timing deficit. *Annals of the New York Academy of Sciences*, 1149, 56–69. doi: 10.1111/nyas.14117

[8] Falk, S., Müller, T., Dalla Bella, S. 2015. Non-verbal sensorimotor timing deficits in children and adolescents who stutter. *Frontiers in Psychology*, 6, 847. doi: 10.3389/fpsyg.2015.00847

[9] Andrews, G., Howie, P., Dozsa, M., Guitar, B. 1982. Stuttering: Speech pattern characteristics under fluency-inducing conditions. *Journal of Speech, Language and Hearing Research*, 25, 208–216.

[10] Frankford, S. A., Heller Murray, E. S. , Masapollo, M., Cai, S., Tourville, J. A., Nieto-Castañón, A., Guenther, F. H. 2021. The Neural Circuitry Underlying the "Rhythm Effect" in Stuttering. *Journal of Speech, Language and Hearing Research*. 18;64(6S), 2325–2346. doi: 10.1044/2021_JSLHR-20-00328. Epub 2021 Apr 22. PMID: 33887150; PMCID: PMC8740675

[11] Parrell, B., Goldstein, L., Lee, S., Byrd, D. 2014. Spatiotemporal coupling between speech and manual motor actions. *Journal of phonetics*, 42, 1–11. https://doi.org/10.1016/j.wocn.2013.11.002

[12] Hulstijn, W., Summers, J. J., van Lieshout, P. H. M, Peters, H. F. M. 1992. Timing in finger tapping and speech: A comparison between stutterers and fluent speakers. *Human Movement Science*, 11(1–2), 113–124. https://doi.org/10.1016/0167-9457(92)90054-F

[13] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.

[14] R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[15] Winter, B. 2020. *Statistics for linguistics: An introduction using R*. New York: Routledge

[16] Lenth, R. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.2-1.https://CRAN.R-project.org/package=emmeans

[17] Pouw, W., Dixon, J. A. 2019. Entrainment and Modulation of Gesture-Speech Synchrony Under Delayed Auditory Feedback. *Cognitive science*, *43*(3), e12721. https://doi.org/10.1111/cogs.12721

[18] Marcus, S. M. 1981. Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, 30(3), 247–256.

[19] De Jong, K. (1992) Acoustic and Articulatory Correlates of P-Center Perception, *UCLA Working Papers in Linguistics*, 81, 66–75.

[20] Schreier, R., Dalla Bella, S., Hoole, P., Falk, S. 2020. Verbal timing deficits in stuttering. https://issp2020.yale.edu/S04/schreier_04_11_083_abstract.pdf. Last viewed: January 5, 2023)

[21] Etchell, A. C., Johnson, B. W., Sowman, P. F. 2015. Beta oscillations, timing, and stuttering. *Frontiers in Human Neuroscience*, 8, 1036. doi: 10.3389/fnhum.2014.01036