Journal of Phonetics 112 (2025) 101432

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

The effect of rhythm on inter-gestural coupling of onset and vowel gestures and predictive timing in stuttering

Mona Franke [a,b,c,d,*], Simone Falk [b,c,d], Nicole Benker [a], Phil Hoole [a]

[a] Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, Germany
[b] Faculté des arts et des sciences – Département de Linguistique et de Traduction, Université de Montréal, Canada
[c] BRAMS, Montréal, Canada
[d] CRBLM, Montréal, Canada

ARTICLE INFO

Article history:
Received 3 October 2024
Received in revised form 25 June 2025
Accepted 28 June 2025
Available online 21 July 2025

Keywords:
Speech motor timing
Inter-gestural timing
Predictive timing
Stuttering
Metronome-paced speech

ABSTRACT

In this study we investigate articulatory timing in fluent speech production in persons who stutter (PWS) and persons who do not stutter (PWNS) by focusing on consonant–vowel (CV)-timing, which refers to the coupling of onset consonant and vowel gestures, as well as on predictive timing, which describes the synchronization of the speech onset to a rhythmic event. These two timing mechanisms are particularly interesting to investigate in relation to stuttering, given that CV-timing is especially challenging for PWS and that they exhibit differences in predictive timing related to speech-motor and manual-motor tasks, suggesting that disturbances in inter-gestural coordination and auditory-motor integration may contribute to stuttering. To shed further light on this, we examine CV-timing and predictive timing under different rhythmic conditions.

Twenty German-speaking adults (10 PWS and 10 PWNS) were recorded using electromagnetic articulography (EMA). Participants produced target words that started with a bilabial onset, followed by a vowel (/a/, /o/, or /u/) and were embedded in a carrier phrase in four different conditions: Unpaced (speaking), Tapping (speaking while concurrently tapping), Metronome (synchronizing speech to a metronome), and Metronome+Tapping (speaking to a metronome while concurrently tapping).

We found evidence for both CV-timing and predictive timing differences between PWS and PWNS. Our results suggest that in general, PWS time CV gestures closer together. However, CV-timing differences were linked to condition in an unexpected way. As to predictive timing, PWS initiated their speech later to a metronome beat than PWNS but they did not differ when timing speech to their own finger tapping, indicating that motor-pacing may stabilize the speech motor system of PWS. In the Metronome+Tapping condition, the groups appeared to rely on different rhythmic cues. While PWNS timed their speech more towards the metronome beat, PWS synchronized their speech onset closer to the finger tap. We discuss that this difference could result from differences in CV-timing. Furthermore, the potential for future research on the interplay of non-verbal and verbal motor systems and the possible benefit for the stuttering population is discussed.

1. Introduction

Producing fluent speech requires finely coordinated timing of movements. Our speech motor system coordinates the complex movements of the lips, tongue, jaw, and larynx to maintain a continuous flow. This process is adaptable, allowing for variations in rhythmic patterns or pace. However, disruptions can occur to the system, for example, when there is a mismatch in timing between articulators, leading to breakdowns in speech.

Stuttering is a good example for such timing differences, but the precise nature of the underlying timing mechanisms remains debated (e.g., Etchell et al., 2014; Olander et al., 2010; Max & Yudman, 2003; Slis et al., 2023). Stuttering is a neurodevelopmental speech motor disorder (Smith & Weber, 2016) that typically emerges in early childhood, often between the ages of 2 and 5 years, and approximately 5 % of all pre-school age children and 1 % of the adult population stutter (Yairi & Ambrose, 2013). It manifests in involuntary disruptions during the initiation and coordination of articulatory

* Corresponding author at: Schellingstr. 3 (Institut für Phonetik und Sprachverarbeitung), 80799 München, Germany.
E-mail address: mona.franke@phonetik.uni-muenchen.de (M. Franke).

https://doi.org/10.1016/j.wocn.2025.101432
0095-4470/

gestures – abstract motor patterns that initiate the building and release of a constriction within the vocal tract (Browman & Goldstein, 1989; Browman & Goldstein, 1992). Gestures involve specific articulators, such as the lips and the jaw, constriction locations and degrees of constriction (see Browman & Goldstein, 1989; Browman & Goldstein, 1992). These disruptions to gestural coordination lead to very specific types of stuttered speech disfluencies such as repetitions, prolongations, and blocks of single sounds, parts of syllables or entire syllables (WHO, 2016). Although the neural origins of stuttering are still under investigation, there is a broad consensus among researchers that stuttering is characterized by atypical processes in the planning and execution of speech movements (Alm, 2021; Chang & Guenther, 2020; Chang et al., 2019; Max & Daliri, 2019; Neef & Chang, 2024; Smith & Weber, 2016).

1.1. CV-timing

The coordination of articulatory gestures can be described within the framework of Articulatory Phonology (Browman & Goldstein, 1989; Browman & Goldstein, 1992) and the timing between two gestures can be expressed as inter-gestural coupling. In the present study, we focus specifically on the inter-gestural timing between consonant (onset) gestures and vowel gestures. In particular, onset-vowel timing (henceforth, CV-timing) is challenging for persons who stutter (PWS). This difficulty is reflected in the fact that the vast majority of stuttered disfluencies occur at the beginning of (stressed) words or syllables (Bloodstein, 1995; Howell & Au-Yeng, 2002; Hubbard, 1998; Natke et al., 2004; Weiner, 1984) and maximally reach the acoustic onset of the vowel (Harrington, 1987). Thus, in the case of a stuttered syllable, differences appear from syllable onset up to the transition to the vowel, particularly in the initial formant transitions following the release of a consonant (Harrington, 1987). This led to the hypothesis of altered gestural coupling between onset consonants (C) and vowels (V) in PWS which we refer to as the “CV-timing hypothesis” (see Harrington, 1988; Wingate, 1988). Harrington (1988) proposed that stuttered speech occurs because individuals who stutter apply incorrect temporal predictions about the moment of occurrence of their own articulatory gestures. According to his approach, PWS expect the time of sensory feedback from their articulatory vowel gesture to occur earlier than it actually does. Thereby, they would correct for the erroneous prediction that their vowel gesture initiation is late and therefore start the gesture too early. This behavior would result in higher-than-usual articulatory CV overlap, leading to higher risk of stuttering (Harrington, 1988). For example, stuttering may occur when there is an attempt to simultaneously close and open the vocal tract. In contrast, Wingate (1988) proposed that a delayed initiation of the vowel (gesture), i.e., less articulatory CV overlap, would destabilize speech production in stuttering.

Evidence for the CV-timing hypothesis is provided by studies on coarticulation, defined as the extent of overlap between (onset and vowel) gestures (Hardcastle & Hewlett, 2006). A lower degree of coarticulation would indicate that there is a greater separation between onset and vowel gestures (as proposed by Wingate, 1988), a higher degree of coarticulation would inversely indicate that gestures overlap more (as proposed by Harrington, 1988). Studies comparing fluent speech of PWS and persons who do not stutter (PWNS) have found mixed results. Some studies report no coarticulatory differences between the groups (Frisch et al., 2016; Maruthy et al., 2018; Sussman et al, 2011). Some studies found a lower degree of coarticulation (Dehqan et al., 2016; Robb & Blomgren, 1997; Verdurand et al., 2020), while others found a higher degree of coarticulation (Klich & May 1982; Lenoci & Ricci, 2018). However, these studies are difficult to compare as they used different methods (e.g. ultrasound, formant-based measures), stimuli (different contexts due to different carrier phrases or isolated productions, CV target words with C corresponding to bilabial, velar, or alveolar plosives, alveolar and glottal fricatives, and different following vowels) as well as different languages (English, Farsi, French, Italian).

While the above-mentioned studies focused on fluent speech, Didirková & Hirsch (2020) examined coarticulation in stuttered speech and found that stuttering was frequently accompanied by a coarticulatory disruption but not always.

To understand the relevance of the CV-timing hypothesis for stuttering, investigating inter-gestural timing in actual articulatory kinematic data is most valuable. However, previous kinematic studies on stuttering focused primarily on the characteristics of disfluencies, speech movement variability, the amplitude and duration of speech movements, and the muscular effort involved in speech production (e.g., Chon et al., 2021; De Nil, 1995; Didirková & Hirsch, 2020; Heyde et al., 2016; Kleinow & Smith, 2000; Loucks et al., 2022; Lu et al., 2022; Usler & Walsh, 2018; Wiltshire et al., 2021; van Lieshout et al., 1996; Walsh et al., 2015; Zimmermann, 1980; for a review, see Wiltshire, 2019). There are very few articulatory studies on inter-gestural timing. Namasivayam & van Lieshout (2008), for example, analyzed inter-gestural timing in the context of motor practice and learning in PWS. Their findings indicated that PWS exhibited stronger inter-gestural coupling. Lu and colleagues (2022) investigated articulatory gestures in stuttered speech of one person who stutters, using real-time MRI. The authors found that disfluencies did emerge when a delayed release and overshoot of consonant gestures happened and not when the initiation of vowel gestures was altered (Lu et al., 2022). In this study, the comparison was only made between the speaker’s disfluent vs. fluent productions and there was no control speaker as a reference production, since the authors were interested in stuttered speech. In a more recent study, Lu et al. (2024) found that the vowel gesture was initiated in the first 50 % of a disfluent labial preceding consonant. Based on their results, the authors suggest that core stuttering does not result from fundamental difficulties in initiating or planning the upcoming vowel gesture, unlike what was proposed by Wingate (1988). However, Lu et al. (2024) did not compare the results to fluent CV productions of PWS to determine if the vowel gesture was actually initiated earlier in stuttered speech, which would be the prediction of Harrington’s (1988) hypothesis. In light of the lack of studies on inter-gestural timing, the present study probes the CV-timing hypothesis of stuttering by examining the kinematics of onset and vowel gestures in perceptually fluent speech of people who do and do not stutter using electromagnetic articulography (EMA).

1.2. Predictive timing

A complementary hypothesis on the role of timing in stuttering comes from brain research. Recent studies support the idea of deficient connectivity among brain areas in PWS that support general timing and rhythm processing, as well as auditory-motor integration (Chang, et al., 2011; Daliri et al., 2017; Jenson, et al., 2020; Lu et al., 2010). In adulthood, speech motor control relies more heavily on *feedforward processing*, that is dynamic interactions between sensory and motor systems via precise predictions of the output states of these systems (e.g., Guenther et al., 2006; Guenther & Vladusich, 2012). These predictions include predictions about future sensory states based on planned and ongoing motor commands (Max & Daliri, 2019). Hence, feedforward processes in motor planning involve both the anticipation and the precise timing of articulatory gestures, which we will henceforth refer to as "predictive timing" (Debarant et al., 2012).

The predictive timing hypothesis on stuttering posits that predictive timing on a neuromotor level is less reliable (Etchell et al., 2014) caused potentially by developmental alterations in prominent neural motor and timing circuits, in particular the basal ganglia-thalamus circuit (Chang & Guenther, 2020; see a summary in Falk, in press). An interesting phenomenon in this respect is that stuttered disfluencies reduce drastically when predictive timing is facilitated by a rhythmic context. Speaking with a metronome can significantly reduce disfluencies, often approaching a (near) 100 percent reduction of stuttering (e.g., Andrews et al., 1982; Davidow et al., 2009; Davidow et al., 2014). Evidence for the fluency-enhancing effect of metronomes has been reported across multiple modalities, including visual, auditory, and tactile (Brady, 1969). The effect is attributed to the fact that the upcoming time of an event can be predicted with very high temporal precision because of the cyclic nature of recurrent rhythmic events (Large & Jones, 1999).

Several studies have found that metronome pacing positively affects speech motor coordination (Davidow, 2014; Franke et al. 2023a; van Lieshout & Namasivayam, 2010; Wiltshire et al., 2023), for example, by reducing articulatory variability to a level of PWNS (Wiltshire et al., 2023) or by reducing durational variability of fricative onsets in a cluster (Franke et al., 2023a), as well as by reducing the amount of short phonated intervals ranging from 30–100 ms (Davidow, 2014). Neurally, metronome pacing has the effect of bypassing some of the malfunctioning neural circuits and reinstates a more stable neural information transfer inside sensory and motor regions of the brain (Frankford et al. 2021; Stager et al., 2003). This supports the conclusion that improved audio-motor coupling is the basis for the fluency-inducing effects in PWS (Stager et al., 2003).

Although PWS's fluency normalizes in metronome speech, timing does not, as some recent results show (Franke et al., 2023b; Schreier et al., 2020; Schreier, 2023). When speaking along with a metronome, PWS showed delayed speech initiation compared to PWNS. This has been demonstrated in children and adolescents who stutter for two measures, the acoustic onset of the syllable initial consonant and the acoustic onset of the vowel (Schreier et al., 2020; Schreier, 2023), as well as in adults who stutter at the articulatory speech onset (Franke et al., 2023b). Furthermore, children who stutter showed more consonant compression in a CC cluster in an unpaced and a metronome-paced condition compared to matched controls, suggesting that children and adolescents who stutter time onset consonants differently, regardless of an external cue (Franke et al., 2023a).

Timing differences have been reported before in non-verbal pacing tasks. Children, adolescents and adults who stutter showed altered timing when tapping with their finger to a metronome (children: Falk et al., 2015, adults: Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017). In these non-verbal tasks, PWS synchronized their manual movements earlier to the beat compared to controls which may be due to higher anticipation of the beat. In paced tapping to a metronome, finger taps typically precede an acoustic rhythmic event. This phenomenon is known as "negative mean asynchrony" which is attributed to strong temporal predictions, leading people to anticipate their movements to align with the rhythmic event (Aschersleben, 2002; Repp, 2005). As a result, PWS might over-anticipate the beat causing their finger taps to occur early in an attempt to align with the expected beat. This effect could derive from increased timing uncertainties and altered auditory-motor coupling in stuttering (Falk et al., 2015). Extending this argument to the verbal domain, it can be suggested that PWS's timing differences in synchronizing speech to a metronome could derive from higher uncertainties about synchronization time points ("the beat") in syllables due to articulatory timing errors. The perceived beat ("perceptual center", Marcus, 1981; Tuller & Fowler, 1980) is hypothesized to be closely tied to the articulatory onset of the vowel gesture.

Thus, it is a possibility that differences in timing speech onsets to rhythmic events (henceforth "onset asynchronies") between PWS and PWNS could result from different inter-gestural timing between consonants and vowels leading to higher uncertainty about the location of the syllabic "beat".

In sum, PWS show predictive timing differences related to speech motor and manual motor timing which suggests that disruptions in both inter-gestural coordination and sensory-motor integration may contribute to stuttering. This makes testing the hypotheses of CV-timing and predictive timing across various rhythmic conditions (metronome and finger tapping) especially intriguing. While auditory-motor integration is a key factor in synchronizing speech to external beats, the tactile and proprioceptive feedback from finger tapping may engage additional sensorimotor pathways, potentially influencing timing patterns differently in PWS compared to PWNS (e.g., sensory accumulation hypothesis [Aschersleben, 2002; Falk et al., 2015]). As it is assumed that proprioceptive tactile feedback is integrated more slowly than auditory information by the central nervous system (Aschersleben, 2002), the timing in self-paced tapping may be linked to a greater anticipatory response in order to integrate tactile feedback on time. Addressing these mechanisms in the context of gestural coordination may help clarify how different sensory feedback modalities affect speech motor control in PWS.

1.3. Aims and hypotheses

Studying the articulatory basis of the metronome effect will enhance our understanding of the underlying speech motor

control mechanisms involved in fluent speech production and shed light on specific articulatory adjustments that contribute to the increased speech fluency in PWS. Therefore, in the present study, we investigate gesture coordination and timing articulatorily in the presence of an auditory pacing stimulus (speaking to a metronome, Metronome condition).

As verbal and non-verbal timing differences in PWS have been reported in several studies (verbal: e.g., Dehqan et al., 2016; Klich & May 1982; Lenoci & Ricci, 2018; Robb & Blomgren, 1997; Verdurand et al., 2020, non-verbal: e.g., Falk et al., 2015; Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017), we also add a motor pacing condition, namely a speech-tapping condition (speaking and tapping at the same time, Tapping condition) which could provide information about general timing mechanisms and how verbal- and non-verbal systems might interact in PWS vs. PWNS.

It is important to note that some studies have not found significant differences between PWS and PWNS across motor domains (e.g., Hilger et al., 2016; Max & Yudman, 2003; Zelaznik et al., 1994). This mixed evidence highlights the need for further investigation into the interplay between timing systems across modalities. Little is known about the intermodal timing of tapping and speaking in stuttering and its impacts on articulation. In contrast, in PWNS, studies on finger tapping and speaking provide evidence for a close linkage between manual and articulatory motor systems, both neurologically (e.g., Meister et al., 2009) and kinematically (e.g., Parrell et al., 2014; Treffner & Peter, 2002). There is evidence that increased task complexity, such as coordinating speech and hand movements to tones, leads to greater variability in PWS (Hulstijn et al., 1992). Therefore, we also add an auditory-motor pacing condition (speaking to a metronome while concurrently tapping, Metronome+Tapping condition) to investigate how task complexity affects timing processes in both PWS and PWNS. Thus, our rhythmic conditions consist of two single pacing (either Tapping or Metronome) and one combined pacing condition (Metronome+Tapping).

In this study, we examine the CV-timing and predictive timing hypotheses for stuttering by investigating inter-gestural timing of onset and vowel gestures, on the one hand, and onset asynchronies, on the other hand, in adults who do and do not stutter in the previously described rhythmic conditions. In addition, we investigate CV-timing in an Unpaced condition.

As to CV-timing, we examine if inter-gestural coupling in perceptually fluent and unpaced speech of PWS differs from PWNS, and whether it is modulated by rhythmic conditions. Thus, the Unpaced condition functions both as a control for evaluating the impact of rhythmic conditions on inter-gestural timing and as a reference point in the study of CV-timing in fluent speech. We hypothesize that PWS have difficulties in generating typical inter-gestural timing in an Unpaced condition (i.e., speaking without a metronome or tapping), but that auditory and motor pacing will reduce or even eliminate these differences. Auditory pacing may positively impact inter-gestural timing by facilitating predictive timing (see above). Motor pacing could enhance speech motor timing through the additional activation of the premotor cortex, which plays a role in integrating verbal and non-verbal gestures (Meister et al., 2009). Given that auditory-motor pacing has been found to elicit more timing variability in PWS (Hulstijn et al., 1992), which could also extend to inter-gestural timing, we hypothesize to find a group difference in the auditory-motor pacing condition. From previous studies, it is not clear whether to expect more or less inter-gestural overlap in PWS. Following Harrington's (1988) model of stuttering, we would expect that PWS show more inter-gestural overlap in the Unpaced condition than PWNS due to predictive timing errors which would result in an earlier vowel gesture initiation and hence, in more overlap between consonant and vowel gesture. While we expect that PWS and PWNS do not differ in the single pacing conditions (auditory pacing and motor pacing), differences in CV-timing are anticipated in the Metronome+Tapping condition due to an increased task complexity. Prior studies suggest that higher task demands can affect motor timing in PWS. For example, increased syntactic complexity has been shown to negatively affect spatial and temporal motor stability (Kleinow & Smith, 2000), and longer vocal and manual reaction times were observed when task demands increased both in verbal and non-verbal conditions (Bishop et al., 1991). Furthermore, PWS show greater variability when synchronizing both speech and hand movements to a metronome, compared to simpler conditions such as synchronizing speech or hand movements alone (Hulstijn et al., 1992). These findings support the idea that increased task complexity, as in the combined Metronome+Tapping condition, may tax general timing mechanisms more strongly in PWS than in PWNS.

As to predictive timing, our first aim is to investigate whether PWS and PWNS differ in timing their speech onset to different rhythmic events, like a metronome beat or a finger tap, in the single pacing conditions. It remains an open question whether a) PWS would synchronize their speech earlier to a metronome and their finger taps than PWNS, matching the over-anticipatory behavior from non-verbal tasks (e.g., Falk et al., 2015; Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017) or b) whether they would show later speech initiation compared to matched control participants as found in metronome speech (Franke et al., 2023b; Schreier, 2020; Schreier, 2023). Furthermore, we are interested in how onset asynchronies are affected when complexity is increased, such as in the auditory-motor pacing condition (speaking to a metronome while concurrently tapping). Therefore, we compare rhythmic events (tap or metronome) in the single pacing vs. the combined pacing condition, without making specific predictions about group differences. However, we hypothesize to observe greater variability in onset asynchronies of PWS in the auditory-motor pacing condition compared to single pacing conditions, relative to PWNS.

2. Methods

2.1. Participants

Ten adults who stutter and ten adults who do not stutter participated in this study. All participants were native speakers of German, and the groups were age- and sex matched (PWS: Mean age = 23.1, SD = 3.18, range 20–30 years; PWNS: Mean age = 23.1, SD = 4.04, range 19–32 years; 5 males, 5 females per group), as well as matched for handedness (8 right-handed and 2 left-handed participants in each group).

One PWS reported having an auditory ossicle replacement in the right ear, with a doctor confirming that the hearing curve is within a normal range. Another PWS reported having ADHD.[1] Aside from stuttering and these reports, no present or past speech or hearing problems were noted. PWS indicated an onset of stuttering between the ages of 3 and 12 years. The mean age of stuttering onset was 6 years (SD 2.67). Out of 10 participants who stutter, 9 reported to have had stuttering therapy during some time of their life. Most of them had various therapies (on and off). One particular participant mentioned fluency shaping as a form of therapy and another reported to still be in therapy.

All procedures were performed in accordance with the Declaration of Helsinki and with institutional protocols. They received approval from the Ethical Committee of the medical faculty, LMU Munich. Every participant provided informed consent before participating in this study. Stuttering severity (disfluencies and physical concomitants) was assessed using the Stuttering Severity Instrument – Fourth Edition (SSI-4, Riley, 2009). Participants who stutter were recorded in person on video prior to the main experiment while doing an interview with the experimenter and while reading a passage. The interview and text reading were recorded approximately one hour prior to the main experiment. Interview questions were intended to get long responses from participants, so the experimenter asked open questions, such as “what do you do in your free-time” and “can you tell me about what you do for a living”. The text passage was chosen from a popular German children’s book, recommended for readers from 8 years on. For technical reasons, one participant did the interview and the reading via teleconference a few months after participating. All recordings were scored off-line by the first author or by a phonetics student who was specifically trained in doing the SSI. Three randomly chosen participants were evaluated by both the first author and the phonetics student. The inter-rater reliability between the two raters was high, evidenced by the same stuttering severity outcome, thereby indicating a strong agreement in their assessments. For one participant, both evaluators assigned the same SSI-4 score. For the other two participants, the ratings differed by only one point. In these cases, the lower SSI-4 score was selected, as it fell within the same stuttering severity category. Stuttering severity ranged from very mild to very severe, as can be seen in Table 1.

Table 1
Stuttering severity.

Participant	SSI-4 score	Stuttering severity
S01	22	mild
S02	16	very mild
S03	31	moderate
S04	25	moderate
S05	14	very mild
S06	19	mild
S07	5	very mild
S08	5	very mild
S09	37	very severe
S10	31	moderate

2.2. Speech material

Participants were asked to produce German mono- and disyllabic nouns (without determiners), embedded in the carrier phrase [ˈzeːə WORD ˈan] (Look at WORD) with the stress on the target word, as described for example in Brunner et al. (2014). Since the testing session included other words that are part of a larger study in addition to the target words for this study, we aimed to create a neutral context for the target words, similar to other studies (e.g., Pouplier et al., 2020). Therefore, the carrier phrase was designed to provide a neutral tongue position prior to the target word due to the schwa.

The target words comprised bilabial onsets ([m], [b]) and three vowels ([a], [o], [u]). Monosyllabic target words had a CVC structure. Apart from one disyllabic word with a CV.CV structure, all other disyllabic words followed a CV.CC pattern (of which the last C is syllabic), differing only in the vowel. In disyllabic words, stress was consistently placed on the first syllable. We chose tense vowels to gain more extreme articulatory movements, given that lax vowels are produced more centralized in stressed syllables (e.g., Fischer-Jørgensen, 1990; Jessen, 1993). The vowels [oː] and [uː] were chosen to detect horizontal tongue movement in a landmark-based approach (see section 2.6.4. *CV-lag*). The vowel [aː] was chosen in order to have an unrounded vowel as well for the trajectory-based analysis (see section 2.6.5. *Tongue Back trajectories over time*).

The final material comprised three target words per vowel forming triplets of words. These word triplets were matched as much as possible in word frequency based on written corpora of German provided by “digital dictionary of the German language” (DWDS, 2024). Table 2 displays the words and their respective frequencies. It can be observed that the target words occur roughly with the same frequency.

2.3. Procedure

Participants were comfortably seated in a sound-attenuated cabin, within the magnetic field of an electromagnetic articulograph (AG501, Carstens Medizinelektronik GmbH, 2014). They were asked to read out words presented in written form, inserting them in the carrier phrase while reading. Stimuli were presented on a monitor positioned in front of the participants that was located outside the magnetic field and at an approximate distance of 80 cm. Target items were organized into two lists, based on syllable length (e.g., all monosyllabic words in one list and all disyllabic words in another list). Monosyllabic words were randomized with 6 and disyllabic words with 5 additional target words that are not relevant to the focus of the present research questions. Note that the present experiment is part of a larger study and, therefore, included also two additional lists with target words with a different syllabic pattern

Table 2
Target words per vowel. Word frequency is given in parenthesis. The frequency scale is a seven-level logarithmic scale, reaching from 1 = rare to 7 = frequent.

/a/	/o/	/u/
Maß [maːs] (5)	Moos [moːs] (4)	Mus [muːs] (3)
Baden [ˈbaːdn] (4)	Boden [ˈboːdn] (5)	Buden [ˈbuːdn] (4)
Mahl [maːl] (3)	Mohn [moːn] (3)	Buhne [ˈbuːnə] (3)

[1] Note that stuttering often co-occurs with comorbidities such as ADHD or dyslexia (e.g., Blood et al., 2003).

(mono- and disyllabic words with onset clusters), each comprising 7 or 8 words. The first and the last word of each list was always a filler word in order to avoid phenomena like phrase-final lengthening in the target words. Hence, the experiment contained 4 different word lists that included 9 to 13 words in total.

To initiate each list, the first word was presented written within the carrier phrase on a white screen. At the same time, the word list arranged vertically appeared at the center of the screen. Therefore, the participants saw all words of one list at the same time, enabling them to establish a reading flow at their own tempo. The text on the screen was initially framed in red when a new list appeared on the screen. Participants were instructed to start reading once the frame turned from yellow to green. The time delay from the yellow to the green frame was identical for all participants and was 0.7 s long. The experimenter manually controlled the duration for which the text remained on the screen using MATLAB version R2017b (MathWorks, 2017), allowing for online monitoring of speech rate differences and disfluencies. Once the participant finished reading a word list, the experimenter closed it, displaying an empty screen, and then opened the next word list, framed in red. Accordingly, the audio recording contained one word list. The experimenter sat outside the cabin, monitoring the participant through a small window and via a video feed that was integrated into the experimenter's workstation.

There were 4 different reading conditions, aiming to investigate the effect of rhythmic triggering on fluent speech production. In the first condition, participants were simply asked to read the words embedded in the carrier phrase as described above (Unpaced condition). In the second condition, participants were asked to tap the index finger of their dominant hand one time per word while reading (Tapping condition). In the third condition, they heard a metronome beep (90 bpm, damped 1000 Hz sinusoid with a total duration of 19 ms) via one in-ear headphone on their right ear[2] and were told to synchronize each word along with the tone (Metronome condition). The metronome volume was adjusted to a comfortable level for each participant. The second and third conditions are referred to as the single pacing conditions. The fourth and final condition combined both of these and is referred to as the combined pacing condition. In this task, participants tapped along with their own speech while synchronizing to the metronome (Metronome+Tapping condition). The first two conditions (Unpaced and Tapping) can thus be classified as self-paced, as participants selected their preferred speech and tapping tempo. In contrast, the Metronome and Metronome+Tapping conditions can be referred to as externally-paced, since participants were asked to synchronize to an external auditory beat. In the self-paced conditions, participants were instructed to read the word lists in their preferred tempo, following a word list pattern style, meaning that they should avoid clear pauses between the end of one carrier phrase and the start of the next one. In the externally-paced conditions, the experimenter directed participants to synchronize each word with one metronome beat. The majority of participants read the word lists without missing a beat, i.e. in most cases there was no pause between sentences. Each word list was followed by a short break of approximately 5 s. In each condition participants were offered a longer break every four word lists to prevent fatigue. However, the majority of participants did not take these breaks and completed the experiment in one go. In cases where a participant needed a break, they could let the experimenter know when they were ready to continue with the experiment.

Each target word was repeated four times per condition in randomized word lists, resulting in a presentation of 16 word lists per condition that appeared in a randomized order. The order of conditions remained the same for all participants: First the Unpaced condition, followed by the Tapping condition, then the Metronome condition, and finally the Metronome+Tapping condition. This order was chosen to avoid a transfer effect of a rhythmic condition to the Unpaced condition and a transfer effect of the external pacing to the self-paced conditions. In total, participants produced a maximum of 144 target words.

The following figure (Fig. 1) provides an overview of the different conditions used in this study and the corresponding terminology that we use when we refer to them.

Before attaching the sensors to the participants' articulators for the main session (as described in the following section), a training session was conducted. This allowed participants to become familiar with the different conditions while also providing a break between the training session and the main experiment to prevent them from becoming too accustomed to the rhythmic conditions. To also get the participants familiarized with speaking with sensors glued on their tongue, one defective sensor was attached to the participant's tongue tip using medical tissue adhesive and one sensor was fixed with medical tape on the index finger of their dominant hand for the tapping conditions.

The training session included two word lists per condition, starting with the Unpaced condition, followed by the Tapping condition, the Metronome condition and lastly, the Metronome+Tapping condition. These word lists were the same as in the main experiment. Participants got feedback from the experimenter whether they were doing the task correctly. By the end of each block of the training session, all participants were performing the task according to the instructions. It is impossible to conduct the experiment without inducing some degree of potential practice-related confound. However, the approximately 30-minute break between the training session and the main session during which sensors were affixed to the participants' articulators, should help minimize the transfer effects of the rhythmic conditions to the main experiment.

2.4. Data acquisition and processing

Articulatory movement was recorded with an electromagnetic articulograph (EMA, AG501 Carstens Medizinelektronik GmbH) sampling at 1250 Hz. Electromagnetic articulography, especially using the AG501, provides reliable tracking of articulatory motion over time (Savariaux et al., 2017) by generating an electromagnetic field via transmitter coils placed around the head.[3] Sensor coils, attached to specific locations in the vocal tract, are then tracked within this field. For the present experi-

[2] Note that the reference sensor was positioned behind the left ear to prevent interference with the in-ear headphone.

[3] For additional comparisons of the AG500 and AG501, see Hoole (2014).

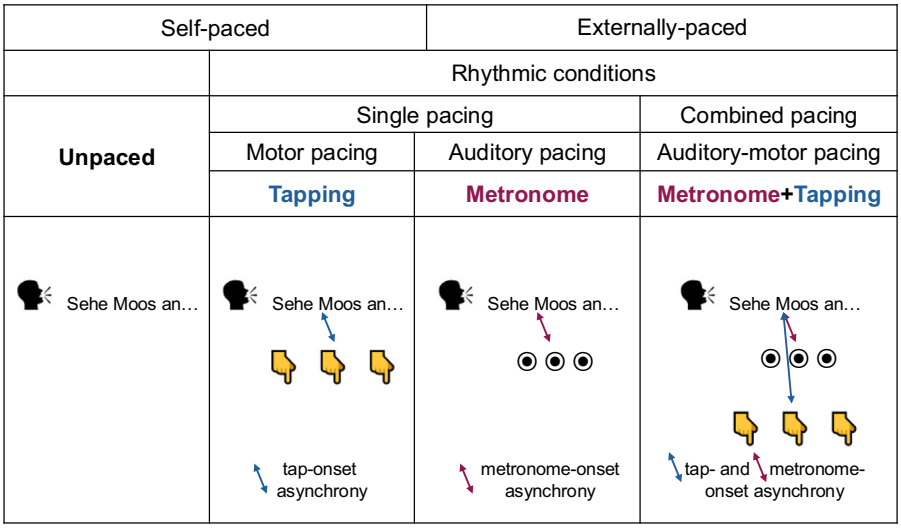

Fig. 1. Sketch of different conditions and respective terminology.

ment, which is part of a larger study, sensors were glued on each of the following articulators:

Lower lip (LL), upper lip (UL), Jaw, tongue tip (TT), tongue mid (TM), and tongue back (TB). The TT sensor was positioned approximately 1 cm behind the actual tongue tip. The TB sensor was placed as far back as the participant's gag reflex permitted. The TM sensor was then positioned midway between the TT and the TB sensors. Furthermore, three reference sensors were placed on the maxilla, the bridge of the nose, and behind the participant's left ear in order to factor out head movement. The following figure (Fig. 2) displays the location of the sensors (except the reference sensor behind the ear).

For all these sensors, a medical tissue adhesive (Cyano Veneer) was used for fixing them on the respective positions. For additional support, dental cement (Ketac) was used for fixating the sensors on the tongue. Both types of adhesives are approved for the use in the oral mucosa area.

In addition, another sensor was glued to the participants' index finger (IF), using medical tape, to capture non-verbal gestural movement. To ensure a high-quality recording of the finger tap movement, a table with a wooden surface was positioned in front of the participants. On this table, a 16.5 cm tall wooden block was added where participants were instructed to perform their finger taps. This elevated, but still comfortable tapping position brought the IF sensor closer to the ideal measuring field, ensuring the acquisition of good-quality non-verbal gestures.

According to the manufacturer, the optimum accuracy within the electromagnetic field is defined as a sphere with a radius of

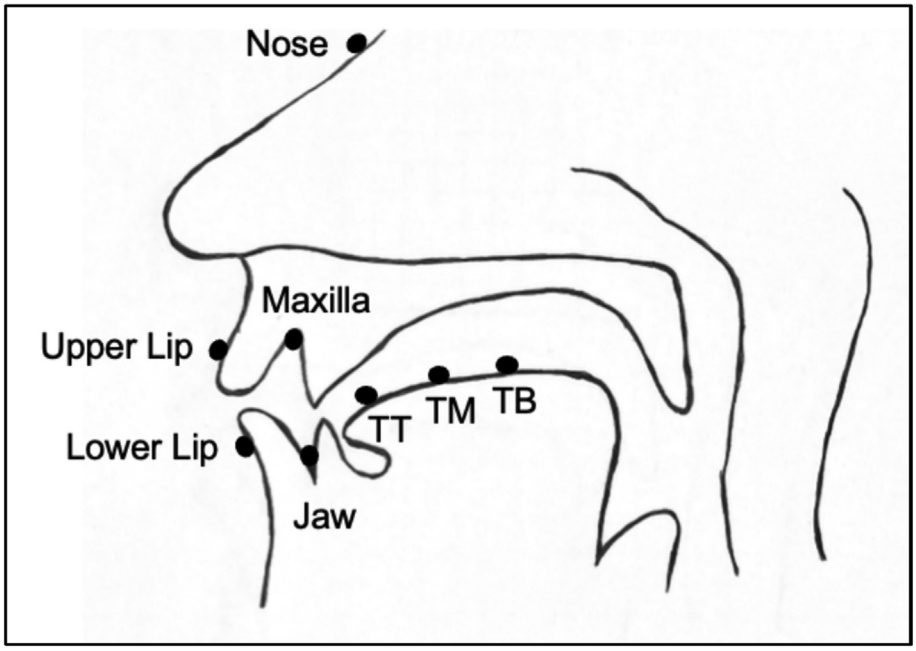

Fig. 2. Illustration of sensor placement.

15 cm, the center of which lies in the middle of the circular measurement plane. All articulatory sensors fall within this range. In addition, the accuracy downwards remains significantly better than in all other directions, which is why the elevated finger tapping position provides reliable data.

For the present study, the sensors LL, UL, TB, and IF are relevant. If a sensor came loose during the experiment – a rare occurrence reported by participants (happening in only 3 out of 20 participants) – the experimenter used the medical tissue adhesive to fixate it again on the same position. Photos taken after the sensors were initially glued to the articulators were used to ensure accurate repositioning.

Simultaneously, acoustic data were recorded at 25.6 kHz with an external floor-standing Sennheiser super-cardioid microphone, placed about 20 cm away from the participants' mouth. On a second channel, the metronome sound was recorded so that both recordings were time-synchronized. Additionally, a video recording of the main session was made from the participants face in order to be able to monitor the participant during the experiment and to evaluate the quality and usability of the data in the post-processing. After the main session, the occlusal plane was determined by having the experimenter place a plastic protractor between the participant's teeth. There were sensors placed on the tip and the center of the longer part of the plastic protractor. To collect a palate trace, the examiner moved a sensor attached to her index finger along the participants palate.

The duration of the experiment (including the glueing part) varied from subject to subject and ranged from approximately 1 h and 30 min to 2 h and 15 min.

2.5. Post-processing

The raw position data were processed using a Kaiser design FIR lowpass filter with a cutoff frequency of 20 Hz for all relevant articulators in this study. Head movements were corrected computationally with reference to the three reference sensors (placed on the maxilla, the bridge of the nose and behind the left ear). The post-processed data underwent a rotational transformation to align the spatial coordinate system with the occlusal plane. Velocities were computed with a three point central difference procedure.

2.6. Analyses

Prior to the analyses, trials were excluded if stuttering occurred within the carrier phrase or the target word, if the target words were mispronounced or if there was a slip of the tongue. In total, 94 trials were excluded (80 in PWS, of which 60[4] trials were removed due to stuttering-like disfluencies, and 14 trials in PWNS).

To support an accurate assessment of onset-vowel timing, we chose to take target word duration into account. This decision was made because speakers might employ different strategies to align their speech to a specific rhythm, such as increasing or decreasing vowel length or prolonging or shortening an onset consonant.

2.6.1. Word duration

An orthographic transcription of each trial (carrier phrase and target word) was semi-automatically generated using MATLAB (MathWorks, 2017). To obtain a phonetic segmentation of the sound signal into words and sounds, the files, together with the corresponding sound file, were processed via "WebMaus Basic", a tool from the Bavarian Archive for Speech Signals (BAS) Services (Kisler et al., 2017; Schiel, 1999). Resulting segmentations were manually checked and, if needed, corrected in Praat (Boersma & Weenink, 2019). From this corrected data, target word duration was extracted in order to account for rate differences between the groups and conditions.

2.6.2. Onset and vowel gesture of the target word and tapping gesture

All articulatory gestures were semi-automatically detected using the MATLAB program mtnew (Hoole, 2012). Lip activity forming the constriction for the bilabial onset was measured using Lip Aperture (LA). This measure was defined as the Euclidean distance between sensors placed on the upper and lower lip in mm.

The vowel gesture of the vowels /u/ and /o/ was segmented based on the anterior-posterior movement of the TBy sensor (we use a coordinate system with x lateral, y anterior-posterior, and z vertical). Given that the carrier phrase ends with a schwa (/zeːə/), the tongue is expected to be in a neutral position before moving backward to articulate the target vowels. Note that the anterior-posterior tongue position should not be much affected by the vertical movement of the lips and the jaw for producing the bilabial onset consonant, as for example demonstrated by Jackson and Singampalli (2009).

The following markers were segmented for the bilabial gesture, the finger tapping gesture, and the vowel gesture (Fig. 3, see panels LipApV, FINGER_zV, TBACK_yV):

A 20 % velocity threshold, referring to 20 % of the peak velocity of the (articulatory) movement, was used to detect the onset and offset of the gestures (see Fig. 3, markers 1 and 6). Additionally, the velocity maxima for the closing and opening movements of the bilabial gestures (Fig. 3, LipApV, markers 2 and 5) were segmented. For the finger-tapping movement, the velocity maxima correspond to the downward and upward movements of the index finger (Fig. 3, FINGER_zV, markers 2 and 5) and for the vowel gesture to the posterior and anterior movement of the TB sensor. Moreover, the onset and end of the gesture nucleus (see Fig. 3, markers 3 and 4) were semi-automatically segmented.

2.6.3. CV-lag

The CV-lag was analyzed as a landmark-based measure for inter-gestural timing. It is defined as the temporal interval between the nucleus onset of the bilabial gesture (see Fig. 3, LipApV, marker 3) and the nucleus onset of the vowel gesture (see Fig. 3, TBACK_yV marker 3). Using the nucleus onset, which can be referred to as target attainment, provided a more reliable measure compared to other landmarks, such as gesture onset-to-gesture onset (e.g., see Svensson Lundmark et al., 2021, for a comparison of different landmarks) as it

[4] Note that 40 trials were excluded from a single participant with very severe stuttering (S09). Specifically, 22 trials were removed from the Unpaced condition, 14 from the Tapping condition, 5 from the Metronome condition, and 1 from the Metronome+Tapping condition.

reduced variability both within individual participants and across participants, and the CV-lag remained more consistent across the vowels (/o/ and /u/). Note that CV-lag could only be calculated for the /u/ and /o/ target words, given that the vowel gesture for /a/ could not be segmented based on the horizontal TB movement.

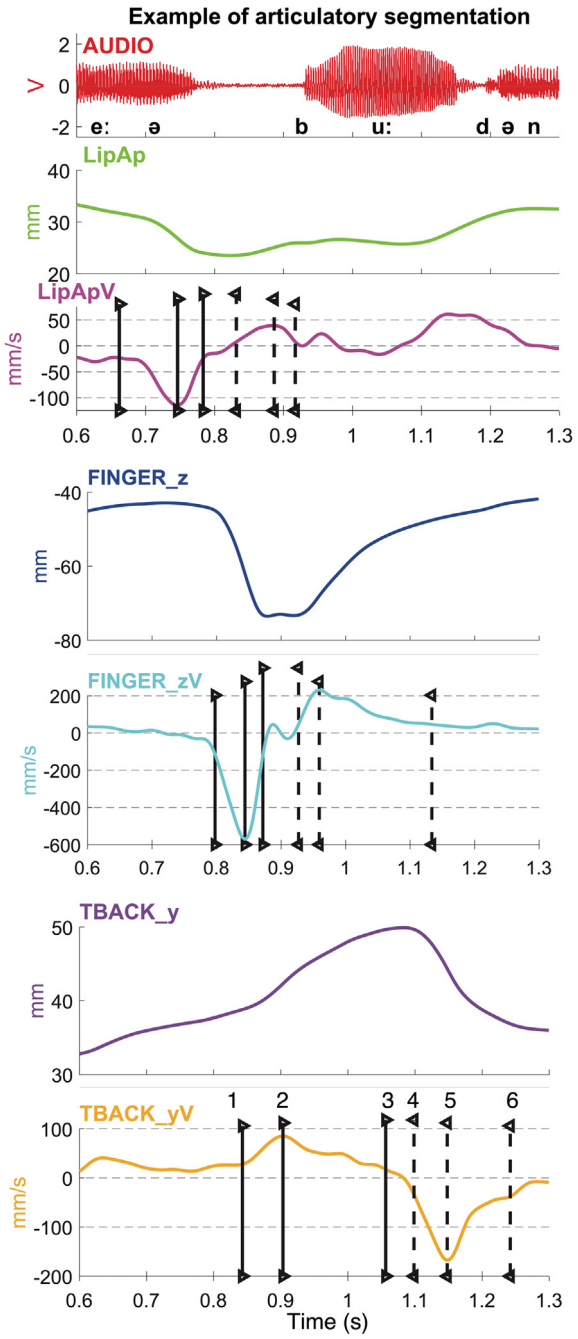

Fig. 3. Example of segmentation for the target word /bu:dən/ in the Tapping condition for the bilabial gesture, the finger tapping gesture, and the vowel gesture. Duration in seconds is displayed on the x-axis. Top panel: Audio signal, voltage (V) displayed on the y-axis, broad phonetic transcription on the x-axis. Lip aperture (LipAp), distance in mm displayed on the y-axis. Velocity of lip aperture (LipApV), velocity in mm/s displayed on the y-axis. Vertical position of the index finger (FINGER_z), distance in mm displayed on the y-axis. Velocity of index finger (FINGER_zV), velocity in mm per seconds displayed on the y-axis. Anterior-Posterior position of Tongue Back (TBACK_y), distance in mm displayed on the y-axis. Velocity of Tongue Back (TBACK_yV), velocity in mm per seconds displayed on the y-axis. Segment markers are displayed as black vertical lines. Numbers (only represented in the TBACK_yV panel) refer to different types of markers. 1 = gesture onset, 2 = maximum velocity closing/downward/backward movement, 3 = nucleus onset, 4 = nucleus offset, 5 = maximum velocity opening/upward/forward movement, 6 = gesture offset.

2.6.4. Tongue back trajectories over time

To incorporate all three target vowels and to ensure that the results were not based solely on one measure (target-to-target attainment), GAMMs were used to compare horizontal TB trajectories (vowel gestures) between PWS and PWNS in each condition. This approach aimed to investigate whether the groups differed in the timing of their vowel gestures in the region of the vowel gesture onset in different rhythmic contexts and is described in the following.

As pointed out by Sóskuthy (2021) GAMMs provide the advantage of modeling non-linear shapes over time while simultaneously accounting for random variability, similar to a generalized linear mixed model. Model predictions of TB contours are then compared between groups across conditions.

For each utterance, the acoustically defined CV portion of the target word was cut out. In order to have comparable time windows between speakers, these CV intervals were time normalized ranging from 0 (acoustic onset of the consonant) to 1 (acoustic offset of the vowel). Additionally, horizontal TB positions were normalized through z-transformation, accounting for individual speaker variations across conditions, following Wieling (2018).

2.6.5. Onset asynchronies

For each target word, two types of asynchronies were determined:

The relative timing of the consonantal onset gesture was calculated both with the metronome beat (metronome-onset asynchrony) as well as with the finger tap gesture nucleus onset (tap-onset asynchrony). The signed asynchrony (positive or negative sign to indicate the direction of the lag between two events, such as between the metronome and the bilabial onset or the finger tap and the bilabial onset) for each target word onset was expressed as the lag between the onset of the bilabial gesture nucleus (i.e. target attainment of lip closure) and the closest acoustic metronome beat *(metronome-onset asynchrony)* and the lag between the onsets of the gesture nuclei of LA and IF (*tap-onset asynchrony*), respectively. This can be thought of as the distance from the moment the lips close to the moment the index finger touches the wooden block (distance between marker 3 of LipApV and FINGER_zV in Fig. 3). Both asynchronies are calculated such that positive values indicate the occurrence of the articulatory onset before the tap or the metronome (and negative values when the articulatory onset occurs after the tap or the metronome). Note that this is the same procedure as described in Franke et al., (2023b). All metronome beats per trial (carrier phrase including the target word) were automatically extracted using a customized MATLAB script. The envelope of the pulses was computed by squaring the raw metronome signal and then smoothing with a cutoff of 50 Hz (non-causal Kaiser FIR filter). Beat location was determined as the time-point at which the pulse envelope first exceeded 50 % of the maximum value of the envelope signal. Beats were constrained to be within a window of +/- 0.002 s around the expected location of 0.6667 s from the previous beat. Typically, there were three metronome

beats per trial, as participants were instructed to align one metronome beat with each word. Therefore, the second metronome beat in a trial was used to calculate the metronome-onset asynchrony. Outliers were detected and excluded based on 3 SD above and below the group mean of the onset asynchronies (metronome vs. tap) for each rhythmic condition. This led to the removal of 47 out of 1404 observations within the metronome conditions (Metronome: PWNS 14, PWS 13; Metronome+Tapping: PWNS 7, PWS 13) which equals about 3 % of the entire data set. For the tapping conditions, there were only 8 out of 1381 observations removed which represents about 0.6 % of the entire data set (Tapping: PWNS 1, PWNS 3; Metronome+Tapping PWNS 2, PWS 2).

2.6.6. Statistics

For statistical analyses, linear mixed effects models (LMM, *lme4* package, Bates et al., 2015) were conducted with R Version 4.0.2 (R Core Team, 2020). To determine p-values for the main effects and interactions between factors, a likelihood ratio test was used to compare a model including the fixed factor/interaction of interest to a simpler model without the fixed factor/interaction (Winter, 2020). Thus, the models differ by only one predictor and any variation in the amount of explained variance is attributable to that predictor (Winter, 2020). Post-hoc Tukey corrected t-tests, using the package *emmeans* (Lenth, 2020), were performed to decompose significant interactions. LMMs were fitted to the data including target word duration, CV-lag, as well as onset asynchronies.

The final models are described in detail in the respective result section. Generally, for each model, we began by including variables of interest (i.e., Group and Condition) and the random factors Participant and (target) Word. Then we added complexity, such as interactions and/or random slopes, where model fit permitted. Likelihood-ratio tests were performed using the R-function *anova*, to compare several models with the intention to find the best fit model. Model fit was assessed using the Akaike Information Criterion (AIC), employing a threshold of 2 AIC units to determine the selection of a more complex model (e.g., Wieling et al., 2014). The explained variance was estimated using the function *r2_nakagawa* from the *performance* package (version 0.12.2, Lüdecke et al., 2021). Residual plots were visually checked for homoscedasticity and normality of residuals before reporting the results.

Type III ANOVAs were performed to assess the variability of onset asynchronies by Group (PWS and PWNS) and by Condition (single pacing conditions vs. combined pacing condition). Details of the analysis can be found in the respective section.

To determine trajectories of vowel gestures, GAMMs were built using the bam() function from the *mgcv* package in R (version 1.8.31, Wood, 2011; Wood, 2017) to analyze the relationship between the horizontal TB trajectory over time and the predictor Condition.Group, which resembles an interaction between the four conditions and the two groups, e.g. Unpaced.PWS or Unpaced.PWNS (procedure following Wieling, 2018). Details on the R syntax can be found in the Appendix. The *itsadug* R package (version 2.4.1, van Rij et al., 2022) was used for visualizing differences. Following Wieling (2018), an autoregressive error model (AR(1)) for the residuals was incorporated in the final model to avoid an overestimation of the effects. A visual method based on the estimated difference between the curves (*diff_plot* function from the *itsadug* package) was used to determine whether PWS show, as hypothesized, a higher value of TBy (i.e. more tongue retraction, increasing values from anterior to posterior), at the beginning of the acoustic CV interval which would indicate an earlier initiation of the vowel gesture and thus, a smaller CV lag. According to Sóskuthy (2021), this is an appropriate procedure for significance testing when there are hypotheses about a specific location.

3. Results

The following results are divided into two main sections, one on CV-timing (section 3.1.), one on predictive timing (section 3.2.). The order of conditions in the following figures corresponds to the sequence in which they were tested: First Unpaced, followed by Tapping, then the Metronome condition, and finally, the Metronome+Tapping condition. In 3.2. only the rhythmic conditions are reported.

Prior to the main analyses, we checked the duration of target words as a proxy for reading tempo to a) show how close spontaneous rate (in the self-paced conditions) was to the metronome rate, b) whether tempo differed between PWS and PWNS across the different conditions (see Fig. 4).

A linear mixed model was run to predict word duration. The final model (conditional $R^2 = 0.57$, marginal $R^2 = 0.12$) included Group (PWS and PWNS) and Condition (Unpaced, Metronome, Tapping, Metronome+Tapping) as fixed factors with a two-way interaction term between them. Random intercepts were specified for Participant and Word with by-Word random slopes for Group.

Firstly, Group was a significant predictor of word duration, $X^2(4) = 71.61$, $p < 0.0001$. Additionally, word duration varied significantly across conditions, $X^2(6) = 313.25$, $p < 0.0001$. Importantly, there was an interaction between Group and Condition, $X^2(3) = 69.86$, $p < 0.0001$. Pairwise comparisons revealed that PWNS slowed down their speech rate in the Metronome condition compared to the Tapping condition ($t(17.9) = 3.85$, $p < 0.0001$), whereas the metronome-paced speech of PWS was similar to their self-paced speech tempo in the Tapping condition. For this reason, target word duration was taken into account when investigating CV-timing. Table 3 shows the mean target word duration and its Standard Deviation (SD) for each group across conditions.

3.1. CV-timing

To investigate CV-timing we used CV-lag as a landmark-based measure of inter-gestural timing. Therefore, the coupling between LA and the horizontal TB movement of /o/ and /u/ target words is expressed as CV-lag. Positive lags indicate that the vowel gesture landmark is located after that of the onset gesture. The smaller the CV-lag on the positive scale, the closer the inter-gestural coupling. As pointed out above, to avoid relying solely on the target-to-target attainment measure and to be able to include all target vowels (/a/, /o/, /u/), we conducted a trajectory-based analysis, investigating the horizontal TB movement of participants over time using GAMMs. We hypothesized to find a group difference in the

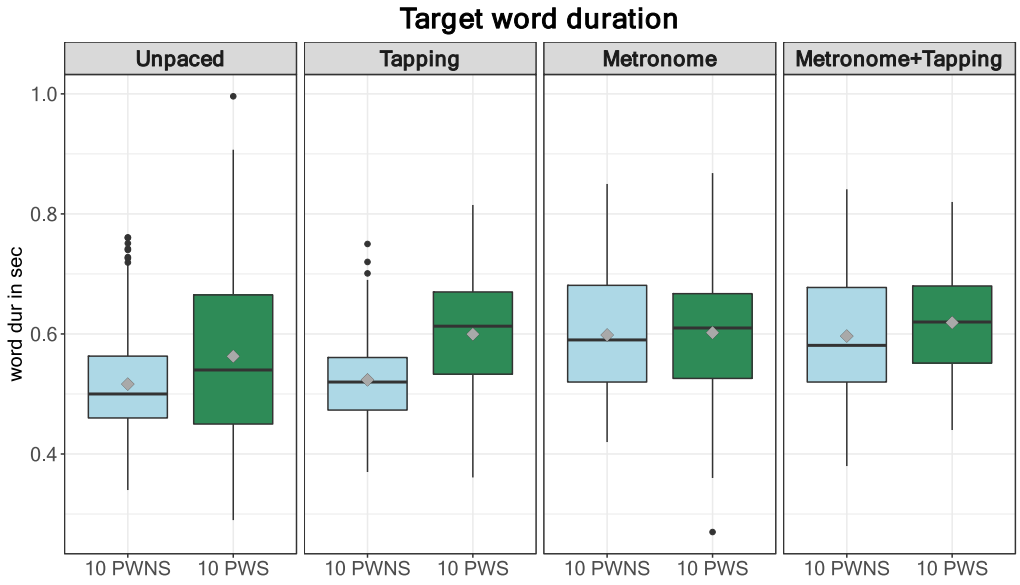

Fig. 4. Target word duration per group and condition. Durations are displayed in seconds on the y-axis. Groups are displayed on the x-axis, PWNS = persons who do not stutter (blue), PWS = persons who stutter (green). Diamonds display the mean. Within each box, the median is denoted with horizontal lines; boxes extend from the 25th to the 75th percentile of each group's distribution of values; the ends of the whiskers denote 1.5 interquartile range beyond the 25th and 75th percentile of each group; dots display observations outside the range of whiskers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Mean target word durations (in s) and Standard deviations (SD, in s) per group across conditions.

	Mean target word duration (SD)	
Condition	**PWNS**	**PWS**
Unpaced	0.52 (0.09)	0.56 (0.14)
Tapping	0.52 (0.07)	0.60 (0.10)
Metronome	0.60 (0.09)	0.60 (0.10)
Metronome+Tapping	0.60 (0.10)	0.62 (0.09)

Unpaced condition and the combined condition (smaller CV-lags in PWS and leftwards shift of the vowel gesture in PWS, indicating an earlier gesture onset).

3.1.1. CV-lag

Since target word duration varied significantly across conditions and groups, CV-lag was normalized based on the target word duration (CV-lag duration/target word duration). Results are visualized in Fig. 5.

A LMM including the fixed effects Group and Condition with an interaction term, as well as intercepts for Participant and Word (conditional R^2 = 0.34, marginal R^2 = 0.02) was run to predict the time-normalized CV-lags. Results should be interpreted with caution as low marginal R^2 indicates that the fixed effects do not explain much of the variance.

While the main effect of group was significant, $\chi^2(4) = 17.64$, $p = 0.0015$, with shorter CV-lags for PWS[5] and also a significant main effect for Condition, $\chi^2(6) = 19.41$, $p = 0.0035$, the most striking effect in the results is the highly significant interaction between Group and Condition, $\chi^2(3) = 16.58$, $p = 0.0009$. This highlights the necessity to look in more detail at pairwise comparisons.[6] In fact, the pairwise comparisons between groups did not actually show a significant difference in any of the conditions. Only suggestive evidence for a difference was observed between groups in the combined condition (estimate$_{PWNS-PWS}$ = 0.0636, $t(21.3) = 1.79$, $p = 0.088$). Pairwise comparisons for conditions within each group revealed that PWS produced shorter CV-lags in the Metronome+Tapping condition compared to the Unpaced condition, $t(1734) = -2.85$, $p = 0.023$. In contrast, PWNS increased their CV-lags in the Metronome conditions compared to the Unpaced condition (Metronome+Tapping: $t(1730) = 2.72$, $p = 0.033$, Metronome: $t(1730) = 2.80$, $p = 0.023$). The strong interaction effect can thus be attributed to this rather different behavior of the groups over the Unpaced vs. the Metronome conditions.

3.1.2. Tongue back trajectories over time

The model included a by-*Condition within Group* smooth function through time to investigate articulatory changes over time, and a random smooth to account for non-linear variation between Participants and Words. The final model explained 67.8 % of the deviance in the data.

Fig. 6 displays model predictions of horizontal TB contours for both PWS and PWNS for the different conditions. The top left panel, which shows the Unpaced condition, indicates that at the acoustic consonant onset, TB was closer to the target position of the vowel (maximum TB position) in PWS compared to PWNS. In no pacing conditions were there any differences between the groups in their TB trajectories over time (see Fig. 7).

This finding is further supported by the visual comparison of the estimated difference in horizontal TB position between the groups, which revealed a significant difference only in the Unpaced condition for the time windows between 0.0 and

[5] An anonymous reviewer suggested that differences in CV-lag might stem from variations in bilabial closure durations. We investigated this possibility and found no significant differences in bilabial closure duration (LipAp nucleus offset – LipAp nucleus onset) between the groups (Mean duration PWNS = 0.074 s, PWS = 0.073 s).

[6] A table with the results of the pairwise comparisons can be found in the Appendix (Table A).

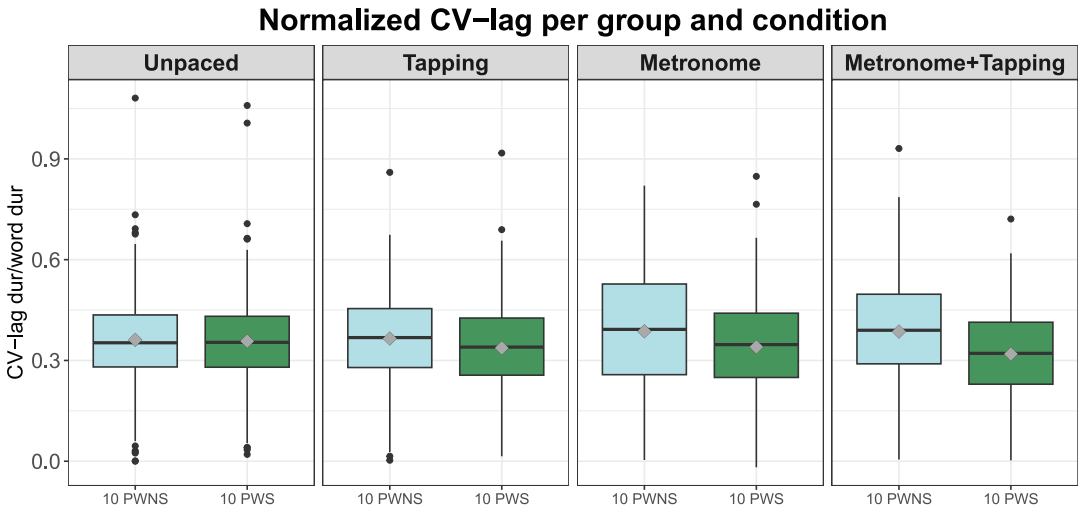

Fig. 5. Time-normalized CV-lags (in s) for each group per condition. Groups are displayed on the x-axis, PWNS = persons who do not stutter (blue), PWS = persons who stutter (green). Positive values indicate that the vowel gesture nucleus onset appeared after the consonant gesture nucleus onset. Details as in Fig. 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

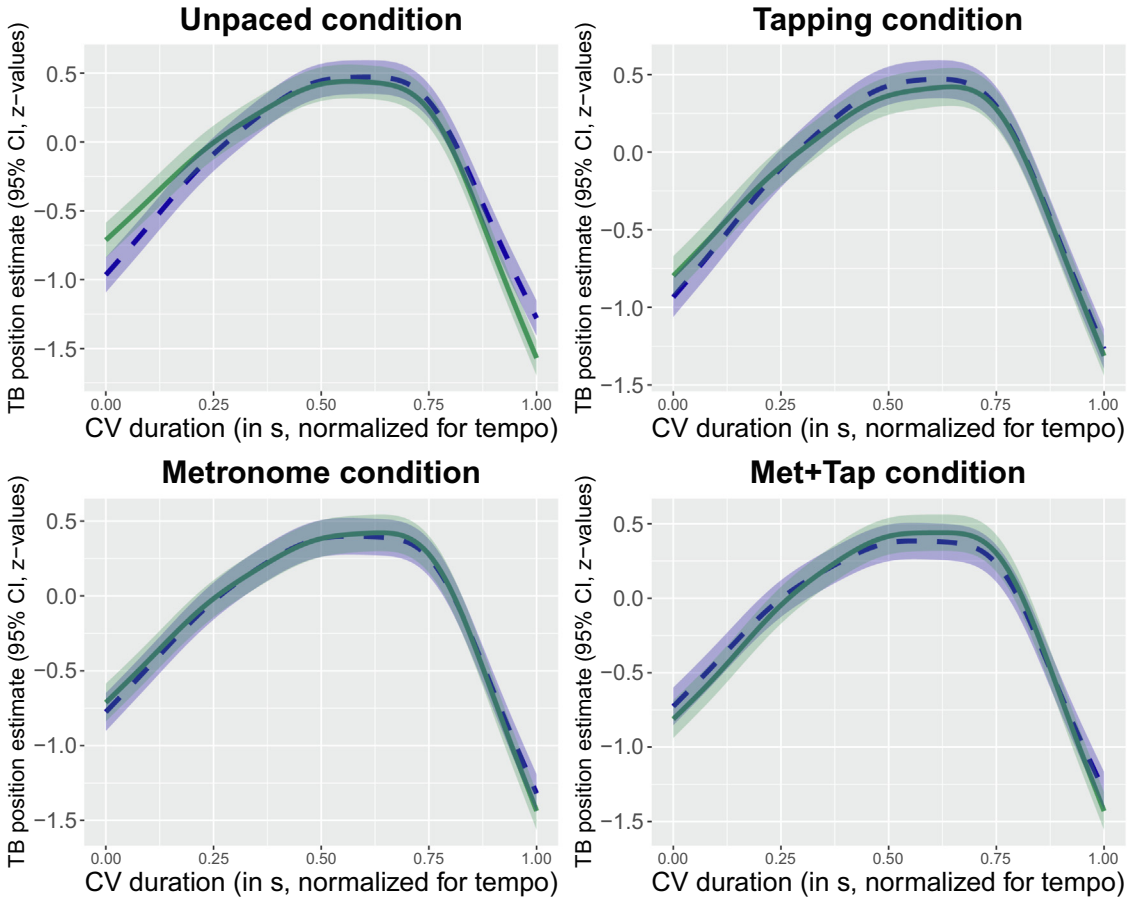

Fig. 6. Model predictions for 10 PWS (green, solid line) and 10 PWNS (blue, dashed line) within 95% pointwise confidence intervals. The x-axis displays the normalized time of the acoustic CV interval, the y-axis displays the estimated z-transformed position of the TB sensor (horizontal movement). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.06 as well as between 0.93 and 1 (see Fig. 7). This indicates that at the acoustic consonant onset, the TB position in PWS was already further back, while by the end of the acoustic vowel, the TB position in PWS had moved further forward.

In sum, the results for CV-timing indicate that CV-lag decreased for PWS in conditions involving auditory pacing (Metronome, or Metronome+Tapping) compared to the Unpaced condition, but PWNS's articulation remained unaf-

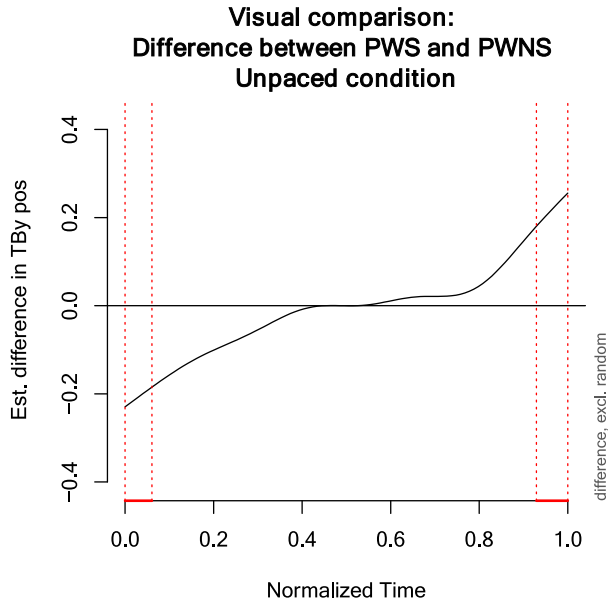

Fig. 7. Estimated difference of the horizontal TB position (Z-scores) between PWS and PWNS in the Unpaced condition within the associated 95% pointwise confidence interval (y-axis) over time (x-axis). The highlighted area in red indicates where the confidence interval excludes zero and the groups differ significantly. Negative values indicate that the TB position for PWS is further back compared to PWNS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fected by conditions. Additionally, the GAMM analyses suggest that PWS have earlier vowel gesture onsets compared to PWNS in the Unpaced condition, but were similar to PWNS in the pacing conditions.

3.2. Predictive timing

To investigate predictive timing, we compare onset asynchronies, defined as the lag between the articulatory speech onset (nucleus onset of the consonantal gesture, see marker 3 in LipApV Fig. 3) and the closest acoustic metronome beat (metronome-onset asynchrony) and the onset of the IF gesture (nucleus onset of the finger tapping gesture, see marker 3 in FINGER_zV Fig. 3) (tap-onset asynchrony), between groups and conditions. Note that positive asynchronies indicate that the rhythmic event (metronome or tap) occurred after speech initiation. Hence, if PWS show over-anticipatory behavior we would expect them to have larger positive onset asynchronies than PWNS, that is, they started speaking before the rhythmic event. Furthermore, it is expected that, compared to the single pacing conditions, the combined pacing condition would elicit higher standard deviations (SDs) of onset asynchronies in PWS compared to PWNS.

Fig. 8 displays the signed asynchrony between the articulatory onset and the metronome beats as well as between the articulatory onset and the finger taps for the respective conditions (panels a, b, and c) for all participants, separated by group.

Variables that were included in the LMM analyses of this section were the fixed factors Group (PWS and PWNS), as well as Condition (Metronome, Tapping, Metronome+Tapping) and Rhythmic event (tap, metronome) with or without a two-way interaction term between Group and one of the latter two factors. As random intercepts we included Participant, Word, and Repetition number. Since Repetition number did not have an effect on any predicting variables, it was excluded

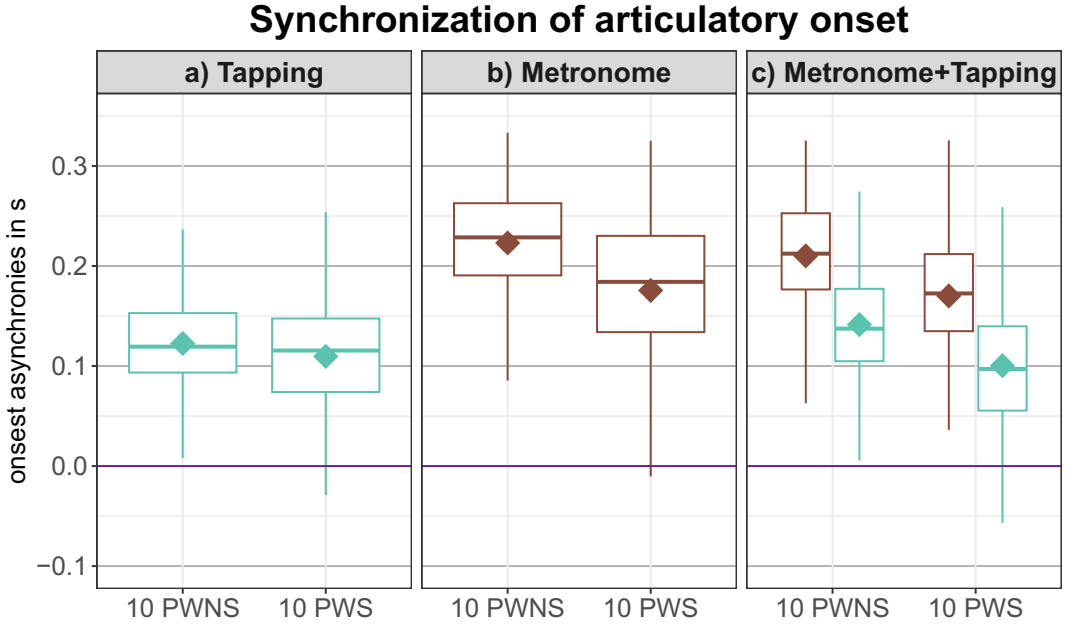

Fig. 8. Signed asynchronies between articulatory onsets of target words and rhythmic events (taps, metronome beats) in three rhythmic conditions (a: Tapping, b: Metronome, c: Metronome+Tapping). Tap-onset asynchronies (turquoise) and metronome-onset asynchronies (brown) expressed in seconds. The horizontal line at 0 s indicates perfect synchronization between the articulatory onset (nucleus onset of the bilabial) and the rhythmic event. Note that positive intervals indicate that the events occurred **after** the articulatory speech onset. Diamonds display the mean. Groups are displayed on the x-axis, PWNS = persons who do not stutter, PWS = persons who stutter. Details as in Fig. 4 with the exception that dots outside the range of whiskers are not displayed. The following outliers were excluded: Tapping: PWNS (n = 1, lower), PWS (n = 6, lower, n = 2, upper). Metronome: PWNS (n = 3, lower), PWS (n = 15, lower). Metronome+Tapping: Metronome-onset asynchronies PWNS (n = 4, lower), PWS (n = 17, lower), tap-onset asynchronies PWS (n = 5, upper).

Table 4
Standard deviations (SD, in s) for metronome onset asynchronies (left part) and tap-onset asynchronies (right part) in the single vs. the combined condition averaged over all participants per group.

	SD metronome-onset asynchrony		*SD* tap-onset asynchrony	
	Met	Met + Tap	Tap	Met + Tap
10 PWS	0.092	0.091	0.062	0.067
10 PWNS	0.056	0.056	0.045	0.053

from all final models. Adding by-Word random slopes for Group either did not improve the model or was not feasible due to model complexity.

To answer our research questions, 4 LMMs were fitted to the data. In the first model (model 1) only the single pacing conditions (see Fig. 8a vs. b) were compared in order to reveal differences between onset asynchronies during auditory vs. motor pacing. To investigate how the combined pacing condition (Metronome+Tapping) affected synchronization performance in PWS and PWNS, we ran three additional models: Model 2 tested the effect of Rhythmic event (i.e., tap vs. metronome) in the Metronome+Tapping condition (Fig. 8c) including potential Group differences (PWS vs. PWNS). Model 3 and model 4 compared the combined pacing condition (Metronome+Tapping) to each of the single pacing conditions to examine how synchronizing speech onsets to finger taps (model 3, Fig. 8c vs. Fig. 8a) and metronome beats (model 4, Fig. 8c vs. Fig. 8b) in the two groups was affected by the complexity of the task.

Model 1 (conditional R^2 = 0.40, marginal R^2 = 0.30) showed (see Fig. 8a + b) that tap asynchronies were shorter than metronome asynchronies (main effect of Rhythmic event, $X^2(2) = 495.32$, $p < 0.0001$). That is, participants aligned their articulatory onset closer with their finger movements than with the beats of an auditorily presented metronome. Furthermore, groups significantly differed in asynchronies, $X^2(2) = 52.48$, $p < 0.0001$, but only when speaking with a metronome ($t(21.4) = 4.71$, $p = 0.0006$) and not when tapping with their own speech (significant interaction between Rhythmic event and Group, $X^2(1) = 44.64$, $p < 0.0001$).

Model 2 (conditional R^2 = 0.33, marginal R^2 = 0.25) did not contain an interaction term between Group and Rhythmic event and included only data from the combined pacing condition (Fig. 8c). Results showed that PWS had overall significantly shorter asynchronies than PWNS in this condition (Group, $X^2(1) = 13.86$, $p = 0.0002$). Moreover, in both groups, finger taps occurred closer to the articulatory onset than the metronome beats (Rhythmic event, $X^2(1) = 16.84$, $p < 0.0001$).

Model 3 compared the tapping results in the combined condition (Fig. 8c) to the simple Tapping condition (Fig. 8a). The model (conditional R^2 = 0.41, marginal R^2 = 0.06), including an interaction term between Group and Condition, revealed a significant effect of Condition, $X^2(2) = 34.39$, $p < 0.0001$, and a significant interaction between Group and Condition, $X^2(1) = 28.80$, $p < 0.0001$. Pairwise comparisons showed that PWNS increased tap-onset asynchronies in the combined condition compared to the single Tapping condition by 19 ms ($t(1341) = 5.46$, $p < 0.0001$). In contrast, PWS showed a non-significant decrease in tap-onset asynchronies by 8 ms in the combined condition. This pattern resulted in a non-significant trend towards a group difference ($t(18.9) = 2.56$, $p = 0.08$). Crucially, however, the highly significant Group × Condition interaction demonstrates that PWS and PWNS responded differently to the shift from the simple Tapping to the combined Metronome+Tapping condition. We will explore the theoretical implications of this differential effect in the Discussion.

Model 4 compared the Metronome results in the combined condition (Fig. 8c) to the simple Metronome condition (Fig. 8b). The model (conditional R^2 = 0.31, marginal R^2 = 0.12) did not include the interaction between Group and Condition. Results revealed that the time points of the articulatory onsets shifted significantly towards the metronome beat in the Metronome +Tapping condition compared to the single Metronome condition, $X^2(1) = 10.29$, $p = 0.0013$. This effect was found independently of Group, $X^2(1) = 9.77$, $p = 0.0017$; PWNS shifted the articulatory word onset 13 ms closer to the beat and PWS 9 ms.

To explore whether task complexity increased timing variability more in PWS compared to PWNS, an additional analysis was conducted. The *SD* of the metronome-onset asynchronies and the tap-onset asynchronies was calculated per participant and condition to examine whether variability differed across conditions between the two groups. Table 4 shows the *SD* for the onset asynchronies per group.

Two-way ANOVAs (type III sums of squares) were performed separately for the two rhythmic events (metronome, tap). Hence, for the dependent variable the models included the *SD* of either the metronome-onset asynchrony or the tap-onset asynchrony, and the between-subject factors Group (PWS vs. PWNS) and Condition (single vs. combined). Results suggest that PWS exhibit more variable speech timing when synchronizing to a metronome compared to PWNS, $F(1, 36) = 4.57$, $p = 0.0394$. However, no significant group differences were found in speech synchronization to self-paced finger tapping, $p = 0.08$. There were no significant differences between the single conditions and the combined condition, and no significant interaction.

In addition to articulatory speech onset timing, we finally tested whether acoustic timing (i.e., using the acoustic vowel onset as another reference point) would yield different results. In previous research, vowel onsets have been pointed out to align quite closely with the moment syllables are perceived as rhythmic events (e.g., Fowler, 1983) and to provide information on how participants synchronize an auditory anchor to a rhythmic cue. As in the articulatory timing analysis above, we used a criterion of 3 SD above and below the group mean of the vowel onset asynchronies (metronome vs. tap) for each rhythmic condition to detect and exclude outliers. Fig. 9 displays the vowel onset asynchrony data without these outliers.

We ran two models to compare the single pacing conditions (see Fig. 9a vs. 9b); their aim was to probe for differences

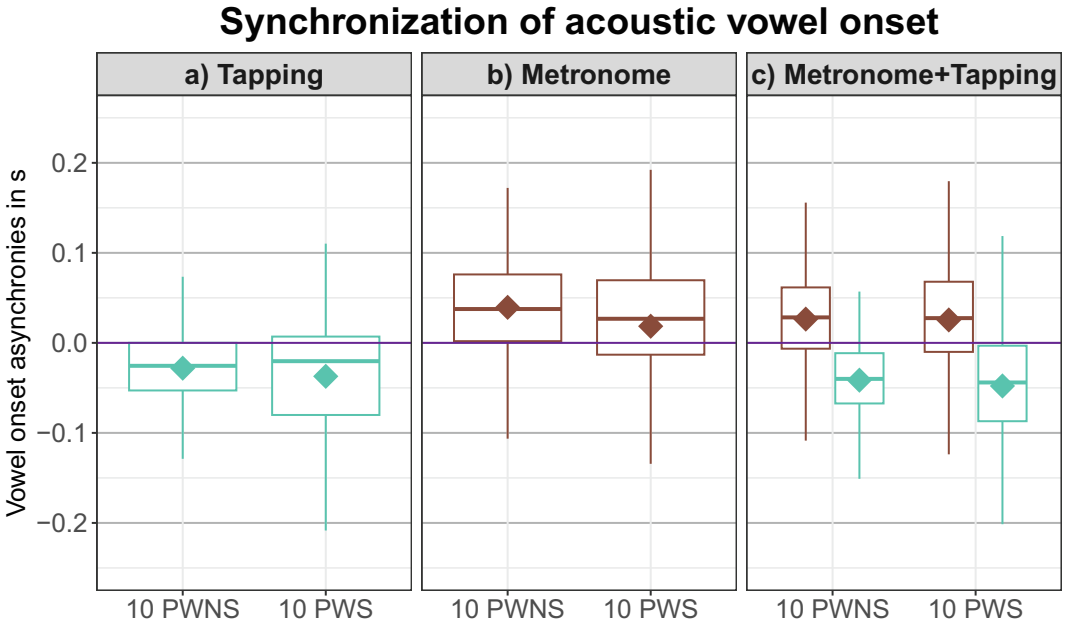

Fig. 9. Signed asynchronies between acoustic vowel onset of target words and rhythmic events (taps, metronome beats) in three rhythmic conditions (a: Tapping, b: Metronome, c: Metronome+Tapping). Tap-onset asynchronies (turquoise) and metronome-onset asynchronies (brown) expressed in seconds. The horizontal line at 0 s indicates perfect synchronization between the acoustic vowel onset and the rhythmic event. Note that positive intervals indicate that the events occurred **after** the acoustic vowel onset. Diamonds display the mean. Groups are displayed on the x-axis, PWNS = persons who do not stutter, PWS = persons who stutter. Other details as in Fig. 4 with the exception that dots outside the range of whiskers are not displayed. The following outliers were excluded: Tapping: PWNS (n = 1, lower, n = 2, upper), PWS (n = 2, lower). Metronome: PWNS (n = 3, lower, n = 2, upper), PWS (n = 25, lower, n = 1, upper). Metronome+Tapping: Metronome-onset asynchronies PWNS (n = 2, lower, n = 2, upper), PWS (n = 17, lower, n = 4 upper), tap-onset asynchronies PWNS (n = 1, lower, n = 2 upper), PWS (n = 3, lower).

between vowel onset asynchronies during auditory vs. motor pacing and groups (PWS vs. PWNS) (model 1, included an interaction term between Condition and Group), as well as to test the effect of Rhythmic event (i.e., tap vs. metronome) in the Metronome+Tapping condition (Fig. 9c) including potential group differences (model 2, no interaction term included). These models were identical to the models that used the articulatory speech onset as a reference point. Model 3 and 4 are not reported for the acoustic vowel onset reference point, due to weak statistical models (marginal R^2 lower than 0.03).

Model 1 (conditional $R^2 = 0.29$, marginal $R^2 = 0.15$) revealed a significant main effect of Rhythmic event, $X^2(2) = 239.28$, $p < 0.0001$ (see Fig. 9a + b), a significant Group effect, $X^2(2) = 22.68$, $p < 0.0001$, and a significant interaction between Rhythmic event and Group, $X^2(1) = 13.25$, $p = 0.0013$. Decomposing the interaction showed that the group difference was marginally significant in the Metronome condition ($t(21.1) = 2.54$, $p = 0.08$), but not in the Tapping condition.

Model 2 (conditional $R^2 = 0.29$, marginal $R^2 = 0.21$) indicates that in the combined condition (Fig. 9c) metronome beats occurred closer to the acoustic vowel onset than finger taps (Rhythmic event, $X^2(1) = 18.84$, $p < 0.0001$). In contrast to the articulatory onset reference point, there was no significant Group effect.

To sum up the main results on predictive timing, PWS show differences in articulatory timing, and a trend towards differences in acoustic timing (aligning the acoustic vowel onset with the metronome beat), compared to PWNS. PWS displayed shorter and more variable articulatory onset asynchronies with metronome beats than PWNS in both externally-paced conditions. As to intermodal effects, across groups, tap-onset asynchronies were shorter than metronome-onset asynchronies indicating potential differences in the articulatory timing mechanisms underlying auditory and motor pacing, as implemented in the present study. Furthermore, PWS and PWNS showed different tapping responses in the combined condition, whereas no group differences were observed in the single Tapping condition.

4. Discussion

With the present study we aimed to shed light on speech motor timing mechanisms in adults who stutter by using direct articulatory measurements to investigate the CV-timing and predictive timing hypotheses for stuttering. Additionally, our study investigates articulatory timing in a multimodal setting, providing novel contributions to the study of speech production timing in general. Therefore, we conducted an EMA study with 10 PWS and PWNS who produced speech in the four different conditions: Unpaced, Tapping, Metronome, Metronome+Tapping. These conditions were chosen to probe into auditory-motor coupling and its effects on predictive timing as well as inter-gestural timing in stuttering and to learn more about the interaction between verbal and non-verbal motor systems. Overall, our results indicate that adults who stutter differ from adults who do not stutter in both CV-timing and predictive timing, ultimately supporting both hypotheses.

4.1. CV-timing

As to the CV-timing hypothesis, we examined if inter-gestural coupling in perceptually fluent and unpaced speech of PWS differs from PWNS either by showing greater or lesser

overlap between consonantal onsets and following vowels. Recall that previous research, based primarily on acoustics, had given a mixed picture about whether to expect more or less overlap between consonant and vowel (CV) gestures (Dehqan et al., 2016; Klich & May 1982; Robb & Blomgren, 1997; Verdurand et al., 2020). With the present study we aimed to provide evidence for the CV-timing hypothesis using an articulatory approach. Moreover, we aimed to shed light on the effect of rhythmic auditory pacing, one of the most striking fluency-inducing effects in persons who stutter, on inter-gestural timing. We hypothesized that auditory pacing, and potentially motor pacing, lead to similar gestural timing between PWS and PWNS in line with previous research on metronome-paced speech (Wiltshire et al., 2023). However, adding complexity to the pacing task (i.e., speaking to a metronome while concurrently tapping) was hypothesized to lead to more variability in PWS, and hence, to a possible group difference. In order to examine inter-gestural timing of CV gestures, two different approaches were used: One landmark-based measure of the CV-lag (target-to-target attainment) as well as GAMMs for analyzing the TB trajectory over time.

Generally, results on CV-timing showed that PWS were producing more gestural overlap or a more posterior vowel position at the acoustic consonant onset, indicating an earlier vowel gesture initiation compared to PWNS in selected conditions. This general result is in line with studies that report a higher degree of coarticulation between consonants and vowels in stuttering (Klich & May 1982; Lenoci & Ricci, 2018). It also supports the version of the CV-timing hypothesis that stipulates higher CV-overlap as a source of stuttering (Harrington, 1988). However, our two approaches to CV-timing produced different results regarding the conditions. Results from the landmark-based approach indicate that the groups behaved differently across conditions. Within-group comparisons suggest that PWS and PWNS shift their lags in opposite directions with respect to the auditory-pacing conditions.[7] PWNS significantly increased their CV-lags from the Unpaced to the Metronome+Tapping condition, while PWS produced significantly shorter CV-lags. PWNS, in addition increased CV-lags from the Unpaced to the Metronome condition, while no significant difference was found in PWS between these conditions. Overall, the findings on CV-lags are rather subtle, as the model only explained a small amount of variance. Therefore, these results should be interpreted with caution. Nonetheless, we want to discuss them as they do not go in the expected direction.

The observation that PWS couple CV gestures closer in the Metronome+Tapping condition and show no difference in the others compared to the Unpaced condition, is contrary to our expectations. Metronome-paced speech is considered a fluency-inducing measure for PWS which is why we would have anticipated CV-lags to become more similar to those of PWNS and thus, rather longer than shorter. The fact that we observed the opposite pattern, namely shorter CV-lags, is also not in line with the results reported by Verdurand and colleagues (2020). They investigated coarticulation acoustically under normal and altered auditory feedback which is another fluency-enhancing condition for PWS and found that PWS show weaker coarticulation in the normal auditory feedback condition, i.e. a greater separation between the CV gestures, that even led to a greater separation under altered auditory feedback (Verdurand et al., 2020). Future research could address this difference by investigating the effect of various fluency-enhancing conditions on inter-gestural timing.

Importantly the GAMMs analysis (which explains more variance than the landmark-based approach), including all three vowels /a/, /o/, and /u/ also points in a different direction. When examining the precise tongue-back (TB) trajectory of the vowel gesture over time, PWS and PWNS differed in the Unpaced condition. Here, differences were evident around the acoustic consonant onset which, according to Articulatory Phonology, should also be in the area of the articulatory vowel onset (e.g., Goldstein et al., 2009; Hall, 2010; Nam et al., 2009). Moreover, differences were found around the acoustic offset of the vowel in the Unpaced condition. As to the initiation of the vowel, it is one possibility that the TB position of PWS was already closer to the target position of TB around the acoustic consonant onset, indicating an earlier initiation of the vowel gesture. Another interpretation could be that PWS had a different starting position of the tongue back, e.g. deriving from a more backwardly produced previous vowel (i.e., schwa) and thus being already closer to the vowel target position. However, PWS and PWNS were reaching the vowel target around the same time (same course in the area of maximum TB position). The groups again differed towards the end of the vowel gesture. This implies that the vowel gestures of PWS and PWNS were not directly shifted, as they did not differ for the central portion of the gesture, but that the initiation and the termination of the vowel gestures were differently timed in PWS, at least in the Unpaced condition.

Stuttering has been primarily associated with problems in the initiation and termination of syllabic onsets due to an altered basal-ganglia-cortical information transfer, as for example modeled with the GODIVA model (e.g., Chang & Guenther, 2020; Civier et al., 2013). According to this model, stuttering occurs because the next syllable program is not activated in time. However, our findings from the GAMM analysis suggest that speech motor differences may manifest themselves not only in syllable onsets but also in vowel gestures or potentially in the rhythmic syllabic “beats” of speech.

According to the GAMM results, all rhythmic conditions led to the mitigation of these differences. This is consistent with previous hypotheses about the fluency-inducing effects of pacing in stuttering (metronome-paced speech: Wiltshire et al., 2023). In general, our results do not support Wingate’s (1988) but rather Harrington's (1988) version of the CV-timing hypothesis, which stipulates that earlier vowel initiation during syllable production could be a general trait in the speech of PWS caused by erroneous temporal feedforward and subsequent error correction processes. Accordingly, even the perceptually fluent speech of PWS could exhibit these timing differences, as evidenced by the divergence in vowel initiation and termination in the Unpaced condition.

To sum up the discussion on CV-timing, the present study provides evidence for the CV-timing hypothesis by showing that PWS and PWNS differ in inter-gestural timing.

[7] For more details, refer to Table A in the Appendix.

4.2. Predictive timing

Regarding the predictive timing hypothesis, our primary goal was to determine whether PWS and PWNS differed in their ability to synchronize their articulatory speech onsets with different rhythmic cues, such as auditory pacing, motor pacing, and combined auditory-motor pacing. While predictive timing deficits in PWS have been previously found in non-verbal synchronization tasks (Falk et al., 2015; Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017), recent results also point towards predictive timing differences in verbal synchronization (Franke et al., 2023b; Schreier et al., 2020; Schreier, 2023). However, to our knowledge, articulatory dynamics have not been studied so far in this context.

Our results on the articulatory onset asynchronies show that PWS and PWNS differed in their synchronization to the metronome beats but not to their self-paced finger taps in the single pacing condition. Specifically, PWS timed their speech onset closer to the metronome beat, resulting in shorter metronome-onset asynchronies. This finding is in line with those reported for children and adolescents in a verbal pacing task (Schreier et al., 2020; Schreier, 2023). Interestingly, when tapping to their own speech, PWS and PWNS did not differ in onset-asynchronies, which we speculate may be due to the fact that PWS benefit from the additional activation of the premotor cortex, which is involved in integrating verbal and non-verbal gestures, leading to more stability (Meister et al., 2009). This idea is also supported by the finding that PWS were only more variable in their asynchronies to metronome beats but not to finger taps. Additionally, considering the sensory accumulation hypothesis (Aschersleben, 2002; Falk et al., 2015), the tapping condition relies more on proprioceptive and tactile feedback, which appears to function in a similar way in both PWS and PWNS. However, note that tapping on the wooden block also generated a subtle form of auditory feedback. The metronome condition, in contrast, requires the integration of solely auditory information (metronome beat) with the tactile information of the lip closure. This difference appears to also trend toward significance when aligning external auditory cues (metronome beats) with internal auditory information (acoustic vowel onsets). Given that the groups differ in the metronome condition, our results therefore suggest that the auditory-motor integration is altered in PWS.

Synchronization with the acoustic vowel onset is the more accurate measure for the synchronization time point, as it led to shorter asynchronies compared to the articulatory word onset. Both metronome beats and taps were closely aligned with the acoustic vowel onset, whereby the metronome beat trails into the vowel and the finger tap precedes the vowel. This close coupling between finger taps and vowel onsets has also been found for tapping with sentences produced by a model speaker (Rathcke et al., 2021). Nevertheless, articulatory onset asynchronies provide more accurate information about speech timing processes, which is why we will focus primarily on them in the discussion.

Contrary to our hypothesis, which predicted that increased task complexity (Metronome+Tapping condition) would lead to more variability in PWS, as observed by Hulstijn and colleagues (1992), neither group showed significantly more variability in onset-asynchronies in the combined condition compared to the single pacing conditions. However, the combined Metronome+Tapping condition did still lead to differences in the tapping behavior of both groups. While PWS shifted their taps more closely toward the articulatory speech onset compared to the single tapping task (this effect was not statistically different), PWNS aligned their finger taps closer to the metronome beat, and hence, further away from the articulatory speech onset. This result indicates that PWNS might prioritize auditory cues for synchronization, while PWS would be more prone to privilege precise inter-gestural timing of verbal and non-verbal gestures, relying more on internal motor timing mechanisms. This could be due to difficulties in generating precise temporal predictions from auditory cues. The neural circuits within the basal ganglia and supplementary motor area are largely involved in internal timing processes (timing movement without an external rhythmic cue) which are suggested to be impaired in PWS (e.g., Etchell et al., 2014). However, our results imply that these circuits function more like those of PWNS when PWS engage in rhythmic non-verbal movements, such as finger tapping, while speaking. Research on PWNS has demonstrated that the basal ganglia and the SMA are particularly active during internally timed movements (such as during a continuation phase of a finger tapping task) as opposed to externally timed ones (such as synchronizing finger taps to an external rhythm) (Rao et al., 1997). Our results suggest that PWS might improve speech motor timing and coordination through tasks that shift reliance more toward internal timing mechanisms, such as finger tapping while speaking. Given this interpretation, it would be compelling to replicate our experiment in a brain imaging setting, to investigate the neural activity underlying these observed differences between conditions. Furthermore, it would be interesting to focus on whether there are differences between the Unpaced and the Tapping condition, given that both are self-paced conditions. It is expected that PWS and PWNS would differ in the Unpaced condition (see for example, Chang & Guenther, 2020) but not in the Tapping condition.

That PWS could have more difficulties in making precise external timing predictions is also supported by the finding that PWS were in general more variable in metronome-onset asynchronies, regardless of condition, in our study. It is important to note that the Tapping condition preceded the Metronome condition in this experiment to avoid transfer effects from the timing of the external auditory stimulus. Furthermore, no transfer effects from the training session to the main experiment were observed, as evidenced by the differences found between the self-paced and externally-paced conditions.

What both groups had in common was that finger-taps were more closely aligned with the articulatory speech onset than with the metronome beat, supporting the notion of a close coupling between verbal and non-verbal motor systems (Meister et al., 2009; Parrell et al., 2014; Treffner & Peter, 2002). In non-verbal sensorimotor synchronization tasks, the gap between finger tap and metronome, known as “negative mean asynchrony”, is a common phenomenon (Repp, 2005). Therefore, it was not surprising to find this gap also in the Metronome +Tapping condition.

In terms of novel results for general speech production, our study highlights that the timing of articulatory gestures is influenced by the nature and combination of sensory inputs. Our results showed that, in multisensory, but not in self-paced speaking, the groups differed in articulatory timing, driven by a different weighting of external auditory cues vs. internal motor cues. This divergence suggests that multisensory integration plays a crucial role in speech timing and that individuals may weight sensory modalities differently based on task demands and underlying sensorimotor processing strategies.

The ability to time speech with cues of different modalities is crucial, for example, for smooth turn-taking in conversations or speech-gesture integration. While the current study focused on more predictable, rhythmic cueing, conversational turn-taking involves a different form of externally-based timing – one that is often less predictable and requires the speaker to time their response in reaction to subtle, multimodal cues. These may include auditory cues like intonation (e.g. phrase-final lengthening, a rising pitch contour or pauses), visual cues (e.g., facial expressions, head nods, body language), but also tactile cues (e.g., physical touch). Importantly, differences in conversational timing between PWS and PWNS have been reported. For instance, Jensen and colleagues (1986) found that PWS with severe stuttering exhibited shorter response latencies compared to PWNS. This result parallels the earlier synchronization to the metronome beat observed in PWS in the present study. It would be an interesting area of future research to investigate whether anticipatory timing patterns may extend to conversational contexts. Investigating turn-taking behavior with a multimodal approach could therefore be an interesting avenue to explore in future research.

To summarize, it can be concluded that our results on metronome-onset asynchronies point towards an alteration of predictive timing in PWS compared to PWNS, while the single motor pacing condition seems to eliminate these differences. The combined pacing condition indicates that PWS and PWNS rely on different cues when synchronizing their speech to rhythmic events.

4.3. The impact of predictive timing on inter-gestural timing in stuttering

Building on the findings related to onset asynchronies, it is plausible that the observed differences in timing speech onsets to rhythmic events between PWS and PWNS (particularly in the Metronome+Tapping condition) may result from differences in inter-gestural timing. As observed in the landmark-based approach, CV-lags of the groups moved in opposite directions, especially in the Metronome+Tapping condition. Whereas PWS produced smaller CV-lags, PWNS produced bigger ones. Thus, both, PWS and PWNS could still align their taps with the same articulatory reference point.

4.4. Limitations

The present study had several limitations. For example, the landmark-based measurement captured only six out of nine target words because the measurement of horizontal TB movement was not suitable for /a/ target words. In addition, the segmentation of the target vowel gesture was challenging as velocity patterns were not always clear enough to distinguish the onset of the target vowel from the preceding schwa vowel. Having a high front vowel instead of a schwa preceding the target word could have led to a clearer distinction in the landmark-based approach. However, the carrier phrase was chosen as part of a larger study and was intended to be as neutral as possible to exclude potential coarticulatory effects. Öhman (1966) noted that there is a continuous vowel gesture overlaid by consonantal gestures, highlighting the difficulty of investigating vowel gestures. Therefore, we chose target-based measures for investigating CV-lags as they were clearly assignable to the corresponding sounds. It remains a topic of debate whether CV coordination is solely anchored around gesture onsets or whether different coordination relations exist, such as the gestural target-coordination or endpoint-coordination (Durvasula & Wang 2023; Kramer et al., 2023; Shaw & Chen, 2019; Turk & Shattuck Hufnagel, 2020).

For this reason, among others, a trajectory-based approach was included to see whether groups differ in the vowel gestures over time. Using a two-dimensional analysis of the vowel gesture over time, focusing on both the horizontal and vertical movement of the TB sensor, could have provided an additional method to detect the actual onset of the vowel gesture (and not the nucleus onset of the vowel gesture), as it could highlight the points of divergence between vowels more clearly. Additionally, the sample size of target words was limited to nine, representing only three different vowels in order to keep the experiment to an acceptable timeframe. To gain a comprehensive understanding of onset-vowel timing in PWS and PWNS, future research should aim to include a broader range of words that cover as many vowels as possible from the phonemic inventory of the language being examined. Furthermore, it should be mentioned that CV timing is affected by the coarticulatory resistance of the vowel to the preceding consonant (Pastätter & Pouplier, 2017) and that there are consonant-specific timing patterns (Brunner et al., 2014). Therefore, conducting the study with different target words could lead to different results.

Another limitation is the small number of participants, which is common in articulatory studies, but may affect the generalizability of the findings.

5. Conclusion

This study aimed to shed light on the underlying mechanisms of speech motor control to contribute to our theoretical understanding of speech production but also to a better understanding of stuttering. It is the first study to investigate multisensory aspects in speech timing by including Metronome and Tapping conditions. In conclusion, this study provides evidence for the CV-timing hypothesis for stuttering as we found differences in inter-gestural timing between adults who stutter and adults who do not stutter, pointing towards closer CV coupling in PWS. Furthermore, we found predictive timing differences in the perceptually fluent speech of adults who stutter since PWS started speaking later when synchronizing to a metronome than PWNS. The groups did

not differ in timing their speech to their own finger tapping but appear to prefer different cues during the auditory-motor pacing condition. We propose that this difference might stem from inter-gestural timing differences. This is a novel aspect, highlighting that there are fundamental differences in how PWS and PWNS integrate sensory information for speech-motor coordination. While PWNS appear to rely more on auditory cues (metronome beat), PWS lean more towards tactile information (finger tapping). Our findings pave the way for future studies that could address the effects of (auditory-)motor-pacing on the speech motor system of PWS on a neural basis.

Conflict of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

CRediT authorship contribution statement

Mona Franke: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Simone Falk:** Writing – review & editing, Supervision, Funding acquisition. **Nicole Benker:** Methodology, Data curation. **Phil Hoole:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Data curation.

Data Availability

Scripts and data can be found at https://osf.io/tvwx6/?view_only=873282b2394040e4b5d7603535ecbbd8.

Acknowledgments

We thank all our participants for their participation in this study. Special thanks go to Charlie Wiltshire for helping with data collection, to Henry Derwanz for helping with data segmentation, and to Magdalena Saumweber for help with assessing stuttering severity. We are also grateful to Michele Gubian for valuable discussions on statistical analyses. Finally, we thank the editor and three anonymous reviewers for their insightful and constructive feedback, which helped improve the manuscript.

Funding

This work was supported by DFG grants [FA 901/4-1 and HO 3271/6-1], the Graduate School "Class of Language" of LMU München, the Bavarian Research Alliance, as well as the German Academic Research Exchange Service.

Appendix

Table A1
Tukey corrected pairwise comparisons of the normalized CV-lag.

Row no.	Comparison	Estimate	Standard Error	Degrees of Freedom	t-ratio	p-value
1	Unpaced PWNS − PWS	−0.003	0.0357	21.6	−0.083	0.9348
2	Tapping PWNS − PWS	0.0283	0.0357	21.6	0.792	0.4370
3	Metronome PWNS − PWS	0.0439	0.0355	21.2	1.237	0.2296
4	Metronome+Tapping PWNS − PWS	0.0636	0.0356	21.3	1.788	0.0881
5	PWS Combined vs. Metronome	−0.0202	0.0118	1730	−1.711	0.3180
6	PWS Combined vs. Tapping	−0.0160	0.0123	1732	−1.274	0.5796
7	PWS Combined vs. Unpaced	−0.0348	0.0122	1734	−2.852	**0.0228**
8	PWS Metronome vs. Tapping	0.0045	0.0123	1731	0.370	0.9828
9	PWS Metronome vs. Unpaced	−0.0143	0.0121	1732	−1.200	0.6269
10	PWS Tapping vs. Unpaced	−0.0191	0.0125	1731	−1.524	0.4233
11	PWNS Combined vs. Metronome	−0.0006	0.0115	1730	−0.053	0.9999

(*continued on next page*)

Table A1 (*continued*)

Row no.	Comparison	Estimate	Standard Error	Degrees of Freedom	t-ratio	p-value
12	PWNS Combined vs. Tapping	0.01961	0.0115	1730	1.703	0.3224
13	PWMS Combined vs. Unpaced	0.03178	0.0117	1730	2.724	**0.0329**
14	PWNS Metronome vs. Tapping	0.02021	0.0114	1730	1.769	0.2886
15	PWNS Metronome vs. Unpaced	0.0324	0.0116	1730	2.797	**0.0267**
16	PWNS Tapping vs. Unpaced	0.0122	0.0116	1730	1.049	0.7203

R syntax for the GAMM

acf_model <- bam(pos ~ ConditionGroup + s(time, by=ConditionGroup) + s(time,Subject,by = word,bs="fs",m = 1), data = data, discrete = TRUE).

autocor_acf <- acf_resid(acf_model).

final_model <- bam(pos ~ ConditionGroup + s(time, by=ConditionGroup) + s(time,Subject,by = word,bs="fs",m = 1), data = data, rho = autocor_acf[2], AR.start = data$begin, discrete = TRUE).

References

Alm, P. A. (2021). The dopamine system and automatization of movement sequences: A review with relevance for speech and stuttering. *Frontiers in Human Neuroscience, 15*. https://doi.org/10.3389/fnhum.2021.661880 661880.

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.

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition, 48*(1), 66–79. https://doi.org/10.1006/brcg.2001.1304.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bishop, J. H., Williams, H. G., & Cooper, W. A. (1991). Age and task complexity variables in motor performance of stuttering and nonstuttering children. *Journal of Fluency Disorders, 16*(4), 207–217. https://doi.org/10.1016/0094-730X(91)90003-U.

Blood, G. W., Ridenour, V. J., Qualls, C. D., & Scheffner Hammer, C. (2003). Co-occurring disorders in children who stutter. *Journal of Communication Disorders, 36*(1), 427–448. https://doi.org/10.1016/S0021-9924(03)00023-6.

Bloodstein, O. (1995). *A handbook on stuttering*. San Diego: Singular.

Boersma, P. & Weenink, D. (2019). Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1. 2019. Available online: http://www.praat.org/ (last accessed 03/01/2024).

Brady, J. P. (1969). Studies on the metronome effect on stuttering. *Behaviour Research and Therapy, 7*(2), 197–204. https://doi.org/10.1016/0005-7967(69)90033-3.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251. https://doi.org/10.1017/s0952675700001019.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155–180.

Brunner, J., Geng, C., Sotiropoulou, S., & Gafos, A. (2014). Timing of German onset and word boundary clusters. *Laboratory Phonology, 5*(4), 403–454.

Carstens Medizinelektronik GmbH (2014). AG501 Manual. Retrieved from http://www.ag500.de/manual/ag501/ag501-manual.pdf, (last accessed 09/29/24).

Chang, S.-E., Horwitz, B., Ostuni, J., Reynolds, R., & Lodlow, C. (2011). Evidence of left inferior frontal-premotor structural and functional connectivity deficits in adults who stutter. *Cerebral Cortex, 21*, 2507–2518.

Chang, S. E., Garnett, E. O., Etchell, A., & Chow, H. M. (2019). Functional and Neuroanatomical Bases of Developmental Stuttering: Current Insights. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 25*(6), 566–582. https://doi.org/10.1177/1073858418803594.

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.

Chon, H., Jackson, E. S., Kraft, S. J., Ambrose, N. G., & Loucks, T. M. (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. https://doi.org/10.1044/2021_JSLHR-20-00606.

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.

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S.-E. (2017). Auditory–motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science, 21*(2). https://doi.org/10.1111/desc.12521 e12521.

Davidow, J. H., Bothe, A. K., Andreatta, R. D., & Ye, J. (2009). Measurement of phonated intervals during four fluency-inducing conditions. *Journal of Speech, Language, and Hearing Research, 52*(1), 188–205. https://doi.org/10.1044/1092-4388(2008/07-0040).

Davidow, J. H. (2014). Systematic studies of modified vocalization: The effect of speech rate on speech production measures during metronome-paced speech in persons who stutter. *International Journal of Language & Communication Disorders, 49*(1), 100–112. https://doi.org/10.1111/1460-6984.12050.

Debrabant, J., Gheysen, F., Vingerhoets, G., & Van Waelvelde, H. (2012). Age-related differences in predictive response timing in children: Evidence from regularly relative to irregularly paced reaction time performance. *Human Movement Science, 31*(4), 801–810. https://doi.org/10.1016/j.humov.2011.09.006.

Dehqan, A., Yadegari, F., Blomgren, M., & Scherer, R. C. (2016). Formant transitions in the fluent speech of Farsi-speaking people who stutter. *Journal of Fluency Disorders, 48*, 1–15. https://doi.org/10.1016/j.jfludis.2016.01.005.

De Nil, L. F. (1995). The influence of phonetic context on temporal sequencing of upper lip, lower lip, and jaw peak velocity and movement onset during bilabial consonants in stuttering and nonstuttering adults. *Journal of Fluency Disorders, 2*, 127–144.

Didirková, I., & Hirsch, F. (2020). A two-case study of coarticulation in stuttered speech. An articulatory approach. *Clinical Linguistics & Phonetics, 34*(6), 517–535. https://doi.org/10.1080/02699206.2019.1660913.

Durvasula, K., & Wang, Y. (2023). Revisiting CV timing with a new technique to identify inter-gestural proportional timing. *Proceedings of the 20th International Congress of Phonetic Sciences*.

DWDS (2024). https://www.dwds.de (last accessed 04/02/24).

Etchell, A. C., Johnson, 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. https://doi.org/10.3389/fnhum.2014.00467.

Falk, S. (in press). Music and stuttering. In: Sammler, D. (Ed.) *The Oxford Handbook of Music and Language*. Oxford University Press.

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

Fischer-Jørgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica, 47*(3–4), 99–140. https://doi.org/10.1159/000261858.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology. General, 112*(3), 386–412. https://doi.org/10.1037/0096-3445.112.3.386.

Franke, M., Hoole, P., & Falk, S. (2023a). Temporal organization of syllables in paced and unpaced speech in children and adolescents who stutter. *Journal of Fluency Disorders, 76*. https://doi.org/10.1016/j.jfludis.2023.1059751. 2. 3 105975.

Franke, M., Benker, N., Falk, S., & Hoole, P. (2023b). Synchronization type matters: Articulatory timing in different rhythmic conditions in persons who stutter. In: Radek Skarnitzl & Jan Volín, *Proceedings of the 20th International Congress of Phonetic Sciences, 3942–3946,* Guarant International.

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, 64*(6S), 2325–2346. https://doi.org/10.1044/2021_JSLHR-20-00328.
Frisch, S. A., Maxfield, N., & Belmont, A. (2016). Anticipatory coarticulation and stability of speech in typically fluent speakers and people who stutter. *Clinical Linguistics & Phonetics, 30*(3–5), 277–291. https://doi.org/10.3109/02699206.2015.1137632.
Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In C. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 239–249). The Commercial Press. ISBN: 978-7-10-006769-0. HAL Id: hal-03127293.
Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96*(3), 280–301. https://doi.org/10.1016/j.bandl.2005.06.001.
Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics, 25*(5), 408–422. https://doi.org/10.1016/j.jneuroling.2009.08.006.
Hall, N. (2010). Articulatory phonology. *Language and Linguistics Compass, 4*(9), 818–830. https://doi.org/10.1111/j.1749-818x.2010.00236.x.
Hardcastle, W. J., & Hewlett, N. (Eds.). (2006). *Coarticulation: Theory, Data and Techniques*. Cambridge, MA: Cambridge University Press.
Harrington, J. M. (1987). Coarticulation and stuttering: An acoustic and electropalatographic study. In H. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering*. New York: Springer Verlag.
Harrington, J. M. (1988). Stuttering, delayed auditory feedback, and linguistic rhythm. *Journal of Speech & Hearing Research, 31*, 36–47.
Heyde, C. J., Scobbie, J. M., Lickley, R., & Drake, E. K. E. (2016). How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound. *Clinical Linguistics & Phonetics, 30*(3–5), 292–312. https://doi.org/10.3109/02699206.2015.1100684.
Hilger, A. I., Zelaznik, H., & Smith, A. (2016). Evidence that bimanual motor timing performance is not a significant factor in developmental stuttering. *Journal of Speech, Language, and Hearing Research, 59*(4), 674–685. https://doi.org/10.1044/2016_JSLHR-S-15-0172.
Hoole, P. (2012). mtnew (https://www.phonetik.uni-muenchen.de/~hoole/articmanual/) (last viewed 01/05/2023).
Hoole, P. (2014). (last viewed 03/09/2025). https://www.phonetik.uni-muenchen.de/~hoole/articmanual/ag501/carstens_workshop_summary_issp2014.pdf.
Howell, P., & Au-Yeung, J. (2002). The EXPLAN theory of fluency control and the diagnosis of stuttering. *Pathology and Therapy of Speech Disorders*. https://doi.org/10.1075/cilt.227.08how.
Hubbard, C. P. (1998). Stuttering, stressed syllables, and word onsets. *Journal of Speech, Language, and Hearing Research, 41*(4), 802–808. https://doi.org/10.1044/jslhr.4104.802.
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.
Jackson, P. J. B., & Singampalli, V. D. (2009). Statistical identification of articulation constraints in the production of speech. *Speech Communication, 51*(8), 695–710.
Jenson, D., Bowers, A. L., Hudock, D., & Saltuklaroglu, T. (2020). The application of EEG Mu rhythm measures to neurophysiological research in stuttering. *Frontiers in Human Neuroscience, 13*, 458. https://doi.org/10.3389/fnhum.2019.00458.
Jessen, M. (1993). Stress-conditions on vowel quality and quantity in German. *Working papers of the Cornell Phonetics Laboratory*, 8, 1–27.
Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language, 45*, 326–347.
Kleinow, J., & Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research, 43*(2), 548–559. https://doi.org/10.1044/jslhr.4302.548.
Klich, R., & May, G. (1982). Spectrographic study of vowels in stutterers' fluent speech. *Journal of Speech, Language, and Hearing Research, 25*, 364–370. https://doi.org/10.1044/jshr.2503.364.
Kramer, B. M., Stern, M. C., Wang, Y., Liu, Y., & Shaw, J. A. (2023). Synchrony and stability of articulatory landmarks in English and mandarin cv sequences. *Proceedings of the 20th International Congress of Phonetic Sciences*.
Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119–159. https://doi.org/10.1037/0033-295X.106.1.119.
Lenoci, G., & Ricci, I. (2018). An ultrasound investigation of the speech motor skills of stuttering Italian children. *Clinical Linguistics & Phonetics, 32*(12), 1126–1144.
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.
Loucks, T. M., Pelczarski, K. M., Lomheim, H., & Aalto, D. (2022). Speech kinematic variability in adults who stutter is influenced by treatment and speaking style. *Journal of Communication Disorders, 96*. https://doi.org/10.1016/j.jcomdis.2022.106194 106194.
Lu, C., Peng, D., Chen, C., Ning, N., Ding, G., Li, K., Yang, Y., & Lin, C. (2010). Altered effective connectivity and anomalous anatomy in the basal ganglia-thalamocortical circuit of stuttering speakers. *Cortex, 46*, 49–67.
Lu, Y., Wiltshire, C. E. E., Watkins, K. E., Chiew, M., & Goldstein, L. (2022). Characteristics of articulatory gestures in stuttered speech: A case study using real-time magnetic resonance imaging. *Journal of Communication Disorders, 97*. https://doi.org/10.1016/j.jcomdis.2022.106213 106213.
Lu, Y., Goldstein, L., & Narayanan, S. (2024). MRI reveals CV coarticulation is preserved in stuttering. *Book of Abstracts of the 13th International Seminar on Speech Production*.
Lüdecke et al. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), 3139 10.21105/joss.03139.
Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics, 30*(3), 247–256. https://doi.org/10.3758/bf03214280.
Maruthy, S., Feng, Y., & Max, L. (2018). Spectral coefficient analyses of word-initial stop consonant productions suggest similar anticipatory coarticulation for stuttering and nonstuttering adults. *Language and Speech, 61*, 31–42. https://doi.org/10.1177/0023830917695853.
MathWorks (2017). *MATLAB (Version R2017.b)*. The MathWorks Inc. https://www.mathworks.com.
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. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0358.
Max, L., & Yudman, E. A. (2003). Accuracy and variability of isochronous rhythmic timing across motor systems in stuttering versus nonstuttering individuals. *Journal of Speech, Language, and Hearing Research, 46*(1), 146–163. https://doi.org/10.1044/1092-4388(2003/012).
Meister, I. G., Buelte, D., Staedtgen, M., Boroojerdi, B., & Sparing, R. (2009). The dorsal premotor cortex orchestrates concurrent speech and fingertapping movements. *The European Journal of Neuroscience, 29*(10), 2074–2082. https://doi.org/10.1111/j.1460-9568.2009.06729.x.
Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. https://doi.org/10.1515/9783110223958.297.
Namasivayam, A. K., & van Lieshout, P. (2008). Investigating speech motor practice and learning in people who stutter. *Journal of Fluency Disorders, 33*(1), 32–51. https://doi.org/10.1016/j.jfludis.2007.11.005.
Natke, U., Sandrieser, P., van Ark, M., Pietrowsky, R., & Kalveram, K. T. (2004). Linguistic stress, within-word position, and grammatical class in relation to early childhood stuttering. *Journal of Fluency Disorders, 29*(2), 109–122. https://doi.org/10.1016/j.jfludis.2003.11.002.
Neef, N. E., & Chang, S. E. (2024). Knowns and unknowns about the neurobiology of stuttering. *PLoS Biology, 22*(2). https://doi.org/10.1371/journal.pbio.3002492 e3002492.
Olander, L., Smith, A., & Zelaznik, H. N. (2010). Evidence that a motor timing deficit is a factor in the development of stuttering. *Journal of Speech, Language, and Hearing Research, 53*(4), 876–886. https://doi.org/10.1044/1092-4388(2009/09-0007).
Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America, 39*(1), 151–168. https://doi.org/10.1121/1.1909864.
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.
Pouplier, M., Lentz, T., Chitoran, I., & Hoole, P. (2020). The imitation of coarticulatory timing patterns in consonant clusters for phonotactically familiar and unfamiliar sequences. *Laboratory Phonology: Journal of the Association for Laboratory Phonology., 11*. https://doi.org/10.5334/labphon.195.
Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz, J. A., Cox, R. W., & Binder, J. R. (1997). Distributed neural systems underlying the timing of movements. *The Journal of Neuroscience, 17*(14), 5528–5535. https://doi.org/10.1523/JNEUROSCI.17-14-05528.1997.
Rathcke, T., Lin, C.-Y., Falk, S., & Dalla Bella, S. (2021). Tapping into linguistic rhythm. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 12*(1), 11.
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/.
Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*(6), 969–992. https://doi.org/10.3758/bf03206433.
Robb, M., & Blomgren, M. (1997). Analysis of F2 transitions in the speech of stutterers and nonstutterers. *Journal of Fluency Disorders, 22*, 1–16. https://doi.org/10.1016/s0094-730x(96)00016-2.
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, 1449*(1), 56–69. https://doi.org/10.1111/nyas.14117.
Savariaux, C., Badin, P., Samson, A., & Gerber, S. (2017). A comparative study of the precision of carstens and northern digital instruments electromagnetic articulographs. *Journal of Speech, Language, and Hearing Research, 60*(2), 322–340. https://doi.org/10.1044/2016_JSLHR-S-15-0223.
Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS99)*, San Francisco, CA, USA, 1–7 August 1999; Ohala, J.J., Hasegawa, Y., Ohala, M., Granville, D., Bailey, A.C., Eds.; University of California: Berkley, CA, USA, pp. 607–610.
Shaw, J. A., & Chen, W. R. (2019). Spatially conditioned speech timing: evidence and implications. *Frontiers in Psychology, 10*, 2726. https://doi.org/10.3389/fpsyg.2019.02726.
Schreier, R., Dalla Bella, S., Hoole, P., & Falk, S. (2020). Verbal timing deficits in stuttering. *Proceedings of the 12th International Seminar on Speech Production (ISSP2020)*.
Schreier, R. (2023). Stuttering and speech-rhythm. Dissertation, LMU München: Fakultät für Sprach- und Literaturwissenschaften. https://doi.org/10.5282/edoc.32998.

Slis, A., Savariaux, C., Perrier, P., & Garnier, M. (2023). Rhythmic tapping difficulties in adults who stutter: A deficit in beat perception, motor execution, or sensorimotor integration? *PLoS One1, 18*(2). https://doi.org/10.1371/journal.pone.0276691 e0276691.

Smith, A., & Weber, C. (2016). Childhood stuttering: Where are we and where are we going? *Seminars in Speech and Language, 37*(4), 291–297. https://doi.org/10.1055/s-0036-1587703.

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics, 84* 101017.

Stager, S. V., Jeffries, K. J., & Braun, A. R. (2003). Common features of fluency-evoking conditions studied in stuttering subjects and controls: An H(2)15O PET study. *Journal of Fluency Disorders, 28*(4), 319–336. https://doi.org/10.1016/j.jfludis.2003.08.004.

Sussman, H. M., Byrd, C. T., & Guitar, B. (2011). The integrity of anticipatory coarticulation in fluent and non-fluent tokens of adults who stutter. *Clinical Linguistics & Phonetics, 25*, 169–186. https://doi.org/10.3109/02699206.2010.517896.

Svensson Lundmark, M., Frid, J., Ambrazaitis, G., & Schötz, S. (2021). Word-initial consonant-vowel coordination in a lexical pitch-accent language. *Phonetica, 78*(5–6), 515–569. https://doi.org/10.1515/phon-2021-2014.

Treffner, P., & Peter, M. (2002). Intentional and attentional dynamics of speech-hand coordination. *Human Movement Science, 21*(5–6), 641–697. https://doi.org/10.1016/s0167-9457(02)00178-1.

Tuller, B., & Fowler, C. A. (1980). Some articulatory correlates of perceptual isochrony. *Perception & Psychophysics, 27*(4), 277–283. https://doi.org/10.3758/bf03206115.

Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech timing: Implications for theories of phonology, phonetics, and speech motor control*. Oxford: Oxford University Press.

Usler, E. R., & Walsh, B. (2018). The effects of syntactic complexity and sentence length on the speech motor control of school-age children who stutter. *Journal of Speech, Language, and Hearing Research, 61*(9), 2157–2167. https://doi.org/10.1044/2018_JSLHR-S-17-0435.

van de Vorst, R., & Gracco, V. L. (2017). Atypical non-verbal sensorimotor synchronization in adults who stutter may be modulated by auditory feedback. *Journal of Fluency Disorders, 53*, 14–25. https://doi.org/10.1016/j.jfludis.2017.05.004.

van Lieshout, P. H., Hulstijn, W., & Peters, H. F. (1996). From planning to articulation in speech production: What differentiates a person who stutters from a person who does not stutter? *Journal of Speech & Hearing Research, 39*(3), 546–564. https://doi.org/10.1044/jshr.3903.546.

van Lieshout, P. H. H. M., & Namasivayam, A. K. (2010). Speech motor variability in people who stutter. In B. Maassen & P. H. H. M. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 191–214). Oxford, England: Oxford University Press.

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2022). itsadug: Interpreting time series and autocorrelated data using GAMMs. *R package version, 2*(4), 1.

Verdurand, M., Rossato, S., & Zmarich, C. (2020). Coarticulatory aspects of the fluent speech of french and Italian people who stutter under altered auditory feedback. *Frontiers in Psychology, 11*, 1745. https://doi.org/10.3389/fpsyg.2020.01745.

Walsh, B., Mettel, K. M., & Smith, A. (2015). Speech motor planning and execution deficits in early childhood stuttering. *Journal of Neurodevelopmental Disorders, 7*(1), 27. https://doi.org/10.1186/s11689-015-9123-8.

Weiner, A. (1984). Stuttering and syllabic stress. *Journal of Fluency Disorders, 9*, 301–305.

WHO (2016). *International Classification of Mental and Behavioral Disorders*. Geneva: WHO (World Health Organisation).

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language, 90*(3), 669–692.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002.

Wiltshire, C. E. E. (2019). *Investigating speech motor control using vocal tract imaging, fMRI, and brain stimulation* [Thesis, University of Oxford].

Wiltshire, C. E. E., Chiew, M., Chesters, J., Healy, M. P., & Watkins, K. E. (2021). Speech movement variability in people who stutter: A vocal tract magnetic resonance imaging study. *Journal of Speech, Language, and Hearing Research, 64*(7), 2438–2452. https://doi.org/10.1044/2021_JSLHR-20-00507.

Wiltshire, C. E. E., Cler, G. J., Chiew, M., Freudenberger, J., Chesters, J., Healy, M., Hoole, P., Watkins, K. E. (2023, April 3). Speaking to a metronome reduces kinematic variability in typical speakers and people who stutter. https://doi.org/10.31219/osf.io/wc29m.

Wingate, M. E. (1988). *The Structure of Stuttering (a Psycholinguistic Analysis)*. New-York, NY: Springer Verlag.

Winter, B. (2020). *Statistics for linguistics: An introduction using R*. New York: Routledge.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B), 73*(1), 3–36.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd edition.). Chapman and Hall/CRC.

Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders, 38*(2), 66–87. https://doi.org/10.1016/j.jfludis.2012.11.002.

Zelaznik, H. N., Smith, A., & Franz, E. A. (1994). Motor performance of stutterers and nonstutterers on timing and force control tasks. *Journal of Motor Behavior, 26*(4), 340–347. https://doi.org/10.1080/00222895.1994.9941690.

Zimmermann, G. (1980). Stuttering: A disorder of movement. *Journal of Speech & Hearing Research, 23*(1), 122–136. https://doi.org/10.1044/jshr.2301.122.