Rhythms of speech
Exploring timing mechanisms
in stuttering

A thesis for a joint doctoral program
submitted for the degree of

Doctor of Philosophy (Dr. phil.) in Phonetics
of
Ludwig-Maximilians-Universität München

and

Doctor of Philosophy (PhD) in Linguistics
of
Université de Montréal

submitted by
Mona Franke

July 2025

Ludwig-Maximilians-Universität München
Fakultät für Sprach- und Literaturwissenschaften

Université de Montréal
Département de linguistique et de traduction, Faculté des arts et des sciences

The cumulative dissertation with the title

Rhythms of speech
Exploring timing mechanisms in stuttering

presented by
Mona Franke

has been evaluated by a jury composed of the following persons

Marianne Pouplier
Président-rapporteur

Simone Falk
Supervisor Université de Montréal

Philip Hoole
Supervisor Ludwig-Maximilians-Universität

Doris Mücke
External examiner

Supervisor Ludwig-Maximilians-Universität: Prof. Dr. Philip Hoole
Supervisor Université de Montréal: Prof. Dr. Simone Falk
Date of Oral Examination: February 6, 2025

Abstract

Smooth articulatory coordination is central to fluent speech, a quality often disrupted in stuttering. Stuttering is a neurodevelopmental speech fluency disorder marked by repetitions, prolongations, and blocks. Interestingly, these disfluencies are reduced when people who stutter (PWS) speak in time with an external rhythm, such as a metronome. However, the role of rhythm in fluent speech production and its effect on speech motor control remains unclear.

This cumulative dissertation explores speech timing mechanisms in PWS and persons who do not stutter (PWNS) across different age groups and rhythmic contexts in perceptually fluent speech. It specifically focuses on articulatory timing at word onsets under paced and unpaced conditions. The work comprises three empirical studies that use acoustic and articulatory data to examine the interaction between rhythmic cues and speech motor control.

Chapter 2 investigates acoustic correlates of a c-center effect—an indicator of onset-vowel timing—in German-speaking children and adolescents with and without stuttering. Participants read monosyllabic words under an unpaced and a metronome-paced condition. Both groups exhibited cues of a c-center effect, but PWS showed greater consonant compression, suggesting differences in the coupling of onsets and vowels. These differences remained in the paced condition, indicating that metronome pacing does not fully normalize articulatory timing in young PWS.

Chapter 3 examines the articulatory timing of verbal and non-verbal gestures in adults who stutter and adults who do not stutter using electromagnetic articulographu (EMA) across three rhythmic conditions: speaking while tapping (Tapping), speaking to a metronome (Metronome), and speaking to a metronome while tapping (Metronome+Tapping). While both groups showed that speech onset preceded rhythmic events (i.e. finger tap, metronome beat), PWS aligned more closely with the metronome, indicating delayed initiation. Interestingly, in the Metronome+Tapping condition, only PWS adjusted their finger tapping to align more closely with speech, suggesting altered integration of auditory and motor cues.

Chapter 4 explores consonant-vowel (CV) timing and predictive timing in adult PWS and adult PWNS via EMA. CV-timing was assessed using both static and dynamic articulatory measures. PWS generally showed greater overlap between CV gestures. Analyses based on the dynamic approach revealed group differences in the unpaced but not rhythmic conditions (Tapping, Metronome, Metronome+Tapping). Moreover, tapping behavior in the Metronome+Tapping condition diverged between groups: PWS tapped closer to speech onset, while PWNS shifted taps toward the metronome beat. These findings support the idea that PWS rely more on internal motor cues (e.g., tapping) than external auditory cues (e.g., metronome) and that their CV-timing aligns more closely with that of PWNS when speaking in rhythmic contexts.

In sum, this dissertation reveals developmental and rhythmic-context-dependent differences in speech motor control between PWS and PWNS. While rhythmic pacing improves CV-timing in adults who stutter, it does not normalize timing in children, pointing to developmental constraints. Furthermore, PWS exhibit altered auditory-motor integration and predictive timing, with a tendency to rely on internally generated cues. These results offer novel insights into articulatory timing in stuttering and contribute to broader models of speech motor control.

Keywords: *Stuttering, articulatory timing, speech motor control, predictive timing, inter-gestural coordination, metronome-paced speech, finger tapping, rhythmic cues, children and adolescents who stutter, adults who stutter.*

Résumé

Une coordination articulatoire fluide est essentielle à la production de parole fluente, une qualité souvent altérée dans le cas du bégaiement. Le bégaiement est un trouble développemental de la fluidité de la parole d'origine neurologique, caractérisé par des répétitions, des prolongations et des blocages. Fait intéressant, ces disfluences diminuent lorsque les personnes qui bégaient (PWS) parlent en synchronie avec un rythme externe, tel qu'un métronome. Cependant, le rôle du rythme dans la production de parole fluide et son influence sur le contrôle moteur de la parole reste mal compris.

Cette thèse cumulative explore les mécanismes du timing de la parole chez les PWS et les personnes qui ne bégaient pas (PWNS) à différents âges et dans divers contextes rythmiques, lors de la production de parole en apparence fluide. Elle porte spécifiquement sur le timing articulatoire au début des mots, dans des conditions avec et sans rythme. Le travail comprend trois études empiriques qui s'appuient sur des données acoustiques et articulatoires pour examiner l'interaction entre les indices rythmiques et le contrôle moteur de la parole.

Le chapitre 2 s'intéresse aux indices acoustiques d'un effet « c-center » – un indicateur du timing entre l'attaque du mot et la voyelle – chez des enfants et adolescents germanophones qui bégaient ou non. Les participants ont lu des mots monosyllabiques avec et sans métronome. Les deux groupes ont présenté des signes d'un effet c-center, mais les PWS ont montré une compression consonantique plus marquée, suggérant des différences dans le couplage entre attaque et voyelle. Ces différences persistent en condition rythmée, ce qui indique que parler avec un métronome ne normalise pas totalement le timing articulatoire chez les jeunes PWS.

Le chapitre 3 examine le timing articulatoire des gestes verbaux et non verbaux chez des adultes PWS et PWNS à l'aide de l'articulographie électromagnétique (EMA), dans trois conditions rythmiques : parler et taper avec le doigt (condition *Tapping*), parler avec un métronome (condition *Metronome*), et parler et taper avec un métronome (condition *Metronome+Tapping*). Dans l'ensemble, le début de la parole précédait les événements rythmiques (tapement de doigt, battement du métronome), mais les PWS alignaient leur parole plus près du métronome, indiquant une initiation plus tardive. Notamment, dans la condition *Metronome+Tapping*, seuls les PWS ajustaient leurs tapements de doigt pour les aligner davantage avec le début de la parole, ce qui suggère une intégration modifiée des indices auditifs et moteurs.

Le chapitre 4 explore le timing consonne-voyelle (CV) ainsi que le timing prédictif chez des adultes PWS et PWNS, toujours via l'EMA. Le timing CV a été analysé à l'aide de mesures articulatoires statiques et dynamiques. Les PWS montraient globalement davantage de chevauchement entre les gestes consonne et voyelle. L'analyse dynamique a révélé des différences significatives entre les groupes dans la condition *Unpaced* (parler seulement), mais pas dans les conditions rythmiques (*Tapping*, *Metronome*, *Metronome+Tapping*). En outre, les comportements de tapement divergeaient dans la condition *Metronome+Tapping* : les PWS tapaient plus près du début de la parole, alors que les PWNS alignaient davantage leurs tapements sur le battement du métronome. Ces résultats soutiennent l'idée selon laquelle les PWS s'appuient davantage sur des indices moteurs internes (comme le tapement de doigt) que sur des indices auditifs externes (comme le métronome), et que leur timing CV devient plus proche de celui des PWNS lorsqu'ils parlent dans un contexte rythmique.

En résumé, cette thèse met en lumière des différences développementales et dépendantes du contexte rythmique dans le contrôle moteur de la parole entre PWS et PWNS. Bien que le rythme améliore le timing CV chez les adultes qui bégaient, il ne suffit pas à normaliser ce timing chez les enfants, ce qui suggère des contraintes développementales. Par ailleurs, les PWS présentent une intégration auditivo-motrice et un timing prédictif altérés, avec une tendance à se fier davantage à des indices internes. Ces résultats offrent un nouvel éclairage sur le timing articulatoire dans le bégaiement et enrichissent les modèles théoriques du contrôle moteur de la parole.

Mots-clés : *Bégaiement, timing articulatoire, contrôle moteur de la parole, timing prédictif, coordination inter-gestuelle, parole avec un métronome, tapement du doigt, indices rythmiques, enfants et adolescents qui bégaient, adultes qui bégaient.*

Zusammenfassung

Eine reibungslose artikulatorische Koordination bildet die Grundlage für eine flüssige Sprachproduktion – eine Fähigkeit, die beim Stottern häufig beeinträchtigt ist. Stottern ist eine neuroentwicklungsbedingte Störung der Sprechflüssigkeit, die durch Wiederholungen, Dehnungen und Blockaden gekennzeichnet ist. Interessanterweise verringern sich diese Unflüssigkeiten, wenn Personen, die stottern (PWS), im Takt mit einem externen Rhythmus, wie einem Metronom, sprechen. Dennoch ist die Rolle des Rhythmus bei der flüssigen Sprachproduktion und seine Auswirkung auf die motorische Sprachsteuerung bislang nicht vollständig geklärt.

Diese kumulative Dissertation untersucht Mechanismen des Sprech-Timings bei PWS und Personen, die nicht stottern (PWNS), in unterschiedlichen Altersgruppen und rhythmischen Kontexten in perzeptiv flüssiger Sprache. Im Fokus steht dabei insbesondere das artikulatorische Timing an Wortanfängen in rhythmischen Kontexten und beim einfachen Sprechen. Die Arbeit umfasst drei empirische Studien, in denen akustische und artikulatorische Daten genutzt werden, um die Wechselwirkung zwischen rhythmischen Kontexten und der motorischen Sprachsteuerung zu analysieren.

Kapitel 2 untersucht akustische Korrelate eines sogenannten *c-center* Effekts – eines Indikators für das Timing zwischen Konsonantenbeginn und Vokal – bei deutschsprachigen Kindern und Jugendlichen mit und ohne Stottern. Die Teilnehmenden lasen einsilbige Wörter sowohl in einer Bedingung mit und ohne Metronom. Beide Gruppen zeigten Hinweise auf einen c-center Effekt, jedoch wiesen PWS eine stärkere Konsonantenkompression auf, was auf Unterschiede in der Kopplung von Wortanfängen und Vokalen hinweist. Diese Unterschiede blieben auch in der Metronom Bedingung bestehen, was darauf hindeutet, dass metronombegleitetes Sprechen das artikulatorische Timing bei jungen PWS nicht vollständig normalisiert.

Kapitel 3 analysiert das artikulatorische Timing verbaler und nonverbaler Gesten bei Erwachsenen, die stottern und nicht stottern mittels elektromagnetischer Artikulographie (EMA) in drei rhythmischen Bedingungen: Sprechen bei gleichzeitigem Fingertappen (*Tapping*), Sprechen im Takt eines Metronoms (*Metronome*) sowie Sprechen im Metronomtakt mit gleichzeitigem Fingertappen (*Metronome+Tapping*). Beide Gruppen zeigten, dass der Sprachbeginn den rhythmischen Ereignissen (z. B. Tap, Metronomschlag) vorausging, jedoch orientierten sich PWS stärker am Metronom, was auf eine verzögerte Initiation hinweist. Bemerkenswert ist, dass in der *Metronome+Tapping* Bedingung nur PWS ihre Fingertaps an den Sprachbeginn anpassten, was auf eine veränderte Integration auditiver und motorischer Hinweise schließen lässt.

Kapitel 4 untersucht das Timing von Konsonant-Vokal-Sequenzen (CV-Timing) sowie prädiktives Timing bei erwachsenen PWS und PWNS mithilfe von EMA. Das CV-Timing wurde anhand statischer und dynamischer artikulatorischer Maße erfasst. PWS zeigten insgesamt eine stärkere Überlappung der CV-Gesten. Die auf dynamischen Analysen basierenden Ergebnisse ergaben Gruppenunterschiede beim einfachen Sprechen, jedoch nicht in den rhythmischen Bedingungen (*Tapping*, *Metronome*, *Metronome+Tapping*). Darüber hinaus unterschied sich das Tapping-Verhalten in der *Metronome+Tapping* Bedingung zwischen den Gruppen: PWS tappten näher am Sprachbeginn, während PWNS ihre Fingertaps eher zum Metronomschlag verschoben. Diese Ergebnisse stützen die Annahme, dass PWS stärker auf intern generierte motorische Hinweise (z. B. Tapping) als auf externe auditive Hinweise (z. B. Metronom) zurückgreifen, und dass ihr CV-Timing in rhythmischen Kontexten dem von PWNS ähnelt.

Zusammenfassend zeigt diese Dissertation entwicklungs- und kontextabhängige Unterschiede in der motorischen Sprachsteuerung zwischen PWS und PWNS auf. Während rhythmische Vorgaben das CV-Timing bei Erwachsenen, die stottern verbessern, reicht dies bei Kindern nicht aus, um das Timing zu normalisieren – was auf entwicklungsbedingte Faktoren hinweist. Darüber hinaus zeigen PWS eine veränderte Integration auditiver und motorischer Informationen sowie ein abweichendes prädiktives Timing, mit einer Tendenz zur stärkeren Nutzung interner Steuerungssignale. Die Ergebnisse liefern neue Erkenntnisse zum artikulatorischen Timing beim Stottern und tragen zu umfassenderen Modellen der Sprechmotorik bei.

Stichwörter: *Stottern, Artikulatorisches Timing, Sprechmotorische Steuerung, Prädiktives Timing, Intergestische Koordination, metronombegleitete Sprache, Finger tapping, rhythmische Hinweise, Kinder und Jugendliche die stottern, Erwachsene, die stottern.*

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

CV	consonant vowel
CVC	consonant vowel consonant
CCVC	consonant consonant vowel consonant
EEG	electroencephalography
EMA	electromagnetic articulography
GAMMs	generalized additive mixed models
Hz	Hertz
IF	index finger
kHz	Kilo Hertz
LA	lip aperture
LL	lower lip
LMM	linear mixed models
PWS	persons who stutter
PWNS	persons who do not stutter
RSD	relative standard deviation
SD	standard deviation
SSI-4	Stuttering Severity Instrument – fourth edition
TB	tongue back
TM	tongue mid
TT	tongue tip
UL	upper lip
VOT	voice onset time

Für Claudio

Acknowledgments

There is this popular saying, “it takes a village to raise a child”. It also takes a village to write a thesis. And I speak from experience that to do both at the same time (at least for a period) it takes a metropolis.

This work would not have been possible without the help and support of so many people, and I would like to specifically thank them.

First of all, I am grateful to my supervisors Simone Falk and Phil Hoole, for their continuous support, guidance, and helpful advice throughout the past years.

Simone, you live up to your role as a “Doktormutter”! Thank you for always being there for me and making me get the best out of myself.

Phil, thank you for being the haven of peace that you are and for your trust in me.

I could not have asked for better doctoral parents for this project!

A big thank you also goes to all my hard-working helpers and a special thanks to Nicole and Charlie. Nicole, without you, I would never have been able to collect so much data. You were a great support and put a lot of time into this project, for which I am incredibly grateful. Charlie, thank you for spending many hours recording participants for my EMA experiment.

Thanks to Ramona, who was always on hand with help and advice when her expertise on stuttering was needed.

Miriam, it was a pleasure sharing an office in Munich with you and spending so much time together – inside and outside the walls of Schellingstraße 3. Thanks for everything! I am looking forward to more adventures to come.

Also, of course, a big thank you to the whole IPS family - I am very grateful to have spent a large part of my PhD (and undergraduate and graduate studies) in such a friendly and supportive institute. I would especially like to mention Jonathan Harrington, who was available for numerous discussions, Michele Gubian, who advised us on statistics, and Klaus Jänsch, who always makes sure that everything technical is running smoothly.

Josie and Chantal, a huge thank you for your support and patience on my journey to learn French.

I would also like to thank everybody at BRAMS who made the 1.5 years I spent in Montréal so special and unforgettable.
Special thanks to Simone Dalla Bella, Johanne Davids and Simone Falk and her lab for warmly welcoming me at BRAMS. Mihaela Felezou for teaching me how to conduct an EEG experiment, Alex Nieva for his help and technical support, and Simon Rigoulout for his helpful advice regarding the EEG experiment.
And special thanks to Agnès and Antoine. You played a big part in making my time there so memorable!
I am also grateful for Marie-Noëlle, Mengwan, Samaneh, and Camille who pushed the EEG experiment forward.

I want to express my gratitude to Daniela Sammler, whose invaluable expertise and enthusiasm greatly enriches the EEG project. I am excited about the prospect of continuing our collaboration.

Furthermore, I would like to thank Doris Mücke for being part of the jury and for her helpful comments on this thesis.

Sincere thanks go to all the participants. This work would not have been achievable without you.

This dissertation journey would not have been possible without the unconditional support of my family. Special thanks go to my mother, my mother-in-law and my husband. Mom, without you, I probably would have never gotten into Phonetics. Thank you for guiding me into this direction, and thank you for stepping in to help with childcare whenever you can. Doris, I do not know how I would have gotten through the final phase without your active support. I truly do not think it would have been possible. Thank you for taking such great care of Jasper and me.
Claudio, words cannot describe how grateful I am that you are by my side and support me in whatever I choose to do. I could not ask for a better partner in life.

Preface

There once was a time, when the world stood still due to a pandemic caused by the virus Covid-19. Shops, restaurants and labs were closed, there was only online teaching, and we had curfews. Traveling was nearly impossible, and planes became a rare sight in the sky. It happened to be during the time of my PhD, which was planned as a cotutelle between the Ludwig-Maximilians-Universität München in Germany (Phonetics) and the Université de Montréal in Canada (Linguistics). I had intended to focus on speech articulation in Munich and on Neurolinguistics in Montreal.

Living in Germany, I was lucky enough to be able to fly to Montreal in December 2020 in one of the rare planes that actually transported passengers to do the linguistics part of my dissertation. I arrived in Montreal in the deepest winter with a plan of a neuro study in mind and 2 suitcases at hand and a 2-week quarantine period ahead to make sure I was not infected. Everything went well.

I worked on an EEG study that aimed to investigate the role of rhythm in the interaction of speech perception and production in persons who stutter and persons who do not stutter.

Unfortunately, given the pandemic-related circumstances and respective delays, it was not possible to prepare and conduct an EEG experiment, analyze the data and write everything up in a paper. But an article is in preparation and I attached the study protocol in the Appendix A to provide an overview of work that I spent a lot of time working on and was a huge part of my PhD time, specifically in Montreal.

Despite these challenges, I am proud to present the following work, which is based on the studies conducted in Munich, with a focus on speech articulation in persons who stutter and persons who do not stutter and embodies the dedication and perseverance that have defined my PhD journey.

Chapter 1

1. General Introduction

Speaking is one of the most accessible and effective tools to communicate, to express our feelings and emotions, to convey information, to persuade others, to entertain audiences, to build relationships, to share knowledge, to create connections in our personal and professional lives and so much more. When we speak, a rhythmic flow arises from various components, such as the rise and fall of pitch, the speech rate or pauses. Thereby, the smooth coordination of articulatory movements plays a primary role in perceiving speech as rhythmic. This becomes particularly evident when listening to the speech of individuals with speech motor disorders, such as persons who stutter (PWS). In stuttering, the natural rhythm of speech is involuntarily disrupted by alterations in coordinating articulatory movements. These alterations manifest themselves as repetitions of sounds or syllables, like "su su super", prolongations of sounds, such as "sssssuper", and blocks which refer to the temporary inability to initiate a speech sound "---super" (WHO, 2016).

Although it is known that certain conditions, such as speaking along with an external rhythm, like a metronome, can tremendously enhance speech fluency in PWS (e.g., Andrews et al., 1982), the role of rhythm during fluent speech production is still poorly understood. Studying populations who have a speech motor disorder, like PWS, provides valuable insights into the motor mechanisms that underlie fluent speech. Investigating the perceptually fluent speech of PWS, offers a promising opportunity to better understand these underlying mechanisms.

Therefore, the substance of the present thesis is to explore articulatory timing in the fluent speech of individuals who stutter and individuals who do not stutter and how articulatory timing is influenced by various rhythmic contexts, such as a metronome and finger tapping. The primary goal of this work is to contribute to the understanding of speech timing mechanisms in fluent speech.

The next section introduces the terminology of rhythm and timing by providing a brief overview and putting the terms in context of the dissertation at hand. The following sections outline the

theories of speech production that are crucial for understanding the temporal structure of speech movements as well as the underlying articulatory and neurological processes of speech production. Subsequently, relevant research on stuttering will be outlined.
Lastly, the objectives of the three empirical studies (*Chapters 2*, *3*, and *4*) of this dissertation will be deduced from the introduced concepts and gaps in research to be addressed. *Chapters 2*, *3*, and *4* all start with a short introduction before the respective article is presented and conclude with a brief discussion following the article.

1.1. Rhythm and Timing

Our everyday life is full of rhythm. From the alarm tone that wakes us, to the repetitive, circular motions of brushing our teeth, to the beating of our hearts as we rush through the day. Though less obvious, rhythm is also central to speech. When we think of "rhythm", most of us likely associate it with music, dancing, or the steady nodding of our heads to a beat. But what happens when rhythm breaks down? In music, a slightly delayed beat can disrupt the flow, making the melody feel off-balance or arhythmic. In this case, the timing of the beat was not right. Hence, even small variations in timing can significantly affect how and whether rhythm is perceived. We probably all know the phrase "perfect timing", referring to events that occur at just the right moment. But what does "perfect timing" mean in relation to speech? In the general discussion of this thesis, I will address this question on the basis of the results presented in *Chapters 2*, *3*, and *4*.
First of all, it should be emphasized that speaking is a highly variable process. No sound or utterance we produce can ever be produced in exactly the same way again. But specific temporal patterns do occur in a regular manner in speech, that lead to the flow of speech sounds and perceiving speech as rhythmic (e.g., Poeppel & Assaneo, 2020). Still, defining rhythm in the context of speech presents a significant challenge due to the multifaceted nature of it (e.g., for an overview, see Turk & Shattuck-Hufnagel, 2013). Therefore, we are aware of the broad scope and complexity of this topic and do not intend to provide a strict definition of rhythm and timing. Instead, this section aims to outline the specific aspect of speech rhythm that this thesis focuses on.

One key element in understanding the rhythm in speech is the amplitude envelope (Giraud & Poeppel, 2012; Hickock & Poeppel, 2007; Kotz et al., 2018; Poeppel & Assaneo, 2020). It represents the continuous curve of the waveform of an acoustic speech signal that is characterized by the rising and falling patterns of the signal amplitude, i.e. sound pressure

(Poeppel & Assaneo, 2020) and reflects temporal variations of spectral energy (Ahissar et al., 2001). Particularly in the frequency range between two and eight Hertz (Hz), the amplitude envelope of speech carries rhythmic aspects of speech by creating robust regularities. Specifically, this frequency range corresponds to the syllabic rate, as well as pauses (Giraud & Poeppel, 2012; Hickock & Poeppel, 2007; Kotz et al., 2018; Poeppel & Assaneo, 2020). Furthermore, it captures stress patterns, highlighting which syllables were produced stressed and unstressed (Giraud & Poeppel, 2012; Hickock & Poeppel, 2007; Kotz et al., 2018; Poeppel & Assaneo, 2020). Together, these features contribute to perceiving languages across the world as rhythmic (Poeppel & Assaneo, 2020). Poeppel and Assaneo (2020) refer to the frequency range between two and eight Hz as "mesoscale of speech" and describe it as an intermediate temporal structure characterized by highly regular temporal patterns that occur across diverse languages. These highly regular patterns result from three domains that are closely intertwined, namely the acoustic, articulatory, and linguistic domain (Poeppel & Assaneo, 2020): the dynamic interplay between different articulators, such as the jaw and the lips, creates a rhythmic pattern by opening and narrowing the vocal tract over time to produce a word or an utterance (articulatory domain). These articulatory movements play a significant role in shaping the speech amplitude (acoustic domain) and the resulting syllable duration and rate (linguistic domain). The precise timing and coordination of articulatory movements can therefore be defined as the basis of speech rhythm (Poeppel & Assaneo, 2020).

This description of the mesoscale aligns with the "B-Prosodie" in Tillmann's & Mansell's (1980), "ABC-Prosodie" framework. Tillmann & Mansell (1980) distinguish between three characteristic timescales: the intonational contour ("A-Prosodie"), the syllabic rhythm ("B-Prosodie"), and the microstructure of the syllable ("C-Prosodie"). "A-Prosodie" refers to the fluctuation of the sound characteristics on a macrolevel, such as the pitch level, which can be tracked continuously. The typical time frame for "A-Prosodie" aligns with that of a breath group, usually lasting well over a second. In contrast, "B-Prosodie" is not continuously trackable, but countable as one can count for example the number of syllables. The time window for "B-Prosodie" corresponds to the cyclic opening and narrowing of the vocal tract during vowel-consonant alternation, occurring a few times per second. "C-Prosodie" pertains to phenomena that, due to their temporal characteristics, create distinct auditory qualities. These phenomena occur approximately 5 to 30 times per second. An example is the production of a trill, which, despite involving cyclic opening and closing of the vocal tract (similar to syllabic structuring), is qualitatively perceived as entirely different from a syllable due to its modulation speed (Pompino-Marschall, 1995; Tillmann & Mansell, 1980).

This thesis is situated within the context of the mesoscale or so-called “B-Prosodie” of speech. Accordingly, we address the rhythmic aspects of speech at the syllable level, encompassing all three dimensions: acoustics, articulation, and linguistics.

In the literature, speech rhythm in this range is often further categorized into two distinct types: Contrastive rhythm and coordinative rhythm (Kotz et al., 2018; Nolan & Jeon, 2014). Contrastive rhythm describes the alternation of strong and weak speech elements, such as stressed and unstressed syllables, without necessarily adhering to an objective temporal pattern, such as the regular occurrence of these elements in equal time intervals that are measurable (Kotz et al., 2018; Nolan & Jeon, 2014; White & Malisz, 2020). Instead, speech elements, such as syllables or feet can be lengthened or reduced to create contrast (White & Malisz, 2020). Coordinative rhythm is often referred to as the temporal or periodic view of rhythm and is therefore intertwined with isochrony (Nolan & Jeon, 2014). Isochrony describes the concept of regularly occurring speech elements (e.g., [stressed] syllables, moras) that approximately occur at equal time intervals (Lehiste, 1977; Nolan & Jeon, 2014; Pike, 1945).

The term isochrony is associated with the idea that languages across the world exhibit diverse rhythms which has been a long-debated topic in linguistics. Early typological models of speech rhythm were established by Pike (1945) and Abercrombie (1967). Pike introduced the concepts of "syllable-timed" and "stress-timed" rhythms. Abercrombie expanded on this by suggesting that linguistic rhythm is based either on the isochrony of syllables where syllables recur at equal time intervals, as in syllable-timed languages like French, Telugu, and Yoruba, or on the isochrony of interstress intervals where stressed syllables occur roughly at equal time intervals, as in stress-timed languages such as Russian, English, and Arabic (Abercrombie 1967). For languages like Japanese or Tamil, the concept of mora-timed rhythm was introduced (Bloch, 1950; Han 1962; Ladefoged, 1975). In these languages, every mora is perceived to take up an equal amount of time, resulting in a consistent rhythmic pattern.

However, Grabe & Low (2002) demonstrated that the distinction between traditional rhythm types such as syllable- and stress-timed languages is more gradual than categorical. Other researchers, such as Nolan and Jeon (2014) and Arvaniti (2009) have questioned the existence of distinct rhythm classes. They noted that evidence for isochrony is lacking, which means that languages do not follow this strict pattern of equally timed intervals. Moreover, metrics, like the pairwise variability index introduced by Grabe & Low (2002), would primarily reflect timing in the sense of durational variability between vowels and consonants rather than rhythm itself (Nolan & Jeon, 2014; Arvaniti, 2009). For a summary on different rhythm metrics see Turk & Shattuck-Hufnagel (2013). However, the observable timing patterns of consonants and vowels on the acoustic level, that are captured with rhythm metrics, represent only an indication of the temporal coordination of the underlying articulatory timing (White & Malisz, 2020). Therefore, timing is a prerequisite of rhythm which is then creating contrasts and durational successions. Here, we focus on articulatory timing. More specifically, the focus will be on the fundamental basis of speech rhythm: the coordination of articulatory movements that are sequentially executed at fairly consistent time intervals (Poeppel & Assaneo, 2020). The following section outlines the framework used to examine these core elements of linguistic rhythm.

1.2. Articulatory Timing

This thesis is grounded in the framework of *Articulatory Phonology* (Browman & Goldstein, 1986; Browman & Goldstein, 1992; Browman & Goldstein, 1995; for a summary see for example, Turk & Shattuck-Hufnagel, 2020) which models speech as a series of coordinated articulatory movements. In the *gestural computational model*, a key component in Articulatory Phonology, these coordinated movements of articulatory organs, such as the tongue, the lips, and the jaw, are referred to as *gestures*. Gestures are abstract entities that initiate the building and release of a constriction within the vocal tract with a set of articulators, a specific constriction location, and constriction degree (Browman & Goldstein, 1990; Browman & Goldstein, 1991; Browman & Goldstein, 1992; Browman & Goldstein, 1995; Saltzman & Munhall, 1989). The dynamic approach in this model combines discrete linguistic categories, such as syllables, with continuous articulatory movements. With a dynamic specification of a gesture for which the start and end-points are predefined, there is no need for a specific timeline to define articulatory movement. Rather, it is specified, how fast the goal position should be reached, but it has not to be planned from the speaker step-by-step. A gestural score contains information about the gestural activation as well as coordinative patterns (Browman & Goldstein, 1991; Browman & Goldstein, 1992; Saltzman, 1991).

The interaction between gestures is known under the term *coarticulation* (Öhman, 1966) which, more specifically, refers to the phenomenon when articulatory movements of one speech sound overlap with those of adjacent sounds, thus creating smooth transitions between them.

The timing between two independent gestures is defined as inter-gestural coupling and the timing within one gesture is characterized as intra-gestural timing. Intra-gestural timing describes, for instance, the coordination of articulatory movements involved in producing a single gesture, like lowering the jaw to produce an open vowel. In contrast, inter-gestural timing refers to the coordination of articulators involved in the production of different gestures (timing between gestures), for example, closing the lips for producing a bilabial consonant and moving the tongue backwards for producing a back vowel. In this thesis, we focus on the timing of separate gestures, hence inter-gestural timing.

Investigating inter-gestural timing in speech is crucial for understanding how the precise coordination of different articulatory movements contributes to the rhythmic flow of speech. But inter-gestural timing between verbal and non-verbal gestures is also an interesting area of investigation, given that non-verbal gestures are closely linked with speech rhythm (e.g., for an overview, Wagner et al., 2014). Exploring this timing relationship can therefore reveal how articulatory timing is affected by non-verbal gestures. For this reason, this thesis also investigates inter-gestural timing between verbal and non-verbal movements. Specifically, *Chapter 2* addresses inter-gestural timing in speech, whereas *Chapter 3* investigates inter-gestural timing between verbal and non-verbal (finger-tapping) gestures. *Chapter 4* combines both and examines inter-gestural timing in speech and inter-gestural timing between verbal and non-verbal (finger-tapping) gestures.

Gestures are coupled to each other in specific ways to describe the timing relationships between consonants and vowels. Assuming that a gesture has an oscillatory structure, having an internal cycle from 0 to 360°, with an onset of the gesture at 0° and the offset of the gesture at 360°, a total of three different phase relationships between gestures can be summarized: in-phase coupling, anti-phase coupling, and eccentric phase coupling (Goldstein, 2011).

In-phase coupling describes the simultaneous initiation of gestures with a phase difference of 0° (just like in clapping, the left and the right hand move towards and away from each other simultaneously). It is assumed to be the most stable phasing relationship and it is associated with onset consonant and vowel timing (Goldstein et al., 2009; Hall, 2010; Nam et al., 2009). Anti-phase coupling refers to gestures that are timed oppositely to each other, meaning that the gestures begin 180° out of phase from each other (like when marching, one foot is up while the other one is down). This timing relationship can be found in the coupling between vowels and coda consonants, but also in onset consonant clusters (Nam et al., 2009; Goldstein et al., 2009; Hall, 2010; Goldstein, 2011). When the gestures' phasal relation is neither completely in-phase nor anti-phase, it is referred to as eccentric phase coupling resulting in a more variable interaction, often found in consonant clusters in syllable onset or offset position (Goldstein, 2011). Thus, syllable onset clusters can be coupled in both ways, anti-phase and eccentric. In summary, coupling relations help explain patterns of speech production and the organization of phonological units, such as the syllable.

Generally, the internal structure of syllables is asymmetrical, due to the coupling relations between the gestures involved. This asymmetry is crucial for inter-articulatory coordination captured, for example, in the Coupled Oscillator Model of Syllable Structure, see e.g., Goldstein

& Pouplier, 2014). In particular, onset consonant gestures and vowel gestures create a more unified structure than coda consonants with preceding vowel gestures (e.g., Hoole & Pouplier, 2015).

The asymmetry of a syllable underlies the *c-center effect* (Browman & Goldstein, 1992; Browman & Goldstein, 2000) according to which there is a constant temporal relationship between the temporal center of the onset and the following vowel regardless of the number of consonants contained in the onset (Browman & Goldstein, 1988). Due to the underlying timing mechanisms that lead to the c-center effect, syllable onsets are particularly interesting to study. However, one should keep in mind that these predictions of consonant(s)-vowel timing are not always supported by studies (e.g., for a critical review see Ikarous & Pouplier, 2022 and Mücke et al., 2020). Instead, timing patterns are variable, which is particularly evident when comparing different groups of speakers (e.g., younger vs. older or with vs. without pathological speech), as pointed out by Mücke and colleagues (2020). Therefore, this thesis does not aim to impose strict categories like in-phase or anti-phase coupling. Instead, we seek to explore timing patterns across various contexts that place differing demands on the speech motor system, such as speaking with and without an external rhythm. Since synchronizing speech to an external rhythm, for instance, stabilizes inter-gestural timing (Tilsen, 2009), this context should place fewer demands on the speech motor system than speaking without an external rhythm.

A particularly intriguing area of investigation is the timing between onset consonants and vowels, which, as discussed above, marks a key point in syllabic organization. Specifically, *Chapter 2* and *Chapter 4* explore this timing relationship using different methods. For example, *Chapter 2* addresses the *c-center effect* and its manifestation in acoustics by analyzing acoustic recordings. In addition, *Chapter 4* employs an articulatory approach to analyze inter-gestural consonant-vowel (CV)-timing on the basis of electromagnetic articulography (EMA) data. Aspects of this section are therefore also repeated and expanded upon in *Chapters 2* and *4*.

1.3. Neurocomputational Speech Production Models and the Role of Timing

While the Articulatory Phonology framework does not account for interactions between the execution of articulatory gestures (feedforward control) and auditory (or somatosensory[1]) feedback, the *Directions Into Velocities of Articulators* (DIVA) model offers a comprehensive and biologically plausible explanation for how cognitive representations of so-called speech segments[2], such as syllables, are built up, modified, and produced (Guenther, 2003; Guenther et al., 2006). The DIVA model is a neurocomputational model which combines feedforward and feedback control systems and associates specific brain regions, neuron types, or synaptic pathways with different components of speech production (see Figure 1). The model has been continually refined and updated in response to findings from imaging studies (e.g., Guenther et al., 2006; Tourville & Guenther, 2011) and was described as being "the most complete computational model of speech motor control" (Parrell et al., 2019, p. 1463).

The DIVA model primarily focuses on the neural control of speech production on the basis of two control systems, *Feedforward control* and *Feedback control.*

[1] See Parrell et al. (2019), for why Articulatory Phonology can also be situated as a somatosensory feedback-driven system.

[2] To avoid misinterpretation, we use the term "speech segment" instead of "speech sound", the term commonly used in the DIVA literature, to refer to a phoneme, a syllable, and an entire word.

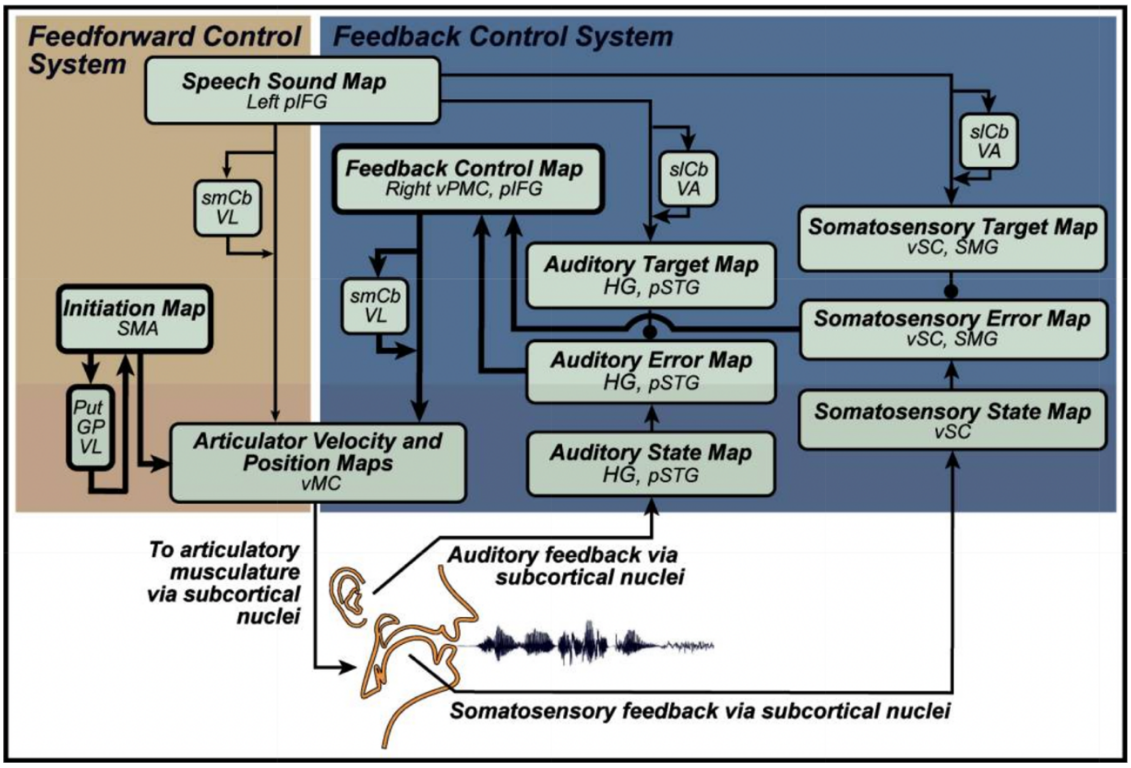


Figure 1: The DIVA model of speech acquisition and production (figure from Tourville and Guenther, 2011 p. 23). GP = globus pallidus; HG = Heschl's gyrus; pIFg = posterior inferior frontal gyrus; pSTg = posterior superior temporal gyrus; Put = putamen; slCB = superior lateral cerebellum; smCB = superior medial cerebellum; SMA = supplementary motor area; SMG = supramarginal gyrus; VA = ventral anterior nucleus of the cerebellum; VL = ventral lateral nucleus of the thalamus; vMC = ventral motor cortex; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex.

The feedforward control system in speech production refers to the generation of motor commands or articulatory gestures for frequently produced speech segments like phonemes, syllables, words, and phrases. These are stored in the *speech sound map* which is hypothesized to be located in the left premotor and adjacent inferior frontal cortex. When the *speech sound map* gets activated in response to the intention to speak, the map sends signals to the primary motor cortex, which controls speech movements. These signals contain instructions on how to execute the speech segments. In other words, the mental representations of speech segments (i.e. phonemes, syllables, words) are turned into a set of *feedforward commands* that produce the respective sound(s) (Civier et al., 2010; Guenther et al., 2006; Meier & Guenther, 2023; Tourville & Guenther, 2011). Additionally, an initiation map, located in the supplementary motor area and involving the basal ganglia, controls the timing of the initiation of the motor program (Tourville & Guenther, 2011). Forward modelling involves both the prediction of articulatory movements and their sensory states, a process known as predictive timing.

Overall, the *feedforward control system* generates speech motor programs for previously learned speech segments. It is suggested that the premotor and primary motor cortex, as well as the basal ganglia and or the cerebellum are involved in encoding the *feedforward motor commands* (Tourville & Guenther, 2011).

Another key component of the DIVA model is *Feedback control.* The *feedback control system* contains information in the auditory and somatosensory target maps about how this sound should sound and feel. When we speak, sensory information is compared with the expected sensory outcomes of the speech sounds that are being produced. If a mismatch occurs between what is intended and what is heard or felt, the system generates adjustments to modify the speech motor commands and improve accuracy (Guenther et al., 2006; Meier & Guenther, 2023; Tourville & Guenther, 2011). To give an example, when a speaker wants to produce the word "mall", and the lips do not fully close to produce the bilabial, thus, not reaching the target position, the incoming sensory feedback (auditory and somatosensory) signals that the upper and lower lips are not in the correct position for producing the intended sound. This deviation from the expected target region (full lip closure) triggers an error signal, which is sent to the feedback control map. Then, motor commands are adapted in order to reach the intended target map, for example, the upper and lower lips should move closer together to achieve the bilabial closure.

Especially the speech motor system of infants relies on the feedback control system as they are in the early stages of learning to coordinate their articulators to achieve the desired sensory feedback (Guenther & Vladusich, 2012). As individuals become more experienced in speaking, there is a shift form relying mainly on feedback systems to depending more on feedforward processing (Guenther et al., 2006; Guenther & Vladusich, 2012). The latter is much faster than feedback control because feedforward control is based on the neural predictions of motor commands that lead to the intended speech segment without waiting for actual feedback. Thus, in feedforward control, the speech motor system relies on learned motor patterns and the anticipation of the outcome of these patterns (Guenther & Vladusich, 2012). This predictive timing process plays a major role in fluent speech production and is discussed in *Chapter 4* in more detail.

With all its components, the DIVA model is able to simulate how sensorimotor interactions are involved in articulator control during speech acquisition and production, encompassing both the simulation of articulatory movements and acoustic characteristics of typical and disordered speech (Civier et al., 2010).

While the DIVA model is only able to simulate speech motor programs for single speech segments, such as phonemes and syllables, an extension of the model, called the Gradient Order DIVA (GODIVA) model, was developed that incorporates mechanisms for the production and parallel planning of multiple speech segments, such as phrases or sentences (Bohland et al., 2010). The GODIVA model places greater emphasis on the temporal coordination of speech movements, ensuring that speech segments are produced in the correct order and with the proper timing. Therefore, a *planning loop* and a *motor loop* were added, whereby the latter is based on the DIVA model. The *planning loop* is responsible for buffering upcoming speech segments enabling parallel planning. The order of a forthcoming speech sequence is selected through an activation gradient. More specifically, through mechanisms of iterative choice and response suppression, the next item for performance is sequentially selected, thereby ensuring a read-out series (Bohland et al., 2010; Meier & Guenther, 2023).

In summary, both the DIVA and the GODIVA model provide a neuro-phonetic framework for understanding the mechanisms involved in speech production, including the timing processes that are crucial for fluent speech, such as relying more on predictive timing mechanisms by pursuing feedforward control instead of relying too much on feedback control. Before introducing a population that struggles with speech fluency, I will conclude this section by outlining the difference in how the neurocomputational model and the Articulatory Phonology framework approach timing.

These two approaches – Articulatory Phonology and (GO)DIVA – differ not only in their use of feedforward and feedback control systems, but also in how they conceptualize speech timing. In Articulatory Phonology, timing is grounded in the coordination of gestures which are continuous and overlapping actions in the vocal tract (e.g., Browman & Goldstein, 1995). In contrast, the (GO)DIVA model conceptualizes speech timing as sequential since speech segments are a activated and produced in a specific order rather than simultaneously (Bohland et al., 2010, Guenther et al., 2006). While the DIVA model also accounts for coarticulation (Guenther, 1995), it does so differently than Articulatory Phonology. Instead of overlapping gestures (Articulatory Phonology), DIVA adjusts target positions of speech sounds with respect to the context, based on experience (Guenther, 1995). In conclusion, Articulatory Phonology has a more continuous view on speech timing, whereas (GO)DIVA treats it as sequential.

1.4. Stuttering

Stuttering is a speech motor and fluency disorder with a neural origin (Smith & Weber, 2016; Watkins et al., 2008). In stuttering, dynamics in speech production are involuntarily disrupted by blocks, repetitions or prolongations of speech sounds, making verbal communication often challenging (WHO, 2016). Characteristic for stuttering, these symptoms are more likely to occur at the onset of words or syllables than at their offset (Bloodstein, 1995; Howell & Au-Yeng, 2002; Hubbard, 1998; Natke et al., 2004; Weiner, 1984), and they are commonly referred to as core stuttering behavior (Van Riper, 1971, 1982). These core symptoms are often accompanied by secondary behavior or associated behavior, including movements such as eye blinks, looking away, muscle tensions, grimacing, or the usage of interjections (Gerlach et al., 2020; Guitar, 2014).

Amongst different types of stuttering, such as neurogenic stuttering (which appears following a brain trauma, caused for example by injuries or diseases of the central nervous system) or psychogenic stuttering (associated with psychological factors like depression or an emotional response to a trauma), developmental stuttering is the most prevalent type arising in early childhood during speech and language acquisition across different cultures and languages (Guitar, 2014). A majority of children are between 24 and 35 months when they begin to stutter (Yairi & Ambrose, 2005) and approximately 75% of children who stutter were younger than six years old at stuttering onset (Andrews, 1985). Stuttering affects about 5-9% of children during childhood and adolescence (e.g., Yairi & Ambrose, 2013) from which many spontaneously recover up to a few weeks to four years after stuttering onset (Yairi & Ambrose, 1999). The opposite of recovered stuttering is persistent stuttering which continues beyond the early years and into adolescence or adulthood (Smith & Weber, 2016). Girls are more likely to (spontaneously) recover from stuttering than boys (Craig & Tran, 2005), leading to a gender bias in the adult population of about four to five men who stutter to one woman who stutters (Bloodstein & Bernstein Ratner, 2008; Guitar, 2014; Smith & Weber, 2017; van Riper, 1971). Approximately 1% of the adult population stutters (Yairi & Ambrose, 2013). According to the International Classification of Diseases (ICD-10), stuttering is a speech fluency disorder characterized by disfluencies that are inappropriate for the individual's age (WHO, 2016). It is classified under behavioral and emotional disorders with an onset typically occurring during childhood or adolescence (WHO, 2016).

Stuttering can tremendously vary between and even within speakers given the variability and heterogeneity of symptoms. It can even vary from day to day or from one conversation to the next, highly depending on social and communicative contexts as well as communicative goals and interlocutors (Bloodstein & Bernstein Ratner, 2008; Gerlach et al., 2020). Moreover, stuttering commonly co-occurs with other speech, language, or non-speech-language disorders, such as a phonological disorder, dyslexia, or ADHD (Arndt & Healey, 2001; Blood et al., 2003; Donaher & Richels, 2012).

Still, despite decades of research, the exact cause and mechanisms of stuttering could not have been identified yet (Smith & Weber, 2016). Nevertheless, numerous studies contribute to a better understanding of the disorder. For example, it has been found that stuttering recurs more frequently within families which raised interest in genetic factors. While it is widely accepted that genetics play a crucial role in the occurrence of stuttering, much work remains to be done to fully understand its genetic basis (for a review, Kang, 2021; Neef & Chang, 2024; Smith & Weber, 2017; Yairi & Ambrose 2013). Furthermore, differences in brain structure and functional patterns as well as brain activity were observed in children and adults who stutter. These neurological alterations are associated with speech motor planning, sensorimotor integration and feedback control, auditory-motor learning, initiation, timing, sequencing, and error monitoring functions (for a review, Chang et al., 2019; Craig-McQuaide et al., 2014; Etchell et al., 2018; Neef & Chang, 2024). Chang and colleagues (2019) summarized that PWS exhibit distinct differences in how they plan and execute self-initiated, intrinsically timed sound sequences. These differences would stem from deficits in neural circuits particularly within the auditory-motor cortical areas of the left hemisphere which are crucial for speech motor planning and execution guided by the sensory context (Chang et al., 2019). Additionally, implicated in these differences, are the basal ganglia-thalamocortical loop and the cerebellum which play a major role in providing the "temporal structure necessary for initiating and timing speech sequences" (Chang et al., 2019).

Given the neurological alterations affecting the speech motor system in PWS, it is not surprising that PWS show kinematic differences compared to persons who do not stutter (PWNS). These kinematic differences suggest that the speech motor system of PWS is less stable, even when their speech appears to be fluent (for a review: Bernstein Ratner & Brundage, 2024; Namasivayam & van Lieshout, 2011; Wiltshire, 2019). For example, PWS show more variability in articulatory movements (e.g., De Nil, 1995; Loucks et al., 2022; Kleinow & Smith, 2000; Smith et al., 2010; Wiltshire et al., 2021) and voice onset times (De Nil & Brutten, 1991; Jäncke, 1994; Max & Gracco, 2005), especially at syllable onsets. There are also coarticulatory differences that either point towards more or less coarticulation in PWS between consonants and vowels (Dehqan et al., 2016; Klich & May, 1982; Lenoci & Ricci, 2018; Robb & Blomgren, 1997; Verdurand et al., 2020). The fact that stuttering symptoms typically occur at word or syllable onsets, and that perceptually fluent speech of PWS also shows temporal differences in these areas compared to PWNS, suggests that syllable onsets pose a significant challenge for PWS.

Chapter 2 highlights the limited research on articulatory characteristics in children and adolescents who stutter. Therefore, the study presented in *Chapter 2* aims to address this research gap by investigating acoustic correlates of the c-center effect in children and adolescents who stutter and children and adolescents who do not stutter. Thus, *Chapter 2* dives deeper into verbal timing differences, with a particular emphasis on the younger population who stutters. In *Chapter 4*, we extend the analysis of verbal timing differences to adults who stutter, focusing specifically on inter-gestural timing of consonant-vowel (CV) gestures.

1.4.1. Sensorimotor synchronization

Timing differences between PWS and PWNS have also been found in the non-verbal domain (i.e., in manual finger tapping tasks), adding to the discussion about whether there are timing alterations which are not speech-specific (Falk et al., 2015; Hulstijn et al., 1992; Olander et al., 2010; Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017). For instance, in an auditory-motor coupling task, PWS synchronized their non-verbal movement such as finger taps earlier to the beat compared to controls (Falk et al., 2015; Olander et al., 2010; Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017). This altered synchronization points towards difficulties with predictive timing, which refers to the anticipation of future events, such as the precise timing of movements (Debarant et al., 2012).

These finger-tapping tasks require the skill of sensorimotor synchronization, which involves the coordination of rhythmic actions with an external auditory or visual stimulus, such as tapping a finger in time to a metronome or a visual cue, dancing to music, or nodding the head to a beat (Repp, 2005; Repp & Su, 2013). Hence, a fine-tuned auditory-motor coupling is necessary for performing these tasks. Studies on sensorimotor synchronization primarily focus on *accuracy* and *consistency* over extended time spans. *Accuracy* refers to how precise a tap (non-verbal task) or the vowel/syllable onset (verbal task) is timed in relation to the pacing event. *Consistency* measures how variable the asynchronies between the synchronization and the pacing event are over multiple repetitions. Studies on non-verbal sensorimotor synchronization report that PWS have a tendency to be less consistent (Falk et al., 2015; Slis et al., 2023) and less accurate in sensorimotor synchronization tasks than PWNS (Falk et al., 2015; Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017).

A verbal sensorimotor synchronization task used in our lab, i.e. synchronizing speech to a metronome, revealed that children and adolescents who stutter initiated their speech later than matched peers relative to the metronome beat (Schreier, 2023; Schreier et al., 2020). This finding supports the hypothesis of altered predictive timing in stuttering as previously found in non-verbal tasks (e.g., Falk et al., 2015). Even though synchronizing speech to a metronome enhances speech fluency for PWS (Andrews et al., 1982) (as further elaborated in the next section), their speech timing remained more variable than that of a matched control group (Schreier, 2023).

Chapter 3 and *Chapter 4* expand on these speech timing differences and explore their articulatory basis. More precisely, we examine speech timing by measuring the asynchrony between the articulatory speech onset and a pacing event under different rhythmic conditions (for more details, refer to *Chapter 3* and *Chapter 4*). By investigating speech production under varying rhythmic conditions, such as externally-paced speech with a metronome, self-paced speech with finger tapping, and a combination of both, we gain valuable insights into general timing mechanisms, including predictive timing and the interaction of the verbal and non-verbal domain.

Importantly, synchronizing speech to rhythmic patterns have long been known to enhance speech fluency in PWS (e.g., Andrews et al., 1982). This creates an opportunity to compare fluent speech across different conditions, such as unpaced speech vs. metronome-paced speech. This comparison offers insights into how these fluency-enhancing conditions affect the speech motor system of PWS, and thereby, how they contribute to fluent speech production. The dissertation at hand leverages this effect. The following section provides an overview of various fluency-enhancing conditions and presents theories that aim at explaining how these conditions improve speech fluency in PWS.

1.4.2. Fluency-enhancing Conditions

There are several conditions that have been reported to enhance speech fluency in PWS, at least for a period of time. For example, whispering (Perkins et al., 1976; Rami et al., 2005), rhythmic speech, often referred to as metronome-paced speech (Brady, 1969; Ingham et al., 2009; Park & Logan, 2015), and singing (for a review, see Falk et al., 2020) have all been shown to reduce stuttering tremendously. Additionally, speaking in unison with another voice, known as choral speech or reading (Ingham et al., 2006, 2009; Park & Logan, 2015; Saltuklaroglu et al., 2009) also decreases stuttering symptoms, even when only the visual feedback of an interlocutor is presented (Kalinowski et al., 2000; Rami et al., 2005; Hudock et al., 2011). Moreover, stuttering is significantly reduced when the voice is masked with white noise (Ingham, 1984, Ingham et al., 2009) or when the auditory feedback is altered, such as delayed or frequency shifted (e.g., Saltuklaroglu et al., 2009; Stuart et al., 2008; or for a review see Howell, 2004). Furthermore, it was also found that fluency-inducing conditions, such as speaking with an external beat, evoke a regulation of neural activity that matches those of PWNS (e.g., see Chang et al., 2019 for a summary).

Building on these observations, various theories have been proposed to explain the underlying mechanisms of fluency-enhancing conditions. For instance, some suggest that external signals, such as chorus reading, singing, and metronome-paced speech, act as an external pacemaker to resynchronize uncoordinated brain activity (Büchl & Sommer, 2004). Others point to the improved efficiency in using auditory information as well as an improved coordination between auditory and motor systems in conditions like singing and synchronizing speech to a metronome (Stager et al., 2003; Frankford et al., 2021).

Despite a long tradition of research investigating the effects of fluency-enhancing conditions on stuttering, a gap remains in understanding their impact on articulatory timing. This thesis addresses this by investigating acoustic and kinematic data of metronome-paced speech and comparing it to unpaced speech. Synchronizing speech to a metronome is one of the most successful fluency-enhancing conditions, with reducing stuttering (almost) completely (e.g., Andrews et al., 1982; Davidow et al., 2009; Davidow et al., 2014). *Chapter 2* and *Chapter 4* specifically examine the effect of metronome-paced speech on articulatory timing.

To better understand the speech motor system of PWS and what leads to the breakdown of speech fluency in PWS, the following section provides an overview of various approaches that have been developed over the past decades.

1.4.3. Theories and models of stuttering

There are two opposing assumptions on the role of sensory feedback in stuttering. On the one hand, some assume that a higher reliance on auditory feedback leads to fluent speech by compensating for a weak feedforward control mechanism (Max et al., 2004, Namasivayam & van Lieshout 2011; van Lieshout et al., 2004). On the other hand, others assume that an overreliance on feedback leads to stuttering, as it causes instabilities in the speech motor system due to the temporal lag between motor commands and the respective sensory feedback (Civier et al., 2013; Civier et al., 2010; Max et al., 2004).

While the interpretation of the role of feedback allows for two positions (maladaptive, which means that altered sensory feedback is contributing to stuttering vs. compensatory, which means that altered sensory feedback is compensating for stuttering) (for an overview, see Bradshaw et al., 2021), most approaches agree on that stuttering is associated with disruptions in the feedforward control.

For example, some research groups (Chang & Guenther, 2020; Civier and colleagues, 2010; Civier and colleagues, 2013; Max and colleagues, 2004; Namasivayam & van Lieshout, 2011; van Lieshout and colleagues, 2004) hypothesize that stuttering is caused by an impairment in the feedforward control. This impairment leads to speech movements that are less automated and efficient suggesting that PWS have limited speech motor skills (Namasivayam & van Lieshout, 2011; van Lieshout et al., 2004). The *Speech Motor Skills* Hypothesis sees speech motor skill as a continuum on which PWS are typically positioned at the lower end, although there is room for variation that allows PWS to exhibit more advanced skills along the continuum approaching the level of PWNS (Namasivayam & van Lieshout, 2011; van Lieshout et al., 2004).

Impairments in the feedforward control could be attributed to a failure of the basal ganglia to initiate the speech sound map cell of the next syllable, caused by elevated dopamine levels, and to some extent also to a failure of the basal ganglia to cancel the activation of the speech sound map choice cell for the current syllable in the case of white matter impairment, as simulations with the GODIVA model revealed (Civier et al., 2013). Thus, a delayed readout of the motor program for the next syllable is hypothesized to lead to stuttering symptoms.

In addition, there are approaches that propose predictive timing difficulties in stuttering. Etchell et al., (2014) for example suggest that brain areas responsible for timing processes are dysfunctional in PWS. Specifically, they note that a deficit in the *Internal Timing Network*, which includes the basal ganglia and the supplementary motor area, contributes to stuttering. An *External Timing Network*, consisting of the cerebellum, the premotor cortex, and the right inferior frontal gyrus, compensates for stuttering by using external timing cues to sequence movements. This approach can therefore also explain why PWS speak more fluently when synchronizing to an external beat as it facilitates predictive timing (Frankford et al., 2021). Particularly, *Chapter 3* discusses the internal and external timing networks with respect to the different rhythmic conditions.

Furthermore, Harrington (1988) proposed a model of stuttering where it is suggested that stuttered speech occurs because PWS make inaccurate temporal predictions about inter-gestural timing in syllable onsets. According to Harrington's (1988) model, PWS anticipate sensory feedback from their articulatory vowel gesture to occur earlier than it actually does. As a result, they attempt to correct this predicted delay by initiating the vowel gesture too early, which would result in stuttering as the CV gestures overlap too much. Accordingly, the feedback-feedforward integration with respect to speech articulation is malfunctioning. Harrington's (1988) model can be translated into DIVA terms by saying that stuttering is caused by erroneous predictions of sensory states. Therefore, the pure anticipation of a mismatch between sensory state and sensory target would cause the feedback controller to send corrective feedback commands. This could in turn lead to prematurely initiating motor commands for the vowel, resulting in a speech fluency breakdown. Even though the erroneous initiation of speech sound sequences in the feedforward control would lead to disruptions in speech, the cause for this lies within the feedback control system.

A related hypothesis stems from Wingate (1988) who suggests that stuttering symptoms occur at the transition from the initial consonant to the stress-bearing vowel. Contrary to Harrington (1988), Wingate (1988) posits that this transition creates a divide and not an overlap, caused by the delayed encoding of the vowel. The approaches by Harrington (1988) and Wingate (1988) are summarized under the term "CV-timing hypothesis" in *Chapter 4*. The "CV-timing hypothesis" therefore refers to an altered timing of onset consonant and vowel gestures. According to Harrington (1988), stuttering occurs because the CV coupling is too tight, whereas according to Wingate (1988), CV coupling is too loose, and therefore, causing stuttering. In *Chapter 4,* we test the CV-timing hypothesis and explore, whether we find evidence for CV-timing differences, even in perceptually fluent speech, by analyzing kinematic EMA data.

Similar to Harrington (1988), Max and Daliri (2019) argue that stuttering is linked to disruptions in sensory prediction processes. The authors outline that typically, during the speech planning phase – before speech movements begin – the central nervous system modulates the auditory system which is crucial for monitoring auditory feedback during speech production. Max and Daliri (2019) found this pre-speech auditory modulation in PWNS but not in PWS when preparing speech or when expecting to hear a playback of their own prerecorded speech. These findings led the authors to hypothesize that in PWS, the challenge may not lie in generating motor commands but in making auditory predictions to effectively prime the auditory system (Max & Daliri, 2019). Therefore, while Harrington's (1988) approach suggests an incorrect predictive timing of auditory events, Max and Daliri (2019) found evidence for an atypical activation of auditory predictions during the speech planning phase which leads to feedback-driven corrections of speech movements and thus results in stuttering.

A more holistic approach to stuttering was developed by Smith and Weber (2017) with their

"Multifactorial Dynamic Pathways Theory". In their theory, the authors suggest that stuttering develops, persists, and changes over time through the complex interaction of multiple factors, including genetic, epigenetic, motor, linguistic, and emotional components. Around stuttering onset, typically between 2 and 5 years of age, the brain goes through an enormous developmental phase involving various neural systems that interact with each other − for example, the speech motor systems with linguistic networks. It is hypothesized that a rapid change in linguistic development, such as when children start to produce longer phrases, may destabilize the developing speech motor system (Smith & Weber, 2017). Other destabilizing factors include an increase in linguistically or emotionally demanding situations, which can lead to more stuttering symptoms. Furthermore, the authors suggest that a key aspect of persistent developmental stuttering is that individuals show an atypical path of developing stable speech networks in the left premotor and primary motor areas. Therefore, even auditorily fluent speech of PWS mostly shows instabilities and when the variability in the speech motor system becomes too large, stuttering occurs (Smith & Weber, 2017).

Summarizing the section on stuttering, it can be concluded that most of the presented models hypothesize that stuttering is associated with problems in the feedforward control of speech (Civier et al., 2010; Civier et al., 2013; Etchell et al., 2014; Namasivayam & van Lieshout, 2011; Max et al., 2004; Smith & Weber, 2017; van Lieshout et al., 2004) while others focus more on impairments in the feedback control system (Harrington, 1988; Max et al., 2004; Max & Daliri, 2019).

In addition, most theories presented here suggest difficulties with speech motor timing, including the initiation and termination of speech sound sequences (Civier et al., 2013; Etchell et al., 2014; Harrington, 1988; Wingate, 1988). This makes the investigation of stuttering especially interesting for gaining insights into the underlying mechanisms of speech timing. Specifically, examining speech under different rhythmic conditions, such as metronome-paced speech—a fluency-enhancing condition for PWS—and self-paced speech using finger-tapping, can provide valuable information about the articulatory basis for fluency-enhancing effects and the interaction of verbal and non-verbal timing networks overall. Investigating these aspects represent a key aspect of the thesis at hand.

Therefore, the present thesis will especially contribute to the approaches about predictive timing and CV-timing proposed by Etchell and colleagues (2014), Harrington (1988) and Wingate (1988), but also to the Speech Motor Skills Hypothesis (van Lieshout et al., 2004; Namasivayam & van Lieshout, 2011). We also discuss our findings with respect to the (GO)DIVA models, presented in section 1.3. in the General Discussion.

1.5. Overview of the thesis

This is a cumulative dissertation that is structured as a series of three original empirical studies corresponding to *Chapters 2*, *3*, and *4*. The main objective of this work is to contribute to the understanding of speech timing mechanisms in PWS and PWNS by examining the effect of rhythmic conditions on fluent speech articulation. We approach this objective by investigating speech in children and adolescents who stutter (*Chapter 2*) and adults who stutter (*Chapters 3* and *4*) which allows us to deepen our understanding of speech motor control development in PWS. The focus thereby lies on articulatory timing at syllable onset position. Word and syllable onsets display a critical point in the speech motor system for PWS, as highlighted in section 1.4. on Stuttering. Furthermore, syllable onsets play a crucial role in syllable organization, as outlined in the section on Articulatory timing. Using a range of experimental methods (acoustics in *Chapter 2*, articulatory via EMA in *Chapters 3* and *4*), this thesis addresses several key questions about speech timing in individuals who stutter, and the relationship between rhythmic conditions and fluent speech production. These questions are outlined in the following.

Chapter 2 focuses on syllabic timing in children and adolescents who stutter and children and adolescents who do not stutter and how it is affected by an external rhythm. More specifically, the study presented in *Chapter 2* explores acoustic cues for a c-center effect in an unpaced condition and a metronome-paced condition by analyzing minimal pairs that differ only in onset complexity.

Previous research suggests that onset-vowel coupling may differ between children and adolescents who stutter and those who do not stutter, even when their speech appears perceptually fluent (De Nil & Brutten, 1991; Smith et al., 2012; Usler et al., 2017; Usler & Walsh, 2018). However, research specifically addressing articulatory properties of children's speech in relation to stuttering is limited. An external rhythm, such as a metronome, is known to enhance speech fluency in a population who stutters (e.g., Andrews et al., 1982) and can positively impact inter-gestural timing leading to greater stability in speech production, as a study on PWNS showed (Tilsen, 2009). But the effect of speaking along with a metronome on syllable timing has not yet been explored in PWS. Therefore, the study presented in *Chapter 2* compares minimal pairs differing in onset complexity in unpaced and metronome-paced speech in children and adolescents who stutter and a matched control group.

The study addresses the following research questions:

i) Do the groups differ in acoustic cues for a c-center effect?

ii) Does an external rhythm eliminate differences between the groups?

iii) Does an external rhythm lead to more stability in timing, particularly in the group who stutters?

This investigation aims to provide valuable insights into speech motor timing, focusing on the temporal syllabic organization and examining the relationship between inter-gestural timing and metronome-paced speech in the developing population.

Chapter 3 is concerned with examining the effect of different rhythmic conditions on articulatory timing and the relationship between verbal and non-verbal gestures in adults. To investigate this, four adults who stutter and four adults who do not stutter were recorded with EMA while engaging in three different conditions: speaking while simultaneously tapping their finger (Tapping condition), speaking in sync with a metronome (Metronome condition), and speaking while tapping along with a metronome (Metronome+Tapping condition). Target words were embedded in a carrier phrase and started with a bilabial onset consonant. The articulatory speech onset of the bilabial consonant and the finger tapping gesture were defined as the start of the gestural plateau. Intervals between the verbal gesture and the pacing events (finger tap and metronome beat) were calculated.

Although stuttering is associated with disruptions in speech timing mechanisms (Etchell et al., 2014) that also extend to the non-verbal domain (e.g., Falk et al., 2015; Slis et al., 2023), articulatory insights into these timing mechanisms remain unexplored. Therefore, this study addresses this gap by exploring

i) whether the groups differ in gestural timing,

ii) whether speech gesture timing is dependent on the rhythmic context (finger tap vs. metronome), and

iii) how the timing of verbal and non-verbal gestures is affected by an external rhythm.

This chapter contributes to understanding how auditory (metronome) and manual rhythms (finger tapping) affect speech timing in PWS and PWNS.

Chapter 4 explores consonant-vowel (CV)-timing and predictive timing across various rhythmic conditions, building on the methodology and research questions introduced in *Chapter 3*. The study presented in *Chapter 4* comprises a larger participant sample of adults who stutter. Furthermore, this chapter situates the questions*from Chapter 3* in the context of predictive timing, an area where PWS face challenges (e.g., Etchell et al., 2014; Falk et al., 2015; Harrington 1988). CV-timing has also been hypothesized to be challenging for PWS, as highlighted by different theories (Harrington, 1988; Wingate, 1988), and supported primarily by acoustic data (Verdurand et al., 2020; Lenoci & Ricci, 2018; Dehqan et al., 2016; Robb & Blomgren, 1997; Klich & May, 1982). While it is known that a metronome can enhance speech fluency significantly in individuals who stutter, the articulatory mechanisms behind this effect remain unexplored. To bridge this gap, the study presented in *Chapter 4* investigates articulatorily via an EMA study whether

i) CV coupling differs between adults who stutter and adults who do not stutter, and

ii) whether inter-gestural timing is affected by different rhythmic pacing conditions.

This study seeks to shed light on the underlying mechanisms of fluent speech motor control and contributes to the broader aim of this thesis: advancing the theoretical understanding of speech production and providing deeper insights into stuttering.

Chapter 5 summarizes the main findings of the three studies and discusses them in the light of speech motor control development, models and theories of stuttering, as well as rhythm and timing.

Each of the three empirical chapters is structured into three parts: (1) an introduction, (2) the paper itself, and (3) a discussion.

Please note that in part (2), the section numbering follows the formatting requirements of the respective journal submission and therefore does not align with the overall section numbering of the thesis.

Chapter 2

2. Temporal organization of syllables in stuttering

2.1. Introduction

Articulatory timing is crucial for the temporal organization of syllables during speech production. The c-center effect is a prominent concept in AP in this respect (Hall, 2010) which posits that consonants in syllable onset position are timed in relation to the vowel, regardless of how many consonants the syllable onset comprises (Browman & Goldstein, 1988). This concept therefore helps to explain how complex syllables are structured in fluent speech. However, little is known about the c-center effect in the developing population. Developing a mature speech motor system is a gradual process and according to Smith and Zelaznik (2004), children and adolescents are still refining the articulatory coordination required for skilled, adult-like speech production. For example, there is a difference in coarticulatory behavior between children and adults, with children displaying greater overlap between gestures (Noiray et al., 2018). Hence, children and adolescents are a particularly interesting participant group for studying syllable timing in relation to articulatory control.

For children and adolescents who stutter, mastering articulatory timing is even more challenging. Several studies indicate that they display an altered articulatory coordination between lip and jaw movements (Smith et al., 2012; Usler & Walsh, 2018; Usler et al., 2017). Lips and jaw are articulators that are also involved in producing bilabial onsets and open vowels, as examined in the study presented in this chapter. Moreover, children who stutter produce more variable voice onset times compared to children and adolescents who do not stutter (De Nil & Brutten, 1991; Dokoza et al., 2011). These results suggest that onset-vowel timing is especially challenging for younger persons who stutter, leading to a potential group difference in temporal syllabic organization.

Speaking with an external rhythm, such as a metronome, has been found to stabilize speech motor coordination (van Lieshout & Namasivayam, 2010, Wiltshire et al., 2023), making it a useful condition to test articulatory timing in both, children and adolescents who stutter and those who do not stutter. Additionally, metronome-paced speech is a fluency-enhancing condition for PWS (Wingate, 1969), providing an opportunity to explore the underlying mechanisms behind this fluency effect.

Therefore, this paper investigates the *c-center effect* in children and adolescents who stutter and children and adolescents who do not stutter under both unpaced and metronome-paced speech conditions. By doing so, it provides insights into the motor patterns involved in fluent speech production at a developmental stage when speech motor control is still maturing.

Acoustic measures offer a practical and efficient way to examine these temporal dynamics. Unlike articulatory measures, such as electromagnetic articulography (EMA), which can be time-consuming and challenging to implement, especially with younger children, acoustic analyses allow for the collection of data from a larger participant sample in a shorter period of time. This method is therefore particularly advantageous when studying speech of children and adolescents, enabling to capture relevant speech patterns in a non-invasive manner.

In this study, acoustic parameters, such as consonant and vowel compression, as well as interval measures serve as correlates for the c-center effect (Katz, 2010). These were examined using German minimal pairs that differed in onset complexity, following the syllabic structures CCVC and CVC. In brief, to support the presence of a c-center effect, consonants and vowels should be shorter in words with a complex onset compared to those with a simple onset (compression effect). Additionally, c-center intervals (midpoint of the onset to the end of the vowel) were compared with right-edge (midpoint of the right-most consonant in the onset to the end of the vowel) and left-edge intervals (left-most consonant in the onset to the end of the vowel) to determine whether German syllables demonstrate the expected c-center organization. The c-center organization would be indicated by minimal or no difference in interval duration between words with a simple and a complex onset. Furthermore, interval stability was assessed using the relative standard deviation, with the hypothesis that the c-center interval would show greater stability than the other intervals, if a c-center effect is present.

The central hypothesis of this chapter is that both children and adolescents, whether they stutter or not, will exhibit acoustic evidence of the c-center effect. However, group differences in compression effects are expected to emerge in the unpaced condition, while no significant differences are anticipated between the groups in the paced condition. Additionally, we hypothesize that participants who stutter will benefit more from the external rhythm than the control group, resulting in a greater increase in interval stability in the paced condition compared to the unpaced condition.

2.2. Paper 1

Paper 1 has been published in a Special Issue of the Journal of Fluency Disorders undergoing a full revision process.

Franke, M., Hoole, P., & Falk, S. (2023). Temporal organization of syllables in paced and unpaced speech in children and adolescents who stutter. *Journal of fluency disorders*, *76*, 105975. https://doi.org/10.1016/j.jfludis.2023.105975

Temporal organization of syllables in paced and unpaced speech in children and adolescents who stutter

Mona Franke[a,b,c,d*], Philip Hoole[a], Simone Falk[b,c,d]

a *Institute for Phonetics and Speech Processing, Ludwig Maximilian University of Munich, Germany*
b *Faculté des arts et des sciences – Département de linguistique et de traduction, Université de Montréeal, Canada,*
c *International Laboratory for Brain, Music and Sound Research (BRAMS), Montréal, Canada*
d *Centre for Research on Brain, Language and Music (CRBLM), Montréal, Canada*

* Correspondence to: International Laboratory for Brain, Music and Sound Research (BRAMS), Pavillon Marie-Victorin, Universit́ e de Montŕ eal, 90 Vincent D'Indy Ave Local A-108, Outremont, Quebec H2V 2S9, Canada.
E-mail address: mona.franke@phonetik.uni-muenchen.de (M. Franke).

Abstract

Purpose: Speaking with an external rhythm has a tremendous fluency-enhancing effect in people who stutter. The aim of the present study is to examine whether syllabic timing related to articulatory timing (c-center) would differ between children and adolescents who stutter and a matched control group in an unpaced vs. a paced condition.

Methods: We recorded 48 German-speaking children and adolescents who stutter and a matched control group reading monosyllabic words with and without a metronome (unpaced and paced condition). Analyses were conducted on four minimal pairs that differed in onset complexity (simple vs. complex). The following acoustic correlates of a *c-center effect* were analyzed: vowel and consonant compression, acoustic intervals (time from c-center, left-edge, and right-edge to an anchor-point), and relative standard deviations of these intervals.

Results: Both groups show acoustic correlates of a c-center effect (consonant compression, vowel compression, c-center organization, and more stable c-center intervals), independently of condition. However, the group who stutters had a more pronounced consonant compression effect. The metronome did not significantly affect syllabic organization but interval stability improved in the paced condition in both groups.

Conclusion: Children and adolescents who stutter and matched controls have a similar syllable organization, related to articulatory timing, regardless of paced or unpaced speech. However, consonant onset timing differs between the group who stutters and the control group; this is a promising basis for conducting an articulatory study in which articulatory (gestural) timing can be examined in more detail.

Keywords: *stuttering, paced speech, timing, syllable organization, c-center effect*

1. Introduction

1.1 Syllable structure and stuttering

Stuttering is a neurodevelopmental speech fluency disorder that approximately 5-8% of preschool children develop (Yairi & Ambrose, 2013). However, the majority of children spontaneously recover from stuttering in childhood, giving a prevalence of approximately 1% in the adult population (Yairi & Ambrose, 2013). Furthermore, girls are more likely to recover from stuttering than boys; while boys and girls are equally affected, there are more men who stutter than women who stutter (ratio approximately 4:1) (Yairi & Ambrose, 2013).

The speech of persons who stutter (PWS) is characterized by involuntary interruptions due to repetitions of sounds, syllables, or words, involuntary blocks, or prolongations of sounds (World Health Organization, 2015). These symptoms do not occur randomly within a syllable. Most often, they occur at the word or syllable onset. Harrington's (1987) auditory, acoustic, and electropalatographic study suggests that stuttering symptoms appear before or within the acoustic onglide of the vowel and never in the rhyme (that is, the nucleus and offset consonants of a syllable). Disfluencies that occur at the end of words or syllables (i.e., echodysphemia) are nowadays considered as a subgroup of developmental dysfluency, in addition to stuttering and cluttering (MacMillan, Kokolakis, Sheedy, & Packman, 2014).

Speech fluency breakdowns in stuttering have been generally linked to an atypical development of speech motor processes (Smith & Weber, 2016). An explanation for the breakdown characteristics of stuttering at syllable or word onsets was proposed by Wingate (1988) over 30 years ago. Wingate postulated the so-called Fault-Line Hypothesis; this hypothesis is based on the idea that stuttering symptoms occur at the transition from the initial consonant to the stress-bearing vowel (Wingate, 1988). Therefore, he posits that this would lead to a divide or "Fault-Line" at the point of syllable onset and rhyme integration which is caused by the delayed encoding of the vowel (Wingate, 1988). Hence, Wingate's Fault-Line Hypothesis claims that the timing relationship between consonants and vowels is atypical in PWS, especially in stressed syllables.

There is evidence from experimental research that onset-vowel timing is particularly challenging for PWS because their speech motor program breaks down at the point where these specific timing relations usually occur, even in perceptually fluent speech. For instance, adults who stutter show more variable vowels when producing monosyllabic nonwords comprising a consonant, vowel, and a consonant (such as in "vep") and more variable fricatives (/s/) in nonwords with the structure: vowel, /s/, vowel (such as in "asi") (Di Simoni, 1974). They also produce more variability in stop gap durations and voice onset time related to syllable onsets in monosyllabic words (Max & Gracco, 2005), as well as more variability in voice onset times of the same stop consonant in syllable initial position in a trisyllabic word (Jäncke, 1994). Furthermore, it was found that when producing word-initial bilabials adults who stutter have longer bilabial closing intervals, measured from the movement onset of the lip closing and the peak velocity of the closing movement to the offset of vocal fold vibration for the preconsonantal vowel (Max & Gracco, 2005). Another study found a similar result, namely overall longer bilabial closing durations and a higher peak velocity when releasing the bilabial onset of nouns (Max, Caruso, & Gracco, 2003). Adults who stutter also have larger amplitudes of upper lip movement when producing syllable onset bilabials (Namasivayam & van Lieshout, 2008). Heyde and colleagues (2016) found that adults who stutter transition from the onset to the vowel with a decreased peak velocity, compared to the control group, which indicates that adults who stutter have lower acceleration/deceleration in releasing the constriction of the consonant. The authors concluded that these differences support Wingate's Fault-Line Hypothesis, since PWS might have difficulty integrating syllable rhymes with their onsets. Another study identified two general strategies that promoted fluency in PWS, namely the reduction of speech rate and the reduction of coarticulation (Zmarich, Balbo, Galatà, Verdurand, & Rossato, 2013). However, in general, adults who stutter are less consistent in inter-articulator coordination across mono- and multisyllabic nonwords (Smith, Sadagopan, Walsh, & Weber-Fox, 2010; Wiltshire, Chiew, Chesters, Healy, & Watkins, 2021).

Less research has been dedicated to the articulatory properties of children's speech in stuttering. A few studies show that children who stutter also display speech motor control differences compared to children who do not stutter. For example, children who stutter show greater articulatory coordination variability across mono- and multisyllabic nonwords (4-5 year-olds, Smith, Goffman, Sasisekaran, & Weber-Fox, 2012) and sentences (5-7 year-olds, Usler, Smith, & Weber, 2017; 6-12 year-olds, Usler & Walsh, 2018). These investigations used the lip aperture variability index, which reflects the inter-articulatory coordination of lip and jaw movements – articulators that were involved in inter-gestural timing between many onset (bilabials) and vowel combinations that were present in the stimuli used by the authors. Interestingly, children who recovered from stuttering did not differ from the control group in terms of articulatory coordination (Usler et al., 2017). Furthermore, children who stutter ranging from 8 to 12 years of age displayed more variable voice onset times compared to a matched control group at all three complexity levels – onsets with a single consonant, a two-consonant cluster, and a three-consonant cluster whereby variability was the largest at the single consonant level, followed by the three-consonant onset, and then the biconsonantal onset (De Nil & Brutten, 1991). Moreover, voice onset time in children who stutter (6.5 - 7.5 year-olds) was more variable in syllable-initial position when produced in a disyllabic word and when included in a sentence, as well (Dokoza, Hedever, & Sarić, 2011). To summarize, these studies lead to the assumption that onset-vowel coupling might differ between children and adolescents who stutter (test group) and children and adolescents who do not stutter (control group), even when considering perceptually fluent speech.

There is evidence that rhythm can affect the timing between two individual articulatory movements (also referred to as inter-gestural timing). For instance, Tilsen (2009) found that the inter-gestural timing between two consonant gestures in a CCV syllable structure was more stable when the phrase in which the word occurred was produced with an easier rhythm (nearer to low-order harmonic ratios). The rhythm was controlled by a metronome. When rhythmic timing was more variable, the inter-gestural timing also became more variable. This is an interesting finding, since an external rhythm works as a fluency-enhancing condition for PWS. For instance, speech fluency tremendously increases in PWS when they synchronize their speech to a metronome (Wingate, 1969; Andrews, Howie, Dozsa, & Guitar, 1982). This fluency-enhancing effect is accompanied by a normalization of hyper- and hypo-activation in neural circuits mediating temporal processing and movement initiation, such as the basal ganglia and the cerebellum (e.g., Toyomura, Fuji, & Kuriki, 2011), indicating better coupling of auditory and motor systems (Stager, Jeffries, & Braun, 2003). Therefore, metronome-paced speech also leads to a more stable speech motor coordination (van Lieshout & Namasivayam, 2010).

The aim of the present study is therefore, to analyze temporal organization of syllables in children and adolescents who stutter and matched control participants, speaking with and without an external rhythm. We examine syllabic timing related to articulatory timing by analyzing the *c-center effect,* which is described in the following section.

1.2 Framework for analyzing temporal syllable structure

From a purely articulatory point of view, an elementary task of fluent speech production is the coordination of movements among groups of articulators (Browman & Goldstein, 1991, 1992). In general, the internal structure of the syllable is asymmetrical, which is crucial for inter-articulatory coordination of the speech gestures involved (captured, for example, in the Coupled Oscillator Model of Syllable Structure, see e.g., Nam & Saltzman, 2003; Goldstein & Pouplier, 2014). Gestures, defined in the framework of Articulatory Phonology, are distinct vocal tract actions (Browman & Goldstein, 1992). In Articulatory Phonology, there are two main gesture phasing patterns that have been suggested to describe the timing relationships between consonants and vowels: In-phase and anti-phase coupling. While onset consonants are supposed to be timed in-phase with the vowel, coda consonants are presumed to be timed anti-phase with the vowel. More specifically, in-phase coupling means that the articulatory gesture of the onset consonant and the articulatory gesture of the vowel start at the same time; by contrast anti-phase coupling means that the coda consonant gesture begins when the vowel gesture reaches its peak (e.g., Hall 2010). However, there are situations of conflict where gestures simultaneously aim to reach opposing goals, e.g. a lip closure to produce a [p] and jaw opening to produce an [a]. In this case, there is a competition between the ideal phasing relationship (in-phase) and reaching the acoustic goal. The consonant will try to get as close to the ideal phasing relationship with the vowel as possible while still reaching the acoustic goal.

Byrd (1996) discovered that there is less variability at the level of gestural overlap in onsets compared to codas. In particular, it appears that onset consonant gestures and vowel gestures together form a much more cohesive unit than vowel gestures and coda consonant gestures (e.g., Hoole & Pouplier, 2015). This asymmetry underlies the so-called *c-center effect* according to which there is a constant temporal relationship between the temporal center of the onset and the following vowel regardless of the number of consonants contained in the onset (Browman & Goldstein, 1988). Describing this phenomenon with the coupling model, consonants forming a complex onset (e.g., CC) are coupled to each other anti-phase, while each consonant in the onset is coupled in-phase with the following vowel (e.g., Browman & Goldstein, 2000). In a complex coda on the other hand, there is only anti-phase coupling between the vowel and the coda consonant as well as between the consonants themselves. Therefore, there is no constant temporal relationship between the vowel and the coda.

The following figure (Figure 1) illustrates the coupling relationships between the consonants (C1, C2) and the vowel in an onset and in a coda organization. The competitive coupling topology in the onset organization (i.e. the combination of in-phase and anti-phase coupling) can be shown to lead to a shift of the rightmost consonant in an onset cluster towards the vowel as more consonants are added to the onset, leading to the *c-center*.

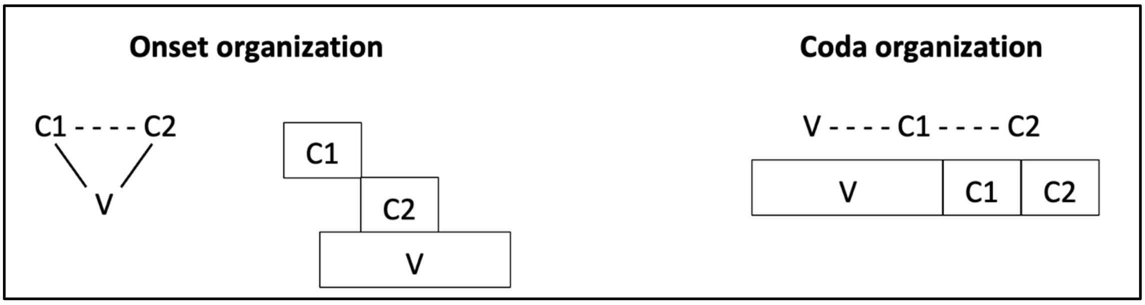


Figure 1: Phasing relations in a complex onset and complex coda. Dashed lines display an anti-phase coupling, solid lines display an in-phase coupling.

While traditional articulatory studies, e.g. using electromagnetic articulography, provide the most direct evidence for gestural organization, they are, of course, time-consuming and hence often go in hand with small participant sample sizes. In fact, the acoustic signal – which is far more efficient to obtain – can also provide evidence for a *c-center effect*. For example, following a gestural approach to syllable organization (Browman & Goldstein, 1988) and summarized by Katz (2010), the (acoustic) duration of the vowel should be shorter in syllables with a complex onset compared to syllables with a simple onset due to the shift of the rightmost onset consonant towards the vowel. Note that we refer to this phenomenon as vowel compression from here on. Moreover, it is suggested that gestural overlap, in general, would cause compression to arise. Hence, acoustic compression should also be observed in the onset, when we compare a single onset to the same consonant (C) in C$_2$ position, e.g. [l] in [klaʊd] vs. in [laʊd]. According to this assumption, [l] in C$_2$ would be acoustically shorter than [l] in C$_1$ because the gesture of [l] in the syllable with the complex onset is more shifted towards the vowel gesture. Henceforth, this will be referred to as consonant compression.

Several studies found vowel compression (e.g., Marin & Pouplier, 2010; Katz, 2012; Brunner, Geng, Sotiropoulou, & Gafos, 2014; Marin & Bučar Shigemori, 2014; Peters & Kleber, 2014), as well as consonant compression (e.g., Katz, 2010; Marin & Bučar Shigemori, 2014; Gibson, Fernández Planas, Gafos, & Remirez, 2015) in complex vs. simple syllables (for instance CCV vs. CV). Moreover, there are studies that showed the usefulness of acoustic methods to measure stability patterns of word-initial consonant clusters (e.g., for English: Selkirk & Durvasula, 2013; for Jazani Arabic: Ruthan, Durvasula, & Lin, 2019). The following figure (*Figure 2*) displays a sketch of two different stability patterns that can be observed for languages allowing complex onsets (left panel) and languages that do not allow complex onsets (right panel).

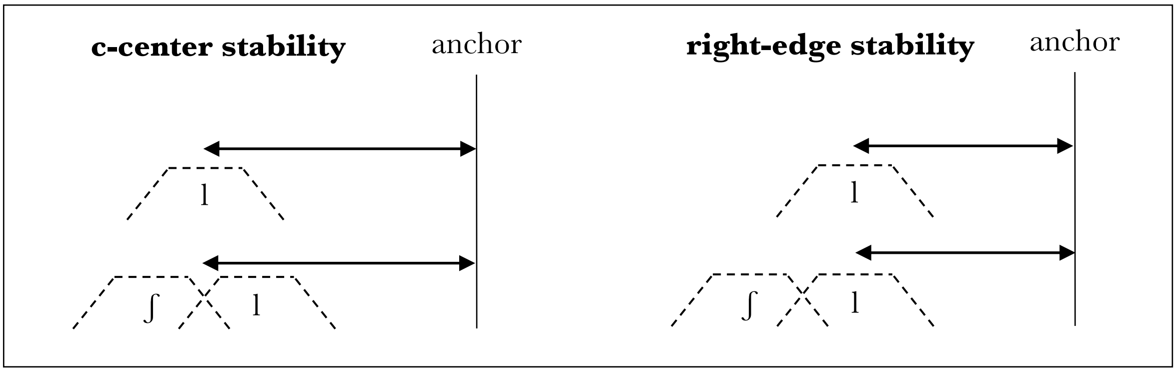


Figure 2: Temporal organization of complex (c-center stability) and simplex (right-edge stability) onset organization.

Selkirk & Durvasula (2013) replicated results on English with acoustic data, showing that there is more timing stability (measured with the relative standard deviation [RSD][1]) in the c-center to anchor interval (c-center in this case was defined as the midpoint of the onset and the anchor was the end of the vowel) compared to other intervals (right-edge and left-edge). For Jazani Arabic it was found that the right-edge (the midpoint of the right-most consonant in the onset) to anchor interval was most stable, which led to the conclusion that this specific dialect has a simple onset organization (Ruthan et al., 2019), contrary to English and German, for example. In summary, this articulatory framework is particularly interesting for studying individuals who stutter, as it provides insights into syllabic organization and associated speech planning. Also, the approach provides clues to articulatory timing processes as well as articulatory control – an area that is particularly challenging for PWS as pointed out in the previous section.

1.3 Hypotheses

Since German admits complex onsets, we would expect to find similar results to the study on English (Pouplier, 2012), where the c-center to anchor interval was most stable. On the one hand, we expect PWS to show c-center stability but overall more variability than persons who do not stutter (PWNS). On the other hand, using a metronome to regulate speech timing could enhance the stability of this interval in PWS.

In particular, we aimed to examine whether syllabic timing related to articulatory timing (c-center) would differ between groups in an unpaced vs. a paced condition.

We hypothesize to find acoustic cues for a *c-center effect* in both groups (by comparing complex and simple syllables), with a difference between the paced and unpaced condition, and between the control group and the test group.

More precisely, we hypothesized that we would find

(1) consonant compression in both groups and

(2) vowel compression in both groups

with a group difference. As for this difference, either direction is possible. Less compression in the test group could derive from less coarticulation, for example, as a strategy to maintain fluency (Zmarich et al., 2013). It is another possibility that compression effects will be more

[1] The RSD is a common measure of interval stability used in studies that examine gestural timing (e.g., Shaw, Gafos, Hoole, & Zeroual, 2011; Brunner et al., 2014; Ruthan et al., 2019), since it takes into account the fact that shorter durational intervals typically have a lower absolute standard deviation. Using the latter would thus bias the results towards greater stability for the right-edge to anchor interval since this is by definition the shortest of the three intervals considered.

pronounced in the test group, if gestures of PWS were to overlap too much (as for example predicted by Harrington, 1988). We hypothesize that

(3) a pacing condition will amplify the compression effects and that the groups will not differ in the paced condition.

Furthermore, we hypothesize to see evidence for a *c-center effect* by finding

(4) more stability in the interval from the acoustic *c-center* to the end of the vowel compared to the left-edge or right-edge interval, together with an increase in stability in the paced condition, whereby the group who stutters will benefit more from the paced condition leading to a higher increase in stability.

Analyzing the perceptually fluent speech of PWS can give us clues on whether PWS might have difficulties in the coordination of onset (consonant) and nucleus (vowel) timing, and thus, syllabic organization. In addition, acoustic analyses enable us to analyze a larger participant sample size and since this is the first study that examines a potential *c-center effect* in PWS and a matched control group, the present study can serve as a basis for further articulatory investigations.

2. Materials and Methods

2.1 Participants

48 PWS (42 males, 6 females, Mean age = 13.03 , SD = 2.55, range = 9-18), and 48 age- and gender-matched controls (Mean age = 12.94 , SD = 2.46, range = 9-18) participated in the experiment. All participants spoke German at a native level. All except five participants reported an absence of language, hearing and/or neurological deficits (other than developmental stuttering which was diagnosed by a speech therapist). These five participants had Dyslexia or ADHD in addition to stuttering. All participants and the parents of the underaged participants signed for informed consent. The participants who stutter were recruited through the intensive therapy course “Stärker als Stottern” (staerker-als-stottern.de), during which their stuttering was assessed by trained speech therapists. The stuttering severity was determined with the SSI-3 by trained speech therapists based on recordings on the day the data were recorded (*Table 1*). All participants who stutter were participating in the intensive therapy course but recordings were done prior to the start of the course. The typically fluent participants were recruited through schools.

Table 1: Distribution of stuttering severity at the recording day across participants.

Stuttering severity	Very mild	Mild	Mild-Moderate	Moderate	Moderate-severe	Severe	Severe - very severe	Very Severe
No. of parti-cipants	2	8	3	11	2	10	1	11
Mean SSI score (SD)	8 (1)	15.25 (2.86)	19.67 (4.5)	22.36 (0.98)	22.5 (0.5)	30.4 (1.56)	35	41.18 (5.11)

2.2. Material

Participants were asked to read two separate wordlists that either contained monosyllabic words with simple onsets (W_{simple}) or monosyllabic words with complex onsets ($W_{complex}$). Words were either nouns or adjectives. The syllable structure of the monosyllabic words was either consonant vowel consonant (CVC) in the W_{simple} wordlist or CCVC in the $W_{complex}$ wordlist, whereby vowels were either a short vowel or a diphthong. Each list contained a set of practice items at the beginning (5 words in total) to familiarize participants with the wordlist reading pattern. These words were followed by 12 nouns and 12 adjectives which occurred twice in the same order in the same wordlist. The words were printed on paper in landscape format (DIN A4, in 14-point Arial font) in rows of 12 words.

As target words, we chose 4 word pairs that only differed in onset complexity (see *Table 2)* and that were suitable for segmentation solely based on the acoustic signal. Since our focus was on the first consonant (C_1) in the onset in words with a single onset (W_{simple}) and the second consonant (C_2) in the onset of words with a complex onset ($W_{complex}$), we chose words with an [l] in these positions as it is fairly easy to detect (in comparison to a plosive, for instance, where we could not detect the closure phase in words with a single onset). Target words were not situated at the margins (beginning or end) of the wordlist to avoid patterns like phrase-final lengthening or a different intonation contour.

Table 2: Target words with simple (left column) and complex onsets (right column). One row displays one word pair.

W$_{\text{simple}}$	**W$_{\text{complex}}$**
Leim [laɪ̯m] (Engl. *glue*)	Schleim [ʃlaɪ̯m] (Engl. *slime*)
Lamm [lam] (Engl. *lamb*)	Schlamm [ʃlam] (Engl. *mud*)
Lauch [laʊ̯x] (Engl. *leek*)	Schlauch [ʃlaʊ̯x] (Engl. *tube*)
lang [laŋ] (Engl. *long*)	Klang [klaŋ] (Engl. *sound*)

2.3 Procedure

Participants were comfortably seated at a table with the experimenter present in the room. The wordlist was placed in front of them on the table by the experimenter who also presented the second wordlist after they finished the first. Participants were recorded with a Zoom H4N recorder (44.1 kHz, 16 bit) via an external headset microphone (beyer dynamic opus 54.16/3) in a quiet room.

Every wordlist was read twice per participant in each condition (unpaced and paced). In the unpaced condition, participants were instructed to read the wordlist in their preferred speech tempo. In the paced condition, they were asked to read the wordlists along with a metronome that had an inter-onset-interval (IOI) of 900ms. In this condition, every word was to be timed with one metronome beep. Half of the participants of each group started with the W$_{\text{simple}}$ wordlist and the other half with the W$_{\text{complex}}$ wordlist (randomized by the experimenter). Paced reading always followed unpaced reading because otherwise, the paced reading could have impacted the speech rate of the unpaced reading.

2.4 Analyses

The segmentation of every word and the corresponding segments was based on the acoustic signal and the oscillogram using Praat (version 6.1) (Boersma & Weenink, 2019) and was done by phonetics students that were trained in acoustic segmentation. Words were excluded from analyses if they displayed a markedly increased tonus or speech rate, a blockade, prolongation or repetition of sounds. Incorrectly read words were also excluded. Hence, only the perceptually fluent speech was analyzed. Exclusions were based on the assessment of the first author and checked by a trained speech therapist. The rules for segmentation were defined as follows:

Laterals were segmented at the first positive zero crossing of the first recognizable period, and fricatives at the beginning of frication in the signal. Note that [klaŋ] is a special case since the start of the plosive is not measurable (the closure duration cannot be detected based on the acoustic recordings) and the [l] may be largely voiceless. For this reason, the word pair *Klang-*

lang was excluded for some analyses. The beginning of a vowel was segmented at the second zero crossing of the first recognizable period and the end of the vowel was determined by the beginning of the coda consonant. In this case, a nasal was segmented at the time when antiresonances were present and/or at the first clearly visible change in the periodic pattern in the oscillogram at the first positive zero crossing. The segmentation of a fricative in coda position was segmented at the beginning of frication, as described for segmentation in onset position. The following figure (Figure 3) displays a segmentation example of the word pair *Schleim-Leim*.

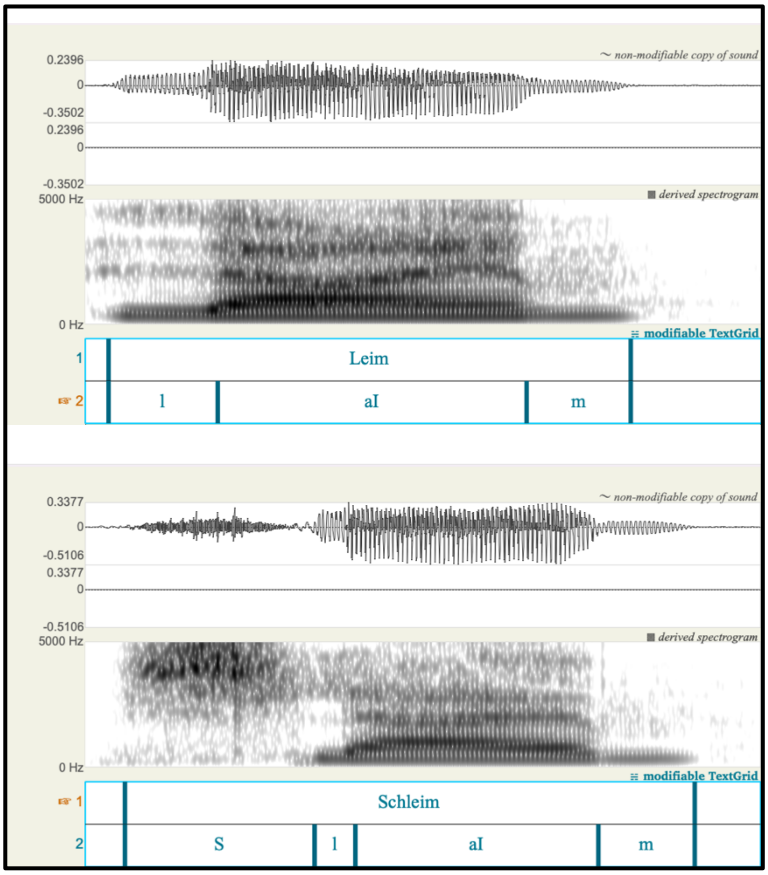


Figure 3: Segmentation example from Praat with oscillogram and spectrogram. Tier 1: Word, tier 2: Phones (Sampa transcription).

The following table displays the number of excluded words out of 384 in total per condition and per group.

Table 1: Target words excluded per group and condition.

	PWNS	**PWS**
W_{simple} unpaced	0	49
W_{simple} paced	4	44
$W_{complex}$ unpaced	2	39
$W_{complex}$ paced	3	29

2.4.1 Acoustic correlates of a c-center effect

We analyzed three different correlates of a *c-center effect*, namely consonant compression, vowel compression (following Katz, 2010) as well as three intervals (left-edge to vowel offset, right-edge to vowel offset, and the c-center to vowel offset) which are associated with (articulatory) syllable organization.

2.4.1.1 Consonant compression

To analyze consonant compression, we extracted [l] durations of W_{simple} and $W_{complex}$ and ran a linear mixed model (see *3.2. Consonant compression*) to compare [l]'s that were produced as C_1 in W_{simple} with [l]'s that were produced in C_2 in $W_{complex}$.

2.4.1.2 Vowel compression

For the analysis of vowel compression, we extracted vowel durations of W_{simple} and $W_{complex}$ and compared them following the procedure for consonant compression.

2.4.1.3 Intervals (acoustic c-center)

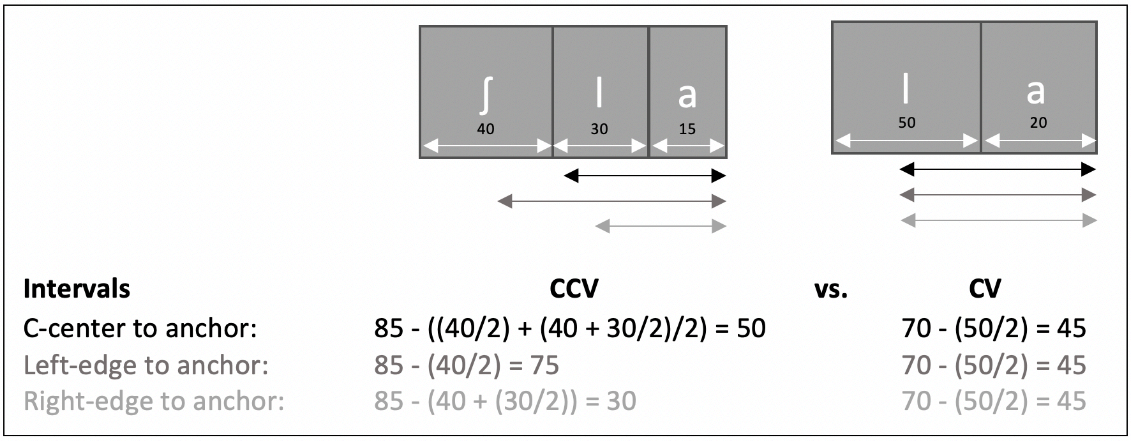
The methodology for this part follows that of Ruthan et al. (2019). They proposed an acoustic method to calculate three intervals that were measured in an articulatory study (e.g., by Shaw et al., 2011). Note that the definition of these intervals was slightly different in articulatory studies (e.g., Shaw et al., 2011; Brunner et al., 2014). The intervals in the present study are defined as follows:

(1) Left-edge to anchor: This interval was calculated as the duration from the midpoint of the left-most consonant (= left-edge) to the end of the vowel (= anchor).

(2) Right-edge to anchor: This interval was calculated as the duration from the midpoint of the right-most consonant (= right-edge) to the end of the vowel (= anchor).

(3) C-center to anchor: This interval was determined by calculating the duration from the mean of the midpoints of the two onset consonants (c-center) to the end of the vowel (=anchor)

For better illustration, *Figure 4* displays the calculation of the intervals.



Intervals	**CCV**	**vs.**	**CV**
C-center to anchor:	85 - ((40/2) + (40 + 30/2)/2) = 50		70 - (50/2) = 45
Left-edge to anchor:	85 - (40/2) = 75		70 - (50/2) = 45
Right-edge to anchor:	85 - (40 + (30/2)) = 30		70 - (50/2) = 45

Figure 4: Example calculation of intervals in ms for a CCV and a CV syllable structure.

If an onset only has a single onset consonant, the three intervals (left-edge to anchor, right-edge to anchor, and c-center to anchor) do not differ, as can be seen in *Figure 4*.

To identify the most stable interval (it is hypothesized that the c-center is the most stable interval in our study), we followed previous studies (Shaw et al., 2011; Brunner et al., 2014; Ruthan et al., 2019) and used the relative standard deviation (RSD), which is defined as *100*(standard deviation of the duration/mean duration)*. It is a measure that refers to the variability of an interval over word pairs (e.g., Brunner et al., 2014; Ruthan et al., 2019) or triads (e.g. Shaw et al., 2011). Therefore, RSDs were calculated across word pairs per group[2].

We also compare the RSD between groups and analyze whether the test group and the control group differ significantly in their stability patterns.

2.4.2 Statistical analyses

All statistical analyses were conducted with R Version 4.0.2 (R Core Team, 2020). We used the package tidyverse (Wickham et al., 2019) for data processing and lme4 (Bates, Maechler, Bolker, & Walker, 2015) to perform linear mixed effects analyses. Linear mixed effects models were calculated to estimate [l] and vowel duration (with regards to compression effects), and a linear mixed effects model was also run to estimate interval duration for the c-center, left-edge, and right-edge interval, as well as the RSD. Variables that were included in the models as random

[2] We calculated the RSD per group and not per participant because we had maximally 2 repetitions per word and condition per participant.

or fixed effects were GROUP (test group vs. control group), CONDITION (paced vs. unpaced), ONSET COMPLEXITY (one or two segment(s)), WORD PAIR (4 in total, see *Table 2*), WORD (8 in total, see *Table 2*) and PARTICIPANT.

We followed the same procedure for all models:

We started all linear mixed models with a full model including a three-way interaction term between the fixed factors (GROUP, CONDITION, ONSEST COMPLEXITY in the case of predicting compression effects and interval durations, and INTERVAL, CONDITION, ONSET COMPLEXITY in the case of predicting RSD) as well as random intercepts and random slopes (intercepts for PARTICIPANT and WORD PAIR, by-participant random slopes for CONDITION and ONSET COMPLEXITY/INTERVAL and by-word-pair random slopes for GROUP and CONDITION). 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 with BIC and the variance that the model explains was estimated using the function *r2_nakagawa*.

The complexity of all models could be reduced to a two-way interaction term between the fixed factors and by excluding all by-word-pair random slopes (because of perfect correlations of WORD PAIR and GROUP, as well as of WORD PAIR and CONDITION and singular fit). Hence, the final models included the fixed effects with a 2-way interaction between all variables, as well as intercepts for PARTICIPANT and WORD PAIR with by-participant random slopes for CONDITION and ONSET COMPLEXITY (in the RSD model only by-participant random slope for CONDITION). All models were first calculated for all participants and then without the five participants with comorbidities to check for the robustness of the results.

Residual plots were visually checked for homoscedasticity of normality before reporting the results. Main effects are reported, using the R-function *anova,* and interactions were analyzed with Post-hoc Tukey corrected t-tests, using the package *emmeans* (Lenth, 2020). Pairwise comparisons were done using the contrast function in the package *emmeans,* and correlations were done using Spearman-rho correlations. Before examining acoustic correlates of a *c-center effect*, word duration was analyzed in order to reveal potential group differences in speaking rate in children and adolescents who stutter and matched peers.

3. Results

3.1 Word duration

Since acoustic correlates of a *c-center effect* are mirrored in durations (durational compression and interval duration), we analyzed word duration beforehand in order to be able to interpret the following results with this information in mind. *Figure 5* displays the word duration grouped by syllable structure (simple vs. complex onsets) and condition (paced vs. unpaced) for the control group (PWNS) and test group (PWS).

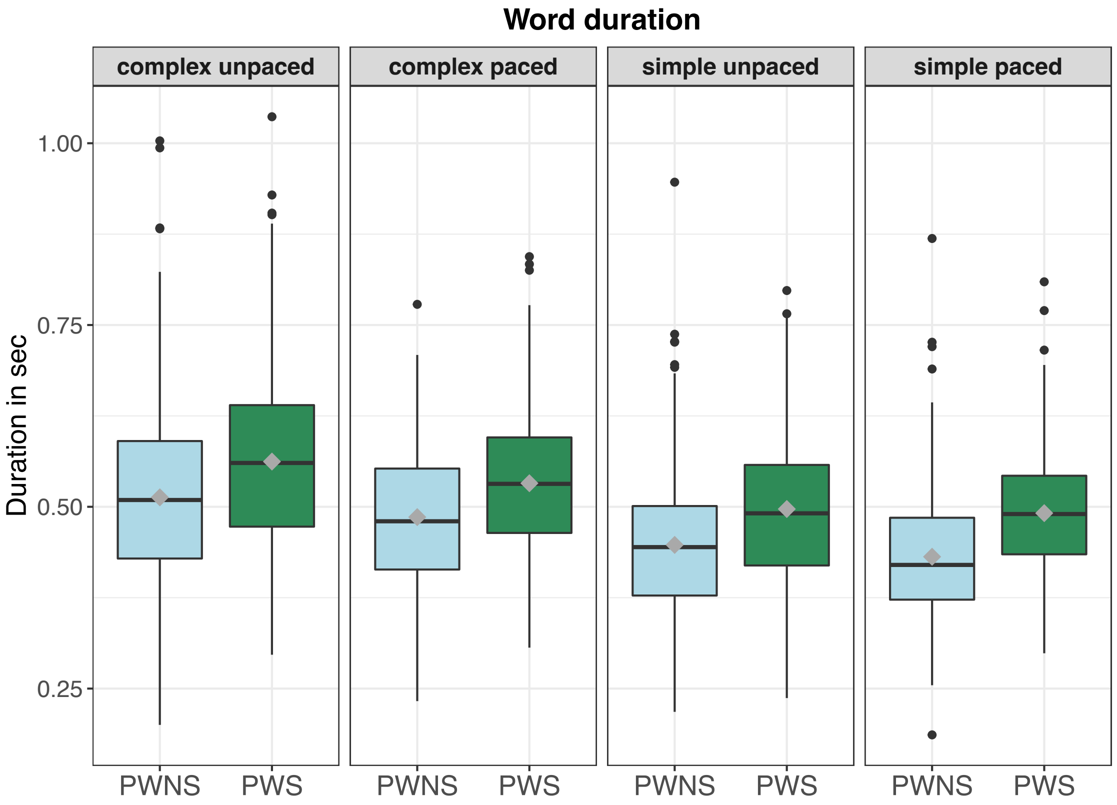


Figure 5: Word duration for each group per condition. Within each box, the median is denoted with black lines; boxes extend from the 25th to the 75th percentile of each group's distribution of values; the ends of the whiskers denote 1.5 interquartile range beyond the 25th and 75th percentile of each group; dots display observations outside the range of whiskers. Participants who do not stutter = blue, participants who stutter = green, mean values per group and condition are displayed with grey diamonds. Words with complex onsets are in the left two columns and words with a simple onset are in the right two columns.

To predict word duration, we followed the procedure described in 2.4.2 Statistical analyses except that we replaced the random intercept WORD PAIR with WORD. Residuals were slightly right-skewed but since the model predicts real word durations of groups who might differ in speech tempo, this is expected. A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 1: Word duration*).

The model (Conditional $R^2 = 0.758$, Marginal $R^2 = 0.130$) revealed that GROUP (F[1, 92.01] = 20.56, p < .001) and CONDITION (F[1, 68.8] = 11.2, p = .01) significantly predicted word duration. As can be seen in *Figure 3*, the group who stutters showed longer word durations than the control group and words were longer in the unpaced condition compared to the paced condition. Moreover, there was a significant interaction between ONSET COMPLEXITY and CONDITION (F[1, 6.09] = 7.06, p = .04). Post-hoc tests showed that the paced condition only affected the duration of words with complex onsets (p < .001) but not words with simple onsets, indicating that word duration significantly decreased in $W_{complex}$ produced along with a metronome but not in W_{simple}.

The results did not change when running the same model without the five participants who had comorbidities. Based on the results regarding word duration, significant main effects of condition (paced vs. unpaced) and group (PWS vs. PWNS) are also expected for the acoustic correlates of a c-center. Therefore, these main effects will not be reported in detail, since they only display the speech rate differences we found in the word duration analysis. Moreover, residuals are also expected to be slightly right and potentially left-skewed. Log transforming the durations ([l] duration, vowel duration, and the interval duration) did not improve the skewness. For this reason, the distribution of residuals for the following statistical models will no longer be reported.

3.2 Consonant compression

If consonant compression takes place, we would expect the duration of [l] to depend on ONSET COMPLEXITY (shorter [l] duration in $W_{complex}$ than in W_{simple}). Moreover, if the groups differ in consonant compression, we would expect to find a significant interaction between GROUP and ONSET COMPLEXITY. Furthermore, we examined potential effects of metronome pacing (CONDITION). *Figure 6* displays [l] duration in seconds for $W_{complex}$ and W_{simple} for each condition per group.

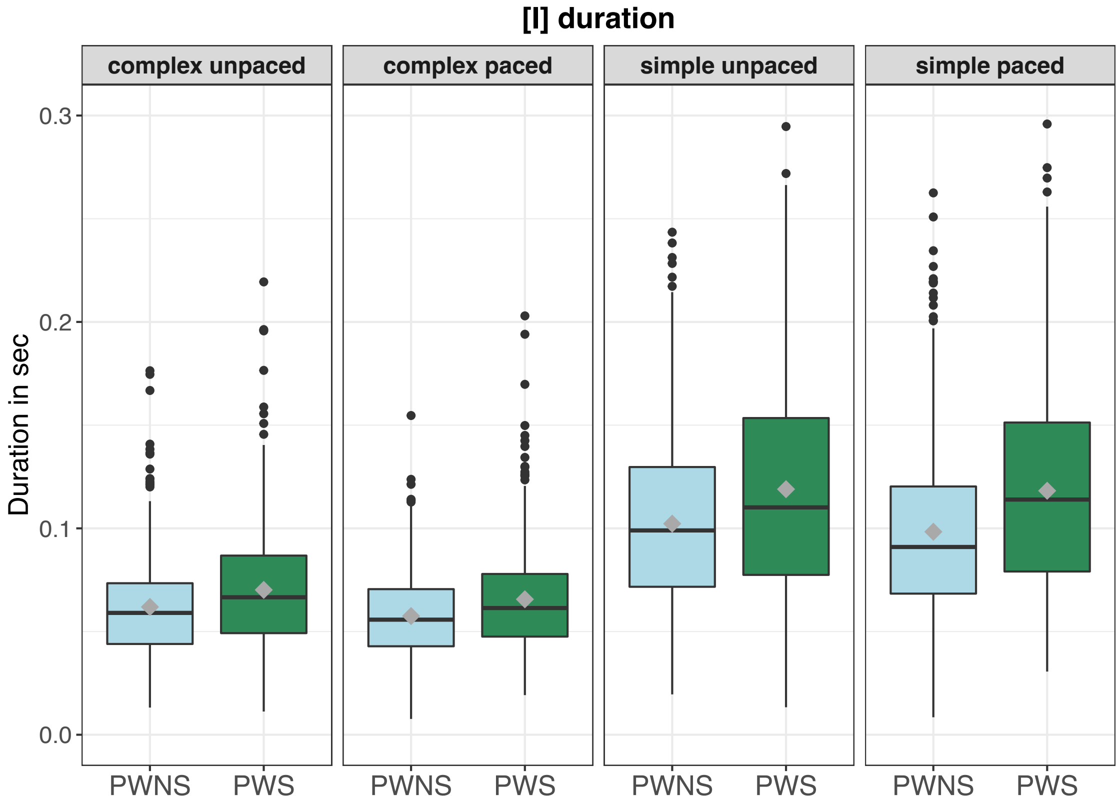


Figure 6: [l] duration for each group per condition. Within each box, the median is denoted with black lines; boxes extend from the 25th to the 75th percentile of each group's distribution of values; the ends of the whiskers denote 1.5 interquartile range beyond the 25th and 75th percentile of each group; dots display observations outside the range of whiskers. Participants who do not stutter = blue, participants who stutter = green, mean values per group and condition are displayed with grey diamonds. Words with complex onsets are in the left two columns and words with a simple onset are in the right two columns.

A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 2: l duration*). The model revealed a strong correlation between the participant intercept and onset complexity (see *Figure 7*).

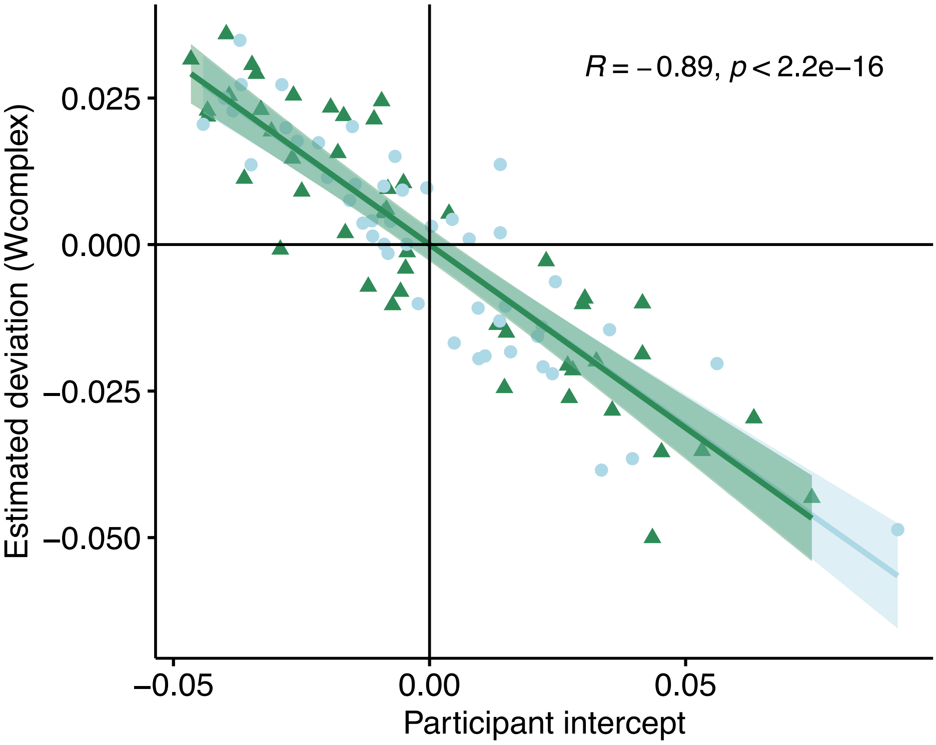


Figure 7: Correlation between participant intercept and estimated deviation for $W_{complex}$ (units in seconds). 0 displays the mean value for all participants. Participants who do not stutter = blue and circles, participants who stutter = green and triangles.

The significant correlation between participant intercept and estimated deviation for the complex onset (R = -0.89, *p* < .001) indicates that participants who produced a long [l] duration in W_{simple}, produced shorter [l] durations in $W_{complex}$. Hence, these particpants produced a bigger change from W_{simple} to $W_{complex}$.

The linear mixed effects model (Conditional R^2 = 0.528, Marginal R^2 = 0.281) showed a significant effect of GROUP (F[1, 89.17] = 13.97, p < 0.001), ONSET COMPLEXITY (F[1, 92.41]= 315.54, p < .001), and a significant interaction between GROUP and ONSET COMPLEXITY (F[1, 92.41] = 4.32, p = .04). Pairwise comparisons between GROUP and ONSET COMPLEXITY revealed that the groups differed significantly in the duration of [l] between W_{simple} and $W_{complex}$ (p = .04). That is, participants who stutter showed a significantly bigger difference between the [l] duration of W_{simple} and $W_{complex.}$

Whether participants read at their preferred tempo or with a metronome did not affect [l] duration significantly (F[1, 79.39] = 3.85, p = .053). No further interactions became significant. These results suggest that, although the test group produced significantly longer [l] durations compared to the control group (given the slower speech rate in this group), they also showed more pronounced consonant compression.

Running the same model without the five participants who stutter who had comorbidities (Conditional R^2 = 0.538, Marginal R^2 = 0.278) did not change the significant main effects. GROUP (F[1, 84.76] = 12.06, $p < .001$) and ONSET COMPLEXITY (F[1, 87.38] = 285.82, $p < .001$) were still significant predictors of [l] duration. However, the interaction between GROUP and ONSET COMPLEXITY was no longer significant, indicating that the groups did not differ in consonant compression anymore.

In conclusion, these results suggest that we observe consonant compression in both groups, whereby the group who stutters produced more consonant compression. Moreover, the results suggest that the pacing condition did not have a significant effect on consonant compression, nor [l] duration in general since the model did not reveal a significant interaction with CONDITION and ONSET COMPLEXITY, nor a significant main effect.

3.3 Vowel compression

For vowel compression to occur, vowels of $W_{complex}$ must be shorter than W_{simple}. If the groups differ significantly in vowel compression, we would expect to find a significant interaction between ONSET COMPLEXITY and GROUP. *Figure 8* displays vowel duration in seconds for $W_{complex}$ and W_{simple} for each condition per group.

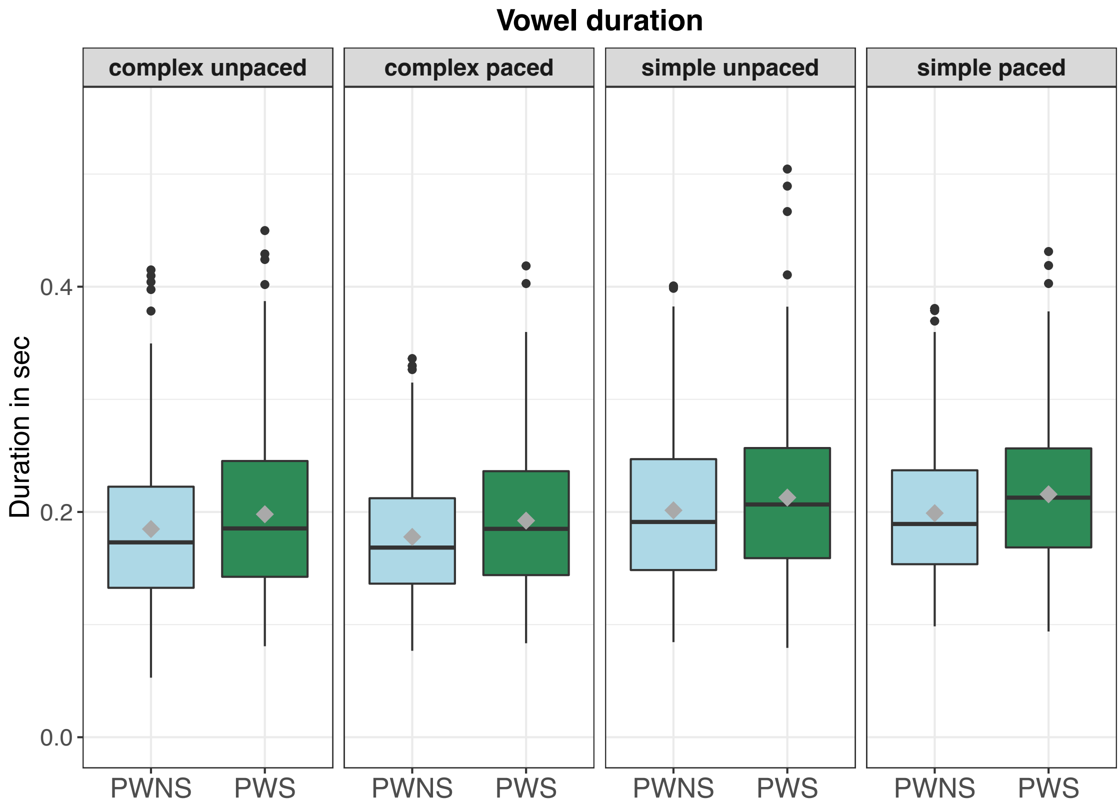


Figure 8: Vowel duration for each group per condition. Within each box, the median is denoted with black lines; boxes extend from the 25th to the 75th percentile of each group's distribution of values; the ends of the whiskers denote 1.5 interquartile range beyond the 25th and 75th percentile of each group; dots display observations outside the range of whiskers. Participants who do not stutter = blue, participants who stutter = green, mean values per group and condition are displayed with grey diamonds. Words with complex onsets are in the left two columns and words with a simple onset are in the right two columns.

The model (Conditional $R^2 = 0.848$, Marginal $R^2 = 0.031$) revealed that GROUP (F[1, 91.72] = 5.20, p = .025) and ONSET COMPLEXITY (F[1, 92.24] = 132.28, p < .001) were significant predictors of vowel duration. As can be seen in Figure 7, vowel durations are longer in W_{simple} compared to $W_{complex}$. Furthermore, the model revealed a significant interaction between ONSET COMPLEXITY and CONDITION (F[1, 2679.48] = 5.45, p = .02). Post-hoc Tukey corrected t-tests support vowel compression as it was found that vowels in words with a simple onset were always longer than in words with a complex onset, regardless of condition (p < .001). However, pairwise comparisons revealed that vowel compression significantly differed between the paced and the unpaced condition, indicating more vowel compression in the paced condition (p = .02). Nonetheless, the low marginal R^2 points out that most of the variance cannot be explained with the fixed effects. Running the same model without the five participants who had comorbidities (Conditional $R^2 = 0.848$, Marginal $R^2 = 0.032$) did not change the results.

Since there was no significant interaction between GROUP and ONSET COMPLEXITY, it can be assumed that the groups do not differ in vowel compression. To summarize, we can conclude that vowel compression occurred in both groups; furthermore, the pacing condition did not affect vowel duration but did affect vowel compression, as the latter was significantly more pronounced in the paced condition.

3.4 Acoustic c-center (intervals and RSDs)

3.4.1 Intervals

In this section we analyze whether the participants had timing patterns characteristic of a complex syllable organization (it is assumed that the intervals of W_{simple} and $W_{complex}$ differ less in the c-center interval compared to the other intervals), whether the groups differ in syllabic organization and whether the pacing condition affects this. Therefore, we compared the intervals of c-center, left-edge, and right edge of $W_{complex}$ with the W_{simple} interval (recall that the latter was the same across intervals; see *2.4.1.3 Intervals (acoustic c-center)*). The results are displayed in *Figure 9*. For this analysis, however, we had to exclude the word pair *Klang-lang*, since the onset of /k/ could not be determined based on the acoustic signal. Thus, the results on this section only include 3 word pairs.

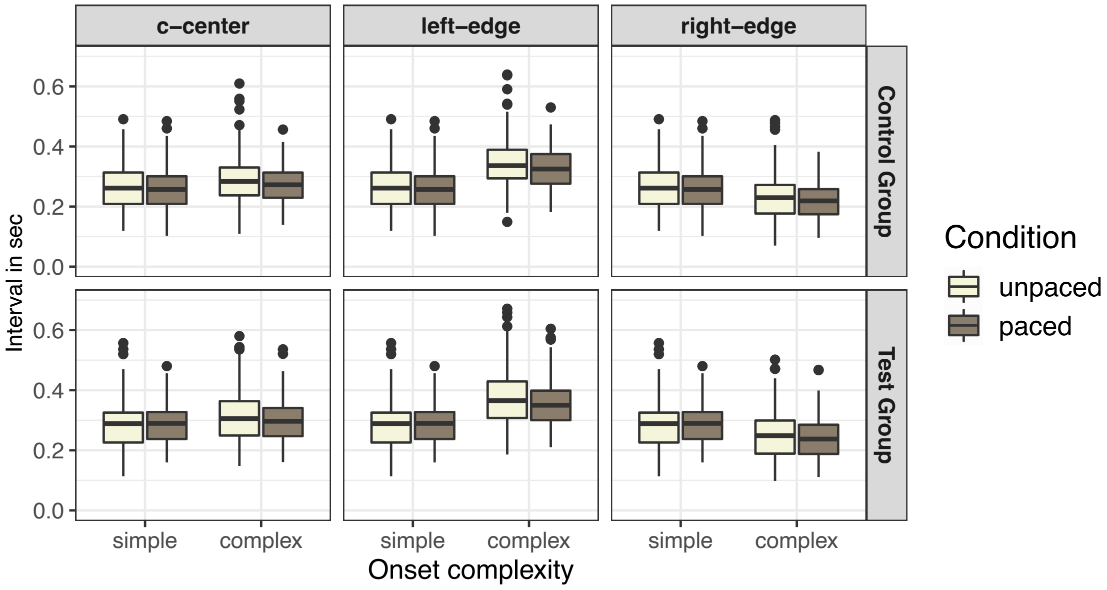


Figure 9: Intervals in sec to the vowel offset anchor point for c-center, left-edge, and right-edge (columns) per condition (unpaced = beige, paced = brown) for different onset complexity (x-axis), separated by group (rows). Within each box, the median is denoted with black lines; boxes extend from the 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.

Note that for W_{simple}, the interval for c-center, left-edge, and right-edge is the same. To examine whether the duration of the interval differs between W_{simple} and $W_{complex}$, whether the groups differ in interval duration, and whether the metronome affects the duration of the intervals, we performed a linear mixed effects analysis for each interval separately. Therefore, the dependent variable varied with respect to the interval (duration of c-center, left-edge or right-edge interval).

3.4.1.1 C-center interval

The model that predicted the duration of the c-center interval had a conditional R^2 of 0.812, and a marginal R^2 of 0.042. From this low marginal R^2 it can be concluded that the fixed effects do not explain much variation. Since the conditional R^2 value is quite high in comparison to the marginal R^2, it can be assumed that most of the variance can be explained with the random effects. A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 4: C-center*).

Nonetheless, the model did show that GROUP (F[1, 89.79] = 10.23, p = .002), ONSET COMPLEXITY (F[1, 89.06] = 46.5, p < .001), and CONDITION (F[1, 91.75] = 5.81, p = .02) were significant predictors of the c-center interval duration, reflecting the speech rate effects found in relation to word duration (see Result section 3.1 Word duration). Moreover, the model revealed a significant interaction between ONSET COMPLEXITY and CONDITION (F[1, 1944.93] = 12.30, p < .001). Pairwise comparisons showed that the c-center interval did differ significantly between W_{simple} and $W_{complex}$ in the unpaced condition (p < .001), and the paced condition (p < .001), whereby $W_{complex}$ intervals were longer than W_{simple} intervals. While speaking along with a metronome did not affect the c-center interval in W_{simple}, the interval became significantly shorter in the paced condition in $W_{complex}$ (p = .004). Thus, the c-center interval did not differ between W_{simple} that were produced in the unpaced condition and $W_{complex}$ that were produced in the paced condition (p = 0.4410) because the interval duration was similar. The results did not change when running the model without the five participants who had comorbidities.

3.4.1.2 Left-edge interval

A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 5: Left-edge*). The model that predicted the duration of the left-edge interval (Conditional R^2 = 0.825, Marginal R^2 = 0.200) revealed that the group who stutters had longer interval durations (F[1, 89.18] = 10.84, p = .001) and that $W_{complex}$ intervals were longer than W_{simple} intervals (F[1, 88.61] = 655.22, p < .001), mirroring the speech rate differences. Moreover, the metronome significantly decreased interval duration (F[1, 91.96] = 7.43, p =

0.008). A significant interaction effect between ONSET COMPLEXITY and CONDITION was also found for this interval (F[1, 1945.25] = 17.87, $p < .001$). The results mirror those of the c-center interval. However, the pairwise comparisons showed that the intervals of $W_{complex}$ are always longer than those of W_{simple} (when comparing W_{simple} unpaced to $W_{complex}$ unpaced ($p < .001$), W_{simple} unpaced to $W_{complex}$ paced ($p < .001$), and W_{simple} paced to $W_{complex}$ unpaced ($p < .001$), as well when comparing W_{simple} paced to $W_{complex}$ paced ($p < .001$)). The pacing condition did not significantly affect the left-edge interval of W_{simple} but the $W_{complex}$ interval ($p < .001$). The results did not change when running the model without the participants who had comorbidities.

The basic pattern for this section is as would be expected for a c-center organization: the left-edge shifts left for words with a complex vs. a simple onsets. The results also suggest that even though participants who stutter had longer left-edge intervals than participants who do not stutter, the groups did not differ in the difference between the W_{simple} and $W_{complex}$, as there was no significant interaction between GROUP and ONSET COMPLEXITY.

3.4.1.3 Right-edge interval

A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 6: Right-edge*). The model to predict the right-edge interval duration (Conditional $R^2 = 0.833$, Marginal $R^2 = 0.0086$) revealed the same main effects as the models for the c-center and the left-edge interval. However, according to the marginal R^2 the fixed effects did not explain much of the variance. Nevertheless, GROUP (F[1, 90.66] = 9.32, $p = .003$), ONSET COMPLEXITY (F[1, 90.03] = 298.87, $p < .001$), and CONDITION (F[1, 91.27] = 4.09, $p = .046$) were significant predictors of the right-edge interval duration pointing towards longer intervals in the group who stutters, longer intervals in W_{simple}, and shorter intervals in the pacing condition (as expected due to the speech rate effects). Furthermore, a significant interaction was found between ONSET COMPLEXITY and CONDITION (F[1, 1944.88] = 6.50, $p = .01$). Pairwise comparisons also showed the same effects but this time, the intervals of W_{simple} were longer than those of $W_{complex}$ (when comparing W_{simple} unpaced to $W_{complex}$ unpaced ($p < .001$), W_{simple} unpaced to $W_{complex}$ paced ($p < .001$), W_{simple} paced to $W_{complex}$ unpaced ($p < .001$), and W_{simple} paced to $W_{complex}$ paced ($p < .001$)). Furthermore, the interval of $W_{complex}$ became shorter in the paced condition ($p = .03$) but the metronome did not affect the interval duration of the W_{simple} interval. When running the model without the five participants who had comorbidities, CONDITION was no longer a significant predictor of the right-edge interval duration. The other results of the main effects did not change. In general, the results suggest that the pacing condition had the least impact on the right-edge interval.

Taking stock of the interval-based measures reported in this section it can be concluded that $W_{complex}$ c-center intervals were the closest to the W_{simple} intervals, pointing towards a complex syllable onset organization in both groups. C-center does shift slightly (about 18ms) from simple to complex onset, but the shifts for right and left edge are larger (39ms and 75ms).

3.4.2 RSDs

The RSD displays the variability of an interval over word pairs, such that the lower the RSD, the lower the variability. A c-center organization would be displayed in a more stable (smaller) RSD for the c-center interval. The following tables display the RSD for the different intervals and the number of word pairs (n) produced per group. The tables are separated by condition.

Table 4: RSD in the unpaced condition

Word pair	n	Group	RSD c-center	RSD left-edge	RSD right-edge
Schlamm -Lamm	198	PWNS	**19.79**	25.70	32.22
Schlauch-Lauch	193	PWNS	**19.78**	21.94	25.20
Schleim-Leim	194	PWNS	**22.30**	24.85	27.82
Schlamm -Lamm	171	PWS	**23.15**	27.49	31.61
Schlauch-Lauch	160	PWS	**21.01**	23.71	26.80
Schleim-Leim	176	PWS	**21.33**	24.09	26.88

Table 5: RSD in the paced condition

Word pair	n	Group	RSD c-center	RSD left-edge	RSD right-edge
Schlamm -Lamm	192	PWNS	**17.85**	22.90	29.99
Schlauch-Lauch	188	PWNS	**17.90**	19.66	23.15
Schleim-Leim	190	PWNS	**17.42**	19.97	23.53
Schlamm -Lamm	172	PWS	**17.48**	21.96	26.46
Schlauch-Lauch	174	PWS	**15.69**	17.33	20.22
Schleim-Leim	177	PWS	**16.32**	18.58	21.75

As can be seen in *Table 4* and *Table 5*, RSD was the lowest in the c-center interval in both groups and in both conditions, indicating that the c-center interval was the most stable one, thus supporting a c-center organization. However, RSD decreased in the paced condition in both groups, pointing towards more stability in the metronome condition. Furthermore, the RSD for all word pairs in the paced condition are less variable in participants who stutter, compared to the control group, indicating that the test group benefits more from the paced condition. While the control group improved stability by only 1.94% in the word pair *Schlamm-Lamm* and 1.84% in *Schlauch-Lauch*, the test group improved stability by almost 4.8% in *Schlamm-Lamm* and 5.3% in *Schlauch-Lauch*. For PWNS, the greatest improvement due to the paced condition in stability happened for the word pair *Schleim-Leim* where they improved by 4.88%. For the same word pair, PWS improved by 5.01%. The biggest difference between groups in RSD was in the unpaced condition for the word pair *Schlamm-Lamm*. The group who stutters was 3.35% less stable than the control group. In the word pair *Schleim-Leim* PWS were even less variable than PWNS in the paced condition (by 0.97%).

In order to test whether the groups differed significantly in the stability of the relevant intervals and if the pacing condition had an effect on stability, a linear mixed effects model was run to predict RSD. The model (Conditional $R^2 = 0.421$, Marginal $R^2 = 0.110$) revealed that INTERVAL (F[1, 1455.23] = 130.46, p < .001) and CONDITION (F[1, 84.06] = 4.95, p = .009) were significant predictors of the RSD, indicating that left-edge and right-edge intervals had a higher RSD compared to the c-center interval and that the paced condition reduced variability significantly. Furthermore, significant interactions were found between GROUP and INTERVAL (F[2, 1455.23] = 4.95, p = .007) and CONDITION and INTERVAL (F[2, 1455.23] = 4.31, p = .01). Post-hoc tests showed that the groups did not differ significantly in interval stability of the same intervals. However there were within group differences, namely that the variability of the left-edge interval did not differ from the variability of the right-edge interval in the group who stutters, while in the control group, there was a significant difference in RSD between these two intervals (p < .001), pointing towards more variability in the left-edge interval compared to the right-edge interval. The post hoc test regarding the significant interaction between CONDITION and INTERVAL showed that the metronome only significantly improved stability in the left-edge interval (p = .02), nearly significantly in the c-center interval (p = .054), and not significantly in the right-edge interval (p = .096).

Hence, it can be concluded that the variability was the lowest in the c-center interval, pointing towards a c-center organization. Moreover, results suggest that PWS did not differ from PWNS in terms of variability and that the paced condition only increased stability in the left-edge interval but not in the c-center, nor the right-edge interval.

4. Discussion

The aim of the present study was to analyze temporal organization of syllables in children and adolescents who stutter and children and adolescents who do not stutter, speaking with and without an external rhythm (metronome). Therefore, participants were asked to read wordlists which contained German monosyllabic words that differed in onset complexity in their own preferred tempo (unpaced condition) and along with a metronome (paced condition). Four minimal pairs (*Klang-lang*, *Schlamm-Lamm*, *Schlauch-Lauch*, *Schleim-Leim*) were analyzed. We focused on syllabic timing related to articulatory timing and analyzed acoustic cues of a *c-center effect* for which our participants indeed showed evidence. Examining the c-center effect in PWS is particularly interesting, as it can provide evidence for difficulties in articulatory control. This study presents novel findings since this is the first study (to our knowledge) that examines these effects in a population who stutters.

We found consonant and vowel compression effects in both groups, as well as support for a complex syllable onset organization as indicated by lower relative standard deviations in c-center intervals compared to left-edge or right-edge intervals. These results point towards a complex syllable onset organization in German children and adolescents who stutter and German children and adolescents who do not stutter, indicating a shift of the rightmost consonant in an onset cluster towards the vowel.

Moreover, speaking along with a metronome did not affect compression effects in consonants, but there was more vowel compression in the paced condition. Furthermore, speaking along with a metronome improved durational interval stability. A group difference was observed with respect to consonant compression indicating that children and adolescents who stutter showed a bigger difference between [l] in words with a simple onset and words with a complex onset than the control group when all participants were included in the analyses.

Our results indicate that the groups do not differ in general (articulatory) syllable organization in perceptually fluent speech, supporting hypothesis (1) and (2), as we did find both consonant and vowel compression in the group who stutters and the control group. These results point towards a c-center organization in syllable articulation. As hypothesized, participants who stutter differed from participants who do not stutter in consonant compression (but not in vowel compression) which suggests that children and adolescents who stutter time onset consonant gestures differently.

According to neurophonetic models of stuttering, such as the GODIVA model (Civier, Bullock, Max, & Guenther 2013), the initiation of the articulatory gestures within a syllable organization is atypical in PWS and thus may lead to stuttering symptoms. Hence, the timing of consonantal onset gestures seems to be particularly challenging for PWS. However, it should be mentioned that the GODIVA model specifically addresses stuttering events while the present study focuses on fluent speech only. In the GODIVA model, stuttering events are interpreted as failures to activate the next syllable's motor program in time (Civier et al., 2013). The neural circuit involved in initiating and terminating syllables consists of basal ganglia, thalamus, and left ventral premotor cortex (Civier et al., 2013). Toyomura and colleagues (2015) found that PWS's basal ganglia activity (which is an indication of motor control) increased to the level of PWNS's after practicing to speak along with a metronome over a period of 8 weeks for at least 15 minutes per day and at least 5 days per week. In our study, the significant difference between groups in consonant compression underlines differences in motor control, particularly gestural timing of syllable onsets. The group difference was present regardless of speaking with or without a metronome, which indicates that differences in motor control might even be present in a fluency-enhancing condition. A future study could analyze the temporal syllabic structure in long-term fluency-enhancing effects. More specifically, it could be investigated whether the group difference regarding consonant compression would be cancelled out after a period of 8 weeks regular practice of speaking along with a metronome. As Toyomura et al. (2015) showed, basal ganglia activity did not differ between PWS and PWNS after this period of time, suggesting that this had led to similar syllable initiation patterns.

Since the groups did not differ in vowel compression, but children and adolescents who stutter showed more consonant compression than the control group, it points towards more gestural overlap between the right-most consonant in an onset cluster and the following vowel in the group who stutters. This could support Harrington's (1988) model of stuttering in which he suggests that stuttered speech occurs because individuals who stutter incorrectly apply their perceptual predictive timing to their own speech production output. According to his theoretical viewpoint, PWS expect the time of sensory feedback to occur earlier than it actually does and, thereby, they would erroneously correct for the moment of their actual segmental production; this then would lead to stuttering because the articulatory movements are too much in conflict with each other (e.g., simultaneous instruction to close the lips and to lower the jaw) (Harrington, 1988). Higher overlap between [l] and the following vowel in words with a complex onset could indicate altered predictive timing mechanisms, resulting in an atypical inter-gestural timing. However, Harrington's theory remains to be tested for syllables with different onset complexities since the model is mainly based on the coupling of one onset consonant and the following vowel. If PWS would coarticulate less, for instance as a strategy to speak more fluently (e.g., Zmarich et al., 2013), this would have led to less consonantal compression in the group who stutters. Another articulatory explanation that would lead to a greater consonant compression is shortening of the [l] gesture in words with a complex onset. This could be either a consequence of altered articulatory timing in PWS or even a strategy for PWS to speak more fluently. Conducting an articulatory study, using for instance electromagnetic articulography, could clarify whether more overlapping between the right-most consonant gesture and the vowel gesture causes more consonant compression in PWS, or whether it is the shortening of the second consonant gesture in a complex onset (CC) that leads to more compression in the group who stutters.

With respect to our results, we should keep in mind that the group difference was not very strong as consonant compression did not differ between groups anymore when the five participants who stutter with comorbidities (Dyslexia, ADHD) were excluded. However, including the five participants who had comorbidities displays a more realistic reflection of the population who stutters as more than 60% of children who stutter have co-occurring speech, language, or non-speech-language disorders, as a study with 2628 American children revealed (Blood, Ridenour, Qualls, & Scheffner Hammer, 2003). Non-speech-language disorders, which affect around 34.3% of the children, include for instance attention deficit disorders (5.9% of the 34.3%) and literacy disorders (8.2% of the 34.3%) (Blood et al., 2003). In our view, an articulatory study (i.e., using articulography) will clarify the strength of consonant compression in young persons who stutter. In addition, it would be especially interesting to look into the phenomenon that participants who produced long [l] durations in words with a simple onset, produced shorter [l] durations in words with a complex onset. These were individuals who displayed greater consonant compression, which was observed independently of group.

The result of more consonant compression in the group who stutters suggests that stuttering may be related to issues with the coordination of speech motor movements, specifically with the timing and sequencing of the movements required for producing onset consonants. This finding may also be relevant for the neurosciences studying the underlying neural mechanisms of stuttering. For instance one might further investigate how the basal ganglia-thalamo-cortical circuit is involved in initiating syllables with different onset complexity in different rhythmic conditions. Clinically, the knowledge of increased consonant compression in individuals who stutter may inform the development of targeted therapy interventions. For example, speech therapists currently teach techniques that prolong the onset of syllables to modify or prevent stuttered disfluencies. It could be a complementary avenue for clinical research to explore the efficiency of techniques that specifically target the coordination of speech motor movements in fluent speech, such as training paced and unpaced fluent speech over a longer period of time (like in the study by Toyomura et al., 2015), to foster objective and subjective articulatory control in individuals who stutter.

Furthermore, contrary to our hypothesis, children and adolescents who stutter were as stable in their syllable organization as children and adolescents who do not stutter since the groups did not differ significantly in interval stability. This means that the group who stutters and the control group had similar durational variability in the c-center to anchor, left-edge to anchor, and right-edge to anchor intervals.

Concerning the effect of a regular rhythm on the temporal organization of speech, speaking along with a metronome did not affect consonant compression but it enlarged vowel compression. This result can be interpreted in the light of the effect that vowel durations in particular are affected by changes in speech rate in PWS (e.g., Davidow, 2014). In the paced condition, participants may have increased their speech rate in words with a complex onset even more than in words with a simple onset, as they had one sound more to produce and wanted to be in time with the metronome. Hence, vowels in words with a complex onset would become even shorter in the paced condition. In this way, paced speech would not only affect speech rate but also vowel timing patterns. To clarify rate variations and their interaction with syllable timing, a future study could focus on investigating compression effects in longer utterances consisting of multiple syllables. Syllable durations will be produced according to the needs of the utterance, where any particular syllable could be expanded or compressed according to those needs. This would allow the study of consonant and vowel compression in context, as well as the contribution of rate variation.

Moreover, the paced condition reduced variability of durational intervals, matching other studies that found reduced variability in fluency-enhancing conditions, such as metronome-paced speech and singing (e.g., a decreased variability of duration of voiced and voiceless segments in metronome-paced speech [Janssen & Wieneke, 1987] and a decreased variability in voice onset time in word-initial stressed positions when PWS were singing [Falk, Maslow, Thum, & Hoole, 2016]). In metronome-timed speech, we found reduced variability especially in the left-edge interval which points towards a reduction in durational variability of the fricative [ʃ]. In addition, the hypothesis that the group who stutters benefits more from the paced condition leading to a higher increase in stability could be confirmed.

The study had limitations with respect to the number and type of word pairs. Our results are mostly based on fricative-lateral [ʃl] onsets in words with a complex onset and on the lateral [l] in words with a simple onset. However, word pairs that differ in onset complexity are not easy to find, especially with the limitation of an acoustic analysis; for instance, onsets with a plosive in C_2 position in words with a complex onset had to be excluded since these would be the onset consonant in words with a simple onset and hence, plosive onsets cannot be determined based on the acoustic signal only. Therefore, an articulatory study would allow to include more diverse word pairs. Onsets that start with a plosive might be more difficult to initiate for PWS than onsets that start with a sibilant or lateral. For English it was found that consonant manner and consonant place were predictors of stuttering rate (Howell, Au-Yeung, Yaruss, & Eldridge, 2006). The first consonant in a cluster could therefore also impact the following consonant(s). This can be addressed in future work where more word pairs of different cluster combinations should be included. Another aspect that should be taken into account when analyzing the *c-center effect* in the future is that c-center stability was found to be more frequent in words that contain tense vowels than lax vowels or diphthongs (Brunner et al., 2014).

Finally, as in other studies on stuttering in a school-age or older population, we had more male than female participants. This imbalance reflects a general bias in the population with persistent stuttering or with a risk of persistence (Yairi & Ambrose, 2013). Our participants were more likely to show persistent stuttering as they were 9 years and older. Hence, we cannot entirely exclude that the articulatory effects found in the current study are more visible in the male population. For example, it is known from the literature on younger children between the ages of 4 and 5 years and 11 months that boys who stutter show greater lags in speech motor development than girls who stutter, compared to their peers (Walsh, Mettel, & Smith, 2015). Moreover, there are differences in speech-related brain regions between boys and girls who stutter between the ages of 3 and 10 years, possibly contributing to the fact that girls are more likely to recover (Chang, Zhu, Choo, & Angstadt, 2016). Overall, it should be an aim for future research to better understand individual patterns (for example age, education, reading skills) in fluent and disfluent speech of stuttering – a general need for studies on rare populations.

In sum, this study is a promising basis for conducting an articulatory study in which articulatory (gestural) timing can be examined in more detail. This is an important step in the investigation of motor control in stuttering and it will greatly help to understand the underlying motor patterns in PWS.

5. Conclusion

The results of our study suggest that the temporal organization of the syllable is similar in children and adolescents who stutter vs. children and adolescents who do not stutter, regardless of speaking in their own preferred speech tempo or along with a metronome. Moreover, paced speech improved durational interval stability in both groups. However, the group who stutters produced more consonant compression than the control group, suggesting differences in articulatory onset timing.

Funding

This work was supported by DFG grants [FA 901/4–1 and HO 3271/6–1], the Graduate School "Class of Language", Ludwig Maximilian University of Munich, and the German Academic Exchange Service (DAAD). It was also supported by the Canada Research Chair on Interdisciplinary studies on rhythm and language acquisition (SSHRC), Canada.

Declaration of Competing 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.

Data Availability Statement

Data will be made available on request.

Acknowledgements

We thank all our participants for their participation, Ingeborg Meyer, Georg Thum, and the entire SAS team for their help with recruiting, as well as Michele Gubian for his support with statistics.

Appendix

Description of Predictors that can be found in the following models:

group[s]: PWS

group[c]: PWNS

nonset[2]: $W_{complex}$

nonset[1]: W_{simple}

met[1]: paced condition

met[0]: unpaced condition

contrastpair: word pair

A-Model 1: Word duration

	Word_dur		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.45	0.38 – 0.51	**<0.001**
group [s]	0.06	0.03 – 0.10	**<0.001**
nonset [2]	0.07	-0.02 – 0.16	0.131
met [1]	-0.01	-0.03 – 0.01	0.158
*group [s] * nonset [2]*	-0.01	-0.02 – 0.01	0.267
*group [s] * met [1]*	-0.00	-0.03 – 0.02	0.834
*nonset [2] * met [1]*	-0.02	-0.03 – -0.00	**0.008**
Random Effects			
σ^2	0.00		
$\tau_{00\ participant}$	0.01		
$\tau_{00\ Word}$	0.00		
$\tau_{11\ participant.nonset2}$	0.00		
$\tau_{11\ participant.met1}$	0.00		
$\tau_{11\ Word.met1}$	0.00		
$\rho_{01\ participant.nonset2}$	-0.20		
$\rho_{01\ participant.met1}$	-0.80		
$\rho_{01\ Word}$	-0.92		
ICC	0.72		
$N_{participant}$	96		
N_{Word}	8		
Observations	2909		
Marginal R^2 / Conditional R^2	0.130 / 0.758		

A-Model 2: l duration

	onset_comp		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.10	0.09 – 0.11	**<0.001**
group [s]	0.02	0.01 – 0.03	**0.002**
nonset [2]	-0.04	-0.05 – -0.03	**<0.001**
met [1]	-0.00	-0.01 – 0.00	0.382
*group [s] * nonset [2]*	-0.01	-0.02 – 0.00	**0.038**
*group [s] * met [1]*	-0.00	-0.01 – 0.01	0.855
*nonset [2] * met [1]*	-0.00	-0.01 – 0.00	0.404
Random Effects			
σ^2	0.00		
$\tau_{00\ participant}$	0.00		
$\tau_{00\ contrastpair}$	0.00		
$\tau_{11\ participant.met1}$	0.00		
$\tau_{11\ participant.nonset2}$	0.00		
$\rho_{01\ participant.met1}$	-0.55		
$\rho_{01\ participant.nonset2}$	-0.85		
ICC	0.34		
$N_{participant}$	96		
$N_{contrastpair}$	4		
Observations	2909		
Marginal R^2 / Conditional R^2	0.281 / 0.528		

A-Model 3: Vowel duration

	Vowel_dur		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.20	0.15 – 0.25	**<0.001**
group [s]	0.01	0.00 – 0.03	0.075
nonset [2]	-0.02	-0.02 – -0.01	**<0.001**
met [1]	-0.00	-0.01 – 0.01	0.557
*group [s] * nonset [2]*	-0.00	-0.01 – 0.00	0.626
*group [s] * met [1]*	0.00	-0.01 – 0.01	0.716
*nonset [2] * met [1]*	-0.00	-0.01 – -0.00	**0.020**
Random Effects			
σ^2	0.00		
$\tau_{00\ participant}$	0.00		
$\tau_{00\ contrastpair}$	0.00		
$\tau_{11\ participant.met1}$	0.00		
$\tau_{11\ participant.nonset2}$	0.00		
$\rho_{01\ participant.met1}$	-0.61		
$\rho_{01\ participant.nonset2}$	-0.35		
ICC	0.84		
$N_{participant}$	96		
$N_{contrastpair}$	4		
Observations	2909		
Marginal R^2 / *Conditional* R^2	0.031 / 0.848		

A-Model 4: C-center

	ccenter		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.27	0.20 – 0.33	**<0.001**
group [s]	0.02	0.00 – 0.05	**0.016**
nonset [2]	0.02	0.02 – 0.03	**<0.001**
met [1]	-0.01	-0.02 – 0.01	0.353
*group [s] * nonset [2]*	-0.00	-0.01 – 0.01	0.775
*group [s] * met [1]*	0.00	-0.01 – 0.02	0.818
*nonset [2] * met [1]*	-0.01	-0.02 – -0.00	**<0.001**
Random Effects			
σ^2	0.00		
$\tau_{00\ participant}$	0.00		
$\tau_{00\ contrastpair}$	0.00		
$\tau_{11\ participant.nonset2}$	0.00		
$\tau_{11\ participant.met1}$	0.00		
$\varrho_{01\ participant.nonset2}$	-0.19		
$\varrho_{01\ participant.met1}$	-0.72		
ICC	0.80		
$N_{participant}$	96		
$N_{contrastpair}$	3		
Observations	2185		
Marginal R^2 / *Conditional* R^2	0.042 / 0.812		

A-Model 5: Left-edge

	left_edge		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.27	0.20 – 0.33	**<0.001**
group [s]	0.03	0.01 – 0.05	**0.014**
nonset [2]	0.08	0.07 – 0.09	**<0.001**
met [1]	-0.00	-0.02 – 0.01	0.435
*group [s] * nonset [2]*	0.00	-0.01 – 0.01	0.667
*group [s] * met [1]*	0.00	-0.02 – 0.02	0.990
*nonset [2] * met [1]*	-0.01	-0.02 – -0.01	**<0.001**
Random Effects			
σ^2	0.00		
$\tau_{00\ participant}$	0.00		
$\tau_{00\ contrastpair}$	0.00		
$\tau_{11\ participant.nonset2}$	0.00		
$\tau_{11\ participant.met1}$	0.00		
$\varrho_{01\ participant.nonset2}$	0.03		
$\varrho_{01\ participant.met1}$	-0.75		
ICC	0.78		
$N_{participant}$	96		
$N_{contrastpair}$	3		
Observations	2185		
Marginal R^2 / Conditional R^2	0.200 / 0.825		

A-Model 6: Right-edge

	right_edge		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.27	0.20 – 0.33	**<0.001**
group [s]	0.02	0.00 – 0.04	**0.020**
nonset [2]	-0.03	-0.04 – -0.03	**<0.001**
met [1]	-0.01	-0.02 – 0.00	0.284
*group [s] * nonset [2]*	-0.01	-0.01 – 0.00	0.266
*group [s] * met [1]*	0.00	-0.01 – 0.02	0.634
*nonset [2] * met [1]*	-0.01	-0.01 – -0.00	**0.011**
Random Effects			
σ^2	0.00		
$\tau_{00\ participant}$	0.00		
$\tau_{00\ contrastpair}$	0.00		
$\tau_{11\ participant.nonset2}$	0.00		
$\tau_{11\ participant.met1}$	0.00		
$\rho_{01\ participant.nonset2}$	-0.41		
$\rho_{01\ participant.met1}$	-0.70		
ICC	0.82		
$N_{participant}$	96		
$N_{contrastpair}$	3		
Observations	2185		
Marginal R^2 / *Conditional* R^2	0.086 / 0.833		

References

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.

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

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.

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

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory. Phonology. *Phonetica*, 45, 140-55.

Browman, C. P., & Goldstein, L. (1991). Gestural structures: distinctiveness, phonological processes, and historical change. In I. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 313-338). New Jersey: Erlbaum.

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

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, 5, 25-34.

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

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24, 209-244.

Chang, S. E., Zhu, D. C., Choo, A. L., & Angstadt, M. (2015). White matter neuroanatomical differences in young children who stutter. *Brain: a journal of neurology*, *138*(Pt 3), 694–711. https://doi.org/10.1093/brain/awu400

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.

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), 100-12. doi: 10.1111/1460 6984.12050. Epub 2013 Aug 24. PMID: 24372888; PMCID: PMC4461240

De Nil, L. F., & Brutten, G. J. (1991). Voice onset times of stuttering and nonstuttering children: The influence of externally and linguistically imposed time pressure. *Journal of Fluency Disorders*, 16, 143-158.

Di Simoni, F. G. (1974). Letter: preliminary study of certain timing relationships in the speech of stutterers. *The Journal of the Acoustical Society of America*, 56(2), 695-696.

Dokoza, K. P., Hedever, M., & Sarić, J. P. (2011). Duration and variability of speech segments in fluent speech of children with and without stuttering. *Collegium Antropologicum*, 35(2), 281-288.

Falk, S., Maslow, E., Thum, G., & Hoole, P. (2016). Temporal variability in sung productions of adolescents who stutter. *Journal of Communication Disorders*, 62, 101-114.

Gibson, M., Fernández Planas, A. M., Gafos, A., & Remirez, E. (2015). Consonant duration and VOT as a function of syllable complexity and voicing in a subset of Spanish clusters. Interspeech 2015. [online] https://www.ling.uni-potsdam.de/~gafos/papers/Gibson-Gafos Interspeech2015.pdf (last accessed: 09/23/22)

Goldstein, L., & Pouplier, M. (2014). The temporal organization of speech. In F. M. Goldrick, & M. Miozzo (Eds), *The Oxford Handbook of Language Production* (pp. 210-227). Oxford: Oxford University Press.

Hall, N. (2010). Articulatory Phonology. *Language and Linguistics Compass*, *4(9)*, 818-830, doi: 10.1111/j.1749-818x.2010.00236.x

Harrington, J. M. (1987). Coarticulation and stuttering: an acoustic and electropalatographic study. In H. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 381-392). New York: Springer Verlag.

Harrington, J. M. (1988). Stuttering, Delayed Auditory Feedback, and Linguistic Rhythm. *Journal of Speech and 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, DOI: 10.3109/02699206.2015.1100684

Hoole, P., & Pouplier, M. (2015). *Interarticulatory coordination - speech sounds*. In M. Redford (Ed.), *The Handbook of Speech Production*, (ch. 7, pp. 133-157). Hoboken, New Jersey: John Wiley & Sons.

Howell, P., Au-Yeung, J., Yaruss, J. S., & Eldridge, K. (2006). Phonetic difficulty and stuttering in English. *Clinical linguistics & phonetics*, *20*(9), 703-716. https://doi.org/10.1080/02699200500390990

Janssen, P., & Wieneke, G. (1987). The effects of fluency inducing conditions on the variability in the duration of laryngeal movements during stutterer's fluency speech. In H.F.M. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (p. 337-344). New York: Springer-Verlag.

Jäncke, L. (1994). Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults. *Journal of Fluency Disorders* , 19(1), 21-37.

Katz, J. (2010). Compression effects, perceptual asymmetries, and the grammar of timing. Dissertation (www.researchgate.net/publication/265231502), last accessed: 09/23/22.

Katz, J. (2012). Compression effects in English. *Journal of Phonetics*, 40, 390-402.10.1016/j.wocn.2012.02.004.

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

MacMillan, V., Kokolakis, A., Sheedy, S., & Packman, A. (2014). End-Word dysfluencies in young children: A clinical report. *Folia Phoniatrica et Logopaedica*, 66, 115-125. doi: 10.1159/000365247

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, 14(3), 380-407.

Marin, S., & Bučar Shigemori, L. S. (2014). Vowel compensatory shortening in Romanian. [online], https://www.phonetik.uni-muenchen.de/universals/pub/ISSP2014_VCS_Marin_BucarShigemori.pdf (last accessed: 09/23/22).

Max, L., Caruso, A. J., & Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech Language, and Hearing Res*earch, 46(1), 215-32. doi: 10.1044/1092-4388(2003/017). PMID: 12647900.

Max, L., & Gracco, V. L (2005): Coordination of oral and laryngeal movements in the perceptually fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research*, 48(June), 524-542.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator of syllable structure. *Proc. XVth International Congress of Phonetic Sciences*, Barcelona, 3-9 Aug 2003.

Namasivayam, A. K., & van Lieshout, P. (2008). Investigating speech motor practice and learning in people who stutter. *Journal of Fluency Disorders*, 33, 32-51.

Peters, S., & Kleber, F. (2014). Articulatory mechanisms underlying incremental compensatory vowel shortening in German. ISSP 2014. [online] https://www.phonetik.uni-muenchen.de/personen/mitarbeiter/kleber_felicitas/publikationen/peters_kleber14_issp.pd (last accessed: 09/23/22)

Pouplier, M. (2012). The gestural approach to syllable structure: universal, language- and cluster-specific aspects. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.) *Speech planning and dynamics* (pp. 63-96). Berlin: Peter Lang.

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/.

Ruthan, M. Q., Durvasula, K., & Lin, Y.-H. (2019). Temporal coordination and sonority of Jazani Arabic word-initial clusters. *Proceedings of the Annual Meetings on Phonology*, 7, 10.3765/amp.v7i0.4485.

Shaw, J., Gafos, A., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: A case study of Moroccan Arabic consonant clusters. *Phonology, 28*(3), 455-490. doi:10.1017/S0952675711000224

Selkirk, E., & Durvasula, K. (2013). Acoustic correlates of consonant gesture timing in English. *The Journal of the Acoustical Society of America*, 134(5), 4202-4202.

Smith, A., Sadagopan, N., Walsh, B., & Weber-Fox, C. (2010). Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. *Journal of Fluency Disorders*, 35, 1-18.

Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: Evidence from behavioral and kinematic indices of nonword repetition performance, *Journal of Fluency Disorders*, 37(4), 344-358. https://doi.org/10.1016/j.jfludis.2012.06.001.

Smith, A., & Weber, C. (2016). Childhood Stuttering: Where Are We and Where Are We Going? *Seminars in Speech and Language*, 37(4), 291-297. doi: 10.1055/s-0036-1587703.

Stager, S. V., Jeffries, K. J., & Braun, A. R. (2003). Common features of fluency-evoking conditions studied in stuttering subjects and controls: an H215O PET study. *Journal of Fluency Disorders*, 28, 219-336.

Tilsen, S. (2009). Multitimescale Dynamical Interactions Between Speech Rhythm and Gesture. *Cognitive Science,* 33, 839-879.

Toyomura, A., Fuji, T., & Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *Neuroimage,* 57, 1507-1516.

Toyomura, A., Fujii, T., & Kuriki, S. (2015). Effect of an 8-week practice of externally triggered speech on basal ganglia activity of stuttering and fluent speakers. *NeuroImage*, *109*, 458–468. https://doi.org/10.1016/j.neuroimage.2015.01.024

Usler, E., Smith, A., & Weber, C. (2017). A Lag in Speech Motor Coordination During Sentence Production Is Associated With Stuttering Persistence in Young Children. *Journal of Speech, Language, and Hearing Research*, 60(1), 51-61. https://doi.org/10.1044/2016_JSLHR-S-15-0367

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, 2157-2167. doi:10.1044/2018_JSLHR-S-17-0435.

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.

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

Wickham H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Milton Bache, S., Müller, K., Ooms, J., Robinson, D., Paige Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. https://doi.org/10.21105/joss.01686

Wiltshire, C. E. E., Chiew, M. Chesters, J., Healy, M., & Watkins, K. E. (2021). Speech movement variability in people who stutter: A coval tract magnetic resonance imaging study. *Journal of Speech, Language, and Hearing Research,* 64, 2438-2452.

Wingate, M. E. (1969). Sound and pattern in "artificial" fluency. *Journal of Speech Language and Hearing Research*, 12, 677-686.

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

World Health Organization (WHO) (2015). The ICD-10 Classification of Mental and Behavioral Disorders; F98.5 Stuttering; Geneva, Switzerland: WHO.

Yairi, E., & Ambrose, N. G. (2013): Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, 38, 66-87.

Zmarich, C., Balbo, D., Galatà, V., Verdurand, M., & Rossato, S. (2013). The production of syllabes in stuttering adults under normal and altered auditory feedback. In V. Galatà (Ed.), *Multimodalità e Multilingualità: la Sfida più Avanzata della Comunicazione Orale* (9th ed., pp. 463-474. Rome, Italy: Bulzoni editore.

2.3. Discussion

The objective of this study was to offer insights into articulatory timing in children and adolescents who stutter and children and adolescents who do not stutter by examining the acoustic manifestation of a c-center effect. Secondly, this study aimed to explore the relationship between inter-gestural timing and metronome-paced speech.

One of the main findings was that children and adolescents who stutter showed greater consonant compression regardless of speaking with or without a metronome, pointing towards differences in the timing of onset consonant gestures.[3] Note that the group differences with regard to consonant compression were considerably more robust than those for vowel compression. This can be most easily observed by inspecting the confidence intervals (see *A-Model 2: l duration* vs. *A-Model 3: Vowel duration* in the Appendix of the paper, presented in *Chapter 2*) where the magnitude of the compression effect is twice as large in consonants than in vowels. The consonant compression finding aligns with previous research indicating that children who stutter display more articulatory coordination challenges across various speech tasks (Usler et al., 2017; Usler & Walsh, 2018; Smith et al., 2012). Increased consonant compression may result from higher gestural overlap between the vowel and the pre-vocalic consonant in words with complex onsets. Notably, higher gestural overlap has been associated with a less mature speech motor system: children who stutter produced more overlap than children who do not stutter (Lenoci & Ricci, 2018) and typically developing children showed more gestural overlap than adults who do not stutter (Noiray et al., 2018). Thus, this interpretation supports the view that PWS are generally at the lower end of a speech motor skill continuum (Namasivayam & van Lieshout, 2011; van Lieshout et al., 2004).

Of course, from our results on perceptually fluent speech of PWS we cannot conclude whether the assumed higher gestural overlap would also be present in their disfluent speech. Importantly, the notion of increased gestural overlap is an inference based on acoustic measures (consonant compression) and remains to be confirmed with articulatory data. Hence, an articulatory study would be needed to verify whether more consonant compression in PWS does indeed reflect greater overlap of articulatory gestures or simply a shorter gesture for the pre-vocalic consonant. The latter scenario could be attributed to biomechanical shortening, as proposed by Mücke et al. (2020). Since our findings are primarily based on the cluster /ʃl/, where the tongue is already in a high position for /ʃ/, the required movement of the tongue tip to reach the alveolar ridge

[3] While PWS showed greater consonant compression overall, a longer /l/ in CV contexts was associated with a shorter /l/ in CCV contexts across groups. However, I do not have an explanation for this pattern.

for /l/ is rather short. This reduced articulatory distance likely contributes to shorter segmental durations of C2 on the acoustic surface.

Harrington's (1988) model of stuttering proposes that excessive gestural overlap between onset consonants and vowels leads to stuttering. In fact, we would expect to find even greater overlap in disfluent compared to fluent speech. While Harrington's (1988) model only addresses syllables with a simple onset, our results suggest that more gestural overlap in PWS is even present in syllables with a complex onset. This raises the question of how such articulatory patterns might manifest across different languages and how it affects stuttering across languages.

Future research could explore whether languages with different syllable complexity, such as Italian, which favors open CV syllables vs. Polish, which features dense consonant clusters in both syllable onset and offset position, would inherently lead to more or fewer stuttering symptoms. Stuttering is observed across languages worldwide (Yairi & Ambrose, 2013). However, studies on bilinguals who stutter have reported language-specific differences in stuttering that are attributed to variations in linguistic and phonetic structures (for a review, Chaudhary et al., 2021). In their comprehensive review, Chaudhary and colleagues (2021) note that while such structural factors are frequently cited as influential, empirical evidence directly linking language structure to stuttering severity remains limited. Although structural differences between languages may plausibly impact stuttering, few studies have experimentally isolated or systematically examined variables such as language complexity and typological features, underscoring a research gap in this area (Chaudhary et al., 2021).

Given the findings of the present study, an interesting next step would be to investigate whether these stronger consonant compression effects persist into adulthood. Adults have a more skilled speech motor control system compared to children (Smith & Zelaznik, 2004) which could reduce variability in syllabic organization in general. In contrast to children who stutter, adults who stutter have lived with the speech motor disorder for a longer period and could have developed compensatory strategies, such as the prolongation of speech sounds, commonly taught in speech therapy (e.g., Georgieva & Stoilova, 2018; O'Brian et al., 2003; Onslow et al., 1996). Such strategies, that are probably not even perceivable when practiced for a long time, could result in a modified syllable organization, potentially leading to greater differences between participants who stutter and those who do not.

Another key finding was that the temporal patterns we found aligned with the predictions of a c-center organization, including acoustic vowel and consonant compression. These patterns emerged in both PWS and PWNS, as well as in paced and unpaced speech, suggesting that gestural organization is robustly established by early adolescence. This finding is particularly valuable, as little is known about the development of syllabic coordination principles. Future research could, for example, compare children, adolescents, and adults to provide critical insights into the developmental trajectory of speech motor patterns related to syllable organization. Additionally, it would also shed light on whether group differences in articulatory timing decrease or intensify as speech motor control matures, potentially revealing a pattern of improvement, specifically in the control group.

While the metronome did not eliminate the group differences found in consonant compression, children and adolescents who stutter exhibited a higher increase in interval stability when speaking along with a metronome compared to the control group. This leads to the assumption that articulatory timing differences persist even in metronome-timed speech.

One interpretation, as discussed in the paper, is that altered predictive timing (Harrington, 1988) could result in greater gestural overlap between the right-most consonant and the following vowel in participants who stutter. Hence, this alteration may impair their ability to anticipate and execute motor gestures with precise timing. Notably, differences in predictive timing have also been observed in non-verbal synchronization tasks, such as finger tapping to an external beat, between individuals who stutter and individuals who do not stutter (Falk et al., 2015; Hulstijn et al., 1992; Olander et al., 2010; Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017). These findings suggest that timing issues in stuttering extend beyond speech to other motor domains.

The next two chapters will delve into the examination of articulatory timing in adults who stutter and adults who do not stutter, under different rhythmic conditions. Specifically, the upcoming chapter aims to determine whether speech motor timing is influenced by the rhythmic context (e.g. metronome vs. finger tap) and to examine whether timing differences emerge between adults who stutter and those who do not stutter across these conditions. More specifically, *Chapter 3* investigates speech onset timing articulatorily with respect to a metronome beat (metronome condition), to finger tapping (tapping condition), and to a combination of both (tapping to a metronome while speaking).

Chapter 3

3. Coupling of auditory, manual, and articulatory rhythms

3.1. Introduction

Coordinating rhythmic movements with an external rhythm is known as a sensorimotor synchronization task (Repp, 2005; Repp & Su, 2013). Typically, sensorimotor synchronization tasks involve the synchronization of non-verbal movements (e.g. finger-tapping as a manual rhythm) with an external rhythm (e.g. a metronome beat as an auditory rhythm) (Repp, 2005; Repp & Su, 2013). Speech movements (as articulatory rhythm) are suitable for investigating verbal sensorimotor synchronization due to their rhythmic nature, characterized for example by the regular opening and closing of the vocal tract to produce words (i.e., articulatory rhythm) (Poeppel & Assaneo, 2020). However, as pointed out by Chow et al. (2015), the difference between a verbal and a non-verbal sensorimotor synchronization task is that the reference point in a verbal task needs to be determined whereas the time the finger hits the surface marks the reference point in a non-verbal sensorimotor synchronization task. Thus, in a verbal task, participants may align different articulatory or acoustic events with an external rhythm (e.g., Chow et al., 2015; Schreier, 2020; Schreier, 2023; Šturm & Violín, 2016).

Sensorimotor synchronization requires not only the perception of the rhythmic beat per se, but also the prediction of when movements would align with the beat (Repp, 2005). The brain relies on predictive timing mechanisms to anticipate the next beat and to coordinate (speech) movements to it (e.g. Avanzino et al., 2016; Schwartze & Kotz, 2015). A cerebellum-cortical loop is primarily engaged while performing a synchronization task, whereas self-paced movements involve increased activity in the basal ganglia and the additional activation of the supplementary motor area which functions as an internal pacemaker (e.g., Avanzino et al., 2016; Repp, 2005; Schwartze et al., 2011). More specifically, the pre-supplementary motor area is active while speaking (Alario et al., 2006; Bohland et al., 2010).

Hence, self-paced movement and externally-paced movement, for example with a metronome, engage two different timing networks – the external and the internal timing network (Etchell et al., 2014). An external timing network, comprised of the cerebellum, the premotor cortex, and the right inferior frontal gyrus, is involved in timing movements with an external cue, such as speaking or tapping to a metronome (Etchell et al., 2014). Conversely, the internal timing network, comprised of the basal ganglia and the supplementary motor area, is active when timing movements without an external cue, such as speaking or tapping (Etchell et al., 2014). These networks can simultaneously be active when an external rhythm is being internalized (Etchell et al., 2014).

Stuttering has been linked to impairments in initiating, sustaining, and or terminating motor programs, that are rooted in disfunctions within the basal-ganglia-thalamo cortical loop (e.g., Chang & Guenther, 2020; Civier et al., 2013). As discussed in section 1.4., even the fluent speech of PWS displays more variability and timing differences compared to the fluent speech of PWNS (e.g., Loucks et al., 2022; Max & Gracco, 2005; Smith et al., 2010; Wiltshire et al., 2021). Accordingly, the internal timing network is hypothesized to be malfunctioning in PWS while an external timing network compensates for stuttering, given that PWS speak more fluently when synchronizing speech with an external rhythm (Etchell et al., 2014).

Further evidence for an impaired internal timing network comes from non-verbal sensorimotor synchronization tasks, where PWS tapped their finger earlier to the beat than PWNS (e.g., Falk et al., 2015; Sares et al., 2019; Slis et al., 2023; van de Vorst & Gracco, 2017). This suggests that PWS may over-anticipate the beat due to difficulties with predictive timing and therefore, synchronized their taps earlier (Falk et al., 2015).

Despite PWS speaking more fluently when synchronizing their speech to a metronome (e.g., Andrews et al., 1982), verbal sensorimotor synchronization tasks still reveal timing differences between PWS and PWNS (Schreier et al., 2020; Schreier, 2023). Particularly, children and adolescents who stutter synchronized their acoustic speech and vowel onset later to the metronome beat than children and adolescents who do not stutter (Schreier et al., 2020; Schreier, 2023). However, how adults synchronize their speech to a metronome and how verbal (speech) and non-verbal movements (finger tapping) interact in PWS has not yet been explored. Investigating inter-gestural timing between verbal and non-verbal movements is especially interesting given that the internal timing network is involved in both, speaking and tapping (Etchell et al., 2014). In PWNS, there is evidence that the timing of articulatory and manual gestures is closely linked (Meister et al., 2009; Parrell et al., 2014; Treffner, 2002), which we hypothesize may also contribute to a close coupling of the gestures in PWS. On the other hand, increased task complexity, such as synchronizing speech to a metronome while tapping, is hypothesized to lead to more variability in PWS, as previously found by Hulstijn and colleagues (1992).

Therefore, the present study investigates how adults who stutter synchronize their speech across three different rhythmic conditions: speaking and tapping (Tapping condition), speaking along with a metronome (Metronome condition) and speaking and tapping to a metronome (Metronome+Tapping condition).

By using electromagnetic articulography (EMA), this study provides a novel insight into articulatory timing across different rhythmic conditions and explores how auditory, manual, and articulatory rhythms interact in PWS and PWNS. EMA is a particularly suitable method for studying articulatory timing, as it tracks the movement of small sensors which are positioned on speech articulators (e.g., jaw, lips, tongue) within an electromagnetic field that is created around the participants' head. Non-verbal gestures, like finger-tapping, can also be tracked by placing a sensor on the participants' index finger. The electromagnetic field also covers a good portion of the participants' upper body to capture good quality recordings of these movements (AG501, Carstens Medizinelektronik GmbH).

The primary goal of this chapter is to examine whether the timing of articulatory speech onsets is influenced by the rhythmic context (e.g. finger tapping, metronome vs. finger tapping to a metronome) and to explore whether adults who stutter and adults who do not stutter differ across these conditions. The following study reports the results of four adults who stutter and four adults who do not stutter.

3.2. Paper 2

Paper 2 has been published in the Proceedings of the 20th International Congress of Phonetic Sciences undergoing a revision process.

Franke, M., Benker, N., Falk, S. & Hoole, P. (2023). 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 (pp. 3942-3946),* Guarant International.

Synchronization type matters: Articulatory timing in different rhythmic conditions in persons who stutter

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

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

*mona.franke@phonetik.uni-muenchen.de

Abstract

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

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

1. Introduction

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

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

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

Studies on joint speech and manual movements indicate that there should be a close coupling between speech and manual motor control systems (e.g., [11]). Hence, tapping and speaking at the same time (in the speech rhythm) could lead to more stable (articulatory) gestures in PWS as it is expected that this inter-gestural timing is closely linked. Moreover, it would be particularly interesting to see how PWS synchronize articulatory gestures and finger-tapping to an external rhythm. This setting can test whether timing information from multiple channels (auditory, manual + articulatory rhythm) is strongly or weakly coupled in PWS and PWNS, which might improve articulatory stability in the former or deteriorate it in the latter case. A study by Hulstijn and colleagues [12] points to weaker integration in PWS. They found that PWS were more variable in coordinating speech and hand movements to tones than PWNS. However, they did not report how exactly the timing occurred nor whether effects on fluency were found.

Therefore, the aim of the current electromagnetic articulography (EMA) study is to shed further light on timing processes in PWS, by analyzing a) the effect of external pacing (metronome) and self-pacing (tapping) on speech gesture timing (where does the beat occur in relation to articulatory gestures) and b) how inter-gestural timing (non-verbal and verbal gestures) is affected by an external rhythm. It is an open question whether PWS synchronize their speech earlier to the metronome than PWNS, which would mirror results on non-verbal tasks (e.g., [7,8]).

Note that, in general, it is unclear whether metronome and finger-tapping synchronization time-points with respect to articulatory gestures differ or coincide. Therefore, it is another aim of our study to compare these conditions. What also remains to be answered is whether PWS differ from PWNS in the time point of synchronizing non-verbal gestures (finger tapping) and verbal gestures (speech). Following the result of Hulstijn et al. [12], we would expect that speaking along with a combination of motor and auditory pacing would lead to greater timing variability in PWS compared to PWNS.

2. Methods

EMA data (AG501, Carstens Medizinelektronik) were collected from 10 adults who stutter and 10 adults who do not stutter. For the present paper, data of 4 persons who stutter (mean age = 24.3, 2 female) and 4 persons who do not stutter (mean age= 24.5, 2 female) were analyzed. All participants were native speakers of German and besides stuttering, no other impairments were reported. PWS and PWNS were matched in pairs having similar musical experience, the same age (±1 year), and sex.

Participants produced mono- and disyllabic German target words (cf. Table 1) embedded in the carrier phrase [ˈzeːə ____ ˈan] (Look at ____).

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

Table 1: Target words

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

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

The metronome tone was presented via an in-ear headphone which participants plugged in their right ear. The onset of the metronome time point closest to the target word was automatically extracted.

For the conditions where tapping was involved, participants were instructed to tap their index finger of the dominant hand on an elevated wooden block that was placed on a table close to the participants. Sensors relevant for the data we report here were placed on the tip of the index finger of the participants' dominant hand and on the upper and lower lip. In addition, we had sensors placed on the tongue and the jaw. Only fluent productions (determined by listening to audio recordings) were analyzed. Therefore, a maximum of 144 target words were analyzed per participant.

2.1. Kinematic measures

Lip activity forming the constriction for the bilabial onset was measured using Lip Aperture (LA), defined as the Euclidian distance between transducers placed on the upper and lower lip. For LA and the finger tap (FT), the gesture onset was semi-automatically detected using a 20% velocity threshold and the onset of the gesture nucleus (start of the plateau) was also semi-automatically detected.

For each target word the relative timing of the consonantal gesture (onset) to the metronome onset and the finger tap onset was calculated as the lag between the onsets of the gesture nuclei of LA and FT and the lag between the onset of the LA gesture nucleus and the acoustic metronome onset.

We will refer to the two resulting intervals as *tap - articulatory onset interval* and *met - articulatory onset interval*. Both intervals are calculated such that positive values result if the articulatory onset is before the tap or the metronome.

2.2. Statistical analyses

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

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

3. Results

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

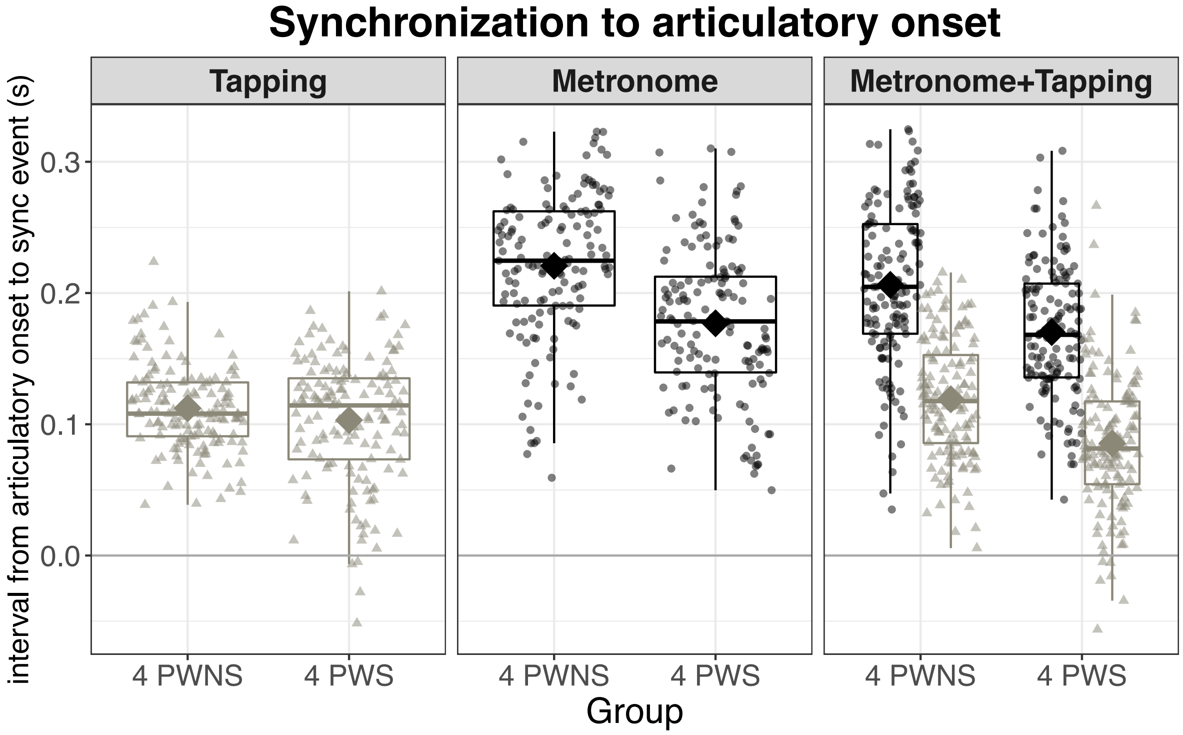


Figure 1: tap - articulatory onset interval (light grey, triangles) and met - articulatory onset interval (dark grey, dots). 0 seconds indicates the articulatory onset (nucleus onset of the bilabial), positive intervals indicate that the event of synchronization was after articulatory speech onset. Each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

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

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

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

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

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

The third model (Conditional $R^2 = 0.36$, Marginal $R^2 = 0.11$) revealed that the time points of the metronome beat shifted significantly towards the articulatory onset in the Metronome+Tapping condition compared to the single Metronome condition ($p = 0.0452$). This effect was found independently of group; no interaction was found.

Finally, in order to test if PWS were more variable than PWNS in synchronizing, the standard deviation was calculated for the different intervals per condition.

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

Table 2: Standard deviations for interval durations

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

4. Discussion

The study revealed both differences in timing when synchronizing speech with an internally generated rhythm (inter-gestural timing) as well as when synchronizing speech with an external rhythm (paced timing). Moreover, the data suggests differences between PWS and PWNS. We will first address differences between conditions and then group differences.

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

In terms of the group effect we found that PWS had shorter intervals between the metronome and the articulatory speech onset. This result indicates that PWS time their speech later to the metronome than PWNS, potentially because of later speech initiation in the group who stutters. This finding would be in line with preliminary results for children who stutter reported by Schreier et al. [20].

However, finger tapping to one's own speech did not differ between PWS and PWNS, indicating similar inter-gestural timing conditions. Interestingly, joint speaking and tapping to an external rhythm (Metronome+Tapping condition) led to a group difference in the interval between articulatory speech onset and finger tap such that PWS have shorter intervals than PWNS. This difference was caused by the fact that compared to the single Tapping condition, PWS decreased the *tap - articulatory onset interval* in the combined condition, whereas PWNS did not change the timing of their finger taps. Therefore, it can be assumed that in PWS inter-gestural timing is more affected by an external rhythmic cue than in PWNS. As previous research showed, PWS engage different timing mechanisms and/or brain circuits to time movements with an external cue [10,21].

Furthermore, it was hypothesized that timing would become more variable in PWS with increasing task complexity, however, this hypothesis was not supported by the results of our study. Despite PWS being more variable than PWNS in general, PWNS have a greater increase in timing variability in the combined condition.

From our study it can be concluded that PWS potentially couple auditory, manual, and articulatory rhythms in a different way, leading to later speech initiation and more temporal variation. This remains to be tested with a greater participant sample, of course. We aim to present data from 10 participants per group at the conference. Finally, our dataset offers the possibility for specific consideration of the vowel gesture, of inter-gestural timing between onset consonants and vowels, as well as on intra-gesture stability in different rhythmic conditions.

5. Acknowledgements

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

6. References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.3. Discussion

The present study investigated how four adults who stutter and four adults who do not stutter synchronized their articulatory speech onset across three rhythmic conditions: self-paced finger tapping (Tapping), externally-paced metronome (Metronome), and a combination of both (Metronome+Tapping). In this discussion, we will focus on group differences and their relation to internal and external timing networks (Etchell et al., 2014).

Since there were no group differences in the Tapping condition, we suggest that articulatory and manual rhythms are similarly coupled in PWS and PWNS. This indicates that the internal timing network of PWS and PWNS probably work in a similar way, when two tasks that require an internal pacemaker are involved. Future research could address this by conducting a neuro study to investigate how the internal timing network of PWS and PWNS behaves during a simple speech task, and when they perform the same task while simultaneously tapping their finger along. We hypothesize that finger tapping may normalize neural activity within the internal timing network in PWS so that it reaches levels compared to those of PWNS, similar to the effect of metronome-paced speech on the brain activity of PWS (e.g., Toyomura et al., 2011; Toyomura et al., 2015). However, even though synchronizing speech with a metronome beat enhances speech fluency in PWS (Andrews et al., 1982; Davidow et al., 2014) and is associated with normalized neural activity (Toyomura et al., 2011; Toyomura et al., 2015), we found that the groups differed in timing their speech to an auditory rhythm. Adults who stutter timed their speech closer to the metronome beat than matched control participants which is consistent with results previously reported for children who stutter (Schreier et al., 2020; Schreier, 2023). One explanation could be that PWS have difficulties in predicting external auditory events, such as metronome beats, which has been observed in non-verbal tasks (e.g., Falk et al., 2015). This may cause challenges in integrating external cues with their own speech production and therefore, leading to a difference in the Metronome but not in the Tapping condition. Indeed, evidence suggests that individuals who stutter perceive rhythm differently, as demonstrated in rhythm discrimination tasks (Wieland et al., 2015; Chang et al., 2016).

In the Metronome+Tapping condition PWS aligned their taps more closely with their speech onset compared to the single Tapping condition, while PWNS did not change their tapping behavior across conditions. This suggests a different integration of auditory, manual, and articulatory rhythms in PWS.

Another explanation for the group differences in the Metronome conditions could be rooted in verbal inter-gestural timing differences, such as an altered consonant-vowel coupling in PWS. While PWS and PWNS could still target the same reference point in the Metronome conditions (e.g., the same articulatory event in their speech), differences in articulatory timing may cause the groups to differ in their synchronization time points with respect to the articulatory speech onset. Alternatively, PWS and PWNS may have different reference points in the Metronome conditions: PWS could aim to align an articulatory event (e.g., articulatory vowel onset) with the metronome beat, whereas PWNS may target an acoustic event (e.g., acoustic vowel onset). Although there were no group differences in the Tapping condition, we cannot be certain that the groups share the same speech reference point. Our results only indicate that, when using the articulatory speech onset as a reference, they do not differ in synchronizing their finger taps to their speech. Future research could therefore focus on the comparison of several reference points, such as the acoustic and the articulatory vowel onset in addition to the articulatory word onset.

The next chapter will shed light on potential articulatory timing differences by analyzing consonant-vowel coupling in adults who stutter and adults who do not stutter across the different rhythmic conditions, that have been introduced in this chapter. In addition, unpaced speech will be investigated. The next chapter also presents data from a larger participant sample of 10 adults who stutter and 10 adults who do not stutter, addressing the research questions of this chapter in greater depth, with a specific focus on predictive timing. Thus, the interplay between inter-gestural timing, predictive timing, and their implications will be discussed in detail in the following chapter.

Chapter 4

4. Consonant Vowel Timing and Predictive Timing

4.1. Introduction

From the previous chapters we have learned that onset consonant vowel (CV-) timing poses a significant challenge for individuals who stutter. Our earlier findings, namely more consonant compression in children who stutter (Franke et al., 2023a; *Chapter 2*) and different synchronization time points when synchronizing speech to a metronome in adults who stutter (Franke et al., 2023b; *Chapter 3*), suggest that inter-gestural coupling, i.e., the coordination between consonant and vowel gestures, may be altered in PWS. In this chapter, we aim to shed more light on this by expanding our participant sample from *Chapter 3* and by analyzing direct articulatory (EMA) data.

Previous research on CV-timing has largely focused on coarticulatory aspects, which give an indication of whether articulatory gestures overlap more or less in PWS compared to PWNS. Most of these studies measured articulation indirectly via acoustic measures (Dehqan et al., 2016; Klich & May, 1982; Maruthy et al., 2018; Robb & Blomgren, 1997; Sussman et al., 2011; Verdurand et al., 2020). Only a few used direct articulatory techniques, such as ultrasound, to explore coarticulatory aspects in individuals who stutter (Frisch et al., 2016; Lenoci & Ricci, 2018). Lenoci & Ricci (2018) found that children who stutter produced more spatiotemporal overlap of CV gestures compared to matched peers, whereas Frisch and colleagues (2016) did not observe coarticulatory differences in adults who stutter.

There are two main theories about how differences in coarticulatory behavior of perceptually fluent speech in PWS can be interpreted. On the one hand, reduced overlap between consonant and vowel gestures is hypothesized to be a strategy to stabilize the speech motor system (Verdurand et al., 2020). And on the other hand, more overlap is an indicator for a less mature speech motor system (Lenoci & Ricci, 2018). The latter theory derives from findings in individuals who do not stutter, where children produce more gestural overlap than adults (Noiray et al., 2018). Reduced overlap between gestures is linked to more precise control over articulators (e.g. Noiray et al., 2018). Since PWS have a less stable speech motor system (e.g., Namasivayam & van Lieshout, 2011; van Lieshout et al., 2004), they might exhibit more gestural overlap compared to their peers, reflecting less mature speech motor control (Lenoci & Ricci, 2018).

In the paper, presented in this chapter, we analyzed EMA data to investigate CV-timing in 10 adults who stutter and 10 adults who do not stutter across different conditions. As introduced in *Chapter 3*, participants read target words embedded in a carrier phrase in four different conditions: Unpaced, Tapping, Metronome, and Metronome+Tapping.

The advantage of EMA is that it captures the precise timing and coordination of different articulators. There are several options to measure CV-timing, encompassing temporal and spatial measures, as reviewed by Svensson Lundmark et al. (2021). In this chapter, we use a rather classical landmark-based approach, and additionally, a trajectory-based approach with which we compare the tongue back (TB) movement of PWS and PWNS at the time of the acoustic onset of the onset consonant to the offset of the acoustic vowel. Our main hypothesis regarding CV-timing is that differences emerge between PWS and PWNS in the Unpaced condition. For the landmark-based approach this would be reflected in greater or smaller CV-lags and in the trajectory-based approach in a shift of the vowel gesture or a different position of TB position at the beginning of the target word. We also investigate how rhythmic conditions affect CV-timing, as external rhythms have a stabilizing effect on the speech motor system (van Lieshout & Namasivayam, 2010; Wiltshire et al., 2023).

In addition to CV-timing, we investigate predictive timing by exploring how adults who stutter and adults who do not stutter synchronize their articulatory speech onsets to rhythmic events. Predictive timing refers to the anticipation and precise timing of (articulatory) movements (Debarant et al., 2012) and differences between PWS and PWNS have been observed in non-verbal (e.g., Falk et al., 2015; Sares et al., 2019) and verbal sensorimotor synchronization tasks (Franke et al., 2023b; Schreier, 2023; Schreier et al., 2020).

The main hypothesis is that PWS differ from PWNS in how they time their speech to rhythmic events and that they exhibit greater variability when task complexity is increased, such as when speaking, tapping and simultaneously synchronizing to a metronome (as found by Hulstijn et al., 1992). Examining both CV-timing and predictive timing with the same participant sample allows us to relate potential differences to underlying mechanisms of speech motor control and timing coordination.

4.2. Paper 3

Paper 3 has been published in the Journal of Phonetics undergoing a full revision process.

Franke, M., Falk, S., Benker, N., & Hoole, P. (2025). The effect of rhythm on inter-gestural coupling of onset and vowel gestures and predictive timing in stuttering. *Journal of Phonetics*, 112, 101432 [PDF, open access], https://doi.org/10.1016/j.wocn.2025.101432

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

Mona Franke[1,2,3,4]*, Simone Falk[2,3,4], Nicole Benker[1], Phil Hoole[1]

[1]*Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, Germany,*

[2]*Faculté des arts et des sciences – Département de linguistique et de traduction, Université de Montréal,* Canada,

[3]*BRAMS, Montréal, Canada,* [4]*CRBLM, Montréal, Canada*

* Correspondence:
Mona Franke
Schellingstr. 3 (Institut für Phonetik und Sprachverarbeitung),
80799 München, Germany
mona.franke@phonetik.uni-muenchen.de

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.

Keywords: *Speech motor timing, inter-gestural timing, predictive timing, stuttering, metronome-paced speech*

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 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 by-passing some of the malfunctioning neural circuits and rein-states 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 flu-ent 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 offline 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 interrater

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

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-Jøergensen, 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.

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 [ˈbaːdn̩] (4)	Boden [ˈboːdn̩] (5)	Buden [ˈbuːdn̩] (4)
Mahl [maːl] (3)	Mohn [moːn] (3)	Buhne [ˈbuːnə] (3)

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
(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.

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

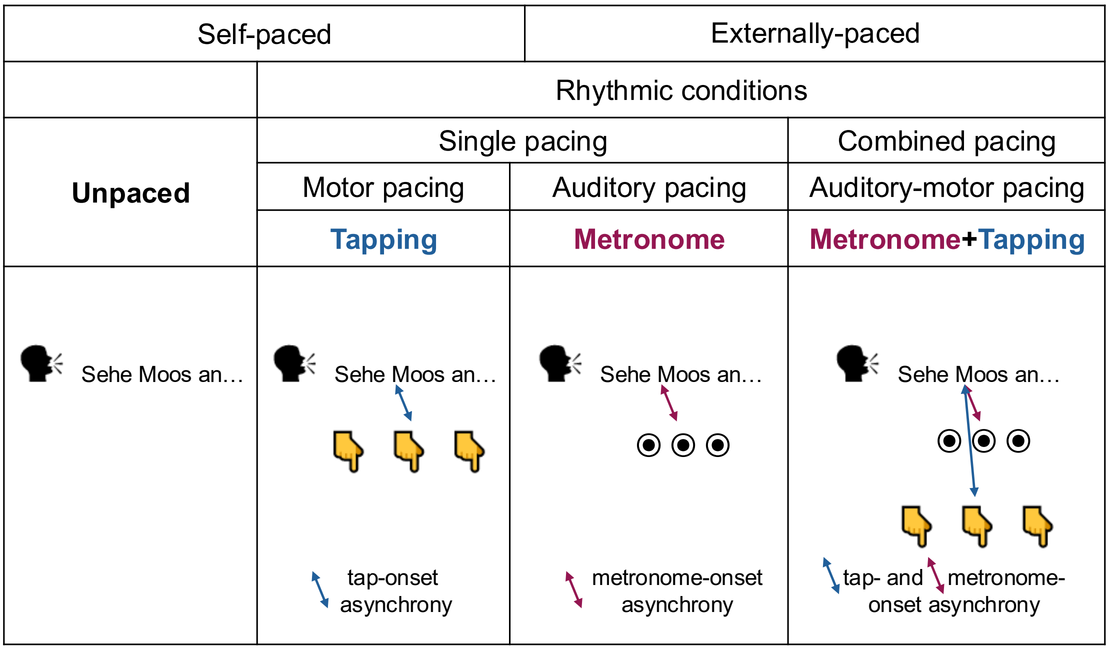


Fig. 1. *Sketch of different conditions and respective terminology.*

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 experiment, 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).

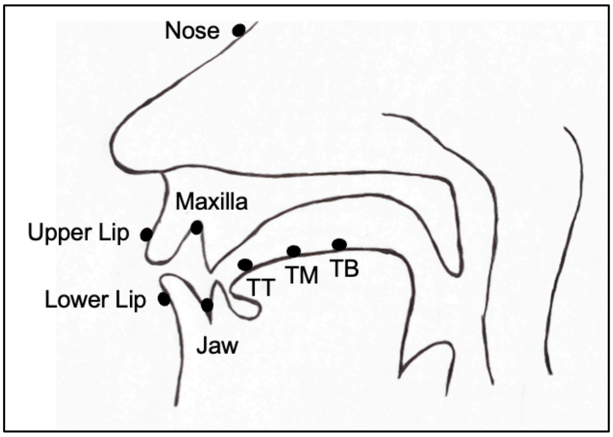


Fig. 2. *Illustration of sensor placement.*

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

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

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 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.

[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.

2.6.2. Onset 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.

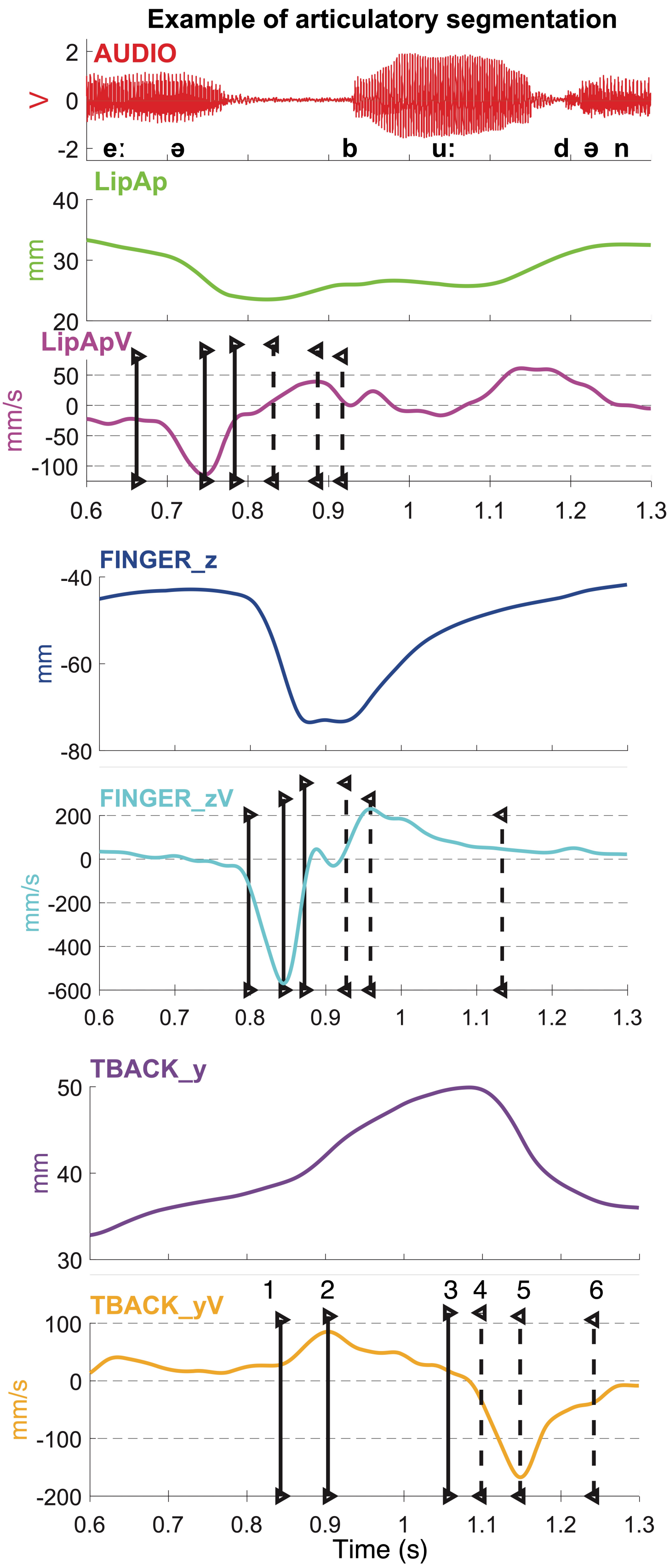


***Fig. 3**. Example of segmentation for the target word /buːdən/ in the Tapping condition for the bilabial gesture, the finger tapping gesture, and the vowel gesture. Duration in seconds is displayed on the x-axis. Top panel: Audio signal, voltage (V) displayed on the y-axis, broad phonetic transcription 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.

2.6.3. 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 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.

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).

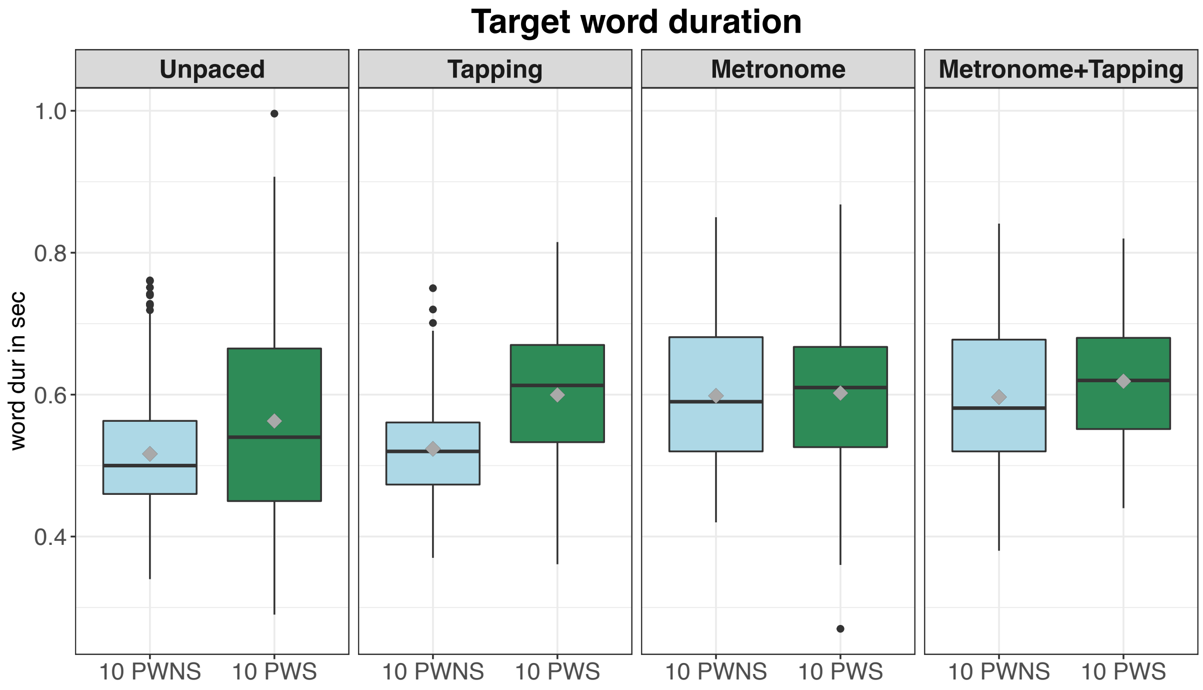


Fig. 4. *Target word duration per group and condition. Durations are displayed in seconds on the y-axis. Groups are displayed on the x-axis, PWNS = persons who do not stutter (blue), PWS = persons who stutter (green). Diamonds display the mean. Within each box, the median is denoted with horizontal lines; 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.)*

A linear mixed model was run to predict word duration. The final model (conditional R2 = 0.57, marginal R2 = 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, X2(4) = 71.61, $p < 0.0001$. Additionally, word duration varied significantly across conditions, X2(6) = 313.25, $p < 0.0001$. Importantly, there was an interaction between Group and Condition, X2(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.

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)

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 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.

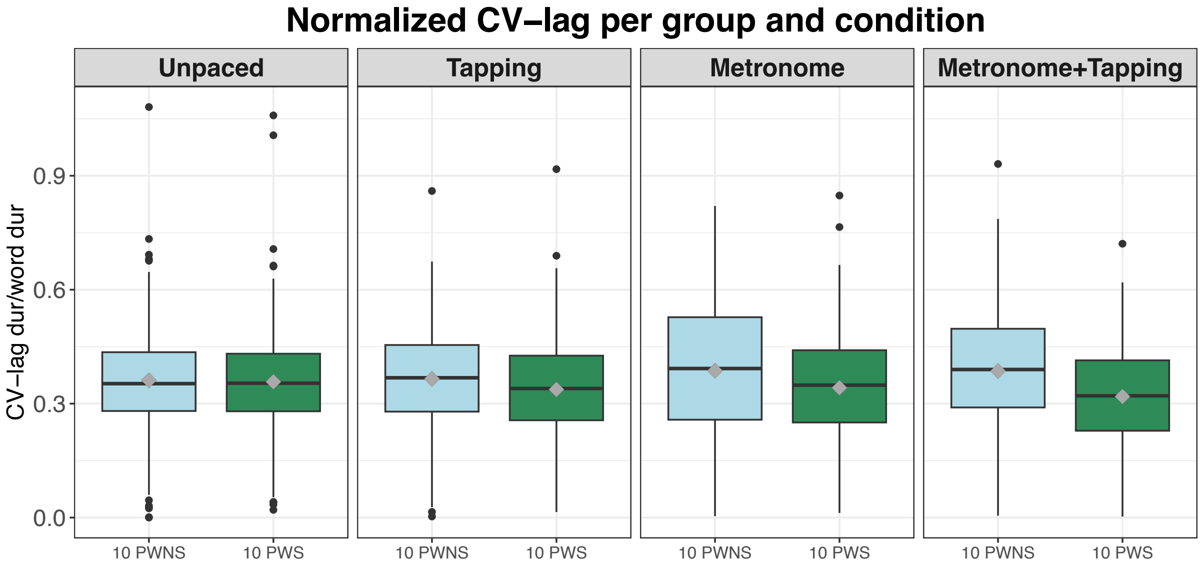


Fig. 5. *Time-normalized CV-lags (in s) for each group per condition. Groups are displayed on the x-axis, PWNS = persons who do not stutter (blue), PWS = persons who stutter (green). Positive values indicate that the vowel gesture nucleus onset appeared after the consonant gesture nucleus onset. Details 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.)*

A LMM including the fixed effects Group and Condition with an interaction term, as well as intercepts for Participant and Word (conditional R2 = 0.34, marginal R2 = 0.02) was run to predict the time-normalized CV-lags. Results should be interpreted with caution as low marginal R2 indicates that the fixed effects do not explain much of the variance.

While the main effect of group was significant, v2(4) = 17.64, p = 0.0015, with shorter CV-lags for PWS[5] and also a significant main effect for Condition, v2(6) = 19.41, p = 0.0035, the most striking effect in the results is the highly significant interaction between Group and Condition, v2(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$_{\text{PWNS-PWS}}$ = 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

[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).

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. 6).

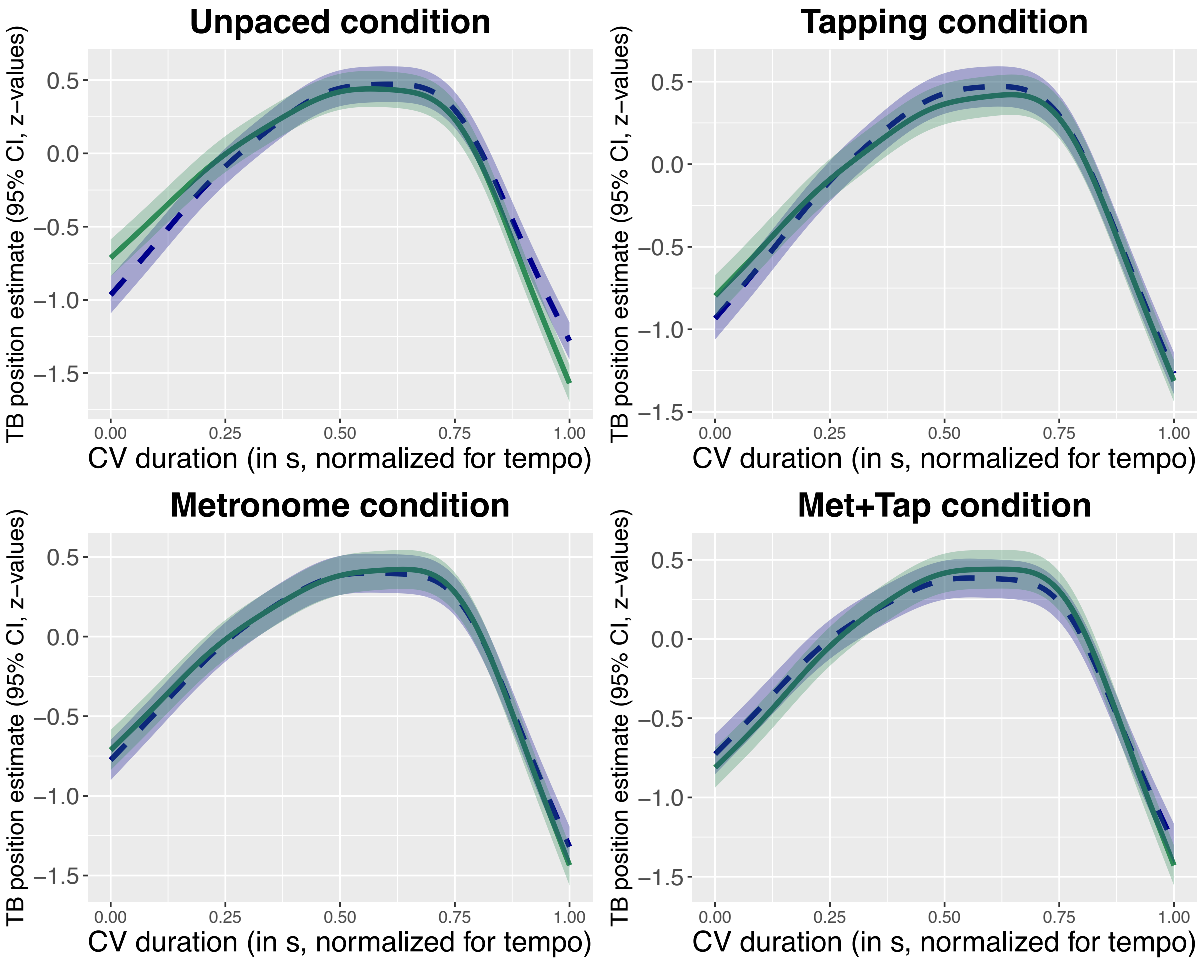


Fig. 6. Model predictions for 10 PWS (green, solid line) and 10 PWNS (blue, dashed line) within 95% pointwise confidence intervals. The x-axis displays the normalized time of the acoustic CV interval, the y-axis displays the estimated z-transformed position of the TB sensor (horizontal movement). (For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

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 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.

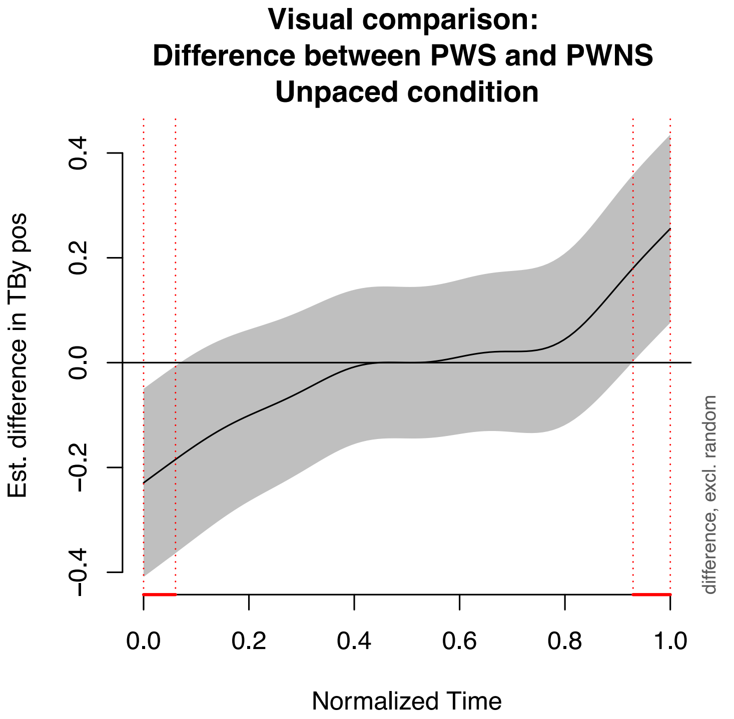


Fig. 7. *Estimated difference of the horizontal TB position (Z-scores) between PWS and PWNS in the Unpaced condition within the associated 95% pointwise confidence interval (y-axis) over time (x-axis). The highlighted area in red indicates where the confidence interval excludes zero and the groups differ 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.)*

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 unaffected 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.

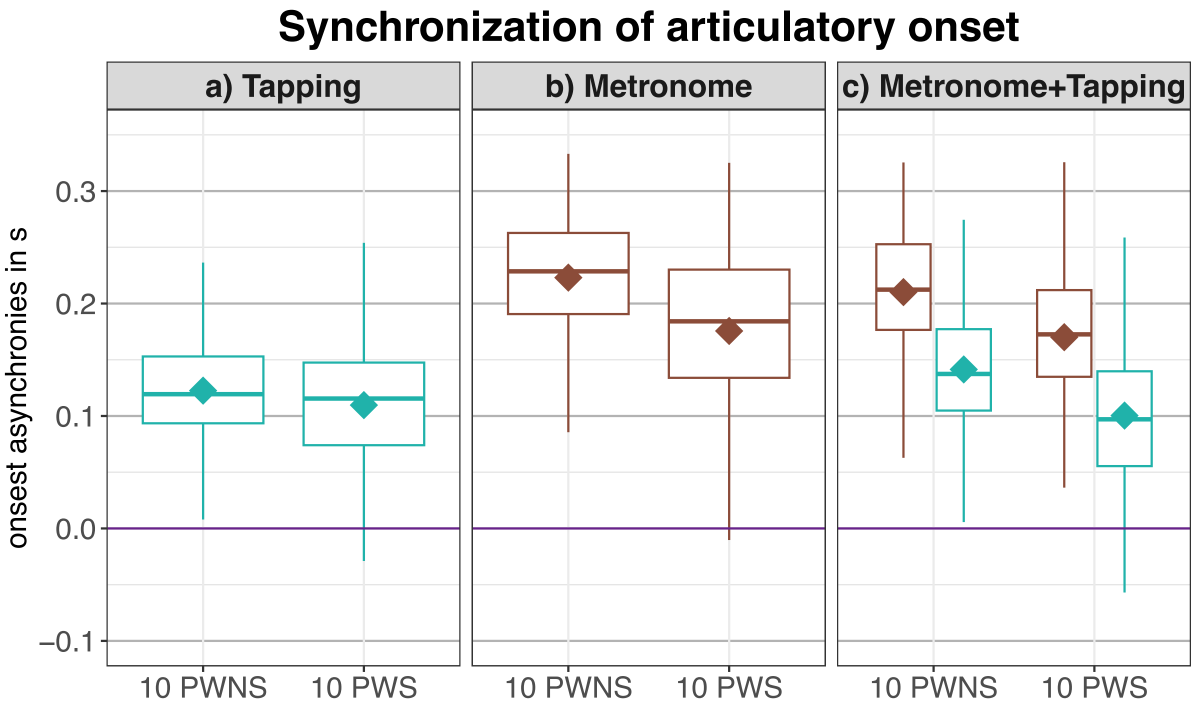


Fig. 8. *Signed asynchronies between articulatory onsets of target words and rhythmic events (taps, metronome beats) in three rhythmic conditions (a: Tapping, b: Metronome, c: Metronome+Tapping). Tap-onset asynchronies (turquoise) and metronome-onset asynchronies (brown) expressed in seconds. The horizontal 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).*

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 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 R2 = 0.40, marginal R2 = 0.30) showed (see Fig. 8a + b) that tap asynchronies were shorter than metronome asynchronies (main effect of Rhythmic event, X2(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, X2(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, X2(1) = 44.64, $p < 0.0001$).

Model 2 (conditional R2 = 0.33, marginal R2 = 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, X2(1) = 13.86, p = 0.0002). Moreover, in both groups, finger taps occurred closer to the articulatory onset than the metronome beats (Rhythmic event, X2(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 R2 = 0.41, marginal R2 = 0.06), including an interaction term between Group and Condition, revealed a significant effect of Condition, X2(2) = 34.39, $p < 0.0001$, and a significant interaction between Group and Condition, X2(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 R2 = 0.31, marginal R2 = 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, X2(1) = 10.29, p = 0.0013. This effect was found independently of Group, X2(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.

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

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.

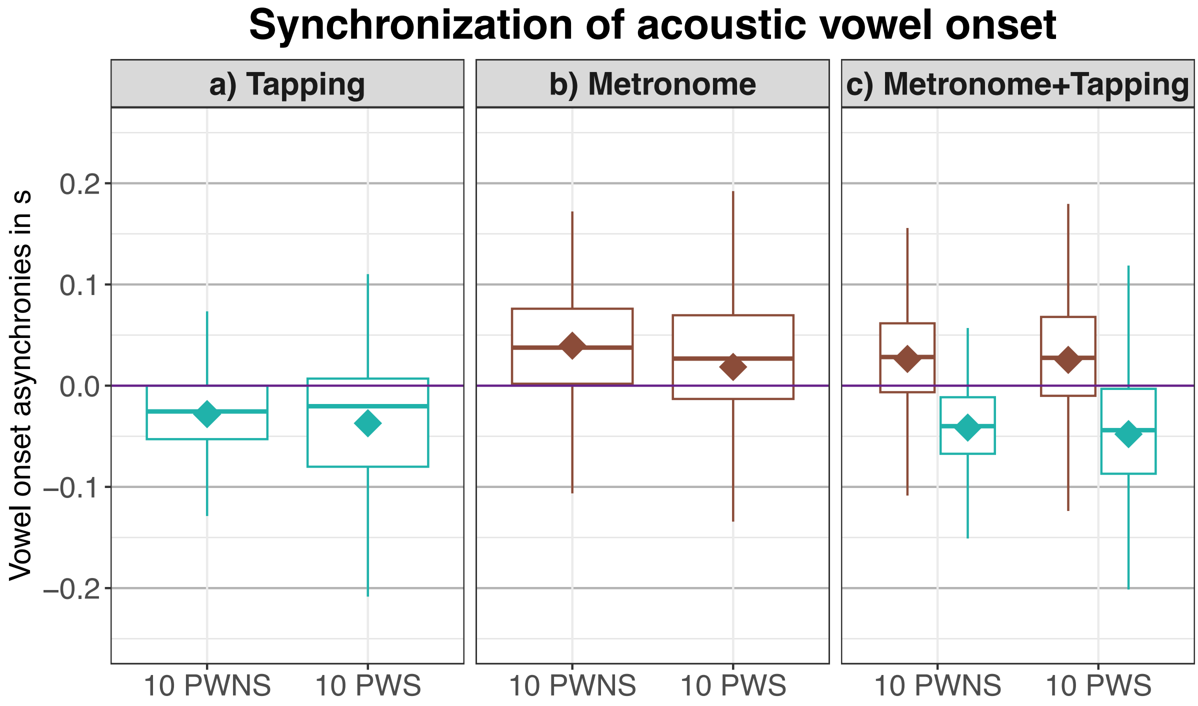


Fig. 9. *Signed asynchronies between acoustic vowel onset of target words and rhythmic events (taps, metronome beats) in three rhythmic conditions (a: Tapping, b: Metronome, c: Metronome+Tapping). Tap-onset asynchronies (turquoise) and metronome-onset asynchronies (brown) expressed in seconds. The horizontal line at 0 seconds 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 Figure 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).*

We ran two models to compare the single pacing conditions (see Fig. 9a vs. 9b); their aim was to probe for differences 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 R2 lower than 0.03).

Model 1 (conditional R2 = 0.29, marginal R2 = 0.15) revealed a significant main effect of Rhythmic event, X2(2) = 239.28, p < 0.0001 (see Fig. 9a + b), a significant Group effect, X2(2) = 22.68, p < 0.0001, and a significant interaction between Rhythmic event and Group, X2(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 R2 = 0.29, marginal R2 = 0.21) indicates that in the combined condition (Fig. 9c) metronome beats occurred closer to the acoustic vowel onset than finger taps (Rhythmic event, X2(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

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

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.

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 (Paststätter & 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 A: 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
12	PWNS	0.01961	0.0115	1730	1.703	0.3224

	Combined vs. Tapping					
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*, 661880. https://doi.org/10.3389/fnhum.2021.661880

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. doi: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), pp. 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., and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251. doi: 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), e12521 https://doi.org/10.1111/desc.12521

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. doi: 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. doi: 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. doi: 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*, 105975. https://doi.org/10.1016/j.jfludis.2023.105975

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, doi: 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, DOI: 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*. 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. doi: 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., and May, G. (1982). Spectrographic study of vowels in stutterers' fluent speech. *Journal of Speech, Language, and Hearing research*, 25, 364–370. doi: 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*, 106194. https://doi.org/10.1016/j.jcomdis.2022.106194

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*, 106213. https://doi.org/10.1016/j.jcomdis.2022.106213

Lu, Y., Goldstein, L., & Narayanan, S. (2024). MRI reveals CV coarticulation is preserved in stuttering. *In Book of Abstracts of the 13th International Seminar on Speech Production*, Autrans, FR.

Lüdecke et al. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, 6(60), 3139. https://doi.org/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., and 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. doi: 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. 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), e3002492. https://doi.org/10.1371/journal.pbio.3002492

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. 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 : the official journal of the Society for 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. doi: 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)*. December 14-18th, 2020.

Schreier, R. (2023). Stuttering and speech-rhythm. Dissertation, LMU München: Fakultät für Sprach- und Literaturwissenschaften. 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 one*, *18*(2), e0276691. https://doi.org/10.1371/journal.pone.0276691

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. doi: 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

4.3. Discussion

The aim of this study was to provide deeper insights into timing mechanisms involved in fluent speech production and to improve our understanding of stuttering by investigating CV-timing and predictive timing in 10 adults who stutter and 10 adults who do not stutter.

The central discovery on CV-timing was that adults who stutter appear to time CV gestures more closely together than adults who do not stutter. Overall, the trajectory-based method produced the more powerful model, making these results more robust and more informative than those from the landmark-based method. Hence, the primary difference between PWS and PWNS emerged in the Unpaced condition. We propose that PWS couple CV gestures more tightly than PWNS because the TB position at the time of the acoustic word onset had already moved further back in PWS than in PWNS. However, we acknowledge that the trajectory-based approach is not a direct measurement of CV-timing, as it only compares the vowel gestures of PWS and PWNS without accounting for the previously uttered consonant. Therefore, both analyses (landmark- and trajectory-based) should be seen as complementary, together providing a coherent picture of PWS showing more –rather than less– overlap between CV gestures compared to PWNS. This finding supports the assumption that adults who stutter have a less mature speech motor system than their peers, which is in line with studies on children who stutter (Lenoci & Ricci, 2018; Franke et al., 2023a) and acoustic research on coarticulation in adults (Klich & May, 1982).

As a next step, it would be interesting to replicate our study with children who stutter to investigate whether they differ more from children who do not stutter than adults who stutter from adults who do not stutter. This would allow us to track speech motor development in both individuals who stutter and individuals who do not stutter.

Interestingly, the trajectory-based method revealed that the rhythmic conditions had a positive effect on CV-timing of PWS, as no group differences were found in the Tapping, Metronome and Metronome+Tapping conditions. This suggests that the metronome stabilizes the speech motor system of PWS (e.g., Wiltshire et al., 2023), and that even an internally generated manual rhythm (finger tapping) has a stabilizing effect. The combined Metronome+Tapping condition also did not reveal any differences between the groups. However, with regard to predictive timing we observed differences in the metronome conditions.

As found in *Chapter 3*, PWS produced smaller asynchronies in the Metronome conditions and did not differ from PWNS in the Tapping condition. These findings are now confirmed with a larger participant sample. We argue that PWS may experience difficulties in auditory-motor integration, leading to differences in how they time their speech to an external rhythm, but not to an internally generated rhythmic event, such as finger tapping. Additionally, PWS showed greater variability in timing their speech to a metronome beat compared to a finger tap, further supporting the hypothesis of weaker auditory-motor integration.

Thus, alterations in inter-gestural timing do not appear to be the cause of the differences in synchronizing speech to a metronome beat, given that the trajectory-based method did not reveal differences between the groups in neither of the rhythmic conditions.

With the larger participant sample in this chapter, we found that PWS and PWNS timed their finger taps in opposite directions in the combined Metronome+Tapping condition. With the smaller sample, this shift was only observed in PWS. Whereas PWS show closer coupling of the verbal and the non-verbal gesture (as observed in the smaller participant sample in *Chapter 3*), PWNS shifted their finger taps closer to the metronome beat, moving away from the articulatory speech onset. This suggests that we obtain a similar picture to that of non-verbal sensorimotor synchronization tasks when comparing only the distance between finger tap and metronome beat - greater negative mean asynchronies in PWS compared to PWNS.

In conclusion, we found evidence for speech motor timing differences in PWS, supporting the hypothesis of a less mature speech motor control system.

In the following, the results of all studies are now placed in a broader context for final discussion.

Chapter 5

5. General Discussion

In the three studies leading to this cumulative dissertation I investigated whether PWS and PWNS differ in articulatory timing at word onset position, both with and without a rhythmic context. This is a relevant question because stuttering, a neurodevelopmental speech motor disorder, is associated with disruptions in speech rhythm at word onsets. As such, this population provides a valuable opportunity to examine the underlying mechanisms of fluent speech production.

The primary aim of this dissertation was to deepen our understanding of speech timing mechanisms in PWS and PWNS by exploring how rhythmic conditions impact speech articulation. In order to do so, we conducted an acoustic study with children and adolescents who stutter (*Chapter 2*), followed by articulatory studies with adults who stutter (*Chapter 3* and *4*).

5.1. Summary of Main Findings

Chapter 2 was concerned with investigating acoustic cues for a c-center effect in unpaced and metronome-paced speech, addressing three key questions: i) whether children and adolescents who stutter differ from their matched peers, ii) whether an external rhythm eliminates potential group differences, and iii) whether an external rhythm leads to more stability in timing, especially in children and adolescents who stutter.

The main finding was that children and adolescents who stutter exhibited greater consonant compression, indicating that they time articulatory gestures differently than their matched peers. Notably, the Metronome condition did not eliminate this group difference, suggesting that even though speaking along with a metronome significantly enhances speech fluency in PWS (Andrews et al., 1982), it does not influence the temporal organization of articulatory gestures in the developing population who stutters. However, there was a trend towards more stable timing between the onset and the vowel in the metronome-paced speech in both groups.

Chapter 3 explored the effect of rhythmic conditions on articulatory timing, as well as the interplay between verbal and non-verbal gestures in adults who stutter and adults who do not stutter in an EMA study. The research question i) of whether the groups differ in gestural timing can be answered with both yes and no. Specifically, no differences where observed, when participants were simultaneously tapping and speaking. However, differences emerged when timing their speech to a metronome. Articulatory gesture onsets in PWS were placed closer to the metronome beat than in PWNS. In addition we can answer the question ii) of whether speech gesture timing is dependent on the rhythmic context with yes. Although the articulatory speech onset preceded both rhythmic events (finger taps and metronome beats), finger taps were more closely aligned with the speech onset than the metronome beat. Lastly, addressing the research question iii) on how an external rhythm affects the timing of verbal and non-verbal gestures, both groups aligned their speech onset more closely to the metronome in the combined Metronome+Tapping condition compared to the single Metronome condition. However, only adults who stutter aligned their finger taps more closely with the articulatory speech onset in the combined condition compared to the single Tapping condition. Hence, the presence of an external rhythm led to a closer coupling of verbal and non-verbal gestures in PWS, but not in PWNS.

Chapter 4 was built on the study presented in *Chapter 3* and focused on CV-timing in unpaced speech and across different rhythmic conditions to shed light on inter-gestural timing dynamics – a relatively unexplored area in PWS. The main finding concerning CV-timing, based on the results of the GAMM analysis (trajectory-based measure), is that adults who stutter showed closer CV coupling than adults who do not stutter in the Unpaced condition and that the groups had a similar CV-timing in all rhythmic conditions (Tapping, Metronome, Metronome+Tapping). This chapter also addressed predictive timing abilities, revealing an alteration in PWS. Specifically, PWS aligned their speech onset more closely to the metronome beat than PWNS when synchronizing their speech onset with a metronome. Furthermore, the groups did not differ in synchronizing their articulatory speech onset with their finger taps which is consistent with the results reported in *Chapter 3* for a smaller participant sample. However, with the larger sample size in this chapter, we found that the external rhythm influenced the coordination of verbal and non-verbal gestures not only in PWS but also in PWNS, although in the opposite direction. While PWS aligned their finger taps more closely with the speech onset in the combined Metronome+Tapping condition than in the single Tapping condition, PWNS shifted their finger taps further away from the articulatory speech onset and more towards the metronome beat. This suggests that PWS and PWNS may rely on different cues to align their speech - PWNS use the metronome and PWS the finger tap. Moreover, PWS were more variable in synchronizing their speech to a metronome beat than to their own finger tapping.

In general, this thesis supports the assumption of PWS having difficulties with inter-gestural timing (*Chapter 2* and *4*) and auditory-motor integration (*Chapters 3* and *4*).

5.2. Development of Speech Motor Control

The thesis at hand sheds light on articulatory timing, the basis of speech rhythm (Poeppel & Assaneo, 2020) in different age groups – from children and adolescents to adults. With investigating acoustic correlates of underlying articulatory processes of the c-center effect (*Chapter 2*), we analyzed a large group of children and adolescents who do and do not stutter (96 participants in total). With the direct measure of articulatory behavior (using EMA), we recorded 20 adults – 10 adults who stutter and 10 adults who do not stutter, which is also a relatively large group for a kinematic study (*Chapter 4*).

The studies presented in *Chapter 2* and *Chapter 4* indicate that PWS time articulatory gestures at the beginning of words differently than PWNS. We hypothesized that more consonant compression in children and adolescents who stutter could derive from greater overlap between the rightmost consonant in a syllable onset cluster with the following vowel. Even though we did not compare syllables with a complex onset and a simple onset in the EMA study (*Chapter 4)*, we found evidence for more gestural overlap between consonants and vowels in adults who stutter. Therefore, it is likely that children and adolescents who stutter also show this pattern, as already suggested by acoustics (*Chapter 2*).

Whereas in children and adolescents who stutter and children and adolescents who do not stutter, group differences remained even in the Metronome condition - a speech fluency-enhancing context for PWS (*Chapter 2*), adults who stutter and adults who do not stutter did not differ in CV-timing across rhythmic conditions (*Chapter 4*, trajectory-based measure). This indicates that adults who stutter may have a more mature speech motor control system than younger people who stutter, allowing adults who stutter to make better use of external rhythmic cues, like a metronome, to adapt their speech timing. The interpretation of a more mature speech motor system in adults who stutter is also consistent with previous kinematic findings on CV coarticulation, where children who stutter were found to exhibit greater gestural overlap than their matched peers (Lenoci & Ricci, 2018), whereas no differences in coarticulatory behavior was found between adults who stutter and adults who do not stutter (Frisch et al., 2016).

Greater gestural overlap has been discussed as reflecting a lack of motor control precision in typically fluent speakers, as it is characteristic for speech in children (Noiray et al., 2018). However, similar to the CV-timing hypotheses on stuttering (Harrington, 1988; Wingate, 1988), there are two opposing hypotheses on coarticulatory behavior in typically developing children, as summarized by Noiray and colleagues (2018). One hypothesis posits, as already mentioned, that children coarticulate more while the other hypothesis states that children coarticulate less than adults (see a summary in Noiray et al., 2018). This also goes in hand with two different views on how children organize speech. According to the latter hypothesis (e.g., Gibson & Ohde, 2007; Green et al., 2002; Katz et al., 1991), children acquire speech in a segmentally driven manner, which describes the maturation of articulatory control as a sequential process. This suggests that inter-articulator coordination for larger units develops at a later stage, resulting in the transition from less to more coarticulation. The opposite approach (e.g., Goodell & Studdert-Kennedy, 1993; Nittrouer et al., 1996; Noiray et al., 2018) suggests that children initially plan speech units in a broader, more holistic manner rather than on a segmental level. As they gain fine-grained control over individual articulators, their speech becomes more precise. Consequently, they progress from more to less coarticulation with rising age (Noiray et al., 2018).

Our findings from *Chapters 2* and *4* support the latter approach of speech motor control maturation; more gestural overlap in children or less mature speech motor control systems, such as in PWS. The GAMM analysis (*Chapter 4*) further supports this view, as we found that adults who stutter and adults who do not stutter differ in the timing of their vowel gestures, indicating that adults who stutter initiate vowel gestures earlier than adults who do not stutter. We propose that the coordination of gestures presents a significant challenge, one that we hypothesize requires a mature speech motor control system.

5.3. Contributions to Theories and Models of Stuttering

The three studies presented in this thesis support several theories and hypotheses on stuttering. For instance, our results are in line with the Speech Motor Skill hypothesis positing that PWS have reduced skill in motor control compared to PWNS (Namasivayam & van Lieshout, 2011; van Lieshout, 2004). This hypothesis is supported by the finding that PWS showed more variability in speech motor timing (*Chapters 3*, *4*), as well as alterations in inter-gestural timing (*Chapters 2*, *4*). Variability among PWS was observed both across and within individuals who stutter as well as across conditions. Some PWS demonstrated high variability across trials or tasks, while others exhibited stable timing patterns that either deviated from or mirrored those of PWNS. This is consistent with the Speech Motor Skill hypothesis, as it presents speech motor skill as a continuum, where PWS can also reach levels of PWNS. However, while the speech motor skill approach offers a useful complementary framework, it is unlikely to provide a causal explanation for stuttering. Individuals with similarly low speech motor skills, such as those with apraxia, do not typically stutter, suggesting that reduced speech motor skill alone is insufficient to account for stuttering.

The observed variability is also consistent with the understanding of stuttering as a heterogenous and variable disorder (Bloodstein & Bernstein Ratner, 2008; Gerlach et al., 2020), and it is noteworthy that such fluctuations are also evident during perceptually fluent speech. These differences may be attributed to factors such as task demands, including for example linguistic complexity and metronome-timed vs. unpaced speech and their effects on the speech motor system of PWS (e.g., Namasivayam & van Lieshout, 2011). Ideally, longitudinal data collected from PWS over extended periods would offer a more comprehensive view of the stability or variability in their speech motor performance, allowing researchers to determine whether observed patterns persist or shift across sessions.

There is the possibility that variability interacts with stuttering severity, where PWS with very mild stuttering might develop speech motor skills comparable to those of PWNS and PWS with more severe stuttering could be positioned at the lower end of the continuum. However, in the studies at hand, no significant correlation was observed between stuttering severity and any of the measured variables: consonant compression, CV-lag, or aligning speech to a metronome beat or a finger tap. This absence of correlation suggests that stuttering severity, at least as typically assessed with the SSI, a clinical tool for stuttering diagnostics, may not reflect underlying differences in speech motor coordination, measured with the above-mentioned variables. Additionally, neither the outliers in beat alignment (visualized in *Figure 9*, *Chapter 4*) nor the high variability in metronome alignment in PWS can be attributed to participants with high stuttering severity.

Nevertheless, the observation that speaking along with a metronome appears to stabilize speech in adults who stutter (*Chapter 4*), but yet not in children who stutter (*Chapter 2*), further suggests that children who stutter are positioned lower on this speech motor skill continuum than adults who stutter.

Additionally, our results support Harrington's (1988) hypothesis of closer CV coupling in PWS and not Wingate's Fault-Line hypothesis (1988). According to Harrington's (1988) hypothesis, closer inter-gestural timing in PWS is caused by the incorrect prediction of the auditory feedback of their own speech, hence a deficit in predictive timing. PWS would therefore correct their articulation due to erroneous predictions of future sensory events in the feedback control system, resulting in greater CV overlap that leads to stuttering, when becoming too large. Wingate (1988), on the contrary, proposed that stuttering arises due to the delayed integration of the syllable nucleus with the onset consonant. According to this view, stuttering arises when there is a divide in the transition between CV gestures, leading to reduced gestural overlap.

An external pacemaker should facilitate the prediction of upcoming speech events (e.g., Jones et al., 2002; London, 2012). Therefore, we hypothesized that the Metronome condition (speaking in time with a metronome) would facilitate predictive timing in PWS. While inter-gestural timing of verbal gestures (CV-timing) in adults who stutter resembled that of adults who do not stutter in the Metronome condition, supporting our hypothesis, this was not observed in children who stutter. Thus, it is possible that children who stutter may face even greater challenges with auditory-motor integration than adults who stutter, as indicated by Kim et al. (2020).

In the following we argue that auditory-motor integration still remains a challenge for adults who stutter, as indicated by the verbal synchronization task consistent with previous studies (e.g., Sares et al., 2019; van de Vorst & Gracco, 2017). We found that adults who stutter timed their articulatory speech onset closer to a metronome beat than adults who do not stutter. Similar findings were reported from our lab for children who stutter (Schreier, 2023; Schreier et al., 2020), with the majority of these children participating in the first study (*Chapter 2*). There are several explanations for these synchronization differences in PWS, which should not be regarded as being the sole causes, but may well be complementary. One possibility is that inter-gestural timing differences, particularly the greater consonant compression observed in children who stutter, could play a role in how they synchronize their speech to a metronome. This may indicate that they are targeting the same reference point in speech to align with the metronome beat than their typically fluent peers. For example, due to more gestural overlap in children who stutter, the targeted reference point (e.g., articulatory vowel onset) occurs earlier. Therefore, they can start speaking later than their matched peers, in order to align with the same reference point.

Regarding the results in adults, there is little evidence that inter-gestural timing differences account for the observed synchronization differences between PWS and PWNS. But, using the landmark-based measure, we found that the Metronome+Tapping condition led to a shift in CV-lags: PWNS exhibited larger CV-lags, while PWS produced smaller ones. If speakers aim to align the articulatory vowel onset with the metronome beat, this suggests that PWS synchronize their articulatory speech onset earlier due to a shorter lag between C and V. However, differences in inter-gestural timing are unlikely the primary factor behind synchronization differences in adults who stutter, as the trajectory-based method revealed no group differences in vowel timing under rhythmic conditions (*Chapter 4*).

Another explanation is that PWS have a delayed syllable selection, as demonstrated with the GODIVA model (Civier et al., 2013). While our results are partially consistent with this account, they allow a more nuanced interpretation that involves the broader architecture of the DIVA/GODIVA framework, particularly the interplay between feedforward control, auditory feedback control, and somatosensory feedback control. Specifically, while the GODIVA model posits timing differences at the level of syllable selection and sequencing, our findings on inter-gestural timing in children and adolescents (*Chapter 2*), as well as in adults (*Chapter 4*), suggest that stuttering may also be linked to differences in how articulators are coordinated over time and not just delays in syllable selection. This is especially evident in the observed differences in CV-timing and consonant compression, which point towards anomalies in how articulatory gestures are coordinated over time, rather than solely delays in syllable selection. Such differences implicate a potential disfunction in the feedforward control system, which is responsible for initiating and coordinating well-learned speech motor commands without relying on feedback during fluent speech.

Furthermore, results from *Chapter 3* and *4*, which examined how PWS respond to external rhythmic cues, point to altered auditory-motor integration. The finding that PWS align their articulatory speech onset more closely with a metronome beat than PWNS, yet show more variability, suggests atypical auditory feedback control. Instead of using external auditory cues as stable timing anchors, PWS may experience instability or reduced gain in their auditory feedback loop, leading to inconsistent alignment with an external rhythm.

That we did not observe a group difference in the finger tapping condition highlights a possible compensatory role of (non-verbal) somatosensory feedback control. Given that target landmarks (used in our study) are more closely tied to proprioceptive and tactile information, the timing patterns observed in PWS suggest a stronger dependence on somatosensory feedback, which could be emphasized by the additional finger tapping due to reduced reliability or efficiency in the auditory feedback or feedforward systems.

Taken together, our results suggest that while delayed syllable activation remains a plausible contributing factor (as outlined in the GODIVA model), stuttering may also involve broader impairments across all three subsystems: 1) weakened feedforward control resulting in alterations in inter-gestural coordination, 2) atypical auditory feedback integration affecting the ability to use external auditory cues for speech alignment, and 3) potentially compensatory reliance on somatosensory feedback to maintain articulatory precision.

Furthermore, PWS could also have an altered perception of where the rhythmic beat of a syllable occurs, in line with studies showing that stuttering is linked to rhythm perception deficits (Wieland et al., 2015; Chang et al., 2016). This is further supported by the synchronization results with the acoustic vowel onset as a reference point, which indicate that the groups align the metronome beat with different targets (*Chapter 4*).

One likely explanation is that auditory-motor integration deficits underlie these findings. When an external rhythm was present, PWS aligned their finger taps more closely with their speech onset compared to the single Tapping condition. In contrast, PWNS shifted their taps further away from the speech onset and more towards the metronome. We hypothesize that PWNS rely more on auditory cues to align their speech, while PWS prioritize internally generated events, such as finger taps. Thus, a more compelling explanation for the verbal synchronization differences between PWS and PWNS is that PWS exhibit altered predictive timing and experience difficulties with auditory-motor integration (as proposed in *Chapter 4*).

We argue that even though the internal timing network of PWS is less reliable (Etchell et al., 2014), leading for example to speech timing differences between PWS and PWNS (e.g. voice-onset-time: De Nil & Brutten, 1991; Jäncke, 1994; Max & Gracco, 2005, articulatory: De Nil, 1995; Loucks et al., 2022; Kleinow & Smith, 2000; Smith et al., 2010; Wiltshire et al., 2021), the additional activation of a self-generated manual rhythm, i.e. finger tapping, may help stabilizing the internal timing network as we did not observe any differences between groups in the single Tapping condition (*Chapters 3*, *4*).

Research on typically fluent persons revealed that the premotor cortex plays a role in integrating verbal and non-verbal movements (Meister et al., 2009). As proposed by Etchell et al. (2014), this area plays also a key role in compensatory processes in PWS, in addition to the cerebellum and the right inferior frontal gyrus. Finger tapping could reinforce sensory feedback, as PWS have to align their speech precisely to the finger tap. Therefore, we hypothesize that finger tapping affects speech motor timing positively in PWS.

To summarize, our findings are in line with the assumption of disrupted processing within internal feedforward mechanisms (i.e., motor plan projections sent to the sensory system to generate expected perceptions based on the planned movement) (Harrington, 1988; Max et al., 2004; Max & Daliri, 2019; for a summary Bradshaw et al., 2021). In *Chapters 2* and *4*, we found evidence for alterations in inter-gestural timing which could be the result of erroneous predictions of future sensory states in adults who stutter (Harrington, 1988), given that they did not differ from their peers when synchronizing speech to a metronome, which is known to facilitate predictive timing. However, since the difference between children who stutter and children who do not stutter persists in the Metronome condition, we assume that they have more difficulties to integrate feedback and feedforward information than adults who stutter.

Moreover, that PWS initiated their speech later than PWNS (*Chapters 3* and *4*) could be rooted in the failure to activate the speech motor program of the next syllable (Civier et al., 2013). This points towards a malfunctioning feedforward system in PWS, as temporal motor control is disrupted. However, we question the assumption that the sequential initiation of syllable motor programs is the sole challenge for PWS (Civier et al., 2013), as our results indicate that the transition between single gestures (the onset and the vowel gesture) is altered in PWS.

Based on the studies at hand it can be concluded that stuttering is related to differences in speech motor control. However, speech motor control differences alone do not fully account for the complexity of stuttering. As discussed in this section, alterations in rhythm perception and timing mechanisms may also contribute to the disorder. Taken together, these findings support the view that stuttering is best understood as a multifactorial neurodevelopmental disorder (Smith & Weber, 2017), involving altered auditory and motor systems and their interaction.

5.4. Rhythm and Timing

Rhythmic timing in speech depends on the precise alignment of articulatory gestures (Poeppel & Assaneo, 2020). Misalignments can affect the rhythmic flow of speech and when they become too large, stuttering arises leading to a disruption of the rhythmic flow.

The thesis at hand revealed that alterations in articulatory coordination, particularly at the level of inter-gestural timing, are central to the perceptually fluent speech of PWS. In children and adolescents who stutter, these timing differences still lead to a similar syllabic temporal organization as in children who do not stutter as both groups show acoustic cues for a c-center effect (*Chapter 2*). Furthermore, adults who stutter showed more gestural overlap between onset consonants and the following vowel, particularly in the Unpaced condition.

Even though not a main result, PWS produced speech at a significantly lower rate than PWNS (*Chapters 2* and *4*), which indicates that the interplay between different articulators results in a slower speech rhythm in PWS.

Furthermore, results of the verbal synchronization task (*Chapters 3* and *4*) also point towards differences in several rhythmic domains, such as the integration of verbal and auditory rhythms (speaking with an external auditory rhythm) and the integration of verbal and non-verbal rhythms with an external auditory rhythm (speaking and tapping in sync with an external rhythm). As discussed in section 1.4.3, several theories can contribute to the explanation of the observation that PWS and PWNS differ in timing their speech to an external rhythm. We proposed that difficulties in predictive timing and auditory-motor integration – relevant mechanisms of fluent speech production – could be the driving factors behind the later speech initiation in PWS compared to PWNS. Hence, speech of PWS displays a different rhythmic pattern when synchronizing speech to an external rhythm, potentially driven by different underlying neural substrates (Frankford et al., 2021).

To draw conclusions about general speech timing mechanisms and rhythm in fluent speech production, it is important to consider how these are influenced by different contextual cues and developmental stages.

Even though articulatory movements are still maturing through childhood and adolescence (Smith, 2010), our results showed that temporal syllabic organization is already well established in children from the age of 9 years on. Future research could explore even younger age groups to gain deeper insights into how this temporal structure develops over time.

In our study with children and adolescents, metronome-paced speech appeared to selectively affect vowel compression without significantly impacting consonant compression. Specifically, vowels in words with complex onsets were shortened more than those in words with simple onsets when speech was synchronized to a metronome, compared to unpaced speech. Interestingly, consonant compression, as examined through the phoneme /l/, remained unaffected, despite the fact that /l/ can typically be lengthened or shortened much like vowels. This suggests that vowel duration may be more flexibly adjusted in response to external timing demands, and that vowels may serve as primary anchors for temporal coordination in rhythmic speech contexts.

Moreover, CV-timing was affected by the metronome – PWNS increased the lag in both metronome conditions compared to the Unpaced condition. The increased CV-lag under metronome conditions emphasizes that inter-gestural timing is flexible and modulated by external rhythmic cues. The CV-lag is a key part of the syllable's internal temporal structure. If metronome pacing leads to a larger CV lag, it may indicate that vowel gestures are being adapted, while the onset consonant remains relatively fixed in time. This reinforces the idea that vowels act as temporal anchors, especially when synchronizing speech to an external rhythm. This also supports the idea that speech timing involves both internal and external timing mechanisms. PWNS adjusting their CV timing in response to the metronome shows that even fluent speakers use external timing cues to modulate their internal speech timing. This points to a hybrid model of speech timing, rather than one governed purely by internal motor control.

The differences observed in how speech onsets were synchronized with different rhythmic cues, namely tapping and metronome, highlight the role of context in shaping speech timing mechanisms. Finger tapping was closely aligned with the speech onset, suggesting a reliance on internal timing mechanisms. In contrast, synchronization with a metronome revealed a stronger influence of acoustic cues (vowels) with participants aligning their acoustic vowel onset very closely to the metronome beat.

A novel contribution of our study is the comparison of three rhythmic conditions: Tapping, Metronome, and Metronome + Tapping. We found that PWNS were more responsive to external auditory cues, as reflected in a shift of their finger taps toward the metronome beat. Conversely, PWS appeared to rely more on internal motor timing, indicated by a shift of their finger taps toward the articulatory speech onset. This finding emphasizes the distinction between internal and external timing mechanisms in speech production and their differential recruitment across populations.

To refer to the question from the introduction about what perfect timing means in relation to speech, we can say that the timing of speech allows for a broad range to be perceived as perceptually fluent while still showing signs of an altered speech rhythm. Fluency is typically defined as smooth, uninterrupted speech that emerges from the effortless coordination of articulatory movements. For PWNS, fluency is generally the product of a stable and automatized speech motor system. In contrast, for PWS, even perceptually fluent speech often differs in its underlying characteristics, as also demonstrated by this thesis. Fluency in PWS may be achieved through compensatory mechanisms, for instance, by circumventing the error-prone basal ganglia–thalamo–cortical loop (Chang & Guenther, 2020; Frankford et al., 2021) or through strategies like altered speaking patterns. According to the Speech Motor Skill Hypothesis, some PWS can reach performance levels comparable to PWNS, but they may do so via different neural and behavioral pathways. This raises the important question of whether fluency, although similar in outward appearance, represents the same phenomenon in PWS and PWNS.

5.5. Outlook

As discussed in the previous chapters, several opportunities for future research can be derived from the present work. For instance, our EMA dataset offers the potential to address further research questions related to articulatory timing in PWS and PWNS.

Specifically, we could investigate CV-timing in disfluent productions and compare it to our findings on perceptually fluent speech in PWS. Notably, our EMA data set contains numerous stuttered target words from two subjects in particular, allowing us to investigate Harrington's (1988) hypothesis further. We would expect to find even more overlap between CV gestures in stuttered speech compared to fluent CV productions. This comparison would contribute even more to our understanding of stuttering and why a breakdown of speech fluency occurs.

In addition, the EMA data set enables us to investigate the c-center effect in adults who stutter and adults who do not stutter across different rhythmic conditions. This would allow us to compare articulatory data of adults to the acoustic correlates found in children, shedding more light on speech motor control development not only in PWS but also in PWNS. Based on our results on inter-gestural timing in *Chapter 4*, we assume to find group differences in the c-center organization in the Unpaced condition in adults who stutter, whereas no differences are expected in the rhythmic conditions.

Another interesting avenue for future research is to investigate onset-vowel timing relations across groups in a wider range of hierarchical prosodic positions, such as at word or phrase edge positions, word-initially or in stressed vs. unstressed words. There is evidence that words at the beginning of a phrase are more likely to be stuttered, as well as words that are less predictable in context (for a review, Brundage & Bernstein Ratner, 2022). Moreover, prosodic boundaries and lexical stress are known to influence articulatory timing and coordination (Byrd et al., 2000; Cho, 2006; Cho et al., 2014), which may interact differently with speech motor planning processes in PWS. Investigating whether onset-vowel timing differs systematically as a function of prosodic structure could provide deeper insight into the conditions under which gestural coordination becomes challenging for PWS and ultimately contribute to more targeted models of stuttering.

Since our results indicate between-group differences in the integration of different sources of feedback (i.e., auditory and somatosensory), and given that the availability of these sources varies considerably over the course of a syllable (as well as between different syllables[1]), it would be valuable to investigate new material that allows for a direct comparison of onset-onset and

[1] An extreme example: for /pa/ in utterance-initial position, auditory feedback is available much later relative to somatosensory feedback (cf. Oschkinat & Hoole, 2020).

target-target coordination precision. Additionally, it would be interesting to examine how precisely metronome beats, taps, or other events align with these different landmarks.

The finding that adults who stutter and adults who do not stutter did not differ in the single Tapping condition (tapping while speaking) appears to be a promising direction for future research. We assume that an internally generated non-verbal rhythmic movement, such as finger tapping, stabilizes the speech motor system of PWS significantly. Also, how for example walking, another rhythmic non-verbal movement, affects speech timing in PWS would be interesting to explore. This research could pave the way for therapeutic approaches integrating facilitating movements with speech therapy (e.g., fluency shaping or stuttering modification techniques).

An interesting approach to further investigate auditory-motor integration in PWS would be to use paradigms that manipulate auditory feedback and/or motor behavior. For example, it would be attractive to explore how PWS respond compared to PWNS when their speech onset gets delayed while producing simple syllables and simultaneously tapping their finger. This experiment would provide insight not only into their ability to integrate auditory feedback from their speech with a self-generated rhythmic cue but also their cue preference. More specifically, are they more likely to adapt verbal or non-verbal gestures, hence, are they more prone to relying on auditory vs. tactile feedback, or do they adapt both? There are three possible responses: 1) shifting the speech onset while tapping remains stable, so the auditory target (e.g. acoustic vowel onset) of the delayed speech signal aligns with the finger tap target (e.g., surface of a wooden block), 2) adapting the tapping while no shift is observed in their speech, so the finger tap target aligns with the auditory speech target, and 3) a shift in both speech and finger tapping.

Recent findings by Lazarri et al. (2024) showed that while delayed auditory feedback of the tapping sound affected non-verbal sensorimotor synchronization (tapping to a metronome) in PWNS, the performance of PWS remained stable. This finding suggests a reduced sensitivity to disruptions in action-perception loops in PWS. Additionally, the authors found that PWS are less able to detect delayed auditory feedback (Lazarri et al., 2024). These results reinforce the idea that PWS may rely less on auditory feedback for motor adjustments. Based on the results of this study, I would hypothesize that PWS probably respond less than PWNS, but that they are more likely to adapt the tapping behavior than their speech, as they prioritize tactile feedback. In contrast, as suggested in *Chapter 4*, I expect PWNS to favor auditory cues for synchronization and therefore, I would hypothesize that they are more likely to adapt their speech.

Expanding on this, another interesting approach would be to alter the timing of the finger tapping movement, for example by adding a resistance that makes it more difficult to reach the target (e.g., the surface of a wooden block) while speaking. This could reveal whether PWS are sensitive to increased non-verbal motor demands and their impact on speech motor coordination. I would expect that PWS respond less adaptively than PWNS, meaning they might struggle to adjust their speech timing in response to the modified finger tapping, suggesting a reduced ability to integrate external motor constraints into their speech production system. This experiment could provide further insight into the ability to integrate sensory and motor information across modalities.

An area that remains unexplored in stuttering is the effect of the temporal aspects of perceived speech on prediction abilities and speech motor planning. Testing neural correlates of stuttering remains one of the most promising topics for future research. As mentioned in the *Preface*, I developed an EEG experiment during my PhD to address this research gap by examining the role of rhythm in the intertwining of speech perception and production in PWS and PWNS. As noted in section 1.4., stuttering is highly variable and is influenced by many factors, such as the communicative context and interlocutors (Bloodstein & Bernstein Ratner, 2008; Gerlach et al., 2020). Furthermore, when PWS speak fluently they show a greater effort in speech motor preparation, which is a compensatory mechanism (Vanhoutte et al., 2016).

It is a possibility that rhythm plays a key role in explaining the variability in speech fluency, typical for stuttering. I hypothesize, for example, that an interlocutor who has a very rhythmic speaking style, facilitates speech motor planning for PWS, leading to more speech fluency and reduced effort in speech motor preparation.

With the EEG study, we investigate the (neural) link between speech perception and speech production in PWS and PWNS. The EEG experiment presents an innovative setup with which we first test to what extent speech motor preparation is influenced by the auditory context (fluent vs. disfluent stimuli) in typically developing adults. A research paper is currently in preparation. Details about the study can be found in the study protocol in Appendix A.

5.6. Conclusion

This dissertation provides important insights into the articulatory timing mechanisms of PWS and PWNS across various rhythmic conditions. By investigating both children and adults who stutter, the studies in this work have revealed that while both populations exhibit differences in the coordination of speech gestures compared to typically fluent peers, these differences vary with age and the presence of rhythmic cues.

Key findings across the studies show that PWS exhibit greater gestural overlap, particularly in unpaced speech, suggesting difficulties in coordinating articulatory gestures. However, the presence of a metronome improved inter-gestural timing in adults who stutter, eliminating group differences but not in children. This suggests that the development of speech motor control plays a key role in how rhythmic contexts are processed by PWS.

The dissertation also revealed that PWS differ from PWNS in a verbal sensorimotor synchronization task, that is synchronizing speech to a metronome, but not when synchronizing their speech to their own finger tapping. Together, these findings support that PWS have difficulties in auditory-motor integration and predictive timing and that speech timing mechanisms are less stable than those of PWNS.

Detailed summary

The smooth coordination of articulatory movements contributes to perceiving speech as rhythmic. This becomes particularly evident when articulatory coordination is disrupted, such as in stuttering. Stuttering is a speech fluency disorder with a neural origin which results in repetitions (“su su super”), prolongations (“sssssuper”), and blocks (“---super”) (WHO, 2016). These disfluencies can be significantly reduced when persons who stutter (PWS) synchronize their speech with an external rhythm, like a metronome (e.g., Andrews et al., 1982), but the role of rhythm during fluent speech production remains poorly understood.

Stuttering is linked with differences in verbal (e.g., Loucks et al., 2022; Smith et al., 2010; Wiltshire et al., 2021) and non-verbal movements (e.g., Falk et al., 2015; Sares et al., 2019; Slis et al., 2023). These differences between PWS and persons who do not stutter (PWNS) are attributed to distinct underlying neural processes (e.g., Etchell et al., 2014).

Word and syllable onsets are particularly critical for PWS, because this is where stuttering typically occurs (Harrington, 1987; Howell & Au-Yeng, 2002) and even speech that appears fluent often differs from that of PWNS (e.g., Dehqan et al., 2016; Max & Gracco, 2005; Verdurand et al., 2020).

Thus, examining the speech of PWS that appears fluent, especially in rhythmic contexts (like speaking along with a metronome or finger tapping while speaking), provides important insights into the mechanisms behind fluent speech production. The main goal of this thesis is to enhance our understanding of speech timing mechanisms in PWS and PWNS by examining the effect of rhythmic conditions on fluent speech articulation. By investigating speech in children and adolescents who stutter (*Chapter 2*) and adults who stutter (*Chapters 3* and *4*), this research also deepens our understanding of speech motor control development in PWS. A specific focus lies on articulatory timing at word onsets and inter-gestural timing, the coordination of two individual (articulatory) movements.

This cumulative dissertation is a compilation of three original empirical studies corresponding to *Chapters 2*, *3*, and *4*. These use different methodological approaches to pursue the main aim of the thesis (acoustics in *Chapter 2*, articulatory [electromagnetic articulography - EMA] in *Chapters 3* and *4*).

Chapter 2 is concerned with investigating acoustic cues for a *c-center effect* in unpaced and metronome-paced speech in children and adolescents who stutter and children and adolescents who do not stutter.

The c-center effect describes the phenomenon that there is a constant temporal relationship between the temporal center of the onset and the following vowel regardless of the number of consonants contained in the onset (Browman & Goldstein, 1988). Acoustically, this effect manifests in shorter vowels (vowel compression) and shorter consonants (consonant compression) in words with a complex onset compared to words with a simple onset (e.g., Katz, 2010). Investigating the c-center effect in young PWS is relevant because there is limited work on articulatory properties of children's speech in relation to stuttering and it allows us to examine whether there are difficulties in gestural timing at a young age.

Therefore, 96 German-speaking children and adolescents in the age range of 9 to 18 – 48 who stutter, 48 who do not stutter – were recorded acoustically while reading monosyllabic words with and without a metronome (unpaced and paced conditions). Analyses were conducted on four minimal pairs that differed in onset complexity (simple vs. complex).

The central hypothesis of this chapter is that both children and adolescents who stutter and who do not stutter will exhibit acoustic evidence of the c-center effect. However, group differences in compression effects are expected to emerge in the unpaced condition, while no significant differences are anticipated between the groups in the paced condition.

We found that both groups showed acoustic cues for a c-center effect. The main finding is that children and adolescents who stutter exhibited greater consonant compression (and no difference in vowel compression), indicating that they time articulatory gestures differently than their matched peers. Notably, the paced condition did not eliminate this group difference, suggesting that even though speaking along with a metronome significantly enhances speech fluency in PWS, it does not influence the temporal organization of articulatory gestures in the developing population who stutters.

The focus of *Chapter 3* is on the effect of rhythmic conditions on articulatory timing, as well as the interplay between verbal (speech) and non-verbal (finger tapping) gestures in adults who stutter and adults who do not stutter.

Although stuttering is associated with disruptions in speech timing mechanisms (Etchell et al., 2014) that also extend to the non-verbal domain (e.g., Falk et al., 2015; Slis et al., 2023), articulatory insights into these timing mechanisms remain unexplored.

Therefore, this study addresses this gap by exploring articulatory timing of four PWS and four PWNS in three different rhythmic conditions: speaking and tapping (Tapping condition), speaking along with a metronome (Metronome condition), and speaking and tapping to a metronome (Metronome+Tapping condition). Using EMA, gestures of the articulatory speech onset and the finger taps were recorded and analyzed.

We were interested in i) whether the groups differ in gestural timing, ii) whether speech gesture timing is dependent on the rhythmic context, and iii) how an external rhythm affects the timing of verbal (speech) and non-verbal (finger tapping) gestures.

In general, the articulatory speech onset preceded both rhythmic events (finger taps and metronome beats), but finger taps were more closely aligned with the speech onset than the metronome beat.

No group differences where observed when participants were simultaneously tapping and speaking, indicating that inter-gestural timing between verbal and non-verbal gestures is similar in PWS and PWNS. However, articulatory gesture onsets in PWS were placed more closely to the metronome beat than in PWNS, pointing towards later speech initiation in PWS.

In the combined Metronome+Tapping condition both groups aligned their speech onset more closely to the metronome compared to the single Metronome condition. However, only adults who stutter timed their finger taps more towards the articulatory speech onset in the combined condition than in the single Tapping condition. Hence, the presence of an external rhythm led to a closer coupling of verbal and non-verbal gestures in PWS, but not in PWNS.

These results suggest that PWS have difficulties in predicting external auditory events which has also been observed in non-verbal tasks (e.g., Falk et al., 2015). This may cause challenges in integrating external cues with their own speech production and therefore, lead to a difference in the Metronome but not in the Tapping condition. Furthermore, differences in the Metronome+Tapping condition suggest that PWS and PWNS differ in integrating auditory (metronome), manual (finger tapping), and articulatory (speech) rhythms.

Chapter 4 investigates consonant-vowel (CV)-timing and predictive timing building on the methodology and research questions introduced in *Chapter 3*.

This chapter situates the questions from *Chapter 3* in the context of predictive timing, an area where PWS face challenges (e.g., Etchell et al., 2014; Falk et al., 2015; Harrington 1988). CV-timing has also been hypothesized to be challenging for PWS, as highlighted by different theories (Harrington, 1988; Wingate, 1988), and supported primarily by acoustic data (Verdurand et al., 2020; Lenoci & Ricci, 2018; Dehqan et al., 2016; Robb & Blomgren, 1997; Klich & May, 1982). While it is known that a metronome can enhance speech fluency significantly in individuals who stutter, the articulatory mechanisms behind this effect remain unexplored.

To bridge this gap, the study presented in *Chapter 4* investigates articulatorily (EMA) data of 10 adults who stutter and 10 adults who do not stutter (age range between 19 and 32 years). Participants were recorded using EMA while producing target words that started with a bilabial onset, followed by a vowel (/a/, /o/, or /u/). These target words were embedded in a carrier phrase. The experiment comprised 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).

The main research questions concerning CV-timing are: i) whether CV coupling differs between adults who stutter and adults who do not stutter, and ii) whether inter-gestural timing is affected by different rhythmic pacing conditions.

To investigate CV-timing, we used two different measures: one measure of the distance between the onset of the consonant gesture and the vowel gesture (CV-lag) and one dynamic measure of the trajectory of the tongue back (GAMM-based measure). Both measures indicate a general pattern where PWS exhibit more overlap between CV gestures than PWNS but the measures produced different results across conditions. The trajectory-based approach revealed the more powerful statistical model, indicating significant group differences in the Unpaced condition, but not in any of the rhythmic conditions.

We could replicate the results from *Chapter 3* with a greater participant sample, with the exception that we found that PWS and PWNS timed their finger taps in opposite directions in the combined Metronome+Tapping condition. Whereas PWS still timed finger taps and speech onsets more closely than in the single Tapping condition, PWNS shifted their taps further away from the speech onset and more toward the metronome beat. Additionally, PWS showed greater variability in timing their speech to a metronome beat compared to a finger tap.

This indicates that PWS may experience difficulties in predictive timing and auditory-motor integration, leading to differences in how they time their speech to an external rhythm, but not to an internally generated rhythmic event, such as finger tapping. Furthermore, results from the Metronome+Tapping condition suggest that PWS and PWNS may rely on different cues to align their speech – PWNS use the metronome and PWS the finger tap.

In conclusion, this cumulative dissertation sheds light on articulatory timing mechanisms of PWS and PWNS across various rhythmic conditions and contributes to various theories and models of stuttering. This work provides evidence for inter-gestural timing differences in children and adolescents, as well as adults who stutter, suggesting difficulties in coordinating articulatory gestures. The presence of an external rhythm improved inter-gestural timing in adults who stutter, eliminating group differences, but not in children. This suggests that speech motor control development plays a key role in how rhythmic contexts are processed by PWS.

The dissertation also reveals that adults who stutter differ from adults who do not stutter in synchronization time points when synchronizing speech to a metronome, but not when synchronizing their speech to their own finger tapping.

In summary, these findings suggest that PWS experience challenges with auditory-motor integration and predictive timing, and that their speech timing mechanisms are less stable compared to those of PWNS, indicating a weaker speech motor control system.

Résumé détaillé

La coordination sans heurt des mouvements articulatoires contribue à ce que la parole soit perçue comme rythmique. Ceci est particulièrement évident lorsque la coordination articulatoire est perturbée, comme dans le cas du bégaiement. Le bégaiement est un trouble de la fluidité de la parole d'origine neuronale qui se traduit par des répétitions (« su su super »), des prolongations (« sssssuper ») et des blocages (« ---super ») (WHO, 2016). Ces disfluences peuvent être considérablement réduites lorsque les personnes qui bégaient (PWS) synchronisent leur parole avec un rythme externe, comme un métronome (Andrews et al., 1982), mais le rôle du rythme lors de la production de parole fluide n'a pas encore été exploré.

Le bégaiement est associé à des différences dans les mouvements verbaux (p. ex. Loucks et al., 2022 ; Smith et al., 2010 ; Wiltshire et al., 2021) et non verbaux (p. ex. Falk et al., 2015 ; Sares et al., 2019 ; Slis et al., 2023), qui sont provoquées par processus neuronaux sous-jacents différents (p. ex. Etchell et al., 2014).

Les débuts de mots et de syllabes sont particulièrement critiques pour les PWS, car le bégaiement se produit typiquement à ces endroits (Harrington, 1987 ; Howell & Au-Yeng, 2002) et même la parole qu'elles semblent produire de manière fluide est souvent différente de celle des personnes qui ne bégaient pas (PWNS) (p. ex. Dehqan et al., 2016 ; Max & Gracco, 2005 ; Verdurand et al., 2020).

Par conséquent, l'exploration de la parole en apparence fluide des PWS, en particulier dans des conditions rythmiques (comme parler avec un métronome ou taper du doigt en parlant), fournit des renseignements importants sur les mécanismes sous-jacents de la production de la parole fluide.

L'objectif principal de cette thèse est d'améliorer notre compréhension des mécanismes du timing de la parole chez les PWS et les PWNS en étudiant les effets des conditions rythmiques sur l'articulation de la parole fluide. En étudiant la parole d'enfants et d'adolescents qui bégaient (chapitre 2) et d'adultes qui bégaient (chapitres 3 et 4), cette recherche approfondit également notre compréhension du développement du contrôle moteur de la parole chez les PWS. Une attention particulière est portée au timing articulatoire au début des mots et au timing inter-gestuel, soit la coordination de deux mouvements (articulatoires) individuels.

Cette thèse cumulative est composée de trois études empiriques originales correspondant aux chapitres 2, 3 et 4. Celles-ci utilisent différentes approches méthodologiques afin de poursuivre l'objectif principal de la thèse (acoustique au chapitre 2, articulatoire [articulographie électromagnétique ; EMA] aux chapitres 3 et 4).

Le chapitre 2 est consacré à l'étude des indices acoustiques d'un effet « c-center » chez les enfants et les adolescents qui bégaient et chez les enfants et les adolescents qui ne bégaient pas, lorsqu'ils parlent avec et sans métronome.

L'effet c-center correspond au phénomène selon lequel il existe une relation temporelle constante entre le centre temporel de l'attaque et la voyelle suivante, indépendamment du nombre de consonnes contenues dans l'attaque (Browman & Goldstein, 1988). Acoustiquement, cet effet se manifeste par des voyelles plus courtes (compression des voyelles) et des consonnes plus courtes (compression des consonnes) dans les mots avec une attaque complexe par rapport aux mots avec une attaque simple (p. ex. Katz, 2010). L'étude de l'effet c-center chez les jeunes PWS est pertinente, car il existe peu de travaux sur les caractéristiques articulatoires de la parole des enfants en lien avec le bégaiement et une telle étude nous permet d'examiner si des difficultés de timing gestuel sont déjà présentes à un jeune âge.

Par conséquent, 96 enfants et adolescents germanophones âgés de 9 à 18 ans - 48 qui bégaient, 48 qui ne bégaient pas - ont été enregistrés acoustiquement pendant qu'ils lisaient des mots monosyllabiques avec et sans métronome. Les analyses ont été effectuées sur quatre paires minimales qui se distinguaient par la complexité de l'attaque du mot (simple versus complexe).

L'hypothèse centrale de ce chapitre est que les enfants et les adolescents qui bégaient, ainsi que ceux qui ne bégaient pas, présenteront des indices acoustiques de l'effet de c-center. On s'attend toutefois à ce que des différences se manifestent entre les groupes dans les effets de compression dans la condition sans métronome, alors qu'aucune différence significative entre les groupes n'est attendue dans la condition avec métronome.

Nous avons constaté que les deux groupes présentaient des indices acoustiques de l'effet c-center. Le résultat principal est que les enfants et les adolescents qui bégaient présentaient une compression consonantique plus grande (mais il n'y avait pas de différence de groupe pour la compression des voyelles), indépendamment de la condition, ce qui indique qu'ils temporisent les gestes articulatoires différemment des enfants et des adolescents qui ne bégaient pas. Cela suggère que, bien que parler avec un métronome améliore significativement la fluidité de la parole des PWS, cela n'affecte pas l'organisation temporelle des gestes articulatoires dans la population en développement qui bégaie.

Le chapitre 3 se concentre sur les effets des conditions rythmiques sur le timing articulatoire ainsi que sur l'interaction entre les gestes verbaux (parole) et non verbaux (taper du doigt) chez les adultes qui bégaient et ceux qui ne bégaient pas.

Bien que le bégaiement soit associé à des perturbations des mécanismes de timing de la parole (Etchell et al., 2014), qui s'étendent également au domaine non verbal (p. ex. Falk et al., 2015 ; Slis et al., 2023), l'aspect articulatoire de ces mécanismes de timing n'a pas encore été étudié.
La présente étude comble cette lacune en examinant le timing articulatoire de quatre PWS et de quatre PWNS dans trois conditions rythmiques différentes : parler et taper avec le doigt (condition Tapping), parler avec un métronome (condition Metronome) et parler et taper avec un métronome (condition Metronome+Tapping). À l'aide de l'EMA, les gestes articulatoires de la parole lors de l'attaque et ceux du tapement de doigt ont été enregistrés et analysés.
Nous étions intéressés de savoir i) si les groupes se différenciaient dans le timing gestuel, ii) si le timing des gestes verbaux dépendait du contexte rythmique et iii) comment un rythme externe influençait le timing des gestes verbaux et non verbaux.
En général, le début articulatoire de la parole précédait les deux événements rythmiques (tapements de doigt et battements de métronome), mais les tapements de doigt étaient plus proches du début de la parole que les battements de métronome.
Aucune différence entre les groupes n'a été observée lorsque les participants tapaient du doigt et parlaient en même temps, ce qui suggère que le timing entre les gestes verbaux et non verbaux est similaire chez les PWS et les PWNS. Cependant, le début des gestes articulatoires était plus proche du battement du métronome chez les PWS que chez les PWNS, ce qui indique une initiation plus tardive de la parole chez les PWS.
Dans la condition combinée Metronome+Tapping, les deux groupes ont aligné le début de leur parole plus près du métronome que dans la condition Metronome. Cependant, seules les PWS ont synchronisé plus étroitement leurs tapements de doigt avec le début de la parole articulatoire dans la condition combinée que dans la condition Tapping. Ainsi, la présence d'un rythme externe a conduit à un couplage plus étroit des gestes verbaux et non verbaux chez les PWS, mais pas chez les PWNS.
Ces résultats suggèrent que les PWS ont des difficultés à prédire les événements auditifs externes, ce qui a également été observé dans des tâches non verbales (p. ex. Falk et al., 2015). Cela peut conduire les PWS à éprouver des problèmes d'intégration d'indices externes dans leurs propres productions verbales et donc à une différence dans la condition Metronome, mais pas dans la condition Tapping. De plus, les différences entre groupes dans la condition Metronome+Tapping indiquent que les PWS et les PWNS diffèrent dans l'intégration des rythmes auditifs (métronome), manuels (taper du doigt) et articulatoires (parole).

Le chapitre 4 examine le timing consonne-voyelle (CV) et le timing prédictif, en s'appuyant sur la méthodologie et les questions de recherche présentées au chapitre 3.

Dans ce chapitre, les questions du chapitre 3 s'inscrivent dans le contexte du timing prédictif, un domaine dans lequel les PWS rencontrent des difficultés (par ex. Etchell et al., 2014 ; Falk et al., 2015 ; Harrington 1988). On a également supposé que le timing CV représentait un défi particulier pour les PWS, ce qui a été souligné par différentes théories (Harrington, 1988 ; Wingate, 1988) et confirmé principalement par des données acoustiques (Verdurand et al., 2020 ; Lenoci & Ricci, 2018 ; Dehqan et al., 2016 ; Robb & Blomgren, 1997 ; Klich & May, 1982). Bien que l'on sache que parler avec un métronome peut améliorer de manière significative la fluidité de la parole des PWS, les mécanismes articulatoires à l'origine de cet effet restent inexplorés.

Pour combler cette lacune, l'étude présentée au chapitre 4 a examiné les données articulatoires (EMA) de 10 adultes qui bégaient et de 10 adultes qui ne bégaient pas (âgés de 19 à 32 ans). Les participants ont été enregistrés avec un système EMA alors qu'ils produisaient des mots cibles commençant par une attaque bilabiale suivie d'une voyelle (/a/, /o/ ou /u/). Ces mots cibles étaient intégrés dans une phrase porteuse. L'expérience comprenait quatre conditions différentes : Unpaced (parler seulement), Tapping (parler et taper avec le doigt), Metronome (synchronisation de la parole avec un métronome) et Metronome+Tapping (parler sur un métronome tout en tapant du doigt).

Les principales questions de recherche concernant le timing CV étaient de savoir si i) le couplage CV est différent entre les adultes qui bégaient et les adultes qui ne bégaient pas, et ii) si le timing inter-gestuel est influencé par des conditions rythmiques différentes.

Pour étudier le timing CV, nous avons effectué deux analyses différentes : une analyse du délai entre le début du geste consonantique et le début du geste vocalique (CV-lag) et une analyse dynamique de la trajectoire du dos de la langue (fondée sur des GAMMs). Les deux analyses indiquent une tendance générale selon laquelle les PWS présentent plus de chevauchements entre les gestes CV que les PWNS. L'analyse dynamique de la trajectoire s'avère le modèle statistique le mieux ajusté et révèle des différences significatives entre les groupes dans la condition sans rythme (condition Unpaced), mais dans aucune des conditions rythmiques.

Les résultats du chapitre 3 ont été reproduits avec un plus grand nombre de participants, à l'exception du fait que nous avons constaté que les PWS et les PWNS synchronisaient leurs tapements de doigt dans des directions opposées dans la condition combinée Metronome+Tapping. Alors que les PWS continuaient à produire des tapements de doigts plus près de l'attaque de la parole que dans la condition Tapping, les PWNS éloignaient davantage leurs tapements du début de la parole et les rapprochaient davantage du battement du métronome. En outre, les PWS ont montré une plus grande variabilité dans la synchronisation de la parole avec le battement du métronome qu'avec le tapement du doigt.

Cela suggère que les PWS peuvent avoir des difficultés de timing prédictif et d'intégration auditivo-motrice, ce qui entraîne des différences dans la synchronisation de leur parole avec un rythme externe, plutôt qu'avec un événement rythmique généré à l'interne, tel que le tapement de doigt. En outre, les résultats de la condition Metronome+Tapping indiquent que les PWNS et les PWNS s'appuient sur des indices différents pour aligner leur parole – les PWNS utilisent le métronome et les PWS le tapement de doigt.

Pour conclure, cette thèse cumulative contribue à la compréhension des mécanismes du timing articulatoire des PWS et des PWNS dans différentes conditions rythmiques et apporte une contribution précieuse aux diverses théories et modèles du bégaiement. Ce travail montre qu'il existe des différences dans le timing inter-gestuel chez les enfants et les adolescents ainsi que chez les adultes qui bégaient, ce qui suggère des difficultés dans la coordination des gestes articulatoires. La présence d'un rythme externe a amélioré le timing inter-gestuel chez les adultes qui bégaient, éliminant ainsi les différences entre groupes, mais pas chez les enfants. Cela suggère que le développement de la motricité de la parole joue un rôle clé dans la façon dont les contextes rythmiques sont traités par les PWS.

La thèse montre également que les adultes qui bégaient diffèrent des adultes qui ne bégaient pas en ce qui concerne le moment de synchronisation de leur parole avec un métronome, mais pas avec le rythme de leur propre doigt (Tapping).

En résumé, ces résultats suggèrent que les PWS ont des problèmes d'intégration auditivo-motrice et de timing prédictif, et que leurs mécanismes de synchronisation de la parole sont moins stables que ceux des PWNS, ce qui indique un système de contrôle moteur de la parole plus faible.

Ausführliche Zusammenfassung

Die reibungslose Koordination von artikulatorischen Bewegungen trägt dazu bei, dass Sprache als rhythmisch wahrgenommen wird. Dies wird besonders deutlich, wenn die artikulatorische Koordination gestört ist, wie zum Beispiel beim Stottern. Stottern ist eine Störung des Redeflusses, die einen neuronalen Ursprung hat und zu Wiederholungen („su su super"), Längungen („sssssuper") und Blockaden („---super") führt (WHO, 2016). Diese Unflüssigkeiten können deutlich reduziert werden, wenn Personen, die stottern (PWS) ihre Sprache mit einem externen Rhythmus, wie zum Beispiel einem Metronom, synchronisieren (Andrews et al., 1982). Aber welche Rolle Rhythmus bei der flüssigen Sprachproduktion spielt, ist noch nicht ergründet worden.

Stottern wird mit Unterschieden in verbalen (z. B. Loucks et al., 2022; Smith et al., 2010; Wiltshire et al., 2021) und nonverbalen Bewegungen (z. B. Falk et al., 2015; Sares et al., 2019; Slis et al., 2023) in Verbindung gebracht. Diese Unterschiede zwischen PWS und Personen, die nicht stottern (PWNS) werden durch unterschiedliche zugrunde liegende neuronale Prozesse hervorgerufen (z. B. Etchell et al., 2014). Wort- und Silbenanfänge sind für PWS besonders kritisch, da das Stottern typischerweise an diesen Stellen auftritt (Harrington, 1987; Howell & Au-Yeng, 2002) und selbst Sprache, die flüssig erscheint, unterscheidet sich oft von der von PWNS (z. B. Dehqan et al., 2016; Max & Gracco, 2005; Verdurand et al., 2020). Daher bietet die Untersuchung der wahrnehmbar flüssigen Sprache von PWS, insbesondere unter rhythmischen Bedingungen, wertvolle Einblicke in die zugrunde liegenden Mechanismen der flüssigen Sprachproduktion.

Das Hauptziel dieser Arbeit ist es, unser Verständnis der Mechanismen des Sprechtimings bei PWS und PWNS zu verbessern, indem die Auswirkungen rhythmischer Bedingungen auf die flüssige Sprechartikulation untersucht werden. Durch die Untersuchung des Sprechens von Kindern und Jugendlichen, die stottern (Kapitel 2), und von Erwachsenen, die stottern (Kapitel 3 und 4), vertieft diese Forschung auch unser Verständnis der Entwicklung der Sprechmotorik bei PWS. Ein besonderer Schwerpunkt liegt dabei auf dem artikulatorischen Timing von Wortanfängen und dem inter-gestischem Timing, der Koordination zweier einzelner (artikulatorischer) Bewegungen.

Diese kumulative Dissertation setzt sich aus drei empirischen Originalstudien zusammen, die den Kapiteln 2, 3 und 4 entsprechen. Diese verwenden verschiedene methodische Ansätze, um das Hauptziel der Arbeit zu verfolgen (Akustik in Kapitel 2, Artikulation [elektromagnetische Artikulographie - EMA] in Kapitel 3 und 4).

Kapitel 2 befasst sich mit der Untersuchung akustischer Hinweise eines *c-center* Effekts bei Kindern und Jugendlichen, die stottern, und Kindern und Jugendlichen, die nicht stottern, wenn sie mit und ohne Metronom sprechen. Der c-center Effekt beschreibt das Phänomen, dass es eine konstante zeitliche Beziehung zwischen dem zeitlichen Zentrum des Anlauts und dem folgenden Vokal gibt, unabhängig von der Anzahl der Konsonanten, die im Anlaut enthalten sind (Browman & Goldstein, 1988). Akustisch manifestiert sich dieser Effekt in kürzeren Vokalen (Vokalkompression) und kürzeren Konsonanten (Konsonantenkompression) in Wörtern mit komplexem Anlaut im Vergleich zu Wörtern mit einfachem Anlaut (z. B. Katz, 2010). Die Untersuchung des c-center Effekts bei jungen PWS ist relevant, da es nur wenige Arbeiten zu den artikulatorischen Eigenschaften der Sprache von Kindern im Zusammenhang mit Stottern gibt und es uns erlaubt zu untersuchen, ob Schwierigkeiten beim gestischen Timing bereits in jungen Jahren vorliegen.

Daher wurden 96 deutschsprachige Kinder und Jugendliche im Alter zwischen 9 und 18 Jahren - 48, die stottern, 48, die nicht stottern - akustisch aufgenommen, während sie einsilbige Wörter mit und ohne Metronom lasen. Die Analysen wurden an vier Minimalpaaren durchgeführt, die sich in der Komplexität des Wortanfangs (einfach vs. komplex) unterschieden.

Die zentrale Hypothese dieses Kapitels ist, dass sowohl Kinder und Jugendliche, die stottern, als auch solche, die nicht stottern, akustische Hinweise eines c-center Effekts zeigen. Es wird jedoch erwartet, dass sich Gruppenunterschiede bei den Kompressionseffekten in der Bedingung ohne Metronom zeigen, während in der Bedingung mit Metronom keine signifikanten Unterschiede zwischen den Gruppen zu erwarten sind.

Wir fanden heraus, dass beide Gruppen akustische Hinweise eines c-center Effekts zeigten. Das Hauptergebnis ist, dass Kinder und Jugendliche, die stottern, eine stärkere Konsonantenkompression aufwiesen (aber es keine Gruppenunterschiede bei der Vokalkompression gab), unabhängig von der Bedingung, was darauf hindeutet, dass sie artikulatorische Gesten anders timen als Kinder und Jugendliche, die nicht stottern. Dies spricht dafür, dass das Sprechen mit einem Metronom zwar den Redefluss bei PWS signifikant verbessert, aber es die zeitliche Organisation der artikulatorischen Gesten bei jungen PWS nicht beeinflusst.

Der Schwerpunkt von Kapitel 3 liegt auf den Auswirkungen rhythmischer Bedingungen auf das artikulatorische Timing sowie auf dem Zusammenspiel zwischen verbalen (Sprache) und nonverbalen (Finger tappen) Gesten bei Erwachsenen, die stottern, und Erwachsenen, die nicht stottern.

Obwohl Stottern mit Störungen von Timing-Mechanismen beim Sprechen einhergeht (Etchell et al., 2014), die sich auch auf den nonverbalen Bereich erstrecken (z.B. Falk et al., 2015; Slis et al., 2023), sind die artikulatorischen Erkenntnisse über diese Timing-Mechanismen noch unerforscht. Die vorliegende Studie schließt diese Forschungslücke, indem sie das artikulatorische Timing von vier PWS, und vier PWNS unter drei verschiedenen rhythmischen Bedingungen untersucht: Sprechen und Finger Tapping (Tapping Bedingung), Sprechen zu einem Metronom (Metronom Bedingung) und Sprechen und Tappen zu einem Metronom (Metronom+Tapping Bedingung). Mit Hilfe von EMA wurden die Gesten des artikulatorischen Sprachbeginns und des Fingertaps aufgezeichnet und analysiert.

Wir waren daran interessiert, i) ob sich die Gruppen im gestischen Timing unterscheiden, ii) ob das Timing von Sprachgesten vom rhythmischen Kontext abhängt und iii) wie ein externer Rhythmus das Timing von verbalen und nonverbalen Gesten beeinflusst.

Im Allgemeinen ging der artikulatorische Sprechbeginn beiden rhythmischen Ereignissen (Fingertap und Metronomschlag) voraus, aber die Fingertaps lagen näher am artikulatorischen Sprechbeginn als der Metronomschlag.

Es wurden keine Gruppenunterschiede beobachtet, wenn die Teilnehmer gleichzeitig tappten und sprachen, was darauf hindeutet, dass das Timing zwischen verbalen und nonverbalen Gesten bei PWS und PWNS ähnlich ist. Allerdings lag der Beginn der artikulatorischen Gesten bei PWS näher am Metronomschlag als bei PWNS, was auf eine spätere Sprachinitiierung bei PWS hindeutet. In der kombinierten Metronom+Tapping Bedingung richteten beide Gruppen ihren Sprechbeginn näher am Metronom aus als in der einzelnen Metronom Bedingung. Allerdings timten nur PWS ihre Fingertaps in der kombinierten Bedingung enger mit dem artikulatorischen Sprechbeginn als in der alleinigen Tapping Bedingung. Somit führte das Vorhandensein eines externen Rhythmus zu einer engeren Kopplung von verbalen und nonverbalen Gesten bei PWS, aber nicht bei PWNS.

Diese Ergebnisse deuten darauf hin, dass PWS Schwierigkeiten bei der Vorhersage externer auditiver Ereignisse haben, was auch bei nonverbalen Aufgaben beobachtet wurde (z. B. Falk et al., 2015). Dies könnte die Integration externer Hinweise in die eigene Sprachproduktion erschweren und somit zu einem Unterschied in der Metronom, jedoch nicht in der Tapping Bedingung führen. Darüber hinaus deuten die Unterschiede in der Metronom+Tapping Bedingung darauf hin, dass sich PWS und PWNS bei der Integration von auditiven (Metronom), manuellen (Finger Tapping) und artikulatorischen (Sprache) Rhythmen unterscheiden.

Kapitel 4 untersucht das Konsonant-Vokal (CV)-Timing und das prädiktive Timing, aufbauend auf der in Kapitel 3 vorgestellten Methodik und den Forschungsfragen. In diesem Kapitel werden die Fragen aus Kapitel 3 in den Kontext des prädiktiven Timings eingeordnet, einem Bereich, der für PWS herausfordernd ist (z. B. Etchell et al., 2014; Falk et al., 2015; Harrington 1988). Zudem wird angenommen, dass das CV-Timing für PWS eine besondere Herausforderung darstellt, was durch verschiedene Theorien (Harrington, 1988; Wingate, 1988) hervorgehoben und insbesondere durch akustische Daten unterstützt wird (Verdurand et al., 2020; Lenoci & Ricci, 2018; Dehqan et al., 2016; Robb & Blomgren, 1997; Klich & May, 1982). Während bekannt ist, dass das Sprechen zu einem Metronom die Sprechflüssigkeit von stotternden Personen signifikant verbessern kann, sind die artikulatorischen Mechanismen hinter diesem Effekt noch unerforscht.

Um diese Lücke zu schließen, untersuchte die in Kapitel 4 vorgestellte Studie artikulatorische Daten von 10 Erwachsenen, die stottern, und 10 Erwachsenen, die nicht stottern (im Alter zwischen 19 und 32 Jahren). Die Proband:innen wurden unter Verwendung von elekromagnetischer Artikulographie (EMA) aufgezeichnet, während sie Zielwörter produzierten, die mit einem bilabialen Anlaut begannen, gefolgt von einem Vokal (/a/, /o/, oder /u/). Diese Zielwörter waren in einen Trägersatz eingebettet. Das Experiment umfasste vier verschiedene Bedingungen: Unpaced (Sprechen), Tapping (Sprechen bei gleichzeitigem Tappen des Fingers), Metronome (Synchronisation des Sprechens mit einem Metronom) und Metronome+Tapping (Sprechen zu einem Metronom bei gleichzeitigem Tappen).

Die wichtigsten Forschungsfragen in Bezug auf das CV-Timing lauten, ob i) die CV-Kopplung zwischen Erwachsenen, die stottern, und Erwachsenen, die nicht stottern, unterschiedlich ist, und ii) ob das intergestische Timing durch unterschiedliche rhythmische Bedingungen beeinflusst wird.

Um das CV-Timing zu untersuchen, haben wir zwei verschiedene Messungen durchgeführt. Eine Messung des Abstandes zwischen dem Beginn der Konsonantengeste und dem Beginn der Vokalgeste (CV-lag) und eine dynamische Messung des Zungenrückenkurvenverlaufs (GAMM-basiert). Beide Maße deuten darauf hin, dass sich die CV-Gesten von PWS mehr überlappen, wobei der Kurvenverlauf-basierte Ansatz das stärkere Modell hervorbrachte. Diesem Maß zufolge unterschieden sich die Gruppen im CV-Timing in der Unpaced Bedingung, jedoch nicht in den rhythmischen Bedingungen.

Wir konnten die Ergebnisse aus Kapitel 3 mit einer größeren Teilnehmerzahl bestätigen, mit der Ausnahme, dass wir feststellten, dass PWS und PWNS ihre Fingertaps in der kombinierten Metronom+Tapping Bedingung in entgegengesetzter Richtung timten. Während PWS ihre Fingertaps nach wie vor näher an den Sprechbeginn platzierten als in der alleinigen Tapping Bedingung, verlagerten PWNS ihre Taps weiter weg vom Sprechbeginn und mehr in Richtung

des Metronomschlags. Darüber hinaus zeigten PWS eine größere Variabilität bei der Synchronisierung ihrer Sprache mit einem Metronom im Vergleich zum Fingertap.
Dies deutet darauf hin, dass PWS möglicherweise Schwierigkeiten beim prädiktivem Timing und bei der auditiv-motorischen Integration haben, was zu Unterschieden bei der zeitlichen Integrierung ihrer Sprache mit einem externen Rhythmus führt, nicht aber mit einem intern erzeugten rhythmischen Ereignis, wie z. B. dem Fingertappen. Darüber hinaus deuten die Ergebnisse der Metronom+Tapping Bedingung darauf hin, dass PWS und PWNS sich auf unterschiedliche Hinweise beziehen, wenn sie ihre Sprache timen - PWNS verwenden das Metronom und PWS den Fingertap.

Zusammenfassend trägt diese kumulative Dissertation zum Verständnis der Mechanismen des artikulatorischen Timings von PWS und PWNS unter verschiedenen rhythmischen Bedingungen bei und liefert wertvolle Beiträge zu verschiedenen Theorien und Modellen des Stotterns. Diese Arbeit belegt, dass Unterschiede im inter-gestischen Timing bei Kindern und Jugendlichen sowie bei Erwachsenen, die stottern vorliegen, was auf Schwierigkeiten bei der Koordination von artikulatorischen Gesten hindeutet. Das Vorhandensein eines externen Rhythmus verbesserte das intergestische Timing bei Erwachsenen, die stottern, wodurch Gruppenunterschiede beseitigt wurden, nicht jedoch bei Kindern. Dies deutet darauf hin, dass die Entwicklung der Sprechmotorik eine Schlüsselrolle dabei spielt, wie rhythmische Kontexte von PWS verarbeitet werden.
Die Dissertation zeigt auch, dass Erwachsene, die stottern, sich von Erwachsenen, die nicht stottern, in dem Synchronisierungszeitpunkt unterscheiden, wenn sie ihre Sprache mit einem Metronom synchronisieren, aber nicht, wenn sie ihre Sprache mit ihrem eigenen Fingertappen synchronisieren. Zusammengefasst deuten diese Ergebnisse darauf hin, dass PWS Probleme mit der auditiv-motorischen Integration und dem prädiktiven Timing haben und dass ihre Mechanismen für das Sprechtiming im Vergleich zu denen von PWNS weniger stabil sind, was auf ein schwächeres sprechmotorisches Kontrollsystem hindeutet.

References

Abercrombie, D. (1967): Elements of General Phonetics. Edinburgh: Edinburgh University Press.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13367–13372. https://doi.org/10.1073/pnas.201400998

Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain research*, *1076*(1), 129–143. https://doi.org/10.1016/j.brainres.2005.11.104

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.

Andrews, G. (1985). Epidemiology of stuttering. In R. F. Curlee & W. H. Perkins (Eds.), *Nature and Treatment of Stuttering: New Directions* (pp. 1-12). San Diego: College Hill Press.

Arndt, J., & Healey, E. C. (2001). Concomitant Disorders in School-Age Children Who Stutter. *Language, speech, and hearing services in schools*, *32*(2), 68–78. https://doi.org/10.1044/0161-1461(2001/006)

Arvaniti A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1-2), 46–63. https://doi.org/10.1159/000208930

Avanzino, L., Pelosin, E., Vicario, C. M., Lagravinese, G., Abbruzzese, G., & Martino, D. (2016). Time Processing and Motor Control in Movement Disorders. *Frontiers in human neuroscience*, *10*, 631. https://doi.org/10.3389/fnhum.2016.00631

Bernstein Rather, N. & Brundage, S. B. (2024). Advances in Understanding Stuttering as a Disorder of Language Encoding. *Annual Review of Linguistics, 10, 127-43. https://doi.org/10.1146/annurev-linguistics-030521-044754*

Bloch, B. (1950). Studies in colloquial Japanese IV: phonemics. *Language*, 26, 86-125.

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.

Bloodstein, O. & Bernstein Ratner, N. (2008). *A handbook on stuttering*, 6th ed. Clifton Park, NY: Thompson/Delmar.

Bohland, J.W., Bullock, D. and Guenther, F.H. (2010). Neural Representations and Mechanisms for the Performance of Simple Speech Sequences. *Journal of Cognitive Neuroscience*, 22 (7), pp. 1504-1529. PMCID:PMC2937837

Büchel, C., & Sommer, M. (2004). What causes stuttering?. *PLoS biology*, *2*(2), E46. https://doi.org/10.1371/journal.pbio.0020046

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

Brady, J. P. (1969). Studies on the metronome effect on stuttering. *Behaviour Research and Therapy*, 7, 197-204.

Browman, C. P. & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3(01), 219-252.

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory. Phonology. *Phonetica*, 45, 140-55.

Browman, C. P. & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18, 299-320.

Browman, C. & Goldstein, L. (1991). Gestural structures: distinctiveness, phonological processes, and historical change. In I. Mattingly and M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception*, pp. 313-338. Erlbaum: New Jersey.

Browman, C. P. & Goldstein, L. M. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4), 155-180

Browman, C. P., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition*, pp. 175–193. The MIT Press.

Browman, C.P., & Goldstein, L.M. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures.

Brundage, S. B., & Ratner, N. B. (2022). Linguistic aspects of stuttering: research updates on the language-fluency interface. *Topics in language disorders*, *42*(1), 5–23. https://doi.org/10.1097/TLD.0000000000000269

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In M. B. Broe & J. B. Pierrehumbert (Eds.). Papers in Laboratory Phonology V: Acquisition and the lexicon (p. 70-87). Cambridge University Press.

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., Chow, H. M., Wieland, E. A., & McAuley, J. D. (2016). Relation between functional connectivity and rhythm discrimination in children who do and do not stutter. *NeuroImage. Clinical*, *12*, 442–450. https://doi.org/10.1016/j.nicl.2016.08.021

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

Chaudhary, C., Maruthy, S., Guddattu, V., & Krishnan, G. (2021). A systematic review on the role of language-related factors in the manifestation of stuttering in bilinguals. *Journal of fluency disorders*, *68*, 105829. https://doi.org/10.1016/j.jfludis.2021.105829

Cho, T. (2005). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. *Laboratory Phonology*, 8, 519–548

Cho, T., Yoon, Y., & Kim, S. (2014). Effects of prosodic boundary and syllable structure on the temporal realization of CV gestures. *Journal of Phonetics, 44,* 96–109. https://doi.org/10.1016/j.wocn.2014.02.007

Chow, I., Belyk, M., Tran, V., & Brown, S. (2015). Syllable synchronization and the P-center in Cantonese. *Journal of Phonetics*, 49, 55-66.

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

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of fluency disorders*, *35*(3), 246–279. https://doi.org/10.1016/j.jfludis.2010.05.002

Craig, A., & Tran, Y. (2005). The epidemiology of stuttering: The need for reliable estimates of prevalence and anxiety levels over the lifespan. *Advances in Speech Language Pathology, 7*(1), 41–46. https://doi.org/10.1080/14417040500055060

Craig-McQuaide, A., Akram, H., Zrinzo, L., & Tripoliti, E. (2014). A review of brain circuitries involved in stuttering. *Frontiers in human neuroscience*, *8*, 884. https://doi.org/10.3389/fnhum.2014.00884

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

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)

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.

De Nil, L. F., & Brutten, G. J. (1991). Voice onset times of stuttering and nonstuttering children: The influence of externally and linguistically imposed time pressure. *Journal of Fluency Disorders, 16*(2-3), 143–158. https://doi.org/10.1016/0094-730X(91)90018-8

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., and Scherer, R. C. (2016). Formant transitions in the fluent speech of Farsi-speaking people who stutter. *Journal of fluency disorders*, 48, 1–15. doi: 10.1016/j.jfludis.2016.01.005

Dokoza, K. P., Hedever, M., & Sarić, J. P. (2011). Duration and variability of speech segments in fluent speech of children with and without stuttering. *Collegium antropologicum*, *35*(2), 281–288.

Donaher, J., & Richels, C. (2012). Traits of attention deficit/hyperactivity disorder in school-age children who stutter. *Journal of fluency disorders*, *37*(4), 242–252. https://doi.org/10.1016/j.jfludis.2012.08.002

Etchell, A. C., Civier, O., Ballard, K. J., & Sowman, P. F. (2018). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. *Journal of fluency disorders*, *55*, 6–45. https://doi.org/10.1016/j.jfludis.2017.03.007

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

Falk, S., Schreier, R., & Russo, F. A. (2020). Singing and Stuttering. In: R. Heydon, D. Fancourt, & A. J. Cohen (Eds.), *The Routledge companion to interdisciplinary studies in singing*, Volume III: Wellbeing (1st Edition). New York, NY: Routledge.

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*, 105975. https://doi.org/10.1016/j.jfludis.2023.105975

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

Georgieva, D., & Stoilova, R. (2018). A clinical training model for students: intensive treatment of stuttering using prolonged speech. *CoDAS*, *30*(5), e20170259. https://doi.org/10.1590/2317-1782/20182017259

Gerlach, H., Subramanian, A., & Wislar, E. (2020). Stuttering and Its Invisibility: Why Does My Classmate Only Stutter Sometimes?. *Frontiers for Young Minds*. 7. 153. 10.3389/frym.2019.00153.

Gibson, T., & Ohde, R. N. (2007). F2 locus equations: Phonetic descriptors of coarticulation in 17- to 22-month-old children. *Journal of speech, language, and hearing research*, 50(1), 97–108.

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of speech, language, and hearing research*, 45(1), 66–79.

Guitar, B. (2014). *Stuttering: An Integrated Approach to its Nature and Treatment* (4. ed.). Philadelphia, PA: Wolters KLuwer/Lippincott Williams & Wilkins.

Giraud, A. L. & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience,* 15, 511–517. https://doi.org/10.1038/nn.3063

Goodell, E. W., & Studdert-Kennedy, M. (1993). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study. *Journal of speech and hearing research*, 36(4), 707–727.

Goldstein, L. (2011). Back to the past tense in English. In R. Gutiérrez-Bravo, L. Mikkelsen, & E. Potsdam (Eds.), *Representing language: Essays in honor of Judith Aissen* (pp. 69–88).

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In C. G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 239–249). The Commercial Press. 978-7-10-006769-0. hal-03127293

Goldstein, L., & Pouplier, M. (2014). The temporal organization of speech. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford handbook of language production*. Oxford University Press.

Grabe, E. & Low, E. (2002). Durational variability in speech and the rhythm class hypothesis. *Laboratory Phonology*, 7, 515-546.

Guenther F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological review*, *102*(3), 594–621. https://doi.org/10.1037/0033-295x.102.3.594

Guenther, F. H. (2003). Neural control of speech movements. In N. Schiller & A. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 209-240). Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110895094.209

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, doi: 10.1111/j.1749-818x.2010.00236.x

Han, M. S. (1962). The feature of duration in Japanese. *Onset no kenkyuu (Phonetics Research)*, 10, 65-80.

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.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews. Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hoole, P., & Pouplier, M. (2015). Interarticulatory coordination: Speech sounds. In M. Redford (Ed.), *The handbook of speech production* (pp. 133–157). John Wiley & Sons.

Howell, P. & Au-Yeung, J. (2002). The EXPLAN theory of fluency control and the diagnosis of stuttering. *Pathology and Therapy of Speech Disorders*. 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

Hudock, D., Dayalu, V. N., Saltuklaroglu, T., Stuart, A., Zhang, J., & Kalinowski, J. (2011). Stuttering inhibition via visual feedback at normal and fast speech rates. *International Journal of Language & Communication Disorders*, 46(2), 169-178.

Hulstijn, W., Summers, J. J., van Lieshout, P. H., & Peters, H. F. (1992). Timing in finger tapping and speech: A comparison between stutterers and fluent speakers. *Human Movement Science, 11*(2-3), 113–124.

Ingham, R. J., Bothe, A. K., Jang, E., Yates, L., Cotton, J., & Seybold, I. (2009). Measurement of speech effort during fluency-inducing conditions in adults who do and do not stutter. *Journal of speech, language, and hearing research*, *52*(5), 1286–1301. https://doi.org/10.1044/1092-4388(2009/08-0181)

Ingham, R. J., Warner, A., Byrd, A., & Cotton, J. (2006). Speech effort measurement and stuttering: investigating the chorus reading effect. *Journal of speech, language, and hearing research*, *49*(3), 660–670. https://doi.org/10.1044/1092-4388(2006/048)

Iskarous, K., Pouplier, M. (2022). Advancements of Phonetics in the 21st Century: A Critical Appraisal of Time and Space in Articulatory Phonology. *Journal of Phonetics*, 95.

Jäncke, L. (1994). Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults. *Journal of Fluency Disorders, 19*(1), 21–37. https://doi.org/10.1016/0094-730X(94)90012-4

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological science*, *13*(4), 313–319. https://doi.org/10.1111/1467-9280.00458

Kalinowski, J, Stuart, A., & Rastatter, M., Snyder, G., & Dayalu, V. (2000). Inducement of fluent speech in persons who stutter via visual choral speech. *Neuroscience letters*, 281. 198-200. 10.1016/S0304-3940(00)00850-8

Kang, C. (2021) Progress, challenges, and future perspectives in genetic researches of stuttering. *Journal of Genetic Medicine*, 18, 75-82. https://doi.org/10.5734/JGM.2021.18.2.75

Katz, J. (2010). Compression effects, perceptual asymmetries, and the grammar of timing. Dissertation (www.researchgate.net/publication/265231502), last accessed: 09/23/22.

Katz, W. F., Kripke, C., & Tallal, P. (1991). Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data. *Journal of speech and hearing research*, 34(6), 1222–1232.

Kim, K. S., Daliri, A., Flanagan, J. R., & Max, L. (2020). Dissociated development of speech and limb sensorimotor learning in stuttering: speech auditory-motor learning is impaired in both children and adults who stutter. *Neuroscience*, 451, 1–21. doi: 10.1016/j.neuroscience.2020.10.014

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. doi: 10.1044/jshr.2503.364

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The Evolution of Rhythm Processing. *Trends in cognitive sciences*, *22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Lazzari, G., van de Vorst, R., van Vugt, F. T., & Lega, C. (2024). Subtle Patterns of Altered Responsiveness to Delayed Auditory Feedback during Finger Tapping in People Who Stutter. *Brain sciences*, *14*(5), 472. https://doi.org/10.3390/brainsci14050472

Ladefoged, Peter. 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5, 253-263.

Lenoci, G. & Ricci, I. (2018). An ultrasound investigation of the speech motor skills of stuttering Italian children. *Clinical Linguistics & Phonetics*, 32(12), 1126–1144.

London, J. (2012). Three Things Linguists Need to Know About Rhythm and Time in Music. *Empirical Musicology Review*, 7(1-2), 5–11. 10.18061/1811/52973.

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*, 106194. https://doi.org/10.1016/j.jcomdis.2022.106194

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., & Gracco, V. L. (2005). Coordination of oral and laryngeal movements in the perceptually fluent speech of adults who stutter. *Journal of speech, language, and hearing research*, *48*(3), 524–542. https://doi.org/10.1044/1092-4388(2005/036)

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders, 31*(1), 105-122.

Meier, A. M., & Guenther, F. H. (2023). Neurocomputational modeling of speech motor development. *Journal of Child Language*, *50*(6), 1318–1335. doi:10.1017/S0305000923000260

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

Mücke, D., Hermes, A., & Tilsen, S. (2020). Incongruencies between phonological theory and phonetic measurement. *Phonology*, *37*(1), 133–170. doi:10.1017/S0952675720000068

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. 10.1515/9783110223958.297.

Namasivayam, A. K., & van Lieshout, P. (2011). Speech motor skill and stuttering. *Journal of motor behavior*, *43*(6), 477–489. https://doi.org/10.1080/00222895.2011.628347

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): e3002492. https://doi.org/10.1371/journal.pbio.3002492

Nittrouer, S., Studdert-Kennedy, M., & Neely, S. T. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of speech and hearing research*, 39(2), 379–389.

Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How Do Children Organize Their Speech in the First Years of Life? Insight From Ultrasound Imaging. *Journal of speech, language, and hearing research*, *61*(6), 1355–1368. https://doi.org/10.1044/2018_JSLHR-S-17-0148

Nolan, F., & Jeon, H. S. (2014). Speech rhythm: a metaphor?. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *369*(1658), 20130396. https://doi.org/10.1098/rstb.2013.0396

O'Brian, S., Onslow, M., Cream, A., & Packman, A. (2003). The Camperdown Program: outcomes of a new prolonged-speech treatment model. *Journal of speech, language, and hearing research*, *46*(4), 933–946. https://doi.org/10.1044/1092-4388(2003/073)

Öhman, S. E. G. (1966). Coarticulation in VCV Utterances: Spectographic Measurements. *The Journal of the Acoustical Society of America*, 39, 151-168.

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, 876–886.

Onslow, M., Costa, L., Andrews, C., Harrison, E., & Packman, A. (1996). Speech outcomes of a prolonged-speech treatment for stuttering. *Journal of speech and hearing research*, *39*(4), 734–749. https://doi.org/10.1044/jshr.3904.734

Oschkinat, M., & Hoole, P. (2020). Compensation to real-time temporal auditory feedback perturbation depends on syllable position. *The Journal of the Acoustical Society of America*, *148*(3), 1478. https://doi.org/10.1121/10.0001765

Park, J., & Logan, K. J. (2015). The role of temporal speech cues in facilitating the fluency of adults who stutter. *Journal of fluency disorders*, *46*, 41–55. https://doi.org/10.1016/j.jfludis.2015.07.001

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

Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019). Current models of speech motor control: A control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America*, *145*(3), 1456. https://doi.org/10.1121/1.5092807

Perkins, W., Rudas, J., Johnson, L., & Bell, J. (1976). Stuttering: discoordination of phonation with articulation and respiration. *Journal of speech and hearing research*, *19*(3), 509–522. https://doi.org/10.1044/jshr.1903.509

Pike, K.L. (1945): *The intonation of American English*. Ann Arbor: University of Michigan Press.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature reviews. Neuroscience*, *21*(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4

Pompino-Marschall, B. (1995). *Einführung in die Phonetik (de Gruyter Studienbuch)*. Berlin/New York: Walter de Gruyter.

Rami, M. & Kalinowski, J., Rastatter, M. Holbert, D., & Allen, M. (2005). Choral Reading with Filtered Speech: Effect on Stuttering. *Perceptual and motor skills*, 100, 421-31. 10.2466/PMS.100.2.421-431.

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

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychonomic bulletin & review*, *20*(3), 403–452. https://doi.org/10.3758/s13423-012-0371-2

Robb, M., and Blomgren, M. (1997). Analysis of F2 transitions in the speech of stutterers and nonstutterers. *Journal of Fluency Disorders*, 22, 1–16. doi: 10.1016/s0094-730x(96)00016-2

Saltuklaroglu, T., Kalinowski, J., Robbins, M., Crawcour, S., & Bowers, A. (2009). Comparisons of stuttering frequency during and after speech initiation in unaltered feedback, altered auditory feedback and choral speech conditions. *International Journal of Language & Communication Disorders*, *44*(6), 1000–1017. https://doi.org/10.1080/13682820802546951

Saltzman, E. L. (1991). The task dynamic model in speech production. In H. F. M. Peters, W. Hulstijn, & C. W. Starkweather (Eds.), *Speech motor control and stuttering* (pp. 47–52). Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering, Nijmegen, the Netherlands, June 13-16, 1990

Saltzman, E. L. & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological psychology*, 1(4), 333-382.

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

Schreier, R. (2023). Stuttering and speech-rhythm. Dissertation, LMU München: Fakultät für Sprach- und Literaturwissenschaften.

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)*. December 14-18th, 2020.

Schwartze, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behavioural brain research*, *216*(2), 685–691. https://doi.org/10.1016/j.bbr.2010.09.015

Schwartze, M., & Kotz, S. A. (2015). The Timing of Regular Sequences: Production, Perception, and Covariation. *Journal of cognitive neuroscience*, *27*(9), 1697–1707. https://doi.org/10.1162/jocn_a_00805

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 one*, *18*(2), e0276691. https://doi.org/10.1371/journal.pone.0276691

Smith, A. (2010). Development of neural control of orofacial movements for speech. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2nd ed., pp. 251–296). Wiley-Blackwell.

Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: evidence from behavioral and kinematic indices of nonword repetition performance. *Journal of fluency disorders*, *37*(4), 344–358. https://doi.org/10.1016/j.jfludis.2012.06.001

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

Smith, A., & Weber, C. (2017). How stuttering develops: the multifactorial dynamic pathways theory. *Journal of Speech, Language, and Hearing Research*, 60(9). 2483–505.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental psychobiology*, *45*(1), 22–33. https://doi.org/10.1002/dev.20009

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

Stuart, A., Frazier, C. L., Kalinowski, J., & Vos, P. W. (2008). The effect of frequency altered feedback on stuttering duration and type. *Journal of Speech, Language, and Hearing Research, 51*(4), 889+.

Šturm, P., & Volín, J. (2016). P-centres in natural disyllabic Czech words in a large-scale speech-metronome synchronization experiment. *Journal of Phonetics, 55,* 38–52. https://doi.org/10.1016/j.wocn.2015.11.003

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

Tillmann, H. G. & Mansell, P. (1980) *Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsproseß*. Stuttgart: Klett-Cotta.

Tilsen, S. (2009). Multitimescale Dynamical Interactions Between Speech Rhythm and Gesture. *Cognitive Science,* 33, 839-879.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and cognitive processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Toyomura, A., Fujii, T., & Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *NeuroImage*, *57*(4), 1507–1516. https://doi.org/10.1016/j.neuroimage.2011.05.039

Toyomura, A., Fujii, T., & Kuriki, S. (2015). Effect of an 8-week practice of externally triggered speech on basal ganglia activity of stuttering and fluent speakers. *NeuroImage*, *109*, 458–468. https://doi.org/10.1016/j.neuroimage.2015.01.024

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

Turk, A. & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, 4(1), 93-118. https://doi.org/10.1515/lp-2013-0005

Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech Timing*. Oxford Studies in Phonology and Phonetics, Vol. 5. Oxford University Press. https://doi.org/10.1093/oso/9780198795421.001.0001

Usler, E., Smith, A., & Weber, C. (2017). A Lag in Speech Motor Coordination During Sentence Production Is Associated With Stuttering Persistence in Young Children. *Journal of speech, language, and hearing research*, *60*(1), 51–61. https://doi.org/10.1044/2016_JSLHR-S-15-0367

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

Vanhoutte, S., Gosyns, M., van Mierlo, P., Batens, K., Corthals, P., De Letter, M., Van Borsel, J., & Santens, P. (2016). When will a stuttering moment occur? The determining role of speech motor preparation. *Neuropsychologica*, 86, 93-102.

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

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 Riper, C. (1971). *The nature of stuttering*. Englewood Cliffs, NJ: Prentice-Hall.

van Riper, C. (1982). *The nature of stuttering* (2 ed.). Englewood Cliffs, NJ: Prentice-Hall.

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

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview [Editorial]. *Speech Communication, 57,* 209–232. https://doi.org/10.1016/j.specom.2013.09.008

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain : a journal of neurology*, *131*(Pt 1), 50–59. https://doi.org/10.1093/brain/awm241

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

White, L. & Malisz, Z. (2020). Speech rhythm and timing. In C. Gussenhoven & A. Chen (Eds.), *Oxford Handbook of Language Prosody* (pp. 167-182). Oxford: Oxford University Press.

WHO (2016). *International Classification of Mental and Behavioral Disorders.* F98.5 Stuttering. Geneva: WHO (World Health Organization)

Wieland, E. A., McAuley, J. D., Dilley, L. C., & Chang, S. E. (2015). Evidence for a rhythm perception deficit in children who stutter. *Brain and language*, *144*, 26–34. https://doi.org/10.1016/j.bandl.2015.03.008

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., & 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, 2438-2452.

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. (1969). Sound and pattern in "artificial" fluency. *Journal of Speech Language and Hearing Research*, 12, 677-686.

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

Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering I: persistency and recovery rates. *Journal of speech, language, and hearing research*, *42*(5), 1097–1112. https://doi.org/10.1044/jslhr.4205.1097

Yairi, E., & Ambrose, N. G. (2005). *Early childhood stuttering: For clinicians by clinicians.* Austin, TX: Pro-Ed.

Yairi, E., & Ambrose, N. G. (2013): Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, 38, 66-87.

Appendices

A: Study protocol for an EEG study

Applicant: Mona Franke, Supervisor: Prof. Simone Falk

Project Title: The neural and physiological correlates of linguistic rhythm

Université de Montréal

Scientific background

Stuttering is defined as a speech fluency disorder characterized by a rhythmic deficit (WHO, 2015) as well as a communication disorder (Etchell et al., 2014). Therefore, it provides a window into the mechanisms underlying the entanglement of communicative intentions with the speech motor control mechanisms. Despite the remarkable progress that has been made in the past decades in research about stuttering, the actual causes still remain unknown.

Recent research supports the approach of a deficient connectivity among brain areas in persons who stutter that support timing and rhythm processing, as well as auditory-motor integration (Jenson, et al., 2020; Chang, et al., 2011; Lu et al., 2010). Hence, stuttering may result from problems with movement preparation and sensory monitoring or sensorimotor integration (Max et al., 2004). One prominent hypothesis is that deficient temporal predictions may be one of the main reasons for stuttering (Etchell et al., 2014). There is also evidence for a non-verbal sensorimotor timing deficit in children and adolescents who stutter (Falk et al., 2015) and adults who stutter (Sares et al., 2019) that support the idea that malfunctioning predictive timing during auditory-motor coupling plays a role in stuttering.

As Max & Daliri (2019) point out, breakdowns in speech fluency in persons who stutter can be attributed to fundamental sensorimotor limitations. The neural control of speech movements depends on dynamic interactions between sensory and motor systems, including the prediction of future sensory and motor systems, which includes the prediction of future sensory states, based on planned and ongoing motor commands (Max & Daliri, 2019).

Studies that examined speech preparation in individuals who stutter concluded that persons who stutter have an atypical feedforward control even when they produce perceptually fluent speech (Sengupta et al., 2019), an enlarged activity during speech motor preparation before fluent words (compensation strategy) (Vanhoutte et al., 2016), and a general auditory prediction deficit caused by inefficiencies in the forward modeling of future auditory inputs (Max & Daliri, 2019).

The recurrence of rhythmic events allows listeners to make predictions about upcoming rhythmic events and the ability of aligning speech motor movement to an external rhythm is also known to lead to a more stable speech motor coordination (Namasivayam & van Lieshout, 2011). Furthermore, an external rhythm, like a metronome, is known to enhance speech fluency in persons who stutter (Andrews et al., 1982; Wingate, 1969). On the other hand, disruptions tend to increase in demanding and stressful situations (Guitar, 2014).

However, what remains still unclear is the effect of the temporal aspects of perceived speech on prediction abilities and speech motor planning that might explain the variability in speech fluency, typical for stuttering.

In order to shed light on this question, we need to take the neural perception of speech into account. Previous research has shown that listeners show specific brain oscillatory patterns related to their interlocutor's speech (Giraud & Poeppel, 2012; Mukherjee et al., 2018). The adjustment of neural oscillations to match the phase of an external stimulus, such as speech, is called (neural) entrainment (Peelle & Davis, 2012). Low-frequency bands between 4 and 8 Hz have been found to elicit robust neural entrainment of speech (e.g. Giraud & Poeppel, 2012) and these bands are also related to syllable production (Chandrasekaran et al., 2009) which is often impaired in speech in persons who stutter but can also be very rhythmical due to speaking techniques. In particular, rhythm tends to increase an alignment of neural oscillations with stimuli (Zoefel et al., 2018; Falk et al., 2017) but the impact of different temporal characteristics of speech (e.g. prosody) on neural entrainment is still inconclusive (see Myers et al., 2019, for a review).

In the present study, we will use a multi-method approach to examine the role of rhythm in the intertwining of speech perception and production.

Aims and research questions

With this study, we want to address the (neural) link between speech perception and speech production in persons who do and persons who do not stutter.

Therefore, we will conduct an EEG experiment examining neural entrainment, as well as speech motor preparation (measured with EEG + EMG) in persons who stutter and persons who speak typically fluent. In addition, this study will also help us to understand the individual adaptation processes that occur in communication situations (e.g. modification of speech rate) and might help explaining the variability in speech fluency, typical for stuttering. Hence, we will measure acoustic adaptation processes and set them in relation with the EEG and EMG results. In sum, we aim

1) to use EEG to understand the neural correlates of different temporal characteristics of speech in persons who stutter and persons who do not stutter,
2) to use EEG and EMG to examine the effect of temporal characteristics of speech on speech motor preparation, and
3) to use a multi-method approach to understand individual differences in acoustic adaptation processes.

The general research question is, whether persons who stutter and persons who do not stutter differ from each other in terms of

- neural entrainment to speech stimuli with different temporal characteristics,
- speech motor planning with respect to the previously perceived stimuli, and
- adaptation abilities to different stimuli.

Methodology and implementation

This study will be conducted in one of the soundproof, electrically shielded Faraday booths at the International Laboratory for Brain, Music, and Sound Research (BRAMS) in Montréal, which also provides the technology for this study. Data collection (piloting) will start in June 2021 and the total duration of the project will be two years.

Pre-study

We will first conduct a pre-study with 20 persons who do not stutter (18-30 years, see exclusion criteria of main study) in order to validate the new EEG paradigm we are using. The neural marker we aim to measure is the contingent negative variation (CNV). This is a slow, negative event-related potential (ERP) known to reflect motor preparation generated by the basal ganglia-thalamo-cortical (BGTC)-loop (for more details, see the section "analyses" below). To date, it is unknown whether the CNV is context-sensitive and varies according to the aims of our main study. Therefore, we want to test the context-sensitivity of the CNV to reveal potential differences in the evoked potential with participants who do not stutter. In two conditions, we will either present an auditory stimulus during the interval between the visual "warning" and the "go" signal (filled condition) or 2.5 seconds of silence (silent condition). We expect to find a CNV for both conditions (silent and filled), with a greater CNV in the filled condition.

Participants

Two groups of participants will be recruited: Participants who stutter and age and gender matched adult control participants. All are healthy adult volunteers. We will recruit up to 30 healthy adult volunteers of ages between 18-30 years who are right-handed and native speakers of French for each group (30 PWS, 30 PTF). Other inclusion criteria are normal hearing and normal or corrected-to-normal vision. Participants will be excluded if they have any speech, language or cognitive disorders (e.g. dyslexia, ADS, neurogenetic stuttering, apraxia, dysarthria). Neurological diseases, psychiatric disorders and medication/drug use affecting cognitive or emotional states are also exclusion criteria. We will only recruit participants who are right-handed to ensure a more homogenous distribution of left-sided language cerebral dominance that is observed in right-handed persons (Kedr, Hamed, Said, & Basahi, 2002). Also, persons with dreadlocks or braided hair are not able to participate in this study, since EEG measures would not be possible. Bearded persons probably have to shave their beards if they want to participate because EMG measures are not possible if facial hair is too long.

Pretests

Questionnaire Data

An in-house questionnaire will be used to gather demographic data (e.g., age, gender), language background and musical abilities. For the participants who stutter, information on the participants' history of stuttering will be collected, too.

Sensorimotor synchronization (SMS)

We will use two tapping tasks from the "Battery for the Assessment of Auditory Synchronization and Timing Abilities" (BAASTA; Dalla Bella et al., 2017). In the first task (unpaced tapping), participants will be asked to tap their finger in their own tempo on a tablet (individual tapping tempo).

In order to assess participants' non-verbal adaptation ability we will use the adaptive tapping task. In this task, participants listen to a sequence of 10 tones, whereby the first six tones have the same inter-onset-interval (600ms) and the remaining four either occur in the same interval, at a slower tempo (IOI of 630 or 670ms), or at a faster tempo (IOI of 570 or 525ms). Participants are asked to synchronize their finger-tapping "to the initial tempo, to adapt to the tempo change, and to continue tapping at the new tempo after the presentation of the last tone" (Dalla Bella et al., 2017:1133) for approximately 10 taps.

Preferred speech tempo in different situations

To identify each participant's individual preferred speech tempo in different situations they will be asked to read a wordlist (normal and fast wordlist reading tempo) and a short text (reading tempo), as well as to give a short interview (spontaneous speech tempo). Participants who stutter will be filmed in order to identify stuttering events and participants who do not stutter will be recorded acoustically only.

Participants will be asked to read a short text for approximately 2-4 minutes (reading passages will be taken from popular sources, such as newspaper articles). Then, they read a wordlist (disyllabic French nouns) for approximately one minute and the same wordlist again as fast as possible. After that, we will record a natural conversation and ask the participant questions, such as 'Where would you like to travel and why?' 'What do you like to do in your spare time?'.

Stuttering assessment (participants who stutter only)

Stuttering severity will be assessed using the SSI-4 protocol (Riley, 1994), which "is the only available standardized measure of stuttering severity that includes the three dimensions frequency, duration and physical concomitants" (Cook, 2013:126). Audiovisual speech recordings will be used in order to quantify the amount of disfluencies in typical speech of the participants. After recording, these speech samples will be scored offline from speech therapists according to the SSI-4 guidelines. In addition, we will ask participants for a self-report on their subjective stuttering experience (Subjective Stuttering Scales, Riley, et al., 2004) and to fill in a questionnaire on the psychosocial impact of stuttering (OASES, Yaruss & Quesal, 2006).

Main experiment

In order to measure speech perception (neural phase synchrony) and production components (neural indices of motor preparation), EEG will be recorded while participants listen to auditory stimuli (lists of words). Participants are prompted to give a verbal response after each stimulus in order to continue the word lists they previously heard. In addition to the EEG, the timing and amplitude of their articulatory response will be recorded via muscle activity (electromyography, EMG). Here, these methods are briefly described alongside their ethical considerations.

EEG (*Electroencephalography*)

While participants are taking part in the experiment, their brain activity will be recorded, using standard EEG equipment (BioSemi, 1020 system with 64 electrodes). We use non-invasive EEG to measure changes in the electrical fields caused by the brain's activity with electrodes placed along the scalp. Non-invasive EEG is a commonly used technique that is safe and well-tolerated by participants (though participants might feel slightly discomforted and though they are asked not to blink their eyes and move as little as possible during the listening part). The procedure of electrode attachment is painless. After the recording session, participants are able to remove the electrolyte gel from their hair with a hair wash (a washbasin is close to the experimental booth). A typical EEG recording session takes about 1.5 hours, 20-30 minutes for the electrode application, followed by the experiment of approximately 1-hour duration. The whole EEG session will be videotaped to be able to inspect the data for exclusion criteria, such as eye blinking during the listening part or inappropriate lip movements within the time of interest for the speech motor preparation analyses.

The ERGO input which is connected to the Neumann microphone and Actiview is used to record speech in synchrony with the EEG signal. Furthermore, the Analog Input Box records the computer sounds (stimuli presentation). The script in Matlab which uses Psychtoolbox also records high-quality audio sound (sampling frequency of 44100Hz) via the Neumann microphone that is placed in front of the participant.

EMG (Electromyography)

To be able to measure physical articulatory activity, we will use surface EMG (Delsys system) on the muscle that encircles the mouth (orbicularis oris muscle). Therefore, surface electrodes will be placed around the lips of the participants; one electrode underneath the left side of the lower lip and one above the right side of the upper lip. Prior to the placement of the electrodes, the participants' skin needs to be prepared to keep the signal-to-noise ratio as minimal as possible. Therefore, the skin is rubbed with gauze soaked in alcohol, which is a painless procedure for the participants. Bearded participants probably have to shave their beards (if it is too long) before skin preparation. EMG electrodes record electrical signals emanating from skeletal muscle contraction.

Procedure

Participants are seated in a sound-proofed booth facing a screen. On each trial, an auditory stimulus will be presented to them consisting of a wordlist of French disyllabic words (city names), read by a native French-speaking man. The words will be presented acoustically over loudspeakers in a wordlist pattern (6-9 words in a row, randomly picked from an overall number of 105 city names). The volume level will be adjusted to the participant's comfort level, prior to the experiment. There will be two different conditions, concerning the temporal structure of the auditory stimulus.

1. Fluent condition: Words are presented in a regular wordlist pattern (interval of 60bpm) with regular pauses of 390ms in between each city name.
2. Disfluent condition: Temporal aspects of speech patterns are altered in an articulatory way, i.e. prolonged stop gap durations, nasals, as well as sibilants, distance between the words varies. (Minimum of 2 disfluent words in a disfluent trial and a maximum of 3 fluent words in a row per trial)

One block will contain 4 different auditory stimuli from the same condition (6 x 6-9 words), viz. participants will listen to stimuli with the same temporal structure (but different words/ different word order) four times in a row. Overall, there will be 40 blocks, 20 blocks per condition. Stimulus duration varies from 5 to 10 seconds, depending on the condition. Each trial will be initiated manually by the experimenter in order to adapt to the individual timing of the participant.

At the end of each auditory trial, there will be two visual prompts presented on the screen in front of the subject indicating that the participant should initiate production.

Following the procedure in Vanhoutte et al. (2015, 2016) these prompts are used to elicit a neural response (CNV), associated with motor preparation of speech production in the EEG signals (more details below). The first visual prompt (a picture which participants have been trained to associate with a word to utter) will occur 2 seconds before the offset of the last word from the auditory stimulus. The picture will stay for 1 second on the screen. Participants will be instructed to say the name of a city associated with the picture out loud when they see a second visual prompt, a big green dot (the "go" signal). The "go" signal will be presented 2 seconds after the onset of the first visual stimulus, so it will occur at the same time as the word offset of the auditory stimulus. Within these two seconds, the disfluent stimuli contain only fluently produced cities.

The first visual stimulus will randomly vary between two pictures participants have been familiarized with before the experiment (e.g., a picture from the Eiffel tower when they shall respond "Paris", and a Pretzel when they shall respond "Munich"). This is done to keep the participants' attention high and to ensure they cannot prepare their speech early.

After the participants listen to each auditory stimulus (6-9 words in a row), they are asked to utter 5 disyllabic city names after the "go" signal as quickly and smoothly as possible. The first word they utter is defined by the picture they see as the first visual prompt. The subsequent 4 words remain the same across the experiment. Participants are instructed to memorize these words before the experiment. They will be given 4 practice trials per condition before the experiment starts in order to make them comfortable with the task. The following combinations will be produced by the participants:

Combination 1:	Paris	Genève	Lyon	Tunis	Québec
Combination 2:	Munich	Genève	Lyon	Tunis	Québec

The participant's response will be recorded acoustically via an external floor-standing microphone. The microphone will be placed approximately 30cm away from the participants' mouth.

Analyses

Pretests

Sensorimotor synchronization (SMS) measures will be calculated with circular statistics. Consistency and accuracy in pacing are most useful for determining individual differences (Sowinski & Dalla Bella, 2013). Results from the SMS tasks and other measures (stuttering severity, age, preferred speech tempo) will be correlated with the CNV, which is also known to reflect sensory anticipation.

Main experiment

Acoustic analyses of speech production

The recorded production data is going to be processed in Praat (Boersma, 2001), where the produced words and the corresponding sounds are segmented and transcribed. We will measure articulation rate, speech rate, and the normalized pairwise-variability-index (Grabe & Low, 2002) to determine individual adaptation ability of the participants with the previous stimulus.

EEG

Time-frequency-analyses

Here, we will analyze cortical phase synchrony of brain oscillations as a measure of "neural entrainment" to the speech signal. We use inter-trial-phase coherence to analyze phase synchrony in low-frequency oscillations (low theta bands, around 5Hz), in the auditory cortex area (e.g. Pefkou et al., 2017; Falk et al., 2017; Giraud & Poeppel, 2012). Furthermore, we want to analyze beta-desynchronization

Speech motor preparation

We will measure the contingent negative variation (CNV) which is a slow, negative event-related potential (ERP) known to reflect motor preparation generated by the basal ganglia-thalamo-cortical (BGTC)-loop. This ERP was found to be a sensitive neural marker for differences in motor preparation in adults who stutter vs typically fluent speakers (Vanhoutte et al., 2015, Vanhoutte et al., 2016). The CNV occurs between a warning stimulus (S1, in our case the picture) which announces that, within a few seconds, a "go"-signal (S2, in our case the black dot) will arrive, asking for a quick motor response (Brunia et al., 2012). Two components/waves of CNV can be distinguished because the interval between the onset of S1 and S2 is 2 seconds. The initial CNV is induced by and related to the warning stimulus. It has its largest amplitude at frontal sites within the first second following S1. The second one (the late CNV) occurs before

S2 and is suggested to represent primarily motor preparation, and, additionally, sensory anticipation for S2 (see Vanhoutte et al., 2015). Therefore, we will focus on analyzing the late CNV, which occurs between 500ms preceding S2 and the onset of speech production (measured using EMG onset, see below). EEG analyses will be done with the MATLAB software toolboxes FieldTrip or EEGlab. For detecting a Baseline (usually a time window of 500ms), we will probably use a Baseline filter instead of using a time window. This is a beneficial approach since our participants will have auditory input during the speech motor preparation time. Note that when using a Baseline of a 500ms window, it can either contain 390ms of silence if there was a pause or a disfluent sound. This would therefore be not a good reference for a Baseline. In addition, we will calculate the slopes in the time window of interest (500ms preceding speech onset) for every electrode to run permutation cluster-based analyses. Another possibility is to build averages per trial and per participant in order to run an ANOVA. The disadvantage of these statistical methods is that we cannot insert, for example, stimulus length as a Fixed Factor. A benefit of recording the speech signal along with the EEG signal is that we can use the envelope curve of the speech signal (ERGO input) to detect the speech onset automatically. Furthermore, the signal from the AIB can be used to detect the onset and the offset of the city name produced by the model speaker in order to track neural entrainment to speech.

EMG

The participants' first response will be a city name with a bilabial onset ([p] in Paris or [m] in Munich), from which the electric activity of the orbicularis oris muscle (a circular muscle that surrounds the mouth) can be measured. Primarily, the EMG measure is going to be recorded to detect the time window of interest for the CNV measure. We will analyze the onset of EMG activity for each lip as an index of articulation initiation. In addition, we will also measure the inter-lip interval (onset value of the upper lip subtracted from the onset value of the lower lip) as a measure for lip-coordination and the intensity of peak amplitudes as a measure of muscle tension.

Analyses will be done with the MATLAB EMG Feature Extraction Toolbox.

Statistical analyses will be performed in RStudio. We will perform parametric statistical tests, such as t-tests (e.g. influence of group on CNV, influence of group on phase locking, influence of group on EMG onset time), between-subject ANOVAS (e.g. influence of group and condition on CNV or phase locking, influence of group on acoustic adaptation ability and non-verbal adaptation ability), and linear mixed models to take the repeated measures into account (e.g. electrode position or EMG activity for the same word onset). Moreover, we will run permutation cluster-based analyses. Further, we will run regression models to evaluate the individual influences, such as stuttering severity, preferred speech tempo, non-verbal adaptive ability on neural and physiological correlates.

Expected results

A higher neural entrainment is expected to occur in the regular condition, compared to the disfluent condition, since rhythm tends to increase the alignment of neural oscillations with speech. If persons who stutter have a general auditory prediction deficit (Max & Daliri, 2019) this might also be mirrored in their neural oscillations. In this case, we would expect to find the absence of phase-locking in the listening phase in participants who stutter in both conditions.

With respect to the speech motor preparation, we hypothesize that participants who stutter and participants who do not stutter have a similar CNV in the fluent condition but differ in the disfluent condition. The regular (fluent) condition exhibits high rhythmicity, and thus, predictability which allows participants to make stable feedforward predictions for their own speech. In the disfluent condition, on the other hand, participants have to make sensory predictions based on a very irregular input. This might lead to one of the following possibilities: Either, the disfluent condition causes a higher demand on the speech motor preparation in both, participants who stutter and participants who do not stutter which will be mirrored in a higher CNV in both groups, or there is no increase in CNV for persons who do not stutter because their speech motor system is stable enough and they can rely on their stable speech motor plans. Another possibility is that, there is no increase in CNV for persons who stutter as their system is more used to disfluent input and thus, their system allows a greater "articulatory error" in the disfluent condition, which would be reflected in a flatter CNV.

As additional information, we will count the number of disfluencies in the participants' responses, respective to the condition (fluent, disfluent) and set it in context with the EEG results. We expect to find a higher amount of stuttering-like disfluencies in the disfluent condition. Furthermore, we suggest that the ability to adapt to the auditory stimuli correlates with the CNV. We hypothesize that an increased CNV makes it difficult to adapt to the temporal characteristics of the stimuli because the participant's focus is more on speech motor preparation. Finally, we hypothesize that persons who stutter will show a lower acoustic adaptation ability, due to speech motor limitations.

This study will help to understand the characteristic variability of speech fluency in stuttering and adaptation processes in general, as well as provide deeper insights into the cause(s) of stuttering. We are going to discuss the findings in light of theories of stuttering and fluent speech production.

References

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.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Brunia, C. H. M., Van Boxtel, G. J. M., & Böcker, K. B. E. (2012). Negative Slow Waves as Indices of Anticipation: The Bereitschaftspotential, the Contingent Negative Variation, and the Stimulus-Preceding Negativity. *The Oxford Handbook of Event-Related Potential Components.* 10.1093/oxfordhb/9780195374148.013.0108.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5, e1000436

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

Cook, S., Donlan, C. & Howell, P. (2013). Stuttering severity, psychological impact and lexical diversity as predictors of outcome for treatment of stuttering. Journal of Fluency Disorders, 38, pp. 124-133.

Cummins, F. & Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics.* 26, 145–171.

Dalla Bella, S., Farrugia, N., Benoit, C. E., Begel, V., Verga, L., Harding, E., & Kotz, S. A. (2017). BAASTA: Battery for the Assessment of Auditory Sensorimotor and Timing Abilities. *Behav Res Methods*, 49, 3, 1128-1145. doi: 10.3758/s13428-016-0773-6. PMID: 27443353.

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. doi: 10.3389/fnhum.2014.00467

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

Falk, S., Lanzilotti, C., & Schön, D. (2017). Tuning Neural Phase Entrainment to Speech. *Journal of Cognitive Neuroscience,*. 29, 8, 1378-1389.

Giraud, A.-L. & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, 15, 4, 511-517.

Grabe, E. & Low, E. (2002). Durational variability in speech and the rhythm class hypothesis. *Laboratory Phonology,* 7, 515-546.

Guitar, B. (2014). *Stuttering: An integrated approach to its nature and treatment* (4. ed.). Baltimore, Philadelphia: Lippincott Williams & Wilkins, Wolters Kluwer.

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. doi: 10.3389/fnhum.2019.00458

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

Khedr, E. M., Hamed, E., Said, A., & Basahi, J. (2002). Handeness and language cerebral lateralization. *European Journal of Applied Physiology*, 87, 469-473. doi: 10.1007/s00421-002-0652-y

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, 3071-3081

Max, L., Guenther, F. H., Gracco, V. L., Gosh, S. S., & Wallace, M. E. (2004). Unstable or Insufficiently Activated Internal Models and Feedback-Biased Motor Control as Sources of Disfluency: A Theoretical Model of Stuttering. *Communication Science and Disorders*, 31, 105-122

Mukherjee, S., Badino, L., Hilt, P. M., Tomassini, A., Inuggi, A., Fadiga, L., Nguyen, N., & D'Ausilio, A. (2018). The neural oscillatory markers of phonetic convergence during verbal interaction. *Human Brain Mapping*, doi.org/10.1002/hbm.24364

Myers, B. R., Lense, M. D., & Gordon, R. L. (2019). Pushing the Envelope: Developments in Neural Entrainment to Speech and the Biological Underpinnings of Prosody Perception. *Brain Sciences*, 9, 70. doi: 10.3390/brainsci9030070

Namasivayam, A. K. & van Lieshout, P. (2011). Speech Motor Skill and Stuttering. *Journal of Motor Behavior*, 43, 6, 477-489.

Peelle, J. E. & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, 3, 320. doi: 10.3389/fpsyg.2012.00320

Pefkou, M., Arnal, L. H., Fontalan, L., & Giraud, A.-L. (2017). θ-Band and β-Band Neural Activity Reflects Independent Syllable Tracking and Comprehension of Time-Compressed Speech. *Journal of Neuroscience*, 37, 33, 7930-7938.

Riley, G. D. (1994). *Stuttering severity instrument for children and adults*, Austin, TX: Pro Ed.

Riley, J., Riley, G. D., & Maguire, G. (2004). Subjective Screening of Stuttering severity, locus of control and avoidance: research edition. *Journal of Fluency Disorders*, 29, 51-62.

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

Sengupta, R., Yaruss, J. S., Loucks, T. M., Gracco, V. L., Pelczarski, K., & Nasir, S. M. (2019). Theta Modulated Neural Phase Coherence Facilitates Speech Fluency in Adults Who Stutter. *Frontiers in Human Neuroscience*, 13, 394. doi: 10.3389/fnhum.2019.00394

Sowinski, J., Dalla Bella, S. (2013). Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychologia*. 51: 1952–1963.

Vanhoutte, S., Santens, P., Cosyns, M., van Mierlo, P., Batens, K., Corthals, P., De Letter, M., & Van Borsel, J. (2015). Increased motor preparation activity during fluent single word production in developmental stuttering: a correlate for stuttering frequency and severity. *Neuropsychologica*, 75, 1-10.

Vanhoutte, S., Gosyns, M., van Mierlo, P., Batens, K., Corthals, P., De Letter, M., Van Borsel, J., & Santens, P. (2016). When will a stuttering moment occur? The determining role of speech motor preparation. *Neuropsychologica*, 86, 93-102

WHO (2015). ICD-10. F98.5 Stuttering. Geneva: WHO.

Wingate, M. E. (1969). Sound and pattern in "artificial" fluency. *Journal of Speech, Language, and Hearing Research*, 12, 677–686.

Yaruss, J. S. & Quesal, R. W. (2006). Overall Assessment of the Speaker's Experience of Stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency Disorders*, 31, 2, 90115.

Zoefel, B., Ten Oever, S., & Sack, A. T. (2018). The Involvement of Endogenous Neural Oscillations in the Processing of Rhythmic Input: More Than a Regular Repetition of Evoked Neural Responses. *Frontiers in Neuroscience*, 12, 95, doi: 10.3389/fnins.2018.00095.

B: List of Publications

Peer-reviewed Journal Articles

Franke, M., Falk, S., Benker, N., & Hoole, P. (2025). The effect of rhythm on inter-gestural coupling of onset and vowel gestures and predictive timing in stuttering. *Journal of Phonetics,* 112, 101432 [PDF, open access], https://doi.org/10.1016/j.wocn.2025.101432

Franke, M., Hoole, P., & Falk, S. (2023). Temporal organization of syllables in paced and unpaced speech in children and adolescents who stutter. *Journal of fluency disorders*, *76*, 105975. https://doi.org/10.1016/j.jfludis.2023.105975

Franke, M., Hoole, P., Schreier, R., & Falk, S. (2021). Reading Fluency and prosodic phrasing in children and adults who stutter. *Brain Sciences*, 11, 1595, https://doi.org/10.3390/brainsci11121595

Aichert, I., Lehner, K., Falk, S., Späth, M., **Franke, M.**, & Ziegler, W. (2021). In Time with the Beat: Entrainment in Patients with Phonological Impairment, Apraxia of speech and Parkinson's disease. *Brain Sciences*, 11, 1524, https://doi.org/10.3390/brainsci11111524

Conference Proceedings

Franke, M., Benker, N., Falk, S., & Hoole, P. (2023). 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 (pp. 3942-3946),* Guarant International

Carlsen, J. M., **Franke, M.**, Huttner, L-H., & Radtke, A. (2018). How to measure a pleasant voice. Eds. Pustka, E., Pöchtrager, M. A., Lenz, A. N., Fanta-Jende, J., Horvath, J., Jansen, L., Kamerhuber, J., Klingler, N., Leykum, H., Rennison, J. In *Proceedings of the Conference "Phonetics and Phonology in the German Language Area (P&P14)"*

The following publications are part of the cumulative dissertation at hand:

- Chapter 2: **Franke, M.**, Hoole, P., & Falk, S. (2023). Temporal organization of syllables in paced and unpaced speech in children and adolescents who stutter. *Journal of fluency disorders*, *76*, 105975. https://doi.org/10.1016/j.jfludis.2023.105975
- Chapter 3: **Franke, M.**, Benker, N., Falk, S. & Hoole, P. (2023). 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 (pp. 3942-3946),* Guarant International.
- Chapter 4: **Franke, M.**, Falk, S., Benker, N., & Hoole, P. (2025). The effect of rhythm on inter-gestural coupling of onset and vowel gestures and predictive timing in stuttering. *Journal of Phonetics,* 112, 101432 [PDF, open access], https://doi.org/10.1016/j.wocn.2025.101432

C: List of Author Contributions

Chapter 2

Franke, M., Hoole, P., & Falk, S. (2023). Temporal organization of syllables in paced and unpaced speech in children and adolescents who stutter. *Journal of fluency disorders*, *76*, 105975. https://doi.org/10.1016/j.jfludis.2023.105975

The author of this dissertation is the first author of this manuscript was the primary contributor to literature review, data analysis, interpretation of the results and writing the manuscript. Philip Hoole and Simone Falk conceived the study and supervised the project. They contributed to the data analysis and advised on the theoretical framing of the study and commented on and helped revise the manuscript.

Chapter 3

Franke, M., Benker, N., Falk, S. & Hoole, P. (2023). 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 (pp. 3942-3946),* Guarant International.

The author of this dissertation is the first author of this manuscript and was primarily involved in study design, literature review, data analyses, interpretation of the results and writing the manuscript. Nicole Benker was the primary contributor to data collection and post-processing. Philip Hoole and Simone Falk supervised the project. Philip Hoole contributed to data analyses. Both supervisors helped revise the manuscript.

Chapter 4

Franke, M., Falk, S., Benker, N., & Hoole, P. (2025). The effect of rhythm on inter-gestural coupling of onset and vowel gestures and predictive timing in stuttering. *Journal of Phonetics,* 112, 101432 [PDF, open access], https://doi.org/10.1016/j.wocn.2025.101432

The author of this dissertation is the first author of this manuscript and was primarily involved in study design, literature review, data analyses, interpretation of the results and writing the manuscript. Nicole Benker was the primary contributor to data collection and post-processing. Philip Hoole and Simone Falk supervised the project. Philip Hoole contributed to data analyses. Both supervisors helped revise the manuscript.