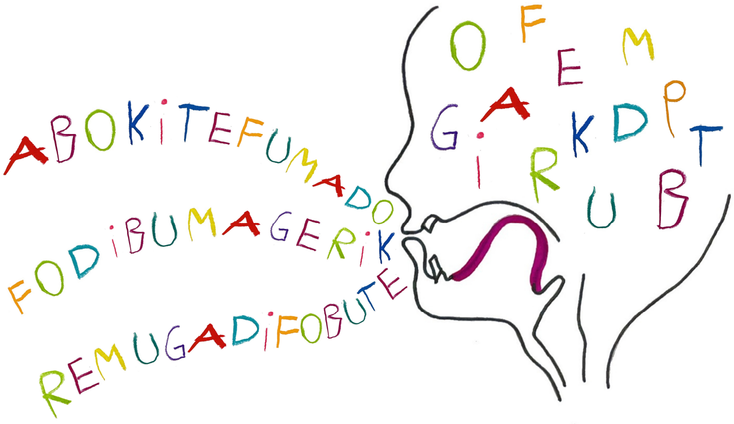
COARTICULATORY CHANGES ACROSS CHILDHOOD

IMPLICATIONS FOR SPEECH MOTOR AND PHONOLOGICAL DEVELOPMENT



DOCTORAL DISSERTATION

submitted to the Faculty of Human Sciences
of the University of Potsdam
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Cognitive Science

by

Elina Rubertus

Defended on March 5, 2024

Supervisors:
Aude Noiray, Ph.D. (HDR), Université Grenoble Alpes
Prof. Dr. Barbara Höhle, University of Potsdam

Reviewers:
Aude Noiray, Ph.D. (HDR), Université Grenoble Alpes
Prof. Dr. Marianne Pouplier, Ludwig-Maximilians-Universität München

https://doi.org/10.25932/publishup-63012
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-630123

DECLARATION OF AUTHORSHIP

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works by any other author, in any form, is properly acknowledged at their point of use.

Potsdam, November 6, 2023

Elina Rubertus

ACKNOWLEDGMENTS

I am very thankful to my supervisors Aude Noiray, Ph.D. (HDR) and Professor Dr. Barbara Höhle, who first piqued my interest in studying language acquisition during my bachelor's degree and supported me on many different levels ever since. Aude, thank you for bringing research on children's articulation to Potsdam, for involving me in all the steps of empirical research from early on, for pushing me into the world of international conferences, for introducing me to so many people who I learnt a lot from, and for your constant faith in me. When I tended to lose my mind in too many details about fine articulatory movements, Barbara, your broader perspective and more general questions regarding the path of language acquisition helped me to relocate my focus. Thank you for providing me with the opportunity to develop and train my teaching skills – the exchange with students surely made me a better researcher as well. Although this project took much longer than originally planned, I felt supported and backed up by both of you the whole time.

At the basis of this work, there are more than 120 participants to whom I am very grateful. Being in a foreign room with strangers who want you to repeat weird words while you have to keep your head still on a slimy machine is certainly a lot to ask of young children (and of adults). Thank you very much for your courage, your endurance, and the joy you brought to the lab.

Dzhuma Abakarova and Stella Krüger, I am so thankful you were on this whole journey together with me. Thank you for being great colleagues and friends at the same time. You provided really helpful feedback whenever I needed it, took time for discussions, and enriched my project with your thoughts. Thank you for adapting to my mensa schedule, for brightening up conference trips, and for your open ears and arms.

Anisia Popescu, thank you for rocking the LesArt study with me – with all those drawbacks, I wouldn't have made it without you. I am very thankful for our frank communication, both in professional as well as in personal contexts. And I think Schlachtensee is calling!

The empiric work conducted for this dissertation wouldn't have been possible without a fantastic lab team. Tom Fritzsche – Thank you for your enormous coordination work in the BabyLAB that made participant recruitment so much easier for us. Jan Ries – Thank you for your help in every single technical question, from writing matlab codes to building a spaceship. The LOLA-Team: Adi, Alma, Cheslie, Clara, Daniela, Gina, Gwen, Jenni, Helene, Lisa, Liuba, Michelle, Noa, Rebecca – Thank you for helping to prepare all the experiments, for patiently calling families to recruit participants, for your engagement in the recordings, and for your endurance in labeling tons of acoustic and articulatory data. But no less for the chatting, the picknicks, for welcoming Caspar, for being much more than just a lab team.

Annika Unger, Antonia Goetz, Anisia Popescu, Evi Fuhrmann, Dzhuma Abakarova, and Stella Krüger – I am grateful for your feedback on different aspects of this dissertation, from first draft to almost-final chapters, from organizational to layout questions. Your support means a lot to me!

I am also thankful to the Potsdam graduate school (PoGs), who supported me both financially via a temporary stipend and various travel grants and personally. Especially the qualification program "Science communication" taught me to look at my research from different angles and improved my visualization, teaching, and presentation skills.

My biggest thanks go to my family. Danke Mama und Papa, dass ihr mich schon immer darin bestärkt habt, meinem Interesse zu folgen, dass ihr mich unterstützt und an mich glaubt. Danke Paula, für ehrliche Nachfragen, für tatkräftige Hilfe, für die Nähe trotz Distanz. Yannick, deine Ermutigung und deine feste Überzeugung, dass ich es schaffe, hat mir durch viele schwierige Phasen hindurchgeholfen – dafür und für so viel mehr danke ich dir. Caspar und Helene, ihr seid die allerbesten Gründe, die Promotionszeit zu verlängern – danke, dass ihr mir geholfen habt, Prioritäten gerade zu rücken. Danke aber vor allem, dass ihr mich täglich daran erinnert, wie faszinierend der Erstspracherwerb ist.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 GENERAL INTRODUCTION

When a young child produces her first words and when at some point in her second year of life the words start to gush out of her, we witness astonishing milestones of the seemingly effortless process of language acquisition. While we were able to figure out her needs more or less easily so far, communication now starts to move towards a new level. With more and more practice and experience, the child will refine her articulation and soon produce segmentally correct words intelligible also by people other than the closest caretakers. However, her language acquisition is not at all completed yet: Besides substantial developments on the lexical, syntactic, and pragmatic levels, her phonology and speech production process undergo important changes as well.

In this dissertation, I investigate one of these changes of speech production after the age of three years and address its implications for speech motor control, phonological development, and characteristics of human's phonological system. The process of interest is coarticulation. Coarticulation refers to articulatory as well as acoustic contextual effects in speech, where characteristics of one phonologically defined segment are reflected in the segments close-by. Language is usually represented by discrete units, traditionally thought of as abstract phonemes, while the speech stream does not provide clear-cut boundaries between those units but is produced in a smooth flow of articulatory movements. These continuous movements lead to contextual effects in both consonants (C) and vowels (V) that have been shown for tongue movements (cf. Recasens, 1999), lip movements (cf. Farnetani, 1999), velar movements (cf. Chafcouloff & Marchal, 1999), as well as laryngeal movements, i.e., voicing coarticulation (cf. Hoole, Gobl, & Ní Chasaide, 1999), and work both from right-to-left (anticipatory coarticulation) as well as from left-to-right (carryover coarticulation). Attempting to maintain the idea of abstract phonological representations despite the challenging findings from newly developed techniques for speech analysis, coarticulatory processes were accused of obscuring the relationship between the assumptive underlying phonemes and the acoustic speech signal considering them destructive (Hockett, 1955) and distorting (Ohala, 1981). However, current theories try not to make a sharp distinction between phonetics and phonology but acknowledge that their interweaving renders a separate understanding of each discipline impossible (Pouplier, 2011). In this sense, coarticulation, a process at the intersection of phonology and phonetics, could instead be evidential for underlying atoms not to have an abstract and timeless character but to be directly related to articulatory properties. This idea of articulatory gestures as atoms of speech is fundamental to Articulatory Phonology (e.g., Browman & Goldstein, 1989, 1992a, 1986; C. A. Fowler, 1980; for a review on phonological atoms: Pouplier, 2011). Gafos and Goldstein (2012) characterize a gesture as a "functional unit of action that achieves a specified task (most often in the case of speech, a vocal tract constriction task)" (p. 222). Importantly, a gesture's constriction task is abstractly defined while there are variable ways to achieve this task via coordinative structures of different cooperating articulators and muscles (i.e., motor equivalence; Kelso & Tuller, 1983; Perrier & Fuchs, 2015). The flexible motion of articulators allows for context-dependent differences in achieving the abstract task. This approach challenged influential views on phonological as well as lexical representations within the generative framework (Chomsky & Halle, 1968; Daniloff & Hammarberg, 1973) while partially reviving earlier approaches that had emphasized the dynamic character of articulation (e.g., Joos, 1948; Öhman, 1966). New value was attached to investigating coarticulatory processes (i.e., context-dependent variation in speech), raising coarticulation to be one of the central issues in contemporary phonetics (Volenec, 2015).

While there is an ever-growing number of partially contradictory coarticulation models, four of which will be summarized in chapter 1.3, we are far from a comprehensive understanding of the various factors contributing to the degree[1] and temporal extent of coarticulatory processes. In this dissertation, I follow the approach that studying the ontogenetic development of speech and language can inform us about the human linguistic system in general. Taking a closer look at ontogenetic changes of coarticulation may, therefore, shed new light on characteristics of the production mechanisms of fluent speech as well as on the nature of phonological representations. Children's speech is not interpreted as a yet imperfect attempt to reach adult targets but as part of a dynamically developing system providing insights into the complexity of human language (e.g., Ferguson, 1986; Redford, 2015, 2019; Vihman & Croft, 2007).

Ongoing changes in the anatomy of children's vocal tract (e.g., Vorperian et al., 2005, 2009), refinements in speech motor control (Green, Nip, & Maassen, 2010; Iuzzini-Seigel, Hogan, Rong, & Green, 2015; Smith, 2006; Smith & Zelaznik, 2004) as well as developing representations for speech and language (A. E. Fowler, 1991; Nittrouer, Studdert-Kennedy, & McGowan, 1989) constitute a complex interplay of possible reasons for developmental changes in speech production fluency, and therefore coarticulation. In addition to expanding our

---

[1] While the synopsis as well as the first two empirical studies use the term *coarticulation degree*, in the third paper, we speak about *coarticulation amount*. Both terms refer interchangeably to the qualitative strength of coarticulatory processes.

knowledge about the human speech production mechanism in general, studying coarticulatory changes in typically developing children, therefore, also enables us to establish norms that may help identify children with atypical speech developments, and get a better understanding of specific clinical pictures (Nijland et al., 2002; Terband, Maassen, Guenther, & Brumberg, 2009; W. Ziegler & Von Cramon, 1985). Especially in children older than three years who produce most words correctly on a segmental level, finer phonetic details like coarticulatory patterns may be relevant for proper diagnoses and clinical descriptions.

Some years later, when entering school around the age of six years, children are formally introduced to written language. The acquisition of literacy constitutes a late milestone in language development that was shown to impact substantially on how language is mentally represented as well as produced (Goswami, 2000; Perre, Pattamadilok, Montant, & Ziegler, 2009). The discrete letters of alphabetic orthographies resemble the discrete phonological units assumed in many theoretical frameworks – reading aloud may, therefore, reveal characteristics of the conversion process from abstract units to fluently coarticulated speech, especially in beginning readers. Within a framework highlighting the articulatory basis of phonology, children's reading fluency could reveal how phonological representations are restructured based on the newly learned many-to-one mappings of variable articulatory movements to phoneme-like graphemes.

As hinted at in the cover picture[2], I explore what children's tongues tell us about the relation between phonological and phonetic information: I capture coarticulatory degree and its systematic patterns in a large cross-sectional set of kinematic data. I look at the development of different aspects of coarticulation across childhood and start to link findings to possible factors responsible for the observations. In particular, I investigate developmental changes in long-distance coarticulation towards the left (anticipatory coarticulation; chapter 2 of this dissertation: Rubertus & Noiray, 2018) as well as towards the right (carryover coarticulation; chapter 3 of this dissertation: Rubertus & Noiray, 2020) field of the utterance in children from three to seven years of age, as well as in adults. My findings provide evidence that articulatory specifications of the segments involved contribute to the degree of coarticulation highlighting that changes in speech motor control may be responsible for some coarticulatory changes. However, they also suggest that the coarticulatory changes across childhood cannot be explained without a representational component. In chapter 4, an empiric comparison of coarticulation in read aloud versus repeated speech shows that early-stage readers' shift of focus from

---

[2] For this figure, my son Caspar helped me out, which not only made the figure nicer but also helped to emphasize the developmental point of view of this thesis. Thank you, Caspar!

the continuous auditory signal to discrete letters may be a driving force for reinterpreting linguistic material and refining their articulation.

The three empirical investigations build the heart of this dissertation in chapters 2 to 4. The remainder of chapter 1 embeds them within a broader context highlighting their corporate value to the field: It first summarizes the most important steps of children's early spoken language acquisition (chapter 1.1) and introduces the long debated question of articulatory unit size (chapter 1.2). Four different approaches on how to model coarticulation within the speech production process are introduced in chapter 1.3. Chapter 1.4 then presents a reasoning for the measure of coarticulation used in the presented studies and finally the aims and objectives in chapter 1.5 lead over to the empirical chapters. In chapter 5, I summarize the three studies and discuss the implications the presented findings have for the development of speech motor control as well as for changes in phonological representations. Importantly, several aspects of the findings in this dissertation challenge speech production models built upon time-less abstract phonological units. Their corporate implications on how to model human speech production and its developmental changes instead, are the central theme of the general discussion.

1.1 Early developments in spoken language acquisition

Soon after birth most healthy infants start to cry (Moyo & Tetsiguia, 2020). While these first vocalizations are not themselves linguistic yet, they were shown to carry intonational information characteristic of the infants' mothers' language (Mampe, Friederici, Christophe, & Wermke, 2009). Oller and Eilers (1992) described the following vocalic development in different stages: The first two months after birth, the so-called *phonation stage*, are characterized by crying and occasional productions of vowel-like sounds. Due to anatomical characteristics of their vocal tract, infants' articulatory potential is drastically limited: Infants' larynx has a high position and their oral cavity is very narrow due to a shallow palate and a proportionately large tongue (e.g., El Mogharbel & Deutsch, 2007). The conversion of the oral cavity to a rectangular two-tubes-system in month one to four allows the transition to the *primitive articulation stage* and enables first articulations like proto-consonantal cooing sounds in the back of the mouth. When the infant is about three to eight months old, she starts to expand and explore her vocal repertoire, modulating pitch, intensity, dynamics, and articulatory positions. The vocalizations in this *expansion stage* usually exceed the length of adults' syllables and do not indicate clear differentiations between consonants and vowels. First syllable-like utterances, the marginal babbling, usually indicate formant transition durations longer than adults'.

More controlled and speech-like productions are expected in the *canonical phase* that starts between five and ten months of age. Infants' babbling now consists of well-formed canonical syllables that contain at least one vowel-like and one consonant-like element and approach adults' typical formant transition timing. At first, infants typically repeat the same syllable multiple times (*reduplicated canonical babbling*), and later, they start to combine different syllables within their babbling (*variegated canonical babbling*), rendering it more and more speech-like. As Lleó, El Mogharbel, and Prinz (1994) showed, canonical babbling reflects language-specific distributions of place of articulation. Because of the high frequency of coronal consonants in German, for example, children often produce sequences like [dada] in German babbling. While canonical babbling may contain sound sequences resembling or even matching real words in the native language, the child is not yet aware of the attached meaning. Only around their first birthday, when they transit from the *pre-lexical* to the *lexical stage*, first intentionally produced words of their native language are expected.

As implemented in the DIVA model (Guenther, 1995; Guenther, Hampson, & Johnson, 1998; Tourville & Guenther, 2011), it is likely that auditory feedback from the infant's own babbling productions plays a crucial role for learning the correspondences between motor actions, somatosensory feedback, and acoustic consequences necessary for intelligible speech production. There is growing empirical evidence for continuity between pre-speech utterances and later word productions. Not only does the quantity of vocalizations at six months predict expressive vocabulary at 12 months of age (Werwach, Mürbe, Schaadt, & Männel, 2021), the specific articulatory patterns of early babbling seem to impact on early word perception and production as well. In reference to Piaget (1952), McCune and Vihman (1987) established the term *vocal motor scheme* to refer to those articulatory patterns the individual child preferably produces: "The more often a baby produces the movements that shape the vocal tract to produce particular sounds and sound sequences, the more automatic those movements become and ultimately the easier it is to execute them in producing words" (Stoel-Gammon, 1998, p. 96). The idea is that frequently produced babbling patterns work like an "articulatory filter" (Vihman, 1993) of the child's input, highlighting those sequences the child has a motoric representation for. Upon detection of a phonetically close word in her input, the child may articulate her vocal motor scheme and via enhancement by supportive external feedback establish word representations. The first words are, therefore, usually those that are very close to the articulatory routines the child knows from babbling (selected words), while in the following stage, adult word forms more distant to babbling patterns are adapted to fit the infant's vocal motor schemes (adapted words). Published examples from observational research and diary studies can be found in Elsen (1996), Vihman (1993, 2016), and Menn and Vihman (2011). In addition, the examples in Table 1.1 are productions I detected in the speech of my daughter Helene. I noticed that she consistently used the syllable /ba/ she had frequently produced in babbling to refer to the phonetically related words Ball ('ball' /bal/) and Bauch ('belly' /baʊx/) when she was around 1;6 (Y;M) years old. While these two may be interpreted as "selected words", she then started to use the same pattern (sometimes extended by a partially retroflex /l/) for the word gefallen ('fallen' /gəfalən/), indicating that she adapts adult words to fit her articulatory routines ("adapted word"). Only with 1;9 years, Helene began to phonetically differentiate the three words Ball, Bauch, and gefallen.

Table 1.1

Examples for Schematic Word Adaptation and its Development Over Time in Helene's Speech

	1;6 years	1;8 years	1;9 years
Ball 'ball' /bal/	[ba]	[ba], [bal]	[bal]
Bauch 'belly' /baʊx/	[ba]	[ba], [bal]	[bax], [baʊx]
gefallen 'fallen' /gəfalən/		[ba], [bal]	[bala], [baln]

Phonological patterns that on the surface look like simplifications can characterize children's speech throughout the preschool years (A. V Fox & Dodd, 1999): Reduction, assimilation, stopping, and epenthesis are frequent processes that may in fact still show the impact of the earliest vocal motor schemes. While these variations in children's early speech often seem to be unsystematic, Studdert-Kennedy and Goldstein (2003) point out, that children's choice of articulatory organs for a given word is remarkably consistent. This finding as well as the evidence for vocal motor schemes, already hint at the importance of articulatory actions when debating phonological representations and their development that is a major aspect of this dissertation. Here, however, the focus lies on segmentally correctly produced words as they still differ from adult speech in fine phonetic details like coarticulatory patterns for many years across childhood.

1.2 Coarticulation as an indicator of articulatory unit size?

One of the goals of research on speech production has been to delimit the domain of coarticulatory influences both in space and time to infer the size of the organizational units the speech motor system takes as input to produce fluent speech (see review in Noiray, Wieling, Abakarova, Rubertus, & Tiede, 2019). Because substantial changes in coarticulatory degree were found across childhood, many developmental studies on coarticulation have focused on the debate about unit size and the implication its development has for the maturing speech production system. An important implication of the assumption of vocal motor schemes (McCune & Vihman, 1987) is that children acquire complexes of combined articulatory gestures as unanalyzed wholes before systematically acquiring single speech sounds (Menn & Vihman, 2011). Those well-practiced, holistic, undifferentiated patterns start to be used in early word production. This view, therefore, predicts a developmental decrease of coarticulatory strength. The child starts out with syllable-, or word-sized articulatory units and only gradually differentiates and tunes specific articulatory gestures – a process that may take years (T. Gibson & Ohde, 2007; Goodell & Studdert-Kennedy, 1993; Studdert-Kennedy, 1987). There is empirical evidence for a decrease of coarticulatory degree from which a decrease in articulatory unit size was inferred, in vowel-to-consonant (CV) coarticulation, that is the anticipation of a vowel during a preceding consonant (Nijland et al., 2002; Nittrouer et al., 1989; Nittrouer & Whalen, 1989), as well as for the degree of long distance vowel-to-vowel (V-to-V) coarticulation across a consonant (Boucher, 2007; Nijland et al., 2002).

However, other studies found evidence for a stable or increasing coarticulation degree across childhood (CV-coarticulation: Katz, Kripke, & Tallal, 1991; Kent, 1983; Zharkova, Hewlett, & Hardcastle, 2012; V-to-V coarticulation: Barbier et al., 2020; Goodell & Studdert-Kennedy, 1993; Hodge, 1989; Nittrouer, 1993; Nittrouer, Studdert-Kennedy, & Neely, 1996; Repp, 1986). These latter findings would instead support the idea that children start out with segmental units and over time learn how to coordinate them to produce fluent speech. This segmental view assumes an incremental development from specific articulatory goals to broader phonological structures requiring precise coordination of multiple articulatory goals in space and time. In Kent's (1983, p.73) words, children would master sequencing of segments before mastering phasing. Support for this hypothesis is also taken from well-established findings from perception studies highlighting the importance of segmental units that are often interpreted to suggest segmental phonological representations during the first year of life: Infants have been shown to be experts in discriminating native as well as non-native speech sounds and, importantly, perceive them categorically from birth on (Eimas, 1975; Kuhl, 1991; Werker & Tees, 1984). In the course of the first year of life, the category boundaries shift towards the phonemic system of their native language leading to a loss of discrimination abilities of non-native speech sounds, often termed perceptual narrowing (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984).

In previous work closely related to this dissertation, however, we have argued that it is not sufficient to characterize developmental changes in speech fluency by either a general decrease or a general increase of articulatory unit size based on coarticulatory degree. Instead, in reference to the influential work of Recasens (e.g., 1984b, 1984a; Recasens, Pallarès, & Fontdevila, 1997), attention must also be given to the specific articulatory demands of the segments under consideration, as maturation of speech motor control may be a considerable factor for the observed developmental changes in coarticulatory processes (Gestural hypothesis; Noiray, Wieling, Abakarova, Rubertus, & Tiede, 2019). In this dissertation, I address changes in coarticulatory degree across childhood while accounting for the specific articulatory requirements of the produced utterances. My empiric investigations support the notion that developmental changes of coarticulation in children's speech are associated with a representational as well as an articulatory/speech motor component. However, whether they imply differences in organizational unit size is critically discussed in chapter 5.2.5. In fact, the relevance of the question about organizational unit size highly depends on the theoretical framework one assumes. The following chapter describes four influential approaches on how to model and, relatedly, where to locate coarticulatory processes within the time course of speech production.

1.3 Modeling coarticulatory processes

To produce fluent speech, a balance between efficient articulatory movements and the achievement of phonetic targets ensuring their intelligibility must be found. Depending on the nature of the underlying primitives a theory of speech production assumes, predictions about how this trade-off is found, differ substantially. Four influential models shall be introduced: The feature-spreading account (Daniloff & Hammarberg, 1973), the window model (Keating, 1988), the directions into velocities of articulators model (DIVA, Guenther, 1995), and the coproduction framework (C. A. Fowler, 1980).

Many models of coarticulation aim to account for the lack of equivalence between the abstract representation and the articulated speech stream, as the latter on the surface contains neither clear-cut nor invariant segments. Standard generative grammar (Chomsky & Halle, 1968), for example, ascribes purely assimilatory processes to phonological rules of featural changes but, importantly, outsources all other aspects of coarticulation to physical processes of speech production. This view, however, supposes a dichotomy between intent and execution that researchers like Daniloff and Hammarberg (1973) aim to overcome. They embed the whole process of coarticulation into the phonological instead of the phonetic/articulatory system arguing for a mechanism of phonological feature spreading between segments that ultimately provides a very detailed input to the speech production mechanism that includes all aspects of allophonic variation. Daniloff & Hammarberg (1973) hence, view coarticulation as a deliberate process that is part of phonological planning. Only for carryover coarticulation, they consider additional mechanic-inertial aspects possible. In line with Henke (1966), coarticulation in the feature-spreading account is analyzed as a look-ahead mechanism that scans upcoming segments for their binary featural specification and spreads a given feature to all preceding segments that are in that respect unspecified. The top row in Figure 1.1 exemplifies this process for the feature [+round] in the word ‘screw’.

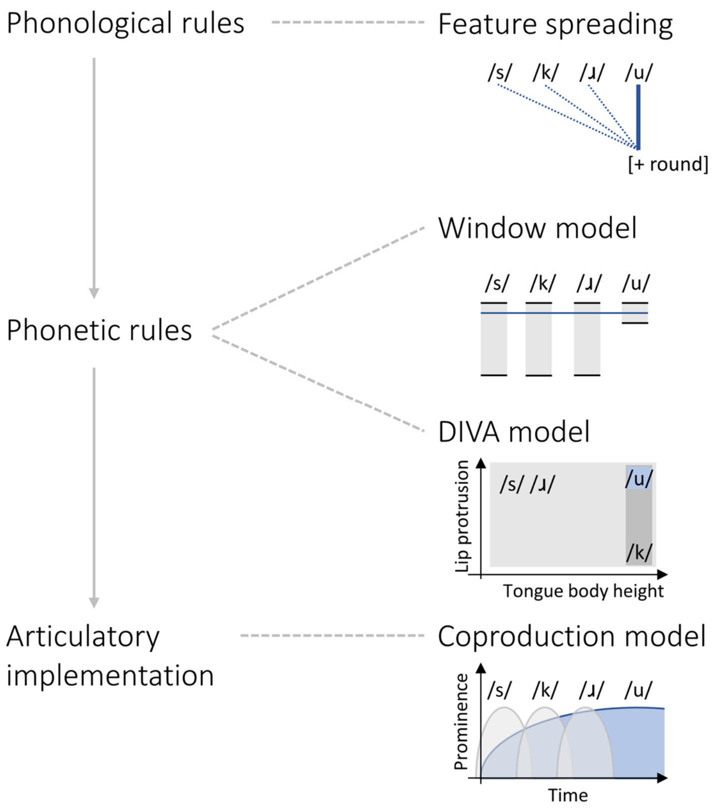
A coarticulatory mechanism built upon assigning binary features to unspecified segments, however, cannot account for the graded nature of coarticulation, shown for example in partially nasalized vowels (Ushijima & Sawashima, 1972) and gradual differences of coarticulatory resistance of liquids (Bladon & Al-Bamerni, 1976). In her window model, Keating (1988) addresses this issue accounting for graded coarticulatory processes with phonetic instead of phonological rules. The phonological representation is specified via (possibly spread) binary features, crucially however, these binary features are not static but are associated with a range of possible values, the window. The width of this window is delimited by the minimum and maximum allowed position of an articulator and, therefore, reflects how contextually variable a given segment is in the physical dimension(s) related to the particular feature. On the phonetic level, a path through these windows that is as smooth as possible, reflecting minimal articulatory effort, is to be constructed. The large windows for segments /s/, /k/, and /ɹ/ in the second row of Figure 1.1 indicate that they can be articulated with a wide range of lip positions, while /u/ has a very narrow window indicating rounded lips. The straight line through all windows (in blue) stands for the least effort, leading to lip rounding in all four segments.

The DIVA model (Guenther, 1995; Guenther et al., 1998; Tourville & Guenther, 2011) elaborated on the idea of partially variable targets by describing and computationally modeling speech production via approximately straight lines through multidimensional convex regions. However, variability in targets is not limited to the dimension of a phonological binary feature. Instead, the convex regions indicate continuous variations in orosensory space. These regions start to be specified during the babbling phase. Similar to the window model (Keating, 1988), the maxim of economy of effort leads to movements from one convex region to the closest acceptable point in the convex region of the following sound, generating coarticulation. As illustrated in the third row of Figure 1.1, for the example word ‘screw’, the segments’ convex regions defined by lip protrusion and tongue body height have very different sizes, indicating how variably they can be produced. Importantly, they all overlap at the location of the vowel’s convex region, which leads to protruded lips and a relatively high tongue body position throughout the whole word.

While the feature spreading as well as the window and the DIVA model assume context-sensitive underlying representations to be changed or specified in the course of the speech production process, Articulatory Phonology assumes invariant underlying atoms, the articulatory gestures (e.g., Browman & Goldstein, 1986). Coarticulation, hence, is no adjustment of ideal canonical segments to their context but happens because of overlapping production of invariant intrinsically-timed gestures (C. A. Fowler, 1980). How these action units are combined from a speech motor perspective, is further described in Task Dynamics (Goldstein, Nam, Saltzman, & Chitoran, 2009; Saltzman, 1986; Saltzman & Byrd, 2000). Here, the apparent dichotomy between static, abstract, phonological representations and the continuous speech signal is not resolved by any kind of translation or specification mechanism but denied by the assumption that continuous speech does in fact consist of discrete units (C. A. Fowler, Shankweiler, & Studdert-Kennedy, 2016). They are just not separable by clear vertical lines perpendicular to the time axis in the acoustic signal because they overlap as their activation is not either on or off, but waxes and veins gradually (Byrd & Saltzman, 1998). The bottom row of Figure 1.1 illustrates how the overlap of gestures' activation or prominence curves can explain the anticipatory lip rounding in the word 'screw'. Articulatory Phonology assumes phonological atoms to be complemented with a timing component, allowing "the ideal or canonical form […] to be executed unaltered in an utterance" (C. A. Fowler, 1980, p.131). In this coproduction framework, coarticulation, therefore, happens only during the implementation in the vocal tract, hence at the last stage of the speech production process. Variations of coarticulatory strength are explained by articulatory specifics of the coproduced segments, that is the strength of gestures and the question of how many articulators they share, i.e., spatial in addition to temporal overlap (C. A. Fowler & Saltzman, 1993). How strongly a vowel or a consonant impedes on others and how likely it is to resist others' articulatory demands depends on the requirements imposed on the active articulator. The Degree of Articulatory Constraint model specifies those demands for the tongue (DAC; Recasens, 2002; Recasens, Pallarès, & Fontdevila, 1997).

Figure 1.1

Models of Coarticulation



Note. Simplified sketches of the main mechanisms accounting for coarticulation in the feature spreading (Daniloff & Hammarberg, 1973), the window (Keating, 1988), the DIVA (Guenther, 1995), and the coproduction model (C. A. Fowler, 1980). In the example word 'screw' /skɹu/, there is anticipation of the vowel's lip rounding. In addition, the left part of the figure indicates at what stage of the speech production process the theory locates coarticulatory processes.

With regard to language acquisition, Browman and Goldstein (1989) point out that the assumption of articulatory gestures as phonological atoms enables the description of infants', children's, and adults' vocal behavior within the same theoretical framework. This is beneficial for the aim to describe speech production development as a continuous process across lifespan. Concerning changes in coarticulation across childhood, Redford (2019) notes that theories assuming discrete, non-overlapping articulatory goals, like the feature-spreading account, the window, and the DIVA model, require computationally intensive mechanisms to account for coarticulation. She argues that this would in turn predict a developmental increase of coarticulation degree, while a theory of overlapping gestural activation may explain a developmental decrease. Changes in coarticulatory degree across childhood are, therefore, not only relevant for the question of articulatory unit size itself but may be indicative of the core functions of the human speech production system.

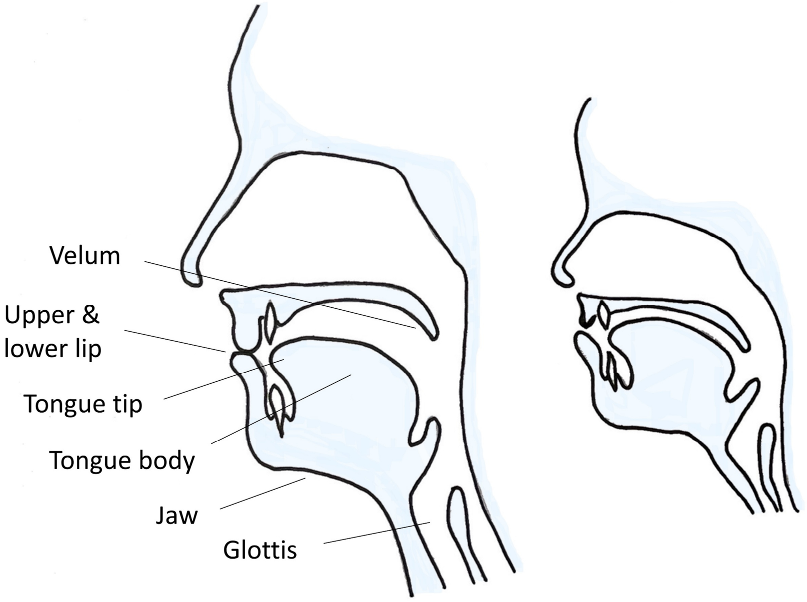
In addition to the diverging predictions for developmental changes in coarticulation degree, predictions regarding analyses of coarticulatory directions differ between the approaches as well. An influential assumption is that anticipatory and carryover coarticulation effects arise from fundamentally different mechanisms: While processes of anticipatory coarticulation are ascribed to planning on the phonological or phonetic level, carryover processes are explained by mechanical constraints and muscle inertia (e.g., Lindblom, 1963). Guenther (1995) emphasizes that in the DIVA model, carryover coarticulation "results solely from the dynamics of moving between targets" (p. 34), neither from pre-planning, nor from purely mechano-inertial factors. However, he still maintains the notion of different underlying processes to be responsible for coarticulatory effects in the two directions. In the coproduction framework, on the other hand, both coarticulatory directions result from the simple overlap of gestural activation. The present dissertation investigates both coarticulatory directions to address possible differences and commonalities in their development across childhood.

1.4 Methodological considerations

To shed light on changes of coarticulatory organization from early childhood to adulthood, we chose to directly investigate movements of the tongue. To produce fluent speech, several articulators need to be precisely coordinated with each other. Mobile articulators that are under active control are the jaw, the lips, the tongue tip, the tongue body, the velum, and the glottis. They are labeled in the vocal tract sketch in Figure 1.2 that illustrates the anatomical differences between adults (on the left) and children (on the right). The tongue is a muscular hydrostat that is "highly mobile and deformable, with a virtually infinite number of mechanical degrees of freedom" (Sanguineti et al., 1997, p.11). It can form constrictions and frictions at different places within the vocal tract and shapes the air chambers to modulate the acoustic signal produced by the vocal cords. Therefore, the tongue is essential for producing all vowels as well as most consonants. This makes the tongue an especially relevant organ for investigations of coarticulation. However, most of the tongue is not visible from the outside but is hidden in the oral cavity. Recording its shapes and movements hence requires more sophisticated techniques than for example the lips.

Figure 1.2

The Vocal Tract



Note. The vocal tract of an adult (left-hand side) with indications of the active articulators. The right-hand side displays the vocal tract of an approximately 4-year-old child.

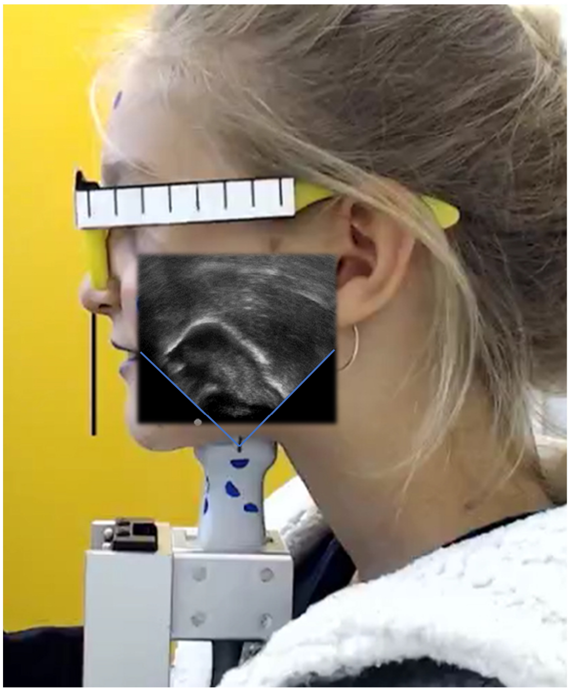
Traditionally, the acoustic speech signal was investigated via spectrographic analysis. Indeed, the first (F1) and the second formant (F2) of the acoustic signal roughly reflect vertical and horizontal tongue movements, respectively (Fant, 1970). Measures of locus equations of F2 (Lindblom, 1963) enabled important findings of CV coarticulation in adult (e.g., Sussman et al., 1991) as well as in child speech (e.g., Gibson & Ohde, 2007). However, inferring the tongue's shapes and positions from acoustics remains an estimation and falls short of the complexity of tongue movements - among others because of equifinality, i.e., the tongue reaching specific goals context-dependently by using a variety of different muscles and possible trajectories (e.g., Kelso & Tuller, 1983; Perrier & Fuchs, 2015). Electropalatography (EPG) provides an opportunity to access previously invisible tongue positions by visualizing the place of tongue-palate contact (Hardcastle, Jones, Knight, Trudgeon, & Calder, 1989). As investigations with EPG are limited to questions of place of contact, though, vocalic tongue movements are out of scope. Electromagnetic articulography (EMA) provides three-dimensional movement patterns of several previously determined points of the tongue surface contour. Although it does not display the whole tongue contour, EMA has the big advantage of being able to follow single parts of tissue precisely through space and time, which neatly enables investigations of the tongue tip and the tongue body separately, for example (e.g., Hoole & Nguyen, 1999; Saito, Tomaschek,

& Baayen, 2020). EMA, however, is not very suitable for investigating the speech of young children as gluing the receptor coils on the tongue can be a time-consuming and unpleasant procedure. For investigating children with language and speech disorders (Goozée et al., 2007; Murdoch & Goozée, 2003; Nip, Arias, Morita, & Richardson, 2017) as well as older typically developing children (Terband, Van Brenk, Van Lieshout, Nijland, & Maassen, 2009), EMA has been successfully applied, though.

In past years, the technique of ultrasound tongue imaging has gained popularity for investigating speech because it is non-invasive, portable, and relatively cheap. When the ultrasound probe is positioned below the chin as visualized in Figure 1.3, pulses of ultrasonic wavelengths are sent out along the midsagittal plane and are reflected when reaching the tissue-air boundary at the tongue surface. Via the temporal difference between sending the pulse and receiving its echo, the ultrasound device calculates the distance to the boundary and visualizes the strength of the reflection by image brightness. This way, the midsagittal tongue surface contour is visualized as a white line on dark background. In Figure 1.3 the resulting ultrasound image is superimposed on a participants' profile to illustrate this mechanism as well as relations between the probe, the tongue, and the face.

Figure 1.3

Ultrasound Tongue Imaging



Note. An ultrasound image of the tongue surface contour superimposed on a participant's profile.

To obtain a reliable image of good quality, a stable relation between the ultrasound probe and the tongue is necessary – otherwise the sound pulses may be sent out along other planes and the resulting image would not display the midsagittal tongue surface contour. For this reason, it is common to use devices fixating the probe to the head via a helmet or use external head stabilization (review in Noiray et al., 2020; Stone, 2005). In the data collection process for this dissertation, however, we abstained from using a helmet or external fixation to allow for maximally natural speech conditions and not constrain oral cavity opening. To make the procedure as pleasant as possible for children, we instead helped participants to maintain a stable head position via visual fixation points, a child-friendly story, reminders, and if necessary, adjusting their head position manually. In addition, our recording platform includes a post hoc movement correction procedure. All the details about our Sonographic and Optical Linguo-Labial Articulation Recording system (SOLLAR) were published in Noiray et al. (2020).

1.5 Aims and objectives of this dissertation

This dissertation dissects the speech production mechanism to investigate developmental changes in coarticulatory patterns. Importantly, I address coarticulation both from a quantitative perspective (how strong are the articulatory context effects?), as well as from a qualitative perspective (what are the specific articulatory requirements of the utterances under investigation?). Developmental changes in coarticulatory degree have often been interpreted to imply changes in the size of organizational units of speech production (i.e., phoneme, syllable, word). My main motivation for studying coarticulatory patterns in children, specifically age-related changes and contrasts to adults' patterns, is twofold: First, considering the articulatory demands of the produced utterances, they can inform us about aspects of speech motor control development across childhood. Second, coarticulation as a process at the intersection between phonology and phonetics can teach us more about the nature of our phonological representations and their changes across development. Are the atoms of speech to be thought of as abstract phonemes? Or are our findings more adequately explained within a theory built upon articulatory gestures? As outlined in chapter 1.3, any model assuming a clear distinction between phonological representations and the continuous speech stream requires a computationally complex translation mechanism which cannot easily explain a developmental decrease of coarticulation degree. Similarly, investigating both anticipatory as well as carryover coarticulation can provide evidence for one or the other model: Clear differences between the coarticulatory directions would speak for different underlying mechanisms (planning vs. mechanics) as assumed in look-ahead models, while similar patterns would provide evidence for coarticulation to be rather a byproduct of articulatory implementation as envisioned in gestural overlap.

Previous studies on developmental changes of coarticulation present us with a puzzling picture of conflicting results calling for a more comprehensive, well-controlled, methodologically sound, and statistically powerful investigation of coarticulation in a wide age range across childhood. My work is part of a research program on speech development within which we developed a platform to record ultrasound imaging data of tongue movements in children from as young as three years of age as well as in adults (Noiray et al., 2020). This allowed us to collect the hitherto largest cross-sectional corpus of kinematic data built of neatly controlled pseudo-utterances by more than 120 speakers in seven age cohorts, enabling the investigation of various research questions on speech production in children and adults. In contrast to many previous studies, I recorded children in tightly sampled age cohorts ranging from three to nine years of age. Moreover, our set of stimuli was thoroughly controlled and the number of participants per age cohort allowed for high statistical power. Using ultrasound tongue imaging, we could track articulatory positions and movements directly instead of inferring them from the acoustic signal.

In all three empirical studies of this dissertation pseudo-utterances of the same format are used, but different coarticulatory processes are addressed. The first study (chapter 2, Rubertus & Noiray, 2018) looks at anticipatory V-to-V coarticulation from a stressed vowel to the preceding schwa. We addressed the change of coarticulatory degree across age as well as impacts of the articulatory demands of the specific segments involved. In the second study (chapter 3, Rubertus & Noiray, 2020), we posed the same questions for coarticulatory processes towards the right instead of the left side, hence carryover V-to-C and V-to-V coarticulation. While both studies on their own allow for inferences about the atoms of speech, the combination of the anticipatory and the rarely considered carryover direction in addition provides new insights into underlying mechanisms responsible for coarticulation as well as their temporal integration within the speech production process.

The third study (chapter 4) addresses one potential reason for the developmental changes observed in the first two studies of this dissertation as well as in other related publications within the research program (Noiray et al., 2018; Noiray, Wieling, Abakarova, Rubertus, & Tiede, 2019). After Popescu & Noiray (2021) found the degree of coarticulation in spoken language to be correlated with participants' level of reading skill, we directly addressed coarticulation in oral reading in German second and third graders as well as in adults. A new dataset allowed us to compare anticipatory coarticulation between read aloud and repeated stimuli in both children and adults. This further advances our knowledge about the developmental change of coarticulation degree and its driving forces and provides insights into mechanisms of literacy acquisition. Figure 1.4 sketches the stimuli used in the three empirical studies. The stressed vowel is indicated in blue while the segments investigated for coarticulatory effects in each study are colored light blue. On the right-hand side, it is indicated that study 1 and 2 dealt with acoustically presented speech only, while study 3 drew comparisons between speech elicited via orthographic (oral reading) and speech elicited via acoustically presented (repetition) stimuli.

Figure 1.4

Schematic Study Overview

Note. Sketch of the stimuli of the three empirical studies of this dissertation. The phonetic material is provided, and blue indicates the stressed vowel while light blue indicates the segments investigated for vocalic coarticulation effects. The soundwaves indicate acoustic and the monitor written stimuli.

2 ON THE DEVELOPMENT OF GESTURAL ORGANIZATION: A CROSS-SECTIONAL STUDY OF VOWEL-TO-VOWEL ANTICIPATORY COARTICULATION[3]

2.1 Abstract

In the first years of life, children differ greatly from adults in the temporal organization of their speech gestures in fluent language production. However, dissent remains as to the maturational direction of such organization. The present study sheds new light on this process by tracking the development of anticipatory vowel-to-vowel coarticulation in a cross-sectional investigation of 62 German children (from three to seven years of age) and 13 adults. It focuses on gestures of the tongue, a complex organ whose spatiotemporal control is indispensable for speech production. The goal of the study was threefold: 1) investigate whether children as well as adults initiate the articulation for a target vowel in advance of its acoustic onset, 2) test if the identity of the intervocalic consonant matters and finally, 3) describe age-related developments of these lingual coarticulatory patterns. To achieve this goal, ultrasound tongue imaging was used to record lingual movements and quantify changes in coarticulation degree as a function of consonantal context and age. Results from linear mixed effects models indicate that like adults, children initiate vowels' lingual gestures well ahead of their acoustic onset. Second, while the identity of the intervocalic consonant affects the degree of vocalic anticipation in adults, it does not in children at any age. Finally, the degree of vowel-to-vowel coarticulation is significantly higher in all cohorts of children than in adults. However, among children, a developmental decrease of vocalic coarticulation is only found for sequences including the alveolar stop /d/ which requires finer spatiotemporal coordination of the tongue's subparts compared to labial and velar stops. Altogether, results suggest greater gestural overlap in child than in adult speech and support the view of a non-uniform and protracted maturation of lingual coarticulation calling for thorough considerations of the articulatory intricacies from which subtle developmental differences may originate.

2.2 Introduction

In spoken language, speech segments overlap with each other. These coarticulatory effects have been detected in the acoustic output of speech as well as in the shapes, positions, and movements of the active articulators of speech, the lips, the tongue, the velum, and the larynx (for a review see Hardcastle & Hewlett, 2006). The present study focuses on lingual coarticulatory

---

[3] Chapter 2 of this dissertation was published as:
Rubertus, E., & Noiray, A. (2018). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. *PloS One*, *13*(9), e0203562. doi: 10.1371/journal.pone.0203562.
Adaptations were made regarding citation style, figure embedding, and cross references.

processes and aims to outline their maturation across childhood. In adults, the positioning and shaping of the tongue is not only determined by the segment currently under production but shows characteristics of neighboring speech segments at the same time. These gestural overlaps exist in heterorganic sequences employing different articulators for achieving the consonantal and vocalic gestures (e.g., /ba/ where the tongue body anticipates a low back position for /a/ during the production of /b/) as well as in homorganic sequences involving the same primary articulator for both gestures (e.g., the point of contact between the tongue body and the palate or velum for /g/ varies with the frontness of the following vowel) (e.g., Recasens, 1985). The domain a vowel can influence this way is not restricted to its adjacent neighbors but can span several segments in an utterance (e.g., Magen, 1997). While a still growing body of literature has described adults' lingual anticipatory processes extensively, similar scrutiny for the maturation of this organizational scheme in childhood has lacked. Most developmental studies have focused on coarticulatory processes within the syllabic domain (intrasyllabic coarticulation; e.g., T. Gibson & Ohde, 2007; Nittrouer et al., 1996; Noiray, Ménard, & Iskarous, 2013; Sussman, Duder, Dalston, & Cacciatore, 1999; Zharkova, 2017). Yet, research on intersyllabic processes is crucial because it tackles a broader organization of speech production processes and therefore addresses questions about the interplay between cognitive (e.g., phonological planning, gestural phasing) and motor domains (the physical implementation).

To begin to fill this gap, the present study tracked the maturation of vowel-to-vowel (V-to-V) coarticulation in four groups of German children (from three to seven years of age) in comparison to adults. Before presenting our data, we first briefly review existing evidence of V-to-V coarticulation in adults as well as suggested implications for planning and motor processes and provide an overview of the existing body of literature in the developmental field that the present study aims to augment. Finally, we relate our findings to previous literature and discuss whether the outcome pattern could be explained by differences between children's and adults' gestural organization.

2.2.1 Vowel-to-vowel coarticulation in adults

Adults begin to produce the vowel for a forthcoming syllable during a preceding syllable. Multiple studies have provided evidence for coarticulatory effects of V_2 in the domain of the transconsonantal vowel V_1 in vowel$_1$-consonant-vowel$_2$ (V_1CV_2) sequences. This lingual anticipation has been either measured in the acoustic signal by comparing formant values (e.g., Beddor, Harnsberger, & Lindemann, 2002; C. A. Fowler, 1981; Modarresi, Sussman, Lindblom, & Burlingame, 2004; Öhman, 1966; Recasens, 1987) or (additionally) in the articulatory signal by directly observing changes in tongue positioning (e.g., x-ray: Browman & Goldstein, 1992; Öhman, 1967; electropalatography: Butcher & Weiher, 1976; Recasens, 1984; electromagnetic articulography: C. A. Fowler & Brancazio, 2000). The magnitude of these V-to-V coarticulatory effects, however, was shown to vary with several factors: Among others, data from Beddor et al. (2002) and Manuel (1990) suggest a language dependency according to which vowels from dense inventories are anticipated to a lower degree than those from relatively sparse inventories. Suprasegmental factors also impact on the degree of vocalic anticipation with stressed vowels being less affected by contextual effects than unstressed vowels (C. A. Fowler, 1981). One of the main characteristics of unstressed vowels being a reduction of articulatory strength approaching schwa (e.g., Lindblom, 1963), it follows logically that schwa is more malleable and therefore affected to a higher degree by coarticulatory processes than full vowels (for a discussion, see Browman & Goldstein, 1992).

The role of the intervocalic consonant. Finally, another influencing factor of particular interest for the current study is the nature of the intervocalic consonant. In measures of vowel anticipation during the preceding consonant itself (i.e., V-to-C-coarticulation), there are consistent effects of a consonant-specific property (Recasens, 1984a, 1985; Recasens & Rodríguez, 2016): "Coarticulatory resistance" (Bladon & Al-Bamerni, 1976) refers to how likely a segment's articulatory gestures are to be coproduced with those of another. As conceptualized in the Degree of Articulatory Constraints model (DAC), the more the tongue dorsum is constrained during the production of a segment, the less likely this segment is to coarticulate with its neighbors (Recasens et al., 1997). Accordingly, labial consonants were shown to display lower coarticulatory resistance and therefore more lingual coarticulation with following vowels than alveolar consonants (Iskarous, Fowler, & Whalen, 2010; Recasens, 1984a; Sussman, Hoemeke, & McCaffrey, 1992). Palatal consonants like /ɲ/ on the other hand, put more constraints on the tongue dorsum and were found to be even more resistant to vocalic influences than alveolar consonants (e.g., Recasens & Rodríguez, 2016). However, velar consonants like /g/ display a rather low coarticulatory resistance despite of employing the tongue dorsum (Recasens, 1985), because the location of tongue body contact with the palate is relatively flexible (Ladefoged & Johnson, 2014). Consequently, /g/'s exact place of articulation usually varies along the front-back dimension according to its vocalic context.

This differing permeability of consonants can be attributed to mechanisms ensuring the achievement of phonetic targets and their intelligibility. However, whether those mechanisms are implemented in the speech production system at a rather early level adjusting the speech plan with regard to contextual variation (e.g., look-ahead models / feature-spreading models) or at a later stage of physical implementation in the vocal tract (e.g., coproduction models) is a matter of dispute (cf. Farnetani & Recasens, 1999 for overview and discussion). While the former theories assume variable gestural plans (e.g., Keating, 1988; Whalen, 1990), the latter build on temporally invariant underlying gestures (C. A. Fowler & Brancazio, 2000; C. A. Fowler & Saltzman, 1993).

Expanding the concept of coarticulatory resistance and context sensitivity to V_1CV_2 sequences, one could hypothesize high resistant consonants to also limit V_2's influence on V_1. Indeed, among others, Recasens (1984b, 1987), and C. A. Fowler & Brancazio (2000) found influences of the intervocalic consonant's resistance on the degree of V-to-V coarticulation. However, in none of these three studies results were entirely consistent. First, within the rather limited sets of participants, there were some speakers whose V-to-V coarticulation was not at all affected by the intervocalic consonant's resistance. And second, instances of V_1CV_2 sequences were found that indicated anticipatory V-to-V coarticulatory effects but at the same time no V-to-C effects (C. A. Fowler & Brancazio, 2000; Recasens, 1984b). Despite high resistant consonants' articulation not being affected by the vocalic gestures themselves, they did thus not always attenuate V_2's influence on the preceding vowels. These occasional findings of discontinuous coarticulatory effects were interpreted as evidence for a speech production model assuming gestural plans of relatively invariant phasing and activation curves to be combined and coproduced in fluent speech (C. A. Fowler & Brancazio, 2000; C. A. Fowler & Saltzman, 1993). According to C. A. Fowler & Saltzman (1993), it is implausible for these discontinuous effects to be part of a speech plan because there is no reason to start, stop, and restart producing a vocalic gesture. Within the coproduction framework the sequencing of consecutive gestures in the planning phase of an utterance is predetermined and quasi blind to contextual variations. Consequently, coarticulatory effects are not part of the speech plan (as contrarily suggested by Whalen, 1990) but occur only during the physical implementation of the gestures in the vocal tract.

Taken together, the literature on V-to-V lingual coarticulation in adults shows that vowels are initiated already during the production of preceding segments. The strength of this vocalic anticipation seems to depend on several factors, one of which is the coarticulatory resistance of the intervocalic consonant. How and in which conditions exactly the consonant's resistance modulates V-to-V coarticulation, however, is not consistently deducible from existing studies yet.

2.2.2 Coarticulatory processes in children

Intrasyllabic coarticulation. Turning to the maturation of coarticulatory processes in children's speech, previous studies have almost exclusively focused on measures of intrasyllabic V-to-C coarticulation. The overarching aim of most of these studies was to infer the unit size of gestural organization and control at different ages. While a low degree of coarticulation between consecutive segments is in that respect interpreted to indicate a segment-driven language organization, a high degree of coarticulation suggests control units larger than the segment. However, diverging results were found: An increasing or stable coarticulation degree across age in some studies (e.g., Katz et al., 1991; Kent, 1983; Zharkova et al., 2012) as well as a decreasing coarticulation degree with age in other studies (e.g., Nijland et al., 2002; Nittrouer et al., 1989, 1996; Noiray et al., 2018). Hence, there is a large discrepancy in the theoretical propositions of researchers ranging from theories suggesting that organizational units grow from the size of a segment to (at least) syllable size with age and language experience, to views suggesting a reduction of unit size with language development such that children initially organize their speech in broad (possibly syllabic) units and develop finer and more differentiated control for single segments only later.

In previous analyses of the present sample of German participants, we noted a decrease of intrasyllabic coarticulation degree from three years of age to adulthood (Noiray et al., 2018). This finding raised the question whether vocalic anticipation in young children extends beyond the syllabic domain. Furthermore, we found consistent effects of the consonant's coarticulatory resistance on the degree of V-to-C-coarticulation with the vowel's tongue position being anticipated most during /b/, to an intermediate degree during /g/, and least during the production of /d/. This result provided a main incentive for the present investigation of consonant-related differences in intersyllabic coarticulation effects.

Intersyllabic coarticulation. In the literature addressing coarticulation across syllable boundaries in child speech, findings are as inconsistent as they are for intrasyllabic coarticulation. The early studies measured second formant frequencies in syllable-final schwas followed by a syllable with a full vowel nucleus (Hodge, 1989; Repp, 1986). Repp (1986) reported V-to-V coarticulation from the full vowel to the preceding schwa in an English-speaking adult as well as in his nine-year-old participant but not in his four-year-old participant. In a more extensive study of 10 participants per age cohort, Hodge (1989) reached similar results with an age-related increase in coarticulatory degree from vowels to preceding schwas in “a stee” and “a stew” utterances: three-year-olds showed a non-significant trend towards V-to-V coarticulation, and five-year-olds anticipated the upcoming vowel to a lesser degree than nine-year-olds who in turn exhibited less coarticulation than adults. While these results suggest that V-to-V coarticulation becomes stronger with age, other studies provided evidence that young children already exhibited a magnitude of V-to-V coarticulation similar to that of adults: three-, five-, and seven-year old children and adults displayed significant effects of the vowel's second formant frequency on that of schwa in English schwa-C-V (əCV) sequences (Nittrouer, 1993; Nittrouer et al., 1996). The magnitude of this V-to-V coarticulation did not vary with age. Interestingly, across Nittrouer's (1993) whole data set the effect of the vowel on the schwa interacted with the factor stop consonant identity. Expanding Recasens' (1984b, 1987) and C. A. Fowler & Brancazio's (2000) findings, her results therefore provide evidence for vowel anticipation during schwa to be stronger in /k/ contexts than in /t/ contexts. In a longitudinal study, Goodell and Studdert-Kennedy (1993) compared acoustic coarticulatory effects of different segments in English CəCV sequences between children at 22 and 32 months of age and adults. While the absolute formant values suggested a decrease of anticipatory V-to-V coarticulation with age, after a normalization procedure accounting for the differences in vocal tract size, group differences disappeared for utterances ending in /i/, and for those ending in /a/ only 22-month-olds remained to show significantly greater V-to-V coarticulation than 32-month-olds and adults.

Contradicting these findings, there is also evidence that children show stronger acoustic effects of vowel anticipatory coarticulation than adults do: In a study comparing typical to atypical speech production development in Dutch, the typically developing five- to seven-year-olds exhibited stronger V-to-V coarticulation than the adult control group (Nijland et al., 2002). Similar to the previously reported studies, they looked at measures of schwa's second formant in əCV utterances. The hypothesis that this pattern could be specific to the Dutch language, is called into question by another study on English-speaking children providing evidence for stronger V-to-V coarticulation in three-, four-, and five-year-olds than in adults as measured in first and second formant frequencies of English əCV sequences (Boucher, 2007).

All developmental studies reported so far have employed acoustic measurements of lingual V-to-V coarticulation. While articulatory data can provide more direct insights into speech production mechanisms, most articulatory data collection techniques are not suitable for young children due to their invasiveness (e.g., articulography, MRI). In the last two decades however, ultrasound imaging has become a popular method for observing and collecting tongue data in the young age (e.g., in kindergarten: Ménard & Noiray, 2011; Noiray et al., 2018, 2013; in toddlers: Song, Demuth, Shattuck-Hufnagel, & Ménard, 2013). Barbier and colleagues (2013) report on one of the few studies investigating the maturation of long-distance coarticulation with articulatory in addition to acoustic measurements. They compared Canadian French four-year-old's articulation of VCV sequences to that of adults. While significant lingual V-to-V coarticulation was observed in adult speakers, only some of the children exhibited vocalic anticipation. The authors therefore concluded that as a group, children were unable to anticipate a vowel's tongue configuration during the production of transconsonantal vowels. It should be noted however, that contrary to the previous studies, they did not investigate the vowel's effect on a preceding schwa but on full vowels (/ɛ/ and /a/).

In summary, while most studies showed that children anticipate a vowel during a preceding schwa at least to some extent, there is conflicting evidence for all three possible scenarios of the V-to-V coarticulation degree's development: A decrease with age, an increase with age, or a similar coarticulation degree throughout development. Several reasons may (in part) explain the discrepancies in results found for both intrasyllabic and intersyllabic coarticulation. First, decisions about the design of the study such as the utterance type and the data collection technique (e.g., method, measurement time point) might be a source of contradiction. In addition, a shortcoming of especially the early studies is the very limited number of participants and its impact on statistical power. Given the fast and multi-faceted developments taking place in the anatomical, cognitive, and speech motor control domains during childhood, the speech of children is known to be highly variable both within and between speakers. It is therefore important to investigate large samples of children and narrow the age range within a cohort to a minimum.

2.2.3 Goal and research questions

The overarching goal of this study is to uncover the development of V-to-V coarticulation in German children. In combination with other studies within our research agenda, we aim to provide insights into the underlying mechanisms of typical speech production to be used for diagnostic and potentially therapeutic purposes among German children with speech impairments. We hope to overcome some of the restrictions of previous research outlined above by: a) Investigating four larger age cohorts across childhood and one cohort of adults. Each age cohort includes at least 13 participants within a narrow age range to minimize age-related variability within the cohorts and therefore increase statistical power. b) Employing a well-controlled set of stimuli varying in place of articulation to investigate differences in coarticulatory degree between phonetic contexts. c) Recording speech material with ultrasound tongue imaging, a non-invasive technique allowing for direct access to tongue positions rather than their estimation via acoustic measures.

To assess whether children differ from adults in how strong vocalic gestures are activated and coproduced with a preceding schwa, measures of the vowel-related change in the horizontal position of the highest point of the tongue body during schwa in schwa-C-V sequences were analyzed according to the following three research questions: First, do we observe anticipatory V-to-V coarticulation in children of every age investigated as well as in adults? If the horizontal tongue body position during schwa varies as a function of tongue position during the following vowel, it will provide evidence for anticipatory V-to-V coarticulation. Although its magnitude varied tremendously in previous studies, evidence for anticipatory V-to-V coarticulation in children was found in most studies. We therefore expect every cohort to anticipate the upcoming vowel during schwa. Second, is the degree of anticipatory V-to-V coarticulation modulated by the resistance of the intervocalic consonant? If so, less V-to-V coarticulation should be found in cases in which consonantal resistance is higher (i.e., alveolar context) than when resistance is lower (i.e., labial context). However, predictions are hard to formulate because this question has been addressed only sparsely in adults providing complicated outcome patterns (C. A. Fowler & Brancazio, 2000; Recasens, 1984b, 1987) and was only investigated on the margins for children so far (Nittrouer, 1993). Based on Nittrouer's (1993) findings, we expect a higher degree of V-to-V coarticulation in sequences with low resistant intervocalic consonants (here /b/ and /g/) than in sequences with high resistant consonants (here /d/). The flexibility of the place of palate contact for /g/ might trigger vowel-related fronting or backing of the tongue during schwa resulting in a high (but 'mediated') V-to-V coarticulation degree. And third, are there developmental changes in terms of coarticulation degree and consonantal effects? This question will be addressed by investigating differences between age cohorts. Again, the conflicting results of previous investigations prevent a clear formulation of predictions. Yet, the considerable decrease of V-to-C-coarticulation degree with age in the previous analysis of this data corpus (Noiray et al., 2018) leads us to predict the same direction for the current investigation of V-to-V coarticulation.

2.3 Method

2.3.1 Participants

In total, 75 participants of five different age cohorts were recorded: 19 three-year-old children (10 females, age range: 3;05 - 3;09 (Y;MM), mean: 3;06), 14 four-year-old children (seven females, age range: 4;04 - 4;08, mean: 4;05), 14 five-year-old children (seven females, age range: 5;04 - 5;07, mean: 5;06), and 15 seven-year-old children at the end of their first or beginning of their second grade in primary school (10 females, age range: 7;00 - 7;06, mean: 7;02). The adult cohort included 13 adults (seven females, age range: 19 - 28 years, mean: 23). All participants were from monolingual German families and none of them reported any language-related, hearing-related, or visual problems. Adult participants as well as the parents of the child participants gave written informed consent for participation in the study, and all were provided with the option to stop participation at any time without negative consequences. The study was approved by the Ethic Committee of the University of Potsdam.

2.3.2 Stimulus material

Trochaic pseudowords of the form consonant$_1$-vowel-consonant$_2$-schwa (C_1VC_2ə) that were recorded by a native German female adult speaker served as model stimuli for a repetition task. The consonants used in both positions were /b/, /d/, and /g/. The three places of articulation were chosen because they vary in coarticulatory resistance. The vowel set consisted of the tense and long vowels /i/, /y/, /u/, /a/, /e/, and /o/ which represent the German vowel space quite adequately. C_1Vs were designed as a fully crossed set of Cs and Vs to which the second syllable was added, C_2 was never the same as C_1. These pseudowords were embedded in a carrier phrase with the German female article /aɪnə/ resulting in utterances such as for example /aɪnə biːdə/. Anticipatory V-to-V coarticulation was measured between the full vowel of the pseudoword and the preceding schwa in the article.

The total number of trials per child varied with group because four- and seven-year-olds' stimulus sets included the additional C_1 /z/ which is not analyzed here. Repeating every word three times, three- and five-year-olds ended up with 108 trials and four- and seven-year-olds with 138 trials. For all cohorts of children, trials were presented in six semi-randomized blocks. Adults' stimulus set included /z/ in both consonant positions adding to a total number of 216 trials, which were presented in nine randomized blocks. Mispronounced trials were noted down by the experimenters and if possible repeated at the end of the block. A table summarizing the number of trials used for the present analyses per consonant context per age cohort is provided in Appendix A. 1.

2.3.3 Experimental procedure

All recordings took place at the Laboratory for Oral Language Acquisition (LOLA) at University of Potsdam (Germany). Participants were asked to repeat a series of pre-recorded auditorily presented stimuli while they were recorded within the SOLLAR-platform (Sonographic and Optical Linguo-Labial Articulation Recording system (Noiray et al., 2015)). This child-friendly setup allows for simultaneous recordings of tongue motion using ultrasound imaging (Sonosite, sr.: 48Hz), labial movement via video recording (camera SONY, sr.: 50Hz) and the audio speech signal (microphone Shure, sr.: 48kHz). For the recording, adult participants sat in a comfortable chair and children in a car-seat adjustable in height. The ultrasound probe was positioned straight below the participant's chin between the maxillary bones to record the tongue surface contour in the midsagittal plane. It is fixed on a custom-made probe holder to be flexible in the vertical dimension following natural speech-related vertical jaw movements but prevents motion in lateral and horizontal translations. Additional head-to-probe stabilization was not employed to maximize the naturalness of speech and make the recording comfortable for young children. Instead, a sparkling golden star conforming to the experimental decoration was placed right above the camera helping the children to keep their head stable and look straight. Trials during which participants moved were discarded subsequent to the recordings via visual inspection of the video data.

Teams of two experimenters conducted the recordings. The first one familiarized the participant with the SOLLAR platform and introduced the children to the story the production task was embedded in. She maintained a face-to-face connection with the participant throughout the recording, controlled for head movement as well as correct pronunciation, and prompted the audio stimuli. The second experimenter operated SOLLAR's recording equipment from a desk not visible to the participant. S/he controlled for the quality of the data collection by thoroughly monitoring both video and audio streams and interrupted if necessary.

The recording room was decorated in a universe theme allowing the experiment to be introduced to children as a spaceship journey during which they had to repeat foreign words from other planets' languages. This stimulated their interest and engagement in the task. Except for the chair, the setup was the same for children and adults, however, the adults were not introduced to the planet story.

2.3.4 Data processing

The acoustic signal was recorded both in relation to the ultrasound device and the video camera, enabling the generation of a common time code for the three streams. A cross-correlation function within MATLAB (2016) was used to synchronize the streams (cf. Noiray et al., 2008; Noiray et al., 2013).

Acoustic data served as a reference to define the relevant time points in the ultrasound signal. Therefore, target utterances and segments were first phonetically labeled using Praat (Boersma & Weenink, 2016). For adults, the detection of target words and segments was done semi-automatically using WebMAUSBasic (Kisler et al., 2012) and manual correction when necessary. For children, native speakers of German identified and manually labeled the target words for subsequent detection and manual labelling of the target segments. A stable periodic cycle in the oscillogram as well as a stable formant pattern, especially a clearly detectable second formant, were used as indices for vocalic segments. The first ascending zero-crossing in the oscillogram at the beginning of the periodicity was accordingly used as schwa and vowel onset, the first ascending zero-crossing after the end of periodicity and disappearance of F2 as the beginning of the following consonant. From the resulting intervals, the relevant time stamps for the current analysis, the temporal midpoint of the schwa and the temporal midpoint of the vowel were automatically extracted.

Repetitions that did not correspond to the model speaker's word were discarded from further analysis, except for those of three-year-olds. Here, the approach was to use as many correctly produced first syllables as possible, so words were kept as long as əC_1V corresponded to the model speaker and C_2 did not differ in place of articulation from the model word (e.g., /aɪnə ba:tə/ was kept for model /aɪnə ba:də/). This way, two instances of words with C_2 = /k/ were kept for /g/, 17 with C_2 = /t/ for /d/, and 10 with C_2 = /v/ for /b/.

Ultrasound frames of interest were selected based on the corresponding time stamps of the acoustic data. For each relevant frame, tongue contours were semi-automatically detected with scripts custom-made for MATLAB (2016) as part of the SOLLAR platform. A spline was automatically fit to reference points that were manually placed on the visible midsagittal tongue surface contour for each frame individually. X- and y-coordinates for each of 100 points of these splines were automatically extracted (see Appendix A. 2 for an illustration). For the present analyses, we used only the x-coordinate, hence the horizontal position, of the highest point of the tongue body surface contour as a representation of frontness of the tongue body.

2.3.5 Data analysis

We used R (2015) and lme4 (Bates et al., 2014) to investigate the three research questions. Our first research question addressed whether children in all age cohorts as well as adults anticipated the lingual position of the vowel during the preceding schwa. Because of previous evidence for the degree of V-to-V coarticulation to be modulated by the intervocalic consonant's resistance, each consonant context was checked separately for each cohort. More specifically, we investigated whether the horizontal position of the highest point of the tongue body during the schwa midpoint ($X_ə$) varied systematically depending on the position of the highest point of the tongue body during the vowel midpoint (X_V).

To address this and the other two research questions, we fitted a linear mixed effects model regressing $X_ə$ on X_V, consonant context (Consonant1), and Cohort with their interactions. The random effect structure was selected following Bates and colleagues' suggestions to use principal component analysis (PCA) for checking the dimensionality of the model and likelihood ratio tests for assessing its goodness of fit (Bates et al., 2015). Starting from the full random effects structure by subject and word, smallest variance components were dropped step by step until convergence was reached and the PCA showed that the number of dimensions was supported by the data. This procedure resulted in a random effect structure including intercepts for subjects and words as well as by-subject random slopes for the effect of the consonant and by-word random slopes for the effect of cohort. The model's assumptions were checked via visual inspection of residual plots and outliers were checked individually to either be removed (experimental errors) or corrected (processing errors). This did not change the outcome pattern.

The second research question focused on possible differences in V-to-V coarticulation degree between the three consonant contexts (/b, d, g/) within each cohort. We applied pairwise comparisons of the interactions between X_V and Consonant1 using generalized linear hypothesis tests with adjusted *p*-values (glht, multcomp package (Hothorn et al., 2008), *p*-value adjustment followed the truncated closed test procedure from Westfall (1997)). All pairwise comparisons for the X_V:Consonant1 interaction were obtained by manually setting the contrast matrix.

Finally, age-related developmental differences in coarticulation degree within the three consonant contexts were addressed using pairwise glht comparisons of the interactions between X_V and Cohort that were again obtained with a manually set contrast matrix using Westfall-adjusted *p*-values. Additionally, the three-way-interactions of X_V, Cohort, and Consonant1 indicated whether the differences of the consonant contexts' effects on coarticulation magnitude (i.e., the coarticulation pattern) vary with age cohort.

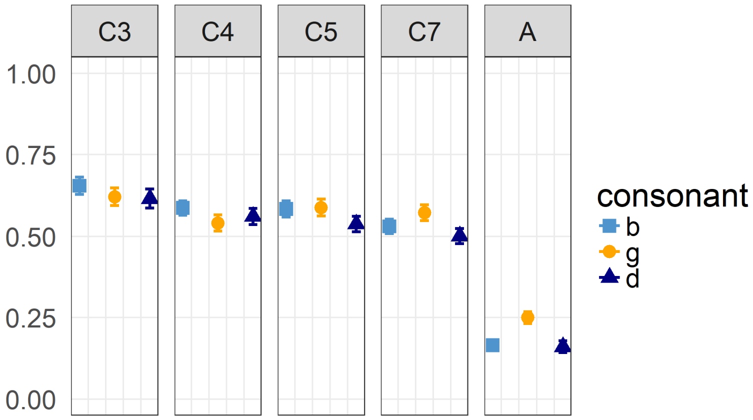
2.4 Results

2.4.1 Vowel-to-vowel coarticulation in every age cohort

The effect of the tongue's horizontal position during the vowel on its position during the preceding schwa is significant for each consonant context in each age cohort ($p < .001$, see Appendix A. 3 for detailed model output). The coarticulation degree, however, differs between the investigated age cohorts and consonant contexts as can be seen in the display of the regression coefficients (Figure 2.1). Statistical relevance of these differences will be addressed in the following sections.

Figure 2.1

Regression Coefficients for the Three Consonant Contexts /b, d, g/ per Cohort



Note. Cohort abbreviations are C3 – three-year-old children, C4 – four-year-old children, C5 – five-year-old children, C7 – seven-year-old children, and A – adults Error bars represent one standard error of the coefficients.

2.4.2 Consonantal impact only in adults

The results of the pairwise comparisons between the consonant contexts within each cohort are summarized in Table 2.1. The intervocalic consonant only has an effect on the V-to-V coarticulation degree in adults with əgV sequences allowing for more V-to-V coarticulation than both əbV and ədV sequences. In none of the cohorts of children does the nature of the intervocalic consonant significantly impact the degree of V-to-V coarticulation. There is only a trend (p = .0834) for /g/-contexts to allow for more coarticulation than /d/-contexts in seven-year-old children.

Table 2.1

Results of the Linear Hypotheses Tests for Consonantal Differences Within Cohort

Cohort	Hypothesis	Estimate	SE	z	*p*-value		*direction*
C3	b-d	0.039210	0.037024	1.059	0.539		
	b-g	0.033378	0.035412	0.943	0.613		
	d-g	-0.005832	0.038511	-0.151	0.987		
C4	b-d	0.02630	0.03214	0.818	0.691		
	b-g	0.04617	0.03162	1.460	0.310		
	d-g	0.01988	0.03453	0.576	0.833		
C5	b-d	0.046438	0.033473	1.387	0.347		
	b-g	-0.004137	0.034414	-0.120	0.992		
	d-g	-0.050574	0.034717	-1.457	0.312		
C7	b-d	0.02949	0.03115	0.947	0.6103		
	b-g	-0.04169	0.03166	-1.317	0.3855		
	d-g	-0.07118	0.03340	-2.131	0.0834	.	
A	b-d	0.00461	0.02317	0.199	0.97836		
	b-g	-0.08467	0.02345	-3.611	0.00087	***	*b < g*
	d-g	-0.08928	0.02565	-3.481	0.00135	**	*d < g*

Note. Results were obtained via glht comparisons with Westfall p-value adjustment. Cohort abbreviations are C3 – three-year-old children, C4 – four-year-old children, C5 – five-year-old children, C7 – seven-year-old children, and A – adults. The last column indicates the direction of significant effects. Significance codes '***': p < .001; '**': p < .01; '*': p < .05; '.': p < .1.

2.4.3 Developmental decrease of V-to-V coarticulation magnitude

To assess differences in coarticulation degree between the cohorts, first, pairwise glht comparisons of the X_V:Cohort interactions were run (see Table 2.2). For every consonant context, the degree of V-to-V coarticulation is significantly lower in the adult cohort than in each of the cohorts of children. In the /b/ and /d/-contexts, there are additional statistically significant differences between the three-year-olds and each of the seven-year-olds: The youngest participants show a higher degree of coarticulation from the vowel to the preceding schwa than the oldest cohort of children for əbV and ədV sequences.

Table 2.2

Results of the Linear Hypotheses Tests for Cohort Differences Within Consonant Contexts

Consonant	Hypothesis	Estimate	SE	z	*p*-value		*direction*
b	A – C3	-0.489537	0.030386	-16.111	<0.001	***	*A < C3*
	A – C4	-0.422136	0.026773	-15.767	<0.001	***	*A < C4*
	A – C5	-0.419119	0.026814	-15.631	<0.001	***	*A < C5*
	A – C7	-0.365634	0.024724	-14.789	<0.001	***	*A < C7*
	C7 – C3	-0.123903	0.031742	-3.903	<0.001	***	*C7 < C3*
	C7 – C4	-0.056502	0.026632	-2.122	0.208		
	C7 – C5	-0.053484	0.027010	-1.980	0.273		
	C5 – C3	-0.070419	0.032696	-2.154	0.196		
	C5 – C4	-0.003018	0.029451	-0.102	1.000		
	C4 – C3	-0.067401	0.030860	-2.184	0.184		
d	A – C3	-0.45494	0.03406	-13.357	<0.001	***	*A < C3*
	A – C4	-0.40045	0.03081	-12.998	<0.001	***	*A < C4*
	A – C5	-0.37729	0.02816	-13.396	<0.001	***	*A < C5*
	A – C7	-0.34075	0.02733	-12.469	<0.001	***	*A < C7*
	C7 – C3	-0.11418	0.03531	-3.234	0.0105	*	*C7 < C3*
	C7 – C4	-0.05970	0.03081	-1.937	0.2946		
	C7 – C5	-0.03654	0.02898	-1.261	0.7127		
	C5 – C3	-0.07765	0.03528	-2.201	0.1773		
	C5 – C4	-0.02316	0.03237	-0.715	0.9524		
	C4 – C3	-0.05449	0.03568	-1.527	0.5416		
g	A – C3	-0.37149	0.03250	-11.429	<0.001	***	*A < C3*
	A – C4	-0.29130	0.03063	-9.511	<0.001	***	*A < C4*
	A – C5	-0.33859	0.02939	-11.519	<0.001	***	*A < C5*
	A – C7	-0.32265	0.02823	-11.431	<0.001	***	*A < C7*
	C7 – C3	-0.04884	0.03395	-1.439	0.601		
	C7 – C4	0.03136	0.03047	1.029	0.841		
	C7 – C5	-0.01593	0.02972	-0.536	0.983		
	C5 – C3	-0.03290	0.03427	-0.960	0.872		
	C5 – C4	0.04729	0.03245	1.457	0.589		
	C4 – C3	-0.08020	0.03337	-2.403	0.113		

Note. Results were obtained via glht comparisons with Westfall p-value adjustment. Cohort abbreviations are C C3 – three-year-old children, C4 – four-year-old children, C5 – five-year-old children, C7 – seven-year-old children, and A – adults. The last column indicates the direction of significant effects. Significance codes '***': p < .001; '**': p < .01; '*': p < .05; '.': p < .1.

In a second step, coarticulatory patterns were compared between cohorts by running three-way-interactions of the effects of X_V, Cohort, and Consonant1. Table 2.3 provides the model output for those interactions that reached significance or indicated a trend. The difference in V-to-V coarticulation degree between /b/- and /g/-contexts is different between adults and each of the three younger cohorts of children (three-, four-, and five-year-olds). It is also different between seven-year-olds and the two youngest age cohorts (only marginally significant between three-year-olds and seven-year-olds). Figure 2.1 visualizes these differences: While for adults /g/-contexts allow for more coarticulation than /b/-contexts, the direction of the (non-significant) difference is the other way around for young children. Regarding the difference between /d/- and /g/-contexts, adults' pattern only differs significantly from four-year-olds' with a trend in comparison to three-year-olds. In addition, four-year-olds differ from seven-year-olds.

Table 2.3

Summary of the Three-Way-Interactions of the Effects of X_V, Cohort, and Consonant1

Consonants	Cohorts	Estimate	SE	t-value	*p*-value	
b / g	A / C3	0.118045	0.042239	2.795	0.005513	**
	A / C4	0.13084	0.03920	3.338	0.001026	**
	A / C5	0.08053	0.03874	2.079	0.037877	*
	C7 / C3	0.075065	0.044536	1.686	0.093453	.
	C7 / C4	0.087860	0.039424	2.229	0.026535	*
g / d	A / C3	0.08344	0.04610	1.810	0.071136	.
	A / C4	0.10915	0.04293	2.542	0.011679	*
	C7 / C4	0.09105	0.04298	2.118	0.0348	*

Note. This table summarizes only the significant and marginally significant three-way-interactions of the effects of XV, Cohort, and Consonant1. Cohort abbreviations are C C3 – three-year-old children, C4 – four-year-old children, C5 – five-year-old children, C7 – seven-year-old children, and A – adults. The last column indicates the direction of significant effects. Significance codes '***': p < .001; '**': p <.01; '*': p <.05; '.': p <.1.

2.5 Discussion

Long-distance coarticulatory processes have been shown to provide valuable information about general speech production mechanisms. However, after Öhman's (1966) work on lingual vowel-to-vowel coarticulation's implications for principles of the speech production process, extensive investigations of the topic have been scarce. Similarly, while a substantial number of studies have compared children's intrasyllabic coarticulation to adults' (e.g., in the acoustic domain: Katz et al., 1991; Nittrouer et al., 1996; in the articulatory domain: Noiray et al., 2018, 2013; Zharkova et al., 2012), coarticulation beyond the syllabic frame has been the topic of only a handful of developmental studies so far. Yet, longer distance coarticulatory processes can help to elucidate what aspects of (co)articulation may be planned while others may rather reflect byproducts of the gestures' implementation in the vocal tract. From a developmental standpoint, this is a highly relevant question because it can shed light on the maturation of spoken language fluency and tease apart the factors that may impact this process.

The current study aimed to contribute to this endeavor by thoroughly investigating lingual vowel-to-vowel coarticulation in a larger participant pool of adults than previously examined as well as in four different age groups across childhood. In addition to testing for the presence of V-to-V coarticulation in each age cohort, we examined the potential impact of intervocalic consonants on the degree of V-to-V coarticulation. Most importantly, we compared coarticulatory patterns (both in terms of degree and consonantal impact) between age cohorts to unveil the maturation of these aspects of the speech production process. The discussion section is framed along these three main questions.

2.5.1 Vocalic gesture's anticipation

Results from this study provide strong evidence that adults anticipate a full vowel's horizontal tongue position during a preceding schwa in əCV sequences. This finding extends previous research (Beddor et al., 2002; Browman & Goldstein, 1992b; Butcher & Weiher, 1976; C. A. Fowler, 1981; C. A. Fowler & Brancazio, 2000; Manuel, 1990; Modarresi et al., 2004; Öhman, 1966, 1967; Recasens, 1987, 1984b) with a larger sample of adult participants and provides insights into V-to-V coarticulatory patterns in German, a language whose coarticulation patterns have not been extensively investigated (e.g., Butcher & Weiher, 1976; Recasens, Fontdevila, & Pallarès, 1995).

A second main finding is that all four cohorts of children exhibited strong vowel anticipation across syllable boundaries as well. This result is in line with the majority of studies addressing children's V-to-V coarticulation (Boucher, 2007; Goodell & Studdert-Kennedy, 1993; Nijland et al., 2002; Nittrouer, 1993; Nittrouer et al., 1996) and augments previous evidence with data from German. However, this result conflicts with those of three existing studies. In particular, Repp (1986) and Hodge (1989) did not find any significant vocalic effect on the preceding schwa in four- and three-year-olds respectively, but only in their older participants (Repp (1986): nine years, Hodge (1989): five & nine years). On the contrary, our data show that at 3.5 years of age, German children do anticipate the tongue body position for target vowels well ahead of their acoustic onsets. Note that Repp's (1986) results are based on a single speaker per age group only, which prevents strong conclusions. Hodge's (1989) sample size of 10 children per age cohort, however, yields greater statistical power and generalizability. Yet, in contrast to other studies including ours, she used utterances containing /st/ clusters (“a stew” versus “a stee”) instead of singleton intervocalic consonants. It is well known that stable productions of consonant clusters are achieved relatively late in childhood (for a review, see McLeod, Van Doorn, & Reed, 2001). For example, Smit and colleagues (1990) reported that English-speaking children do not reach 75% production accuracy for /st/ clusters before the age of 4;06. Given this protracted maturation, studies addressing coarticulatory degree in sequences containing clusters and those testing singleton consonants are not directly comparable.

In a more recent study using ultrasound tongue imaging, Barbier et al. (2013) reported neither acoustic nor articulatory evidence for V-to-V coarticulation in four-year-old children. This strong contradiction with our finding may stem from substantial methodological differences between the two studies (e.g., V_1 being a full vowel versus a schwa, using the whole tongue contour versus a point measure). Furthermore, the authors found a significant effect of vowel anticipation in the acoustic (effect of V_2 on V_1's second formant) as well as in the articulatory data (vocalic anticipation in the front-back dimension) of some four-year-old children. It is therefore surprising that they did not elaborate on these results but instead suggested an "inability to anticipate V_2 in V_1 during the production of V_1-C-V_2 sequences" for four-year-olds (p. 4).

From our results, it is clear that like adults, children from at least 3.5 years of age anticipate the horizontal tongue position of a full vowel during the production of a preceding schwa across an intervocalic consonant. Whether this process should be interpreted as an "ability" or rather as an inevitable byproduct of continuous speech will be discussed in more detail in the following sections.

2.5.2 The impact of the intervocalic consonant

In line with previous evidence (Cole, Linebaugh, Munson, & McMurray, 2010; C. A. Fowler & Brancazio, 2000; Öhman, 1966), we found a significant impact of the consonant context on the degree of vocalic anticipation in adults: In əgV sequences, vowel anticipation was stronger than in əbV and in ədV sequences.

Both əgV and ədV are homorganic sequences because the tongue provides the primary articulators involved in the production of both consonantal and vocalic gestures. However, while the location of tongue body contact with the palate for /g/ is relatively flexible without affecting intelligibility, the contact point for /d/ is more constrained in the alveolar region (Ladefoged & Johnson, 2014). In a previous investigation of intrasyllabic coarticulation in our cohort of adults, this strong difference in coarticulatory resistance between /g/ and /d/ was replicated (Abakarova, Iskarous, & Noiray, 2017). The present finding of more vocalic anticipation in əgV than in ədV sequences is therefore neatly in line with the idea that the consonant's resistance not only accounts for the degree of coarticulation during the consonant production but also for the degree of interference with transconsonantal coarticulation processes. Yet, if the consonant's resistance were the only factor here, one would expect əbV sequences to exhibit the highest degree of lingual V-to-V coarticulation because the tongue body is not recruited for the labial occlusion gesture and can therefore anticipate the upcoming vowel's gestures freely. Many studies including our previous analyses found the predicted high degree of lingual anticipation during /b/ in intrasyllabic coarticulation (C. A. Fowler & Brancazio, 2000; Iskarous et al., 2010; Noiray et al., 2018; Recasens, 1985; Sussman et al., 1992). The present findings in intersyllabic coarticulation however, provide evidence for /b/ to allow V-to-V coarticulation (only) to the same extent as the high resistant /d/ instead of being very permeable for trans-consonantal vowel anticipation as expectable for low resistant consonants like /b/ and /g/. A closer look at C. A. Fowler and Brancazio's (2000) data also reveals less V-to-V coarticulation in /b/ than in /g/ sequences for tongue fronting in one of two speakers and for F2 changes, both speakers exhibited less V-to-V coarticulation in /b/ contexts than in /g/ and /d/ contexts.

However, the origins of /b/'s and /g/'s low resistance are certainly distinct and must be acknowledged in order to understand their contrasting impact on V-to-V coarticulation: While /g/ engages the same primary articulator as following vowels (the tongue body), əbV sequences are heterorganic with /b/ not actively recruiting the tongue body. So, while for gV sequences, the position of the primary articulator is changed by coproduction with the following vowel, gestural blending does not affect the primary articulator of /b/ (the lips) but an articulator that is not actively controlled for the production of /b/. Although both consonants are classified as low resistant because of the flexibility of the tongue body's horizontal position, the sources of this high degree of coarticulation are thus very different in nature.

Looking only at the change of the tongue body's position during the consonant, this difference results in more coarticulation during /b/ than during /g/ because an unspecified or inactive articulator can be changed most flexibly. However, in long-distance processes like vowel-to-vowel coarticulation across these consonants, the picture changes: The primary articulator of the consonant must start moving towards its target during the schwa to ensure the correct place of contact. For /g/ this means that during schwa, the tongue body moves towards a position in the velar or palatal region that will be more front in the case of following front vowels or back for following back vowels. The process of vowel-to-vowel coarticulation in əgV sequences could therefore be understood as being reinforced by /g/: because of the coproduction with the vocalic gesture, the contact point of the tongue body and the palate or velum is changed for /g/; In addition to the direct vocalic anticipation, the initiation of /g/ therefore increases the strength towards a front or back positioning of the tongue body during schwa.

Yet, for /b/ the primary articulators are the lips, so they are the ones starting to move towards each other during schwa in a əbV sequence. There is no consonant-induced need however for the tongue back to start moving towards a specific position because it is unspecified for /b/. While during /b/ the vowel's tongue position is thus anticipated, there is only a weaker vowel-related movement of the tongue towards that target during schwa.

Turning to children, in none of the investigated age groups did the nature of the intervocalic consonant influence the degree of V-to-V coarticulation significantly. There is only a marginally significant trend for seven-year-olds to coarticulate more in /g/- than in /d/-contexts similar to adults (p = .0834). Because previous developmental research has not focused on consonantal effects on vowel anticipation, the lack of consonantal impact is an important finding, especially given the sizeable difference found in comparison to adults. Although Nittrouer (1993) examined the intervocalic consonant's effect on V-to-V coarticulation in her data set of three-, five-, and seven-year old children and adults and found stronger vowel anticipation in /k/ compared to /t/ contexts, her study was not designed to address developmental differences of this effect. The age-related differences in the consonant's impact that we found in our study as well as its implications for our understanding of the development of spoken language fluency will be discussed in the following section.

2.5.3 Developmental differences

The overarching aim of the present study was to investigate the development of intersyllabic V-to-V lingual coarticulation. Expanding on our earlier findings regarding the organization of intrasyllabic V-to-C-coarticulation (Noiray et al., 2018), the present results provide strong evidence for children to exhibit a much higher degree of V-to-V coarticulation than adults (cf., Figure 2.1). Children therefore seem to exhibit a larger extent of gestural overlap not only between the consonant and the following vowel but also earlier during the preconsonantal schwa. This suggests that children initiate vocalic gestures earlier in comparison to adults. Kent (1983) described developing (as well as impaired) speech production to follow a principle of "everything moves at once" (p.70). A conceivable reason for this greater gestural overlap in children compared to adults might be the lack of inhibitory control that is well attested in various cognitive domains for young children (e.g., Bjorklund & Harnishfeger, 1990). A lower inhibition level might accordingly lead to more simultaneously activated gestures and hence more articulatory overlap (cf. Tilsen, 2013).

Among the different cohorts of children, we also noticed a trend towards a developmental decrease in coarticulation degree from three to seven years of age, but it only yields significance for the alveolar and bilabial context, not for the velar one. Indeed, for sequences involving the resistant consonant /d/ the youngest group of children at three years of age exhibits significantly more coarticulation than the oldest group. Both the alveolar stop /d/ and the bilabial stop /b/ requires a very fine spatiotemporal coordination of different articulators: The tongue's subparts (e.g., the tongue tip and the tongue body) in ədV sequences, the lips and the tongue body in əbV sequences. Whether a maturation of this coordination between three and seven years of age is the reason for our preliminary finding should be investigated more thoroughly with a larger set of consonants requiring fine lingual coordination (e.g., /d, t, z, s, l, n, f, w, m, n/). Yet, this result accords well with previous reports on the non-uniform development of articulatory controls for speech (lips and jaw: e.g., Noiray et al., 2010; Smith & Zelaznik, 2004; for the tongue: e.g., Noiray et al., 2013; Zharkova, Hewlett, & Hardcastle, 2011). It further suggests that the developmental spurts and plateaus often reported for other articulators in the literature (e.g., great change in lip movements variability between two and six years: Green, Moore, & Reilly, 2002; variability plateau between seven and 12 years: Smith & Zelaznik, 2004) should be carefully interpreted in relation to the speech material investigated and the complexity of the gestural coordination involved. In practice, the differences in V-to-V coarticulatory degree within childhood certainly call for more scrupulous investigations of coarticulatory patterns in tightly clustered age groups. Such research would provide a description of gestural control development across childhood preventing important transitions from remaining unnoticed. It would further provide much needed normative data to disentangle coarticulatory differences that reflect typical trajectories from those that may predict later articulatory disfluencies. This may for instance be particularly relevant for the early assessment of children with developmental apraxia of speech known to show impairments of speech motor control (see review in Terband, Maassen, et al., 2009).

Interestingly, the developments of V-to-C and V-to-V coarticulation do not seem to go uniformly hand in hand. While the present study unveiled a change in V-to-V coarticulation degree during childhood only between the youngest and the oldest children for sequences involving the alveolar stop /d/ and the bilabial /b/, our earlier results on V-to-C coarticulation provided evidence for significant differences between cohorts of children for /b/ (C3 > C7) and /g/ (C3 > C7, C5 > C7) but not for /d/ (Noiray et al., 2018). This finding again highlights the very different role the articulatory properties of a consonant play for inter- and intrasyllabic coarticulation processes outlined above.

In our study, the gap in the magnitude of coarticulation between seven-year-olds and adults remains tremendous across all consonants. Children at the beginning of primary school are therefore still developing the organization of lingual gestures for articulatory fluency. This result supports previous research pointing at the protracted development of spatiotemporal control of speech gestures (Smith & Zelaznik, 2004). Despite an increasing interest for addressing early language development in recent years, future research should include late childhood investigations to locate transitions towards adult-like patterns of coarticulation and identify the factors responsible for developmental differences across childhood. Note that the nature of those factors may change over time. While age-related differences might initially be driven by discrepancies in lexical knowledge (e.g., Edwards, Beckman, & Munson, 2004; Nicholson, Munson, Reidy, & Edwards, 2015) and/or speech motor control, coarticulatory differences between older groups of children may be affected by the acquisition of new skills (e.g., inhibitory control: Tilsen, 2013) or consolidation of recently acquired ones.

The second developmental difference found in our data is that the consonant's identity impacts on the degree of V-to-V coarticulation in adults but (except for a marginally significant trend for seven-year-olds) not in children. While we found adult-like patterns of consonants' coarticulatory resistance in our previous analyses on children's intrasyllabic coarticulation (Noiray et al., 2018), the strong discrepancy between ages in intersyllabic coarticulation seems surprising at first glance. However, as predicted, adults' effects of the consonant's identity seem to be stronger in intra- than in intersyllabic coarticulation. Being relatively subtle, consonantal effects might therefore be concealed by the higher variability in children's intersyllabic coarticulation. Taking the conducted three-way-interactions of the factors X_V, Cohort, and Consonant1 into account, it becomes obvious however, that the different behavior of adults and children cannot solely result from too high variability: The /g/>/b, d/ pattern observed in adults seems only to develop across the investigated age cohorts. While for three- and four-year-olds both the relations between the /b/- and /g/-context and that between the /d/- and /g/-context differ from that of adults, it is only the b-g relation that is different between five-year-olds and adults. Seven-year-olds pattern in the same way as adults. Albeit the coarticulation degree of seven-year-olds is still very different from that observed in adults, the coarticulatory pattern regarding the relation of /b/, /d/, and /g/-contexts therefore is already approximating that of adults.

Linking these findings, we see hints for the hypothesis that the diverging roles of the intervocalic consonant in children's and adults' V-to-V coarticulation is based in their gestural organization. As Öhman (1966) and C. A. Fowler and Brancazio (2000) argued, vocalic gestures may be phased relatively invariantly with each other while the consonantal gesture occurs as a temporally limited event during the broad vocalic movements. Given that young German children were reported to focus on stressed syllables (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009) and that due to their acoustic and prosodic properties vowels seem to have a special status in an utterance, functioning as attractors and being very prominent for children (Cutler & Mehler, 1993), articulatory gestures relating to V_2 (the stressed full vowel in our stimuli) might be hardest to inhibit for children and therefore show especially broad overlap with preceding gestures. Any subtle effect of the intervocalic consonant's resistance might therefore not (only) be concealed by a high variability but by an underlying outstandingly high prominence of vowels in child speech. In contrast to the idea of consonant-mediated V-to-V effects in adult speech, V-to-V coarticulation in children would therefore be interpreted as a pure coproduction of V_1 and V_2 because of the greater gestural overlap. Accordingly, while the V-to-V coarticulation degree in velar contexts is especially high presumably because of the stop /g/ mediating coarticulation, this context does not promote stronger coarticulation than the others in young children's speech. The strong coproduction of V_1 and V_2 itself, results in approximately the same coarticulation degree in all contexts.

Another factor possibly influencing the maturation of coarticulatory processes during childhood is the anatomical development of the vocal tract. While it is well known that physiological characteristics can affect articulation (e.g., Vorperian et al., 2005, 2009), evaluating the precise impact of those anatomical changes on developmental differences in lingual coarticulation remains an empirical challenge. To overcome difficulties in anatomical measurements of the vocal tract or the tongue, the growing research on articulatory modelling may provide better estimates (e.g., Ménard, Schwartz, & Boë, 2004; Story, 2005).

2.6 Limitations and perspectives

Investigating lingual coarticulation in the first years of life has become increasingly significant for the early detection of spoken language deviancies. However, collecting quantitative tongue data from young children is not exempt of methodological challenges (e.g., limited attention span, intolerance to invasive methods and too long data collection sessions). In this study as in previous research, a few compromises were therefore necessary to meet our research goals.

First, we used a customized probe holder designed to not impede natural jaw movements and collect data that approximate natural speech conditions more faithfully than if we had used a helmet (e.g., Zharkova et al., 2012). We employed three strategies to prevent head movement artefacts: 1) the SOLLAR recording platform included a car seat with seatbelts; 2) a bright star was positioned in front of the child as a visual fixation point; 3) one of the experimenters sat in front of the child to maintain visual contact and monitor the child's position.

Finally, post-recording examination using video data was conducted to discard data in which children moved.

Second, following up on previous research (e.g., Noiray et al., 2018, 2013), this study employed measurements of the highest point on the tongue body to assess variation in the gestural organization of V-to-V coarticulation. While the approach to use a single point measure is certainly convenient for the investigation of large samples, it is not optimal for fine-grained distinctions between the subparts of the tongue (e.g., tongue root) as in studies considering the full tongue contour (e.g., Recasens & Rodríguez, 2016). However, it is important to acknowledge that the reliability of the latter approach highly depends on the quality of the tongue imaging at the two ends of the tongue contour (cf. Noiray et al., 2018 for a more detailed discussion). In previous studies the measure employed here has provided meaningful results as to developmental differences in coarticulatory overlap (e.g., Iskarous et al., 2010; Noiray et al., 2018, 2013). Most acoustic studies used measurements of F2 as an estimate of the tongue position along the antero-posterior dimension and the resulting cavities (e.g., adults: Beddor et al., 2002; C. A. Fowler, 1981; Modarresi et al., 2004; Öhman, 1966; Recasens, 1987; children: Boucher, 2007; Goodell & Studdert-Kennedy, 1993; Hodge, 1989; Nijland et al., 2002; Nittrouer, 1993; Nittrouer et al., 1996; Repp, 1986). The highest point on the tongue body is the most salient for vowel constriction and therefore provides a more direct access to those parameters. Future studies of lingual coarticulation will gain in designing methodologies that integrate measurements of fixed-point parameters and of the full tongue contour. With such a combinatorial approach, it will be possible to unveil subtle developmental differences in coarticulatory patterning, due for instance to discrepancies in coordinative control of the tongue's functional subparts.

Finally, we are well aware that assessing vocalic anticipation via single time point analyses is not optimal because the method does not fully capture coarticulation dynamics (e.g., Scobbie, Lawson, & Stuart-Smith, 2012). The optimization of analytical approaches assessing change over time to ultrasound research will be necessary to unveil the complexity of gestural dynamics (e.g., Winter & Wieling, 2016).

2.7 Conclusion

This study was the first which addressed the maturation of lingual long-distance coarticulatory processes in a cross-sectional investigation of five age cohorts using articulatory measurements. Taken together, our findings provide evidence for children to exhibit stronger vocalic anticipation than adults suggesting a maturational decrease of gestural overlap with age. Across the period from three to seven years of age, no general, but a consonant context-specific decrease of vocalic anticipation was found, which is a sign of non-uniform maturation of gestural organization possibly driven by differences in articulatory complexity. The tremendous disparity in coarticulation degree between the oldest children investigated and the adults indicates that the development of adult-like gestural organization continues during late childhood. Our study therefore highlights the importance of investigations of older children's and adolescents' speech to uncover factors that might lead to a compression of articulatory gestures, hence an adult-like lower gestural overlap.

3 VOCALIC ACTIVATION WIDTH DECREASES ACROSS CHILDHOOD: EVIDENCE FROM CARRYOVER COARTICULATION[4]

3.1 Abstract

This study is the first to use kinematic data to assess lingual carryover coarticulation in children. We investigated whether the developmental decrease previously attested in anticipatory coarticulation, as well as the relation between coarticulatory degree and the consonantal context, also characterize carryover coarticulation. Sixty-two children and 13 adults, all native speakers of German, were recruited according to five age cohorts: three-year-olds, four-year-olds, five-year-olds, seven-year-olds, and adults. Tongue movements during the production of ə.CV.Cə utterances (C = /b, d, g/, V = /i, y, e, a, o, u/) were recorded with ultrasound. We measured vowel-induced horizontal displacement of the tongue dorsum within the last syllable and compared the resulting coarticulatory patterns between age cohorts and consonantal contexts. Results indicate that the degree of vocalic carryover coarticulation decreases with age. Vocalic prominence within an utterance as well as its change across childhood depended on the post-vocalic consonant's articulatory demands for the tongue dorsum (i.e., its coarticulatory resistance): Low resistant /b/ and /g/ allowed for more vocalic perseveration and a continuous decrease, while the highly resistant /d/ displayed lower coarticulation degrees and discontinuous effects. These findings parallel those in anticipation suggesting a similar organization of anticipatory and carryover coarticulation. Implications for theories of speech production are discussed.

3.2 Introduction

The investigation of coarticulatory effects, that is, the overlap of articulatory units in spoken language, served as a window to speech planning and execution mechanisms in adults over the last 60 years (for a review see Recasens, 2018). However, only in the last decade, non-invasive measurement techniques such as ultrasound tongue imaging were administered to young children and hence shed new light on speech motor developments as well as their interactions with cognitive aspects relevant for speech production (e.g., Barbier et al., 2020; Ménard & Noiray, 2011; Noiray et al., 2013; Song et al., 2013; Zharkova, 2017; Zharkova et al., 2011). The present study focuses on the development of lingual carryover coarticulation across childhood, the

---

[4] Chapter 3 of this dissertation was published as:
Rubertus, E., & Noiray, A. (2020). Vocalic activation width decreases across childhood: Evidence from carryover coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1), 7. doi: http://doi.org/10.5334/labphon.228.
Adaptations were made regarding citation style, figure embedding, and cross references.

overlap of a speech segment with following ones after its target was reached. While anticipatory coarticulation has often been described as a sign of speech planning, carryover coarticulation was ascribed to mechanical inertia constraints (e.g., Recasens, 1984b) and was therefore largely understudied. We suggest that both anticipatory and carryover coarticulation are the consequence of the overlap of gestural activation. The parallelism of the development of carryover coarticulation as found in the present study and anticipatory coarticulation as previously reported, provides evidence for this hypothesis.

3.2.1 The development of anticipatory coarticulation

In our previous kinematic analyses of lingual anticipation of a stressed vowel in German children, we provided evidence for a developmental decrease of coarticulation degree in intrasyllabic vowel-to-consonant coarticulation (Noiray et al., 2018), in intersyllabic vowel-to-vowel coarticulation (Rubertus & Noiray, 2018), as well as in the temporal unfolding of the vocalic gesture within the left field of an utterance of the form ə.CV.Cə (C-consonant, V-vowel; Noiray, Wieling, et al., 2019). This finding is in line with several previous investigations on intra- (e.g., Katz et al., 1991; Kent, 1983; Zharkova et al., 2012) and intersyllabic coarticulation (e.g., Goodell & Studdert-Kennedy, 1993; Nijland et al., 2002; Nittrouer, 1993; Nittrouer et al., 1996), but contrasts with others that found an increasing degree of coarticulation with age (intrasyllabic: Nijland et al., 2002; Nittrouer et al., 1989, 1996; intersyllabic: Barbier et al., 2020; Hodge, 1989; Repp, 1986).

3.2.2 The development of carryover coarticulation

Carryover coarticulation in children's speech has been investigated in only very few studies that focused on different speech articulators. Neither Flege (1988), who examined nasal coarticulation, nor Goffman, Smith, Heisler, and Ho (2008), who focused on labial coarticulation, provide systematic evidence for a developmental decrease in carryover coarticulation degree. The only study addressing children's lingual carryover coarticulation we know of is Baum and Waldstein (1991). Using three different types of measures, they compared coarticulation degree in VC syllables (/iʃ, uʃ, it, ut, ik, uk/) between English-speaking hearing-impaired and age-matched normally hearing children in two age groups: six to seven and nine to 10 years of age. No difference between the age groups was found within the cohorts in any of the measures, so they were grouped in the analysis. The first measure of consonant durations did not differ significantly between the normally hearing and the hearing-impaired group. The measure of mean centroid values (in fricatives and stop bursts) demonstrated stronger carryover coarticulation in normally hearing as compared to hearing-impaired children at consonant onset. At consonant midpoint, however, both cohorts exhibited the same degree of coarticulation based on this measure. In the syllables /iʃ/ and /uʃ/ the third measure of F2 peaks at vowel offset and fricative revealed a higher coarticulation degree in normally hearing than in hearing-impaired children again. The authors concluded that it is not the temporal domain of carryover coarticulation but its magnitude within this time frame that differs between the two cohorts. Interestingly, measures of anticipatory coarticulation in the same group of children (Waldstein & Baum, 1991) had indicated shorter temporal domains of anticipation for hearing-impaired than normally hearing children. Baum and Waldstein (1991) interpreted this discrepancy as well as an overall larger degree of carryover compared to anticipatory coarticulation as evidence for different mechanisms underlying the two coarticulatory directions. According to the authors, a significant age difference found in anticipatory but not in carryover coarticulation may either be due to the close and relatively advanced ages studied or provide additional support for carryover coarticulation to depend on mechanical-inertial properties that need not be learned.

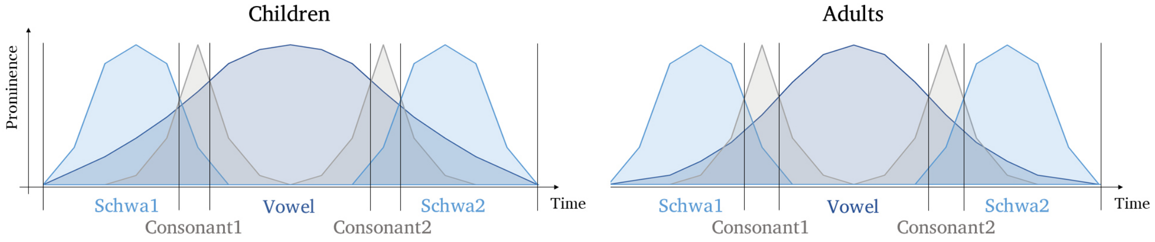
3.2.3 Decrease of coarticulation as compression of vocalic activation curves

According to the broad framework of articulatory phonology (Browman & Goldstein, 1986), articulatory gestures have invariant goals and are planned and phased to each other context-independently. In contrast to suggestions that context-dependency is part of the speech plan and actively changes articulatory goals (e.g., Henke, 1966; Keating, 1988; Wickelgren, 1969), articulatory phonology interprets contextual variation to be introduced only upon execution by the blending of individual gestures' influences on the vocal tract with those of other ongoing ones (e.g., C. A. Fowler, 1980; C. A. Fowler & Saltzman, 1993; Gafos & Goldstein, 2012). Here, coarticulation is seen as the coproduction of invariant articulatory gestures. The more the activation of gestures overlaps, the more coarticulation may take place. The higher degree of anticipatory coarticulation in children than in adults (Noiray et al., 2018; Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2018) can therefore be interpreted as greater overlap of vocalic gestures with preceding ones in the young age. The developmental decrease in coarticulation would in turn be a developmental compression of vocalic activation curves (cf. Nittrouer, 1993; Noiray, Wieling, et al., 2019).

Following Nittrouer (1993, p. 961), the sketch of the prominence, that is, the strength of activation, of an utterance's segments over time in the style of C. A. Fowler and Smith (1986), in Figure 3.1 illustrates the larger overlap of articulatory gestures for neighboring segments that would result from broader vocalic activation curves in children's (left side) than adults' speech (right side). The segment with the highest prominence at a given time point dominates the acoustic signal. Changes in the dominance and therefore acoustic segmentations within the utterance are indicated by vertical lines.

Figure 3.1

Segments' Hypothesized Prominence over Time in Utterances of the Form əCVCə



A reason for children's vocalic activation curves to be broader than adults' may be the attractor or anchor function that multiple findings in language development ascribed to stressed vowels. Cutler and Mehler (1993) for example, suggested that infants have a periodicity bias leading them to attend more to vowels than to consonants in the acoustic signal. This could in turn be one reason why native phonological categories for vowels are constituted earlier in development than for consonants (Kuhl et al., 1992; Werker & Tees, 1984). The information carried by vowels and consonants was also suggested to differ: While vowels carry phonetic as well as prosodic information relevant for rhythm and syntax, consonants' information is mainly lexical (Nespor, Peña, & Mehler, 2003). Young children were shown to focus on the vowel-inherent prosodic information to bootstrap the segmentation of first words (Gleitman & Wanner, 1982). Höhle et al. (2009) for example, provided evidence that young German-learning infants scan their input for stressed syllables to find trochaic patterns as a first strategy to detect words in the continuous signal. Also, in speech production, young children tend to reduce first words to the stressed CV syllable or a trochaic pattern (in German, e.g., A. V Fox & Dodd, 1999). C. A. Fowler (1980) highlights the role of stressed vowels in the coproduction of speech segments and claims that not only consonants but also unstressed vowels are "superimposed on a trajectory of the shape of a vocal tract from one stressed vowel to another" (p. 131). This subsumption of segments in frames of stressed vowels might be responsible for the stress-timed speech rhythm in languages like for example English and German (C. A. Fowler, 1981) – a property of the speech signal that already newborns are very attentive to (Nazzi, Bertoncini, & Mehler, 1998).

An important consequence that the hypothesis of generally broader vocalic activation curves in the young age bears, is that children's vowels would not only overlap more with preceding speech segments but, as visualized in Figure 3.1, larger overlaps would be predicted in the right field of the utterance as well. Whether this is the case has never been explicitly tested because the focus of coarticulation development studies remained in the anticipatory direction.

3.2.4 The role of the articulatory demands of combined segments

In addition to the general decrease of gestural overlap, the role of articulatory parameters of the combined segments for coarticulation development must be considered. The simple overlap of gestural activation does not correspond one-to-one to the degree of coarticulation found in spoken utterances. During the execution of the context-insensitive speech plan, the different parameters of the coproduced gestures, most importantly their degree of coarticulatory resistance (Bladon & Al-Bamerni, 1976; Recasens, 1984a) and the corresponding vocal tract configurations affect the degree of gestural blending. A consonant that is highly resistant to coarticulation, for example the alveolar plosive /d/, employs the tongue body that is relevant for vowel production in a rather constrained way for its own production and therefore interrupts vocalic movements. A bilabial, on the other hand, does not share the primary articulator with vowels and is therefore produced without affecting the tongue body movement necessary for the vowel trajectory. A third case is comprised of velar consonants that share the primary articulator with the vowel but are blended with the vocalic production requirements resulting in different points of palatal contact depending on the frontness of the surrounding vowels. Effects of consonants' coarticulatory resistance on vocalic coarticulation were widely demonstrated in adults (e.g., C. A. Fowler & Saltzman, 1993; Iskarous et al., 2010; Recasens, 1985; Recasens & Rodríguez, 2016). During language development, how strongly a given consonant clamps the tongue dorsum and how much coarticulation it therefore allows, may change with a growing control over the functional subparts of the tongue. For intrasyllabic anticipatory coarticulation, Noiray et al. (2018) found adult-like coarticulation hierarchies of /b/>/g/>/d/ in children from three to seven years of age. However, CV syllables are the fundamental syllables that are best practiced in early childhood (e.g., Fikkert, 1994); different patterns may therefore be found in gestural combinations other than CV syllables and in carryover coarticulation.

3.2.5 A dichotomy of underlying mechanisms?

Many authors describe a dichotomy of underlying mechanisms for anticipatory and carryover coarticulation: While the former is described as part of a speech plan, the latter is attributed to mechanical inertia constraints. Recasens (1984b, 1987) and Parush et al. (1983) for example, provided data from Catalan and English VCV sequences, respectively, suggesting that while the consonant's coarticulatory resistance affects the temporal extent of anticipatory coarticulation, it is the spatial extent of carryover coarticulation that is affected. They interpreted this as evidence for active speech planning controlling the degree of anticipatory but not that of carryover coarticulation with reference to the articulatory requirements of the intervocalic consonant. In German, Hertrich and Ackermann (1995) found that vocalic carryover but not anticipatory effects were smaller in slower speaking rates. According to the authors, stable or increased anticipatory effects in slow speaking rates are not compatible with a view of simple coproduction but indicate a (speaker-specific) planning component in anticipatory coarticulation. The decrease of carryover coarticulation however, was interpreted to suggest that planning processes might be less relevant for this coarticulatory direction.

In a pure coproduction framework on the other hand, the overlap of context-independent articulatory gestures can account for both anticipatory and carryover coarticulatory effects. In their comparison of empirical and modeling data, Ostry, Gribble, and Gracco, (1996) for example, provide evidence that coarticulation in jaw movements is not centrally planned but arises as a by-product of execution. If there is no active planning of context effects, there is no reason to assume different mechanisms underlying the two coarticulatory directions.

3.2.6 What we can learn from carryover coarticulation development

If anticipatory and carryover coarticulation embody a common organization and it is indeed the width of activation curves that changes across childhood, carryover coarticulation should develop in parallel with anticipation and decrease with age. If, however, different mechanisms underlie the two coarticulatory directions, the developmental differences found in lingual anticipatory coarticulation may be absent in lingual carryover coarticulation. Under the hypothesis of a dichotomy of origins for the two coarticulatory directions, it was for example suggested that inertial properties of muscles in contrast to planning processes need not to be learnt (Baum & Waldstein, 1991; Flege, 1988). Following this idea, a developmental change of coarticulation degree would be expected for anticipation but not for perseveration. Investigating the development of carryover coarticulation across childhood may therefore provide additional support for one or the other assumption on the speech production mechanism.

3.2.7 *Research questions and predictions*

There is growing evidence that in the course of speech development, children's degree of lingual anticipatory coarticulation progressively decreases. Regarding carryover coarticulation, however, data are scarce, and predictions differ based on the theoretical framework. The present study aims to provide the first large-scale kinematic investigation of children's carryover coarticulation development. It builds upon previous findings of the same research group to test the hypothesis of a developmental decrease in lingual carryover coarticulation. Our goal was to provide answers to the following two questions:

1. Does the degree of carryover coarticulation decrease with increasing age as we found for anticipatory coarticulation?

Contextualizing the principles of articulatory phonology and the coproduction model of adult speech production to children's development of coarticulation, we hypothesized the underlying vocalic activation curves of children's speech to be generally broader than adults' which results in more overlap of a vowel with preceding as well as following gestures. We therefore predicted a decreasing degree of carryover coarticulation with increasing age. In light of differing findings of non-linear developments across age depending on the type of coarticulation studied, we did not make specific predictions about plateaus or spurts within this decrease.

2. Do the articulatory demands of the following consonant impact the perseveration of the vocalic gesture?

Based on previous findings including ours, we hypothesized that the degree of consonants' coarticulatory resistance affects the degree of vocalic carryover coarticulation significantly. Consonants posing strong and specific articulatory demands on the tongue dorsum (e.g., /d/) were therefore predicted to be more intrusive on the vocalic gesture than those that can blend their gestural goals with the vowels' (e.g., /g/) and those that do not employ the tongue dorsum as a primary articulator (e.g., /b/). Since the balance between clamping and blending depends on a fine speech motor control, we expect developmental changes in the role of the consonant for coarticulation degree.

3.3 Method

3.3.1 *Participants*

Possibly non-linear cognitive as well as speech motor control developments occur in children before they enter school (Green et al., 2010; Noiray, Popescu, et al., 2019). Therefore, we tested three age cohorts of preschool children in yearly increments, a cohort of first graders, and a group of adults, summing up to a total of 75 participants: 19 three-year-old children (10 females [f] and nine males [m], age range: 3;05–3;09 [Y;MM], mean: 3;06), 14 four-year-old children (seven f and seven m, age range: 4;04–4;08, mean: 4;05), 14 five-year-old children (seven f and seven m, age range: 5;04–5;07, mean: 5;06), 15 seven-year-old children at the end of their first or beginning of their second grade in primary school (10 f and five m, age range: 7;00–7;06, mean: 7;02), and 13 adults (seven f and six m, age range: 19–28 years, mean: 23). None of the participants reported any language-, hearing-, or vision-related problem and all were monolingual German. Adult participants and parents of child participants gave written informed consent for participating in the study while children gave oral consent. It was emphasized that they could interrupt or abort the recording session for any reason at any time. The study was approved by the Ethic Committee of the University of Potsdam (DFG project 1098).

3.3.2 Stimulus material

Previously recorded disyllabic pseudowords with a trochaic stress pattern spoken by a native German female adult speaker served as model stimuli for a repetition task. They consisted of the consonants /b, d, g/ and the vowels /i, y, e, a, o, u/ in the form consonant$_1$- vowel-consonant$_2$-schwa (C_1VC_2ə) where C_2 never equaled C_1. Consonants were chosen to bear different degrees of lingual coarticulatory resistance. Vowels were chosen to represent the full front-to-back range of the German vowel space. Each pseudoword was recorded together with the German female article /aɪnə/ resulting in short utterances such as /aɪnə bi:də/. Vocalic carryover effects were measured at four different time points within C_2 and the final schwa.

The crossed set of consonants and vowels resulted in 36 target words that were repeated at least three times in the test phase summing up to a total of 108 trials. For four- and seven-year-olds, and adults, additional stimuli were recorded that are not part of the present analysis. An overview of each cohort's stimulus sets is presented in Table 3.1. The age cohorts will be referred to as C3 (three-year-olds), C4 (four-year-olds), C5 (five-year-olds), C7 (seven-year-olds), and A (adults) in the rest of the paper.

Table 3.1

Overview of Each Age Cohort's Stimulus Sets

	C3	C4	C5	C7	A
Consonant$_1$	/b, d, g/	/b, d, g, z/	/b, d, g/	/b, d, g, z/	/b, d, g, z/
Consonant$_2$	/b, d, g/	/b, d, g/	/b, d, g/	/b, d, g/	/b, d, g, z/
Nr. of stimulus words	36	46	36	46	72
Total nr. of trials	108	138	108	138	216

Note. Stimulus words had the form C$_1$VC$_2$ə with V=/i, y, e, a, o, u/. The total number of trials results from at least three repetitions of each stimulus word during the recording.

For all children, stimuli were presented in six blocks while adults' increased stimulus set required nine blocks. The order of blocks was randomized for each participant, and trials within each block appeared in one of three random but pre-specified orders. We opted for this semi-randomization to be able to quickly take notes on specific trials, for a better synchronization between both experimenters and to allow a semi-automatic phonetic labeling procedure for adults' data. During the recording, the experimenter made a note of mispronounced trials and played those again at the end of the block. Table 3.2 summarizes the number of trials used for the present analysis per C$_2$ per age cohort.

Table 3.2

Summary of the Number of Analyzed Trials per Consonant Context and Age Cohort

Consonant Context	Number of trials				
	C3	C4	C5	C7	A
Vbə	516	555	529	674	477
Vdə	463	542	503	647	479
Vgə	526	571	540	624	483
Total	1505	1668	1572	1945	1439

3.3.3 Experimental procedure

Participants, both children and adults, were asked to repeat acoustically presented stimuli within the SOLLAR platform (Sonographic and Optical Linguo-Labial Articulation Recording system, Noiray et al., in press) at the Laboratory for Oral Language Acquisition at the University of Potsdam (Germany). The SOLLAR platform provides a child-friendly environment allowing for simultaneous recordings of tongue motion (ultrasound imaging: Sonosite, sampling rate: 48Hz), labial movement (video recording: SONY camera, sampling rate: 50Hz), and the acoustic signal (Shure microphone, sampling rate: 48kHz). The ultrasound probe was fixed in a custom-made probe holder providing flexibility in the vertical dimension to follow the natural jaw movements but being rigid in lateral and horizontal translations. Participants sat in a comfortable chair adjustable in height and their head was positioned such that the probe touched their chin between the maxillary bones to record the tongue surface contour in the midsagittal plane. Intending to make the platform as child-friendly as possible and to allow relatively natural speech, no additional head-to-probe stabilization was employed. Instead, a visual attention-getter (a glittering golden star) and if necessary, the experimenter, helped especially the young participants keep their head stable and look straight towards the camera. Trials during which participants moved their head were discarded post-hoc via visual inspection of the video data.

During each recording session two experimenters were present. One experimenter's first task was to make the participant feel comfortable. She familiarized the participant with the SOLLAR platform and introduced the children to a universe-themed story the repetition task was embedded in. Children were told to fly from one planet to the other in the SOLLAR spaceship and repeat foreign words from other planets' languages. Between the blocks, children took a break and were distracted with a little sticker task. This game and the decoration stimulated their interest and engagement in the task. Adult participants were not introduced to the planet story but fulfilled the same task on the same type of stimuli in the same setup as the children to ensure comparability. During each recording block, the experimenter prompted the audio stimuli while maintaining a face-to-face connection with the participant and controlling for head stability and correct pronunciation. The second experimenter operated SOLLAR's recording equipment from a desk not visible to the participant. S/he thoroughly monitored both video and the audio streams to control the data quality.

3.3.4 Data processing

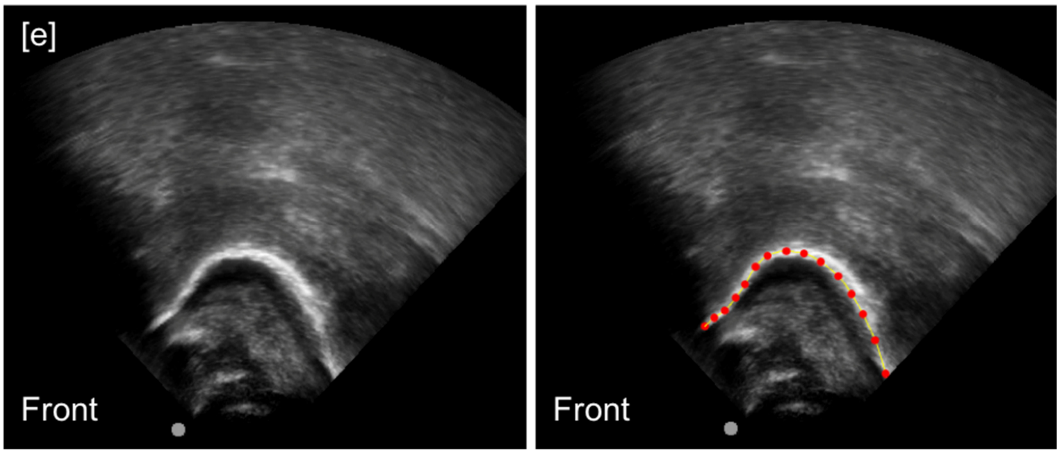
The acoustic signal was recorded both in relation to the ultrasound signal and the video, enabling the generation of a common time code for the three streams. Using a cross-correlation function within MATLAB (2016), the streams were then synchronized (cf. Noiray et al., 2011; Noiray et al., 2013).

Correctly pronounced target utterances were first phonetically labeled in the acoustic signal using Praat (Boersma & Weenink, 2016). In adults' data, WebMAUSBasic (Kisler et al., 2012) detected target words and segments semi-automatically with manual correction when necessary. Child data was labeled completely manually. A stable periodic cycle in the oscillogram as well as a stable formant pattern (especially a clearly detectable second formant) were used as indices for vocalic segments. The first ascending zero-crossing in the oscillogram at the beginning of the periodicity was accordingly set as vocalic onset, the first ascending zero-crossing after the end of periodicity and disappearance of F2 as the beginning of the following consonant. From the resulting intervals, the relevant time stamps for the analysis, the temporal midpoint of the vowel (V50), the end of the vowel (V100), the temporal midpoint of the consonant (C50), the end of the consonant (C100), and the temporal midpoint of the final schwa (schwa50) were automatically extracted.

Via these time stamps from the acoustic signal, ultrasound frames of interest were selected and the corresponding tongue contours were detected semi-automatically with custom-made scripts for MATLAB (2016) as part of the SOLLAR platform (see Figure 3.2). For each individual frame of interest, a spline (yellow line) was automatically fit to manually placed reference points (red dots in Figure 3.2) on the visible midsagittal tongue surface contour.

Figure 3.2

Raw and Tracked Ultrasound Tongue Image



Note. The figure shows an example of an ultrasound tongue image of a five-year-old boy's [e] recorded within SOLLAR. The left panel presents the raw ultrasound image, the right panel shows the highlighted tongue contour resulting from SOLLAR's semi-automatic tracking. In each image, the front part of the tongue is depicted towards the left.

X- and y-coordinates for each of the 100 points of these splines were automatically extracted. For the present analysis, we used the x-coordinate, hence the horizontal position, of the highest point of the tongue dorsum surface contour as a representation of frontness of the tongue body. To prevent taking measures into account where the highest point on the tongue surface contour was on the tongue tip and not on the tongue body, we visually inspected those contours for the /d/ closure that had relatively low x-values for the highest point. This way, eight contours (four in C5, and one in each of the other cohorts) were identified and the corresponding trials removed from the analysis.

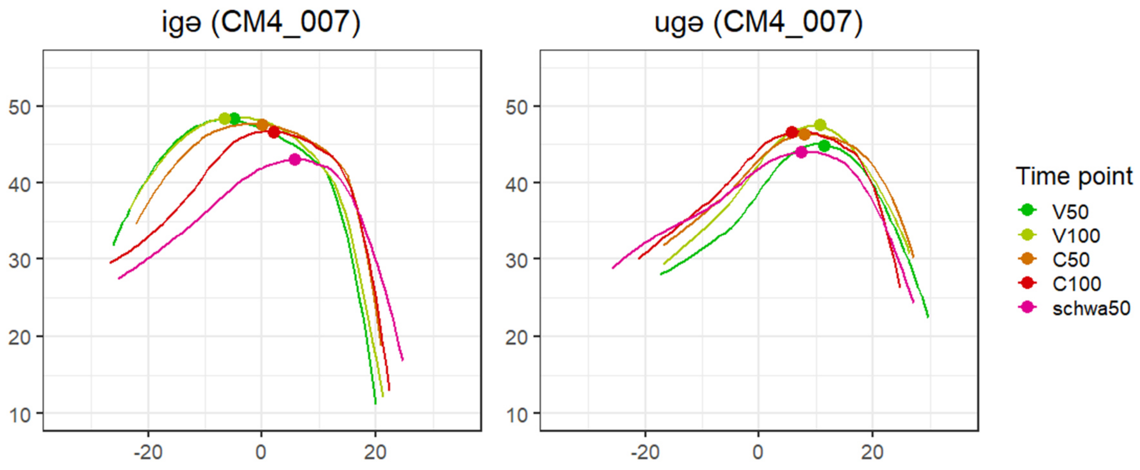
To compare coarticulatory behaviors across participants, we normalized each participant's horizontal tongue dorsum positions on the same scale. Among all of a speaker's trials, the most anterior tongue dorsum position during V50 was set to zero and the most posterior tongue dorsum position at V50 to one. His/her tongue dorsum positions at all time points were then scaled in relation to this range.

3.3.5 Data analysis

We measured coarticulatory patterns as the horizontal positions of the highest point of the tongue dorsum during the consonant and the schwa depending on the frontness of the tongue dorsum position during the preceding vowel and compared these trajectories between consonant contexts and age cohorts. Figure 3.3 presents an example of tongue movement trajectories for 'eineb**ige**' (left) and 'eineb**uge**' (right) illustrated by the tongue contours of a four-year-old boy at the five time points of interest. The highest point of each tongue contour is highlighted by a dot. The contours are presented in a coordinate system in millimeters where x = zero, y = zero is the position of the center of the ultrasound probe. X-values below zero indicate the area in front of the center of the probe (displayed towards the left), x-values above zero the back (displayed towards the right). For /igə/ (left plot), the tongue starts in a front position for /i/ at V50 (in green) and moves back towards a relatively central schwa (in pink) in the course of the utterance. For /ugə/ on the other hand (right plot), the tongue has a relatively back position during V50 and moves forward in the course of the utterance. The figure shows that 1) the horizontal position of the highest point of the tongue represents the frontness of the whole tongue, and 2) not only the tongue dorsum position at V50, but also the positions at later time points differ depending on the vowel. In the present study, the goal was not to illuminate the impact of specific vowels but rather to investigate context-induced spatial changes in tongue dorsum positions within an utterance. Vowel information is therefore not considered categorically but as a continuous variable ranging from front to back tongue dorsum positions. Most front positions correspond to phonologically front vowel categories (/i/, /y/, /e/) and back positions usually express phonologically back vowel categories (/o/, /u/).

Figure 3.3

Temporal Development of Tongue Surface Contours in Trials ‘einebige’ and ‘einebuge’



Note. The figure displays whole tongue surface contours of participant CM4_007, a four-year-old boy, for trials 'einebige' (left) and 'einebuge' (right) at time points V50 (green), V100 (light green), C50 (orange), C100 (red), and schwa50 (pink). The dots highlight the highest points of the respective contours. The front of the tongue is displayed towards the left of each plot.

To statistically assess these vowel-dependent tongue dorsum frontness trajectories, we used generalized additive modelling (GAM). A generalized additive model is a mixed effects regression model that, in contrast to the more familiar linear mixed effects model, also includes non-linear terms similar to polynomial curves, for example. GAMs can therefore detect linear as well as non-linear patterns in dynamically varying data while also taking into account subject- and item-related variability, as known from linear mixed effects models. This approach was previously applied to ultrasound data acquired from adults (Strycharczuk & Scobbie, 2017) and used for the analysis of anticipatory coarticulation in the present developmental data set by (Noiray, Wieling, et al., 2019).

We fit our models using the function *bam* of the *mgcv* package in R (version 1.8–28; Wood, 2011, 2017). For each model, the function *gam.check* was used to examine the normality of residuals’ distribution, heteroscedasticity, and adequacy of the k-parameter. This parameter specifies the maximal non-linearity by setting the size of basis dimensions for each predictor. It is limited to the number of the predictors’ unique points. For more detailed information on the application of GAMs on articulatory data, we recommend Wieling's (2018) tutorial.

For the current analysis, we tested whether the horizontal position of the highest point on the tongue dorsum depended on the horizontal position of the tongue dorsum during the stressed vowel (V50) at the four target time points V100, C50, C100, and schwa50. To include both *time* and *tongue dorsum position at V50* as well as their interaction as predictors, a tensor product (te) was used. It captures changes in the shape of the tongue dorsum frontness trajectory over time as a function of the frontness of the tongue dorsum during the stressed vowel separately for each age cohort and consonant context. In the random effects structure of the model, defined in two factor smooth terms (s), we included potentially non-linear patterns for each participant and consonant over time and for the different horizontal tongue dorsum positions at V50. The complete code for this model with explanations of single parameters can be found in the Appendix B. 1 (model m).

This first model detected the frontness trajectories of the tongue dorsum and tested whether the patterns found are significantly different from zero, i.e., non-linear, for each age cohort and consonant context. To answer our two research questions, however, direct comparisons of these patterns between 1) age cohorts, and 2) consonant contexts are necessary. Within GAMs, binary difference tensors need to be included to assess the statistical significance of comparisons between two dynamical patterns. To answer our first research question addressing developmental differences, we therefore included binary difference tensors capturing whether the age cohorts differed significantly with respect to the influence of the horizontal tongue dorsum position during the vowel on the frontness trajectory of the tongue dorsum during the following segments. An example of a code for a corresponding model including binary difference tensors can be found in Appendix B. 1 (model mb7).

Consonantal differences in vocalic carryover effects within age cohorts, the core of research question two, were assessed similarly: The models here included binary difference tensors capturing whether the consonant contexts /b/, /d/, and /g/ differed significantly with respect to the influence of the horizontal tongue dorsum position during the stressed vowel on the frontness trajectory of the tongue dorsum during the following segments within each cohort.

Because a total of six models was necessary to address all relevant comparisons (by fitting the models with differing reference groups), we Bonferroni-corrected our significance threshold to .008 to account for multiple comparisons.

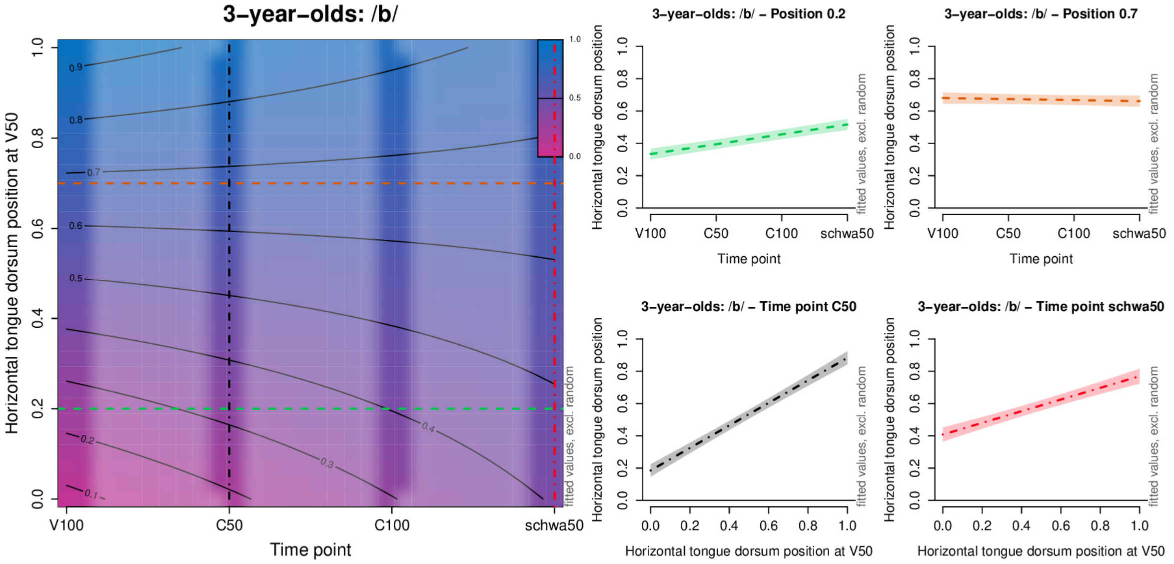
3.4 Results

3.4.1 Patterns of carryover coarticulation

For each age cohort and consonant context, the pattern of carryover coarticulation is described according to three parameters: the dependent variable *horizontal tongue dorsum position* in the course of the utterance, and the two independent variables *time point* and *horizontal tongue dorsum position at the midpoint of the stressed vowel.* To visualize these three dimensions, we present all 15 patterns (five age cohorts x three consonant contexts) in contour plots. Because these have not yet become a standard way of presenting data, we first explain how to read them with the example of 3-year-olds' coarticulatory pattern in /b/ contexts (Figure 3.4).

Figure 3.4

Explanation of the Visualization of Results in Contour Plots



Note. The figure shows an example of a contour plot that visualizes horizontal tongue dorsum positions over time (based on the four time points V100, C50, C100, and schwa50 that are represented on the x-axis) depending on the tongue dorsum position during the midpoint of the vowel (V50, y-axis). Tongue dorsum positions are indicated by color coding as shown in the small legend in the top right corner: from pink for front positions (values close to zero) to blue for back positions (values close to one). The dashed horizontal lines in the contour plot correspond do the two-dimensional graphs in the top row of the right side of the figure display and refer to the tongue dorsum positions over time for a specific V50 position (0.2 and 0.7, respectively). The dashed vertical lines correspond to the lower two graphs that visualize the tongue dorsum position depending on the V50 position for a specific time point (C50 and schwa50, respectively).

In the contour plot on the left side of Figure 3.4, the predictors *time point* and *horizontal tongue dorsum position at the vowel midpoint* are presented on the x- and y-axis, respectively. Values close to zero on the y-axis correspond to anterior tongue dorsum positions, values closer to one to posterior positions. The horizontal position of the highest point of the tongue dorsum at a given time point for a given V50 frontness value is depicted by color shades from pink for anterior positions (values close to zero) to blue for posterior positions (values close to one) as indicated in the small legend in the top right corner of the plot. Black contour lines connect points with the same value to support legibility. The vertical bands at the four different time points are the actual data while the slightly shaded areas in between are what the model predicts on the basis of this data.

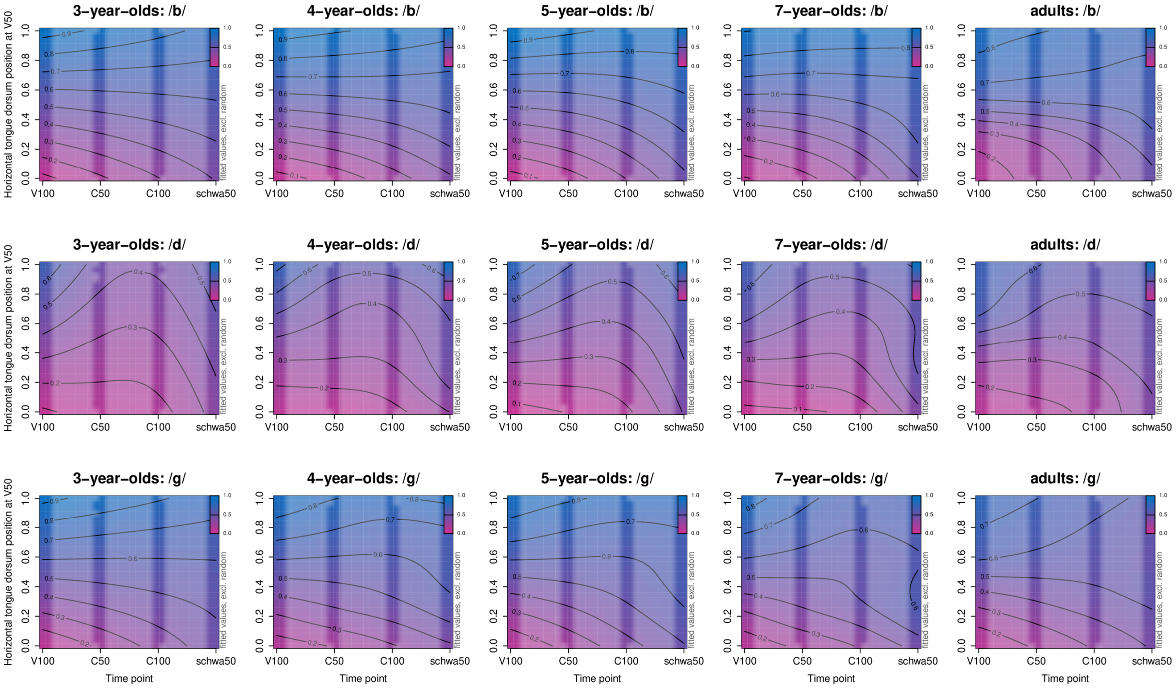
The contour plot in Figure 3.4 presents a pattern resembling a fan getting wider towards the right side with a variety of color shades at V100 but mostly purple shades at schwa50. What this implies is that at V100 there is a broad range of horizontal tongue dorsum positions (i.e., values spread from just above zero to just below one) reflecting roughly the position at the temporal midpoint of the vowel (y-axis). The further you get away from the vowel on the x-axis, however, the less color shades referring to extreme tongue dorsum positions (i.e., far front or far back) are found. Instead, we note more central positions regardless of the previous V50 tongue dorsum positions.

Each of the four small graphs on the right-hand side of Figure 3.4 isolates an independent variable to illustrate relations in a more familiar two-dimensional plot. The two plots in the top row represent the horizontal green and orange dashed lines in the contour plot and depict the horizontal motion of the tongue dorsum over time when the tongue dorsum position at V50 is prespecified at 0.2 (left) and 0.7 (right). Starting from different positions, both lines move towards the center over time. In the two bottom plots, we fixed the time points C50 (black dashed vertical line) and schwa50 (red dashed vertical line). They depict how the tongue dorsum position at a given time point changes with the tongue dorsum position at V50. The fan pattern is reflected here by a stronger relationship at C50 than at schwa50.

Figure 3.5 presents the full matrix of contour plots for all age cohorts (from left to right: C3, C4, C5, C7, and A) and consonants (from top to bottom: /b/, /d/, and /g/). Three main observations can be drawn from this matrix: First, the fan-like gradual shift from vowel-specific to overall more central positions over time described above for three-year-olds' /b/ context, is found in all cohorts' /b/ and /g/ contexts. However, the temporal development towards more central tongue dorsum positions happens faster for older participants than for younger ones: The pattern is compressed in adults' plots as compared to young children's, indicating earlier central positions and therefore shorter vocalic impacts. Second, this developmental trend is more prominent in /g/ compared to /b/ contexts. And third, the pattern in /d/ contexts differs drastically from /b/ and /g/ contexts': The pink hill in the plots indicates a forward movement of the tongue dorsum during the consonant. This is more salient in younger than in older participants. Our first model revealed that all of these carryover coarticulatory patterns are significantly different from zero ($p < .00017$).

Figure 3.5

Contour Plots Illustrating the Horizontal Movement of the Tongue Dorsum over Time



Note. The x-axis indicates the four time points V100, C50, C100 and schwa50, the y-axis represents the tongue dorsum's horizontal position during V50. Color gradients as defined in the upper right corner indicate anterior (pink) to posterior (blue) positions. Patterns are plotted separately for each age cohort and consonant context (/b/, /d/, /g/). In each plot, the bright vertical bands indicate the time points we collected data for on the basis of which the model estimated the shaded areas.

Because in Figure 3.5, the coarticulatory patterns of seven-, five-, and maybe four-year-olds appeared visually similar, we ran a model including binary difference tensors for age cohort comparisons to check whether we should group them. Using the seven-year-olds as reference, results did not reveal any differences between their coarticulatory patterns and those of age cohort C5. A difference was found, however, in comparison to cohort C4 in the /b/ and /g/ contexts as well as to cohorts C3 and A in all three consonant contexts. Cohorts C5 and C7 were therefore grouped for the subsequent analysis (C57 henceforth).

3.4.2 *Comparison of coarticulatory patterns across age cohorts*

To assess the statistical significance of the developmental differences in coarticulatory patterns as impressionistically displayed in Figure 3.5, we fit three binary difference models with varying reference groups. Table 3.3 - Table 3.5 present the outcomes of the models with the reference groups C3, C4, and A respectively. In every table, the first three lines refer to the coarticulatory pattern (i.e., the interaction between *time* and *tongue dorsum position at V50* [V50pos]) of the reference group for the three consonant contexts. Lines four to 12 present the differences between the reference and the indicated age cohort for a given consonant. In all output tables, the asterisks indicating significance levels adhere to the Bonferroni-corrected thresholds.

Table 3.3

Age Differences Within Consonant Context With Reference Group C3

Tensor product	edf	F-value	p-value	
(time, V50pos) : b	9.932	68.618	<0.00017	***
(time, V50pos) : d	23.804	23.693	<0.00017	***
(time, V50pos) : g	10.977	54.796	<0.00017	***
(time, V50pos) : C4 /b/	6.207	1.688	0.11224	
(time, V50pos) : C57 /b/	7.385	9.751	<0.00017	***
(time, V50pos) : A /b/	18.635	3.307	<0.00017	***
(time, V50pos) : C4 /d/	5.739	0.712	0.67056	
(time, V50pos) : C57 /d/	6.323	3.577	0.00058	**
(time, V50pos) : A /d/	18.654	2.341	0.00024	**
(time, V50pos) : C4 /g/	9.796	3.335	<0.00017	***
(time, V50pos) : C57 /g/	11.753	5.259	<0.00017	***
(time, V50pos) : A /g/	9.951	5.384	<0.00017	***

Note. The table displays the output of the binary difference smooth model testing for age differences within each consonant context with reference group C3. Significance codes '***': p < .00017; '**': p < .0017; '*': p < .008; '.': p < .017.

Table 3.4

Age Differences Within Consonant Context With Reference Group C4

Tensor product	edf	F-value	p-value	
(time, V50pos) : b	13.942	50.662	<0.00017	***
(time, V50pos) : d	24.782	22.063	<0.00017	***
(time, V50pos) : g	12.884	41.321	<0.00017	***
(time, V50pos) : C3 /b/	7.309	2.079	0.03133	
(time, V50pos) : C57 /b/	9.292	3.198	0.00029	**
(time, V50pos) : A /b/	12.142	6.163	<0.00017	***
(time, V50pos) : C3 /d/	6.340	1.663	0.11784	
(time, V50pos) : C57 /d/	9.439	1.882	0.03155	
(time, V50pos) : A /d/	12.329	3.580	<0.00017	***
(time, V50pos) : C3 /g/	11.907	2.427	0.00144	**
(time, V50pos) : C57 /g/	7.680	11.743	<0.00017	***
(time, V50pos) : A /g/	7.452	5.906	<0.00017	***

Note. The table displays the output of the binary difference smooth model testing for age differences within each consonant context with reference group C4. Significance codes '***': p < .00017; '**': p < .0017; '*': p < .008; '.': p < .017.

Table 3.5

Age Differences Within Consonant Context With Reference Group A

Tensor product	edf	F-value	p-value	
(time, V50pos) : b	21.156	44.753	<0.00017	***
(time, V50pos) : d	22.815	33.594	<0.00017	***
(time, V50pos) : g	13.446	57.036	<0.00017	***
(time, V50pos) : C3 /b/	9.939	2.830	0.00069	**
(time, V50pos) : C4 /b/	4.001	6.231	<0.00017	***
(time, V50pos) : C57 /b/	5.165	3.222	0.00416	*
(time, V50pos) : C3 /d/	6.447	4.356	<0.00017	***
(time, V50pos) : C4 /d/	8.582	2.638	0.00198	*
(time, V50pos) : C57 /d/	5.810	1.832	0.07487	
(time, V50pos) : C3 /g/	11.800	3.864	<0.00017	***
(time, V50pos) : C4 /g/	7.089	9.661	<0.00017	***
(time, V50pos) : C57 /g/	7.690	12.660	<0.00017	***

Note. The table displays the output of the binary difference smooth model testing for age differences within each consonant context with reference group A. Significance codes '***': p < .00017; '**': p < .0017; '*': p < .008; '.': p < .017.

Table 3.3 shows that three-year-olds differed from all other cohorts in the /g/ context, while they did not differ significantly from four-year-olds in the /b/ and /d/ contexts. Four-year-olds differed from adults in every consonant context but did not differ from the five- and seven-year-olds in the /d/ context (cf. Table 3.4). Table 3.5 finally completes the group comparisons by indicating that adults did not differ from cohort C57 in the /d/ context but did differ in /b/ and /g/ contexts. Figure 3.6 visualizes these age differences in two-dimensional plots. Similar to the black and red graphs in Figure 3.4, we fixed the independent variable *time point* for each plot (from top to bottom: V100, C50, C100, and schwa50). The horizontal tongue dorsum position at V50 is represented on the x-axis and its position at the indicated time point on the y-axis (both from zero = anterior positions to one = posterior positions). Each plot depicts every age cohort's results for one consonant context (from left to right: /b/, /d/, /g/).

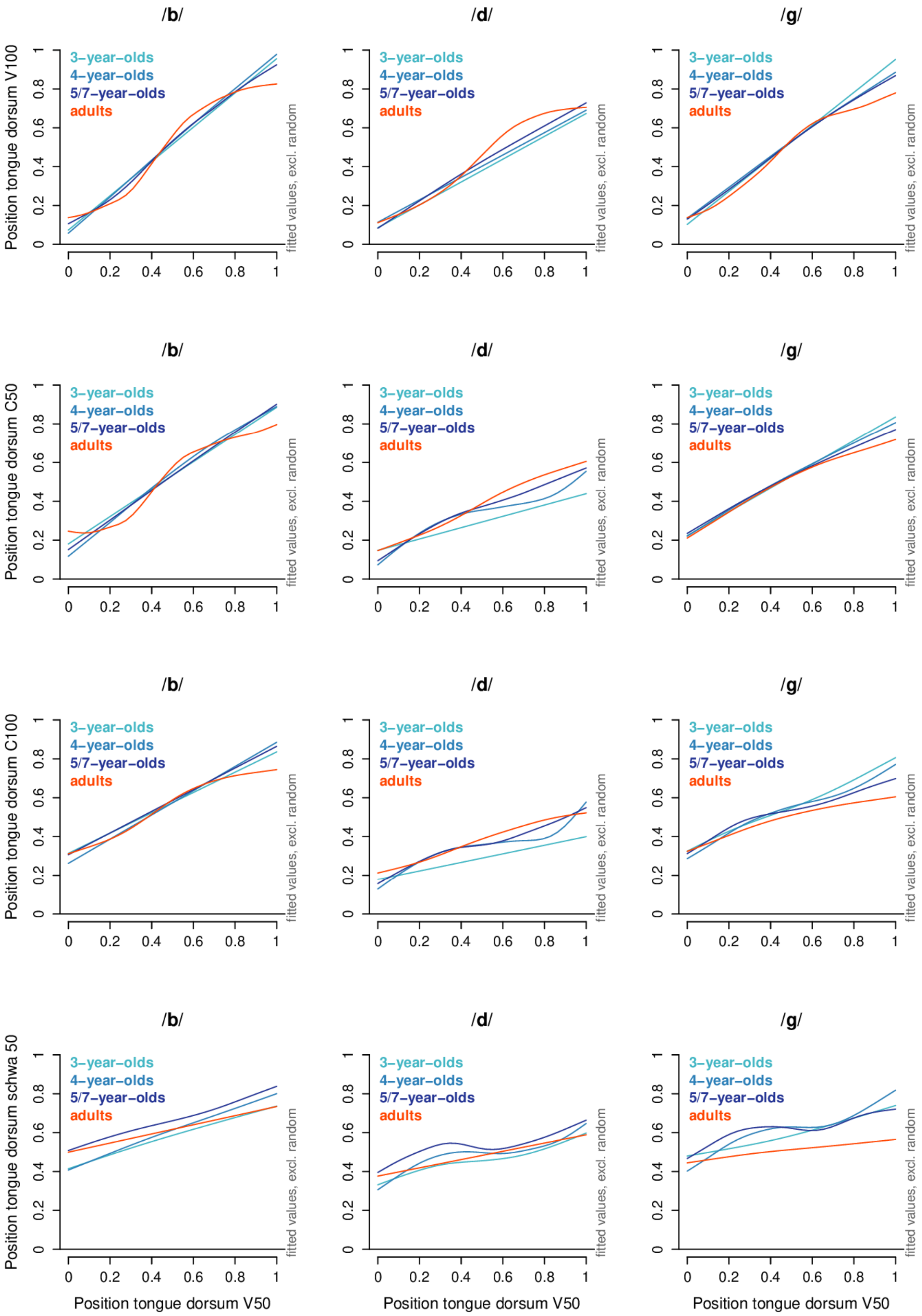
How to read the plots is demonstrated by means of the coarticulatory pattern in /g/ contexts at C100 (fourth plot in the right column). Let us focus on the light blue line that represents three-year-olds' coarticulatory patterns. For each horizontal tongue dorsum position at the midpoint of the vowel (x-axis), the corresponding horizontal tongue dorsum position at the endpoint of the following /g/ (C100) is plotted (y-axis). For vowels produced with relatively front tongue dorsum positions, for example 0.2, which could characterize the categories /i/ or /e/, the tongue dorsum is at about 0.4 at C100. For posterior vowels, for example positions of 0.8 (i.e., /o/ or /u/), the tongue dorsum is at about 0.6 at C100. The tongue dorsum position at C100 therefore depended on the tongue dorsum position at V50. The steeper a cohort's line in a specific consonant context is, the closer the tongue dorsum position during the investigated time point resembles that during the midpoint of the previous vowel, i.e., the higher is the coarticulation degree. The example plot (/g/ at C100) illustrates a higher coarticulation degree for younger than for older speakers, since lines flatten with increasing age.

GAMs allow us to detect linear as well as non-linear patterns in this relationship. Cohorts C4 and C57 in the same plot for example, display a slowly increasing line with higher slopes towards both ends of the V50 continuum. This implies that the tongue dorsum is in a central position (approximately 0.5) following vowels with all positions from 0.3 to 0.7; the correlation between V50 position and C100 position is therefore relatively weak here. When following vowels with extremely front or extremely back positions, however, the tongue dorsum position at time point C100 mirrors this direction resulting in higher correlations towards the edges. In adults' /b/ context at C50 on the other hand, the non-linear pattern is reversed, with flat edges and a strongly increasing part in the middle. This implies that the tongue dorsum moved towards more central positions following extremely front and back vowel positions, while remaining approximately at the vocalic position for V50 values of 0.3 to 0.6.

Figure 3.6 illustrates that the development of coarticulation degree is consonant specific. The least age differences in coarticulation degree were found in the /b/ context. Adults' lines representing /b/ contexts stand apart from those of the child cohorts at every time point in being flatter and less linear. It also becomes apparent that the statistically confirmed difference between the five- and seven-year-olds and the younger children mainly results from a difference in coarticulation degree at schwa50 where cohort C57 displays a lower slope. For the /g/ context, Figure 3.6 illustrates a growing differentiation between the cohorts across time points in the utterance: While at V100 and C50, adults only differed from children in slightly more central productions following back vowel positions; at C100 and schwa50 lines spread apart. Adults produced the schwa in /g/ contexts with a relatively central tongue dorsum position independent of the previous vowel, whereas children's tongue dorsum positions still resembled that of the preceding vowel. Finally, in /d/ contexts, tongue dorsum positions at C50 were relatively front, especially for young children, as already indicated by the pink hills in their contour plots (Figure 3.5), and coarticulation degree increased with age. Interestingly, however, children's patterns at schwa50 suggest a higher coarticulation degree than found in adults again. Similar to adults, their tongue positions were mostly central, but the observed non-linearity displays vowel-induced shifts towards more anterior and posterior positions respectively, that were not found in adults.

Figure 3.6

Relation Between the Tongue Position at V50 and at the Four Investigated Following Time Points per Age Cohort



Note. The figure displays the relation between the tongue dorsum position at V50 (x-axis) and that at each of the four investigated time points (y-axis, per row: V100, C50, C100, schwa50) per age cohort. Each consonant context is plotted separately.

3.4.3 Comparison of consonantal impact within age group

To assess consonant-induced differences in coarticulatory patterns within age cohorts, another binary difference smooth model was fit. Because Figure 3.5 suggested a high similarity between coarticulatory patterns in /b/ and /g/ contexts while /d/ contexts seemed to stand apart, the /b/ context was used as a reference. The output of the model comparing the interaction between time and the horizontal position of the tongue dorsum during V50 between consonant contexts within age cohorts is displayed in Table 3.6. While the first four lines represent the interaction between *time* and *tongue dorsum position at V50* in the /b/ context (reference level) for each cohort, lines five to eight provide information on the difference between /b/ and /d/, and lines nine to 12 between /b/ and /g/ contexts within each cohort.

For three-year-olds, there was no significant difference between the /b/ and /g/ context patterns. In all other cohorts however, both the coarticulatory pattern for /d/ and for /g/ contexts differed significantly from that in the /b/ context. Similar to Figure 3.6, Figure 3.7 visualizes these consonantal differences in two dimensions. Each plot depicts one age cohort's tongue dorsum positions in all three consonant contexts.

Figure 3.7 illustrates that /b/ and /g/ did not differ in their coarticulatory pattern for three-year-olds as was indicated by the model. Here, we get the additional information, that the coarticulation degree of these consonants was higher than that of /d/. In addition to this difference, older cohorts' /b/ contexts allowed an even higher coarticulation degree than /g/ contexts. For adults however, the consonantal differences do not seem as pronounced as for children. In each cohort, V100 was characterized by a high vowel dependency in each consonant context; the consonant-related differences of the vowel's coarticulation degree were strongest at C50 and C100 and decreased again at schwa50.

Table 3.6

Consonantal Differences Within Age Cohorts

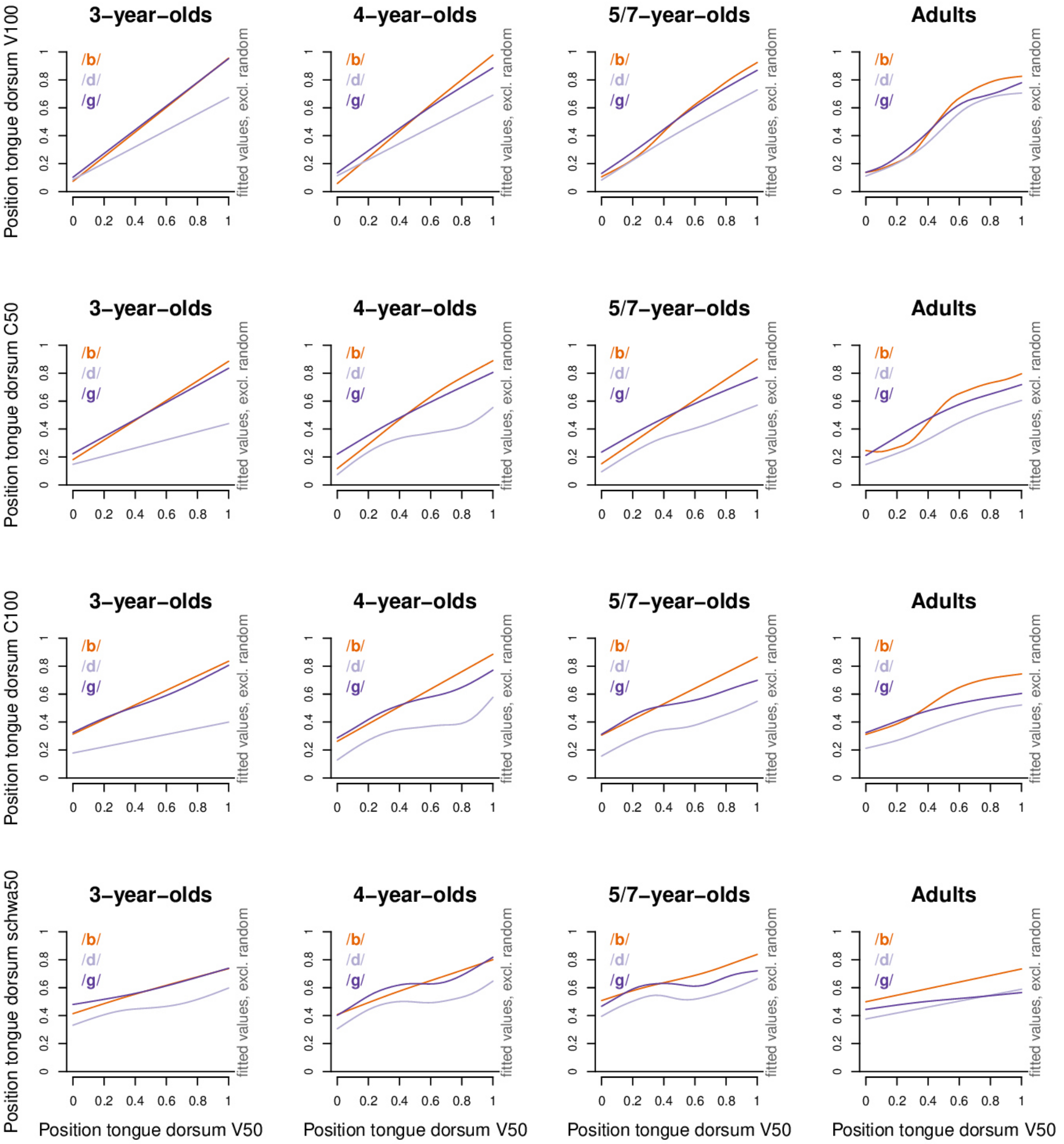
Tensor product	edf	F-value	p-value	
(time, V50pos) : C3	3.001	243.207	<0.00017	***
(time, V50pos) : C4	18.465	34.896	<0.00017	***
(time, V50pos) : C57	23.248	75.983	<0.00017	***
(time, V50pos) : A	19.750	47.268	<0.00017	***
(time, V50pos) : C3 /d/	7.830	32.389	<0.00017	***
(time, V50pos) : C4 /d/	8.876	13.653	<0.00017	***
(time, V50pos) : C57 /d/	7.826	26.629	<0.00017	***
(time, V50pos) : A /d/	8.253	9.348	<0.00017	***
(time, V50pos) : C3 /g/	7.783	1.118	0.34066	
(time, V50pos) : C4 /g/	6.886	3.655	0.00042	**
(time, V50pos) : C57 /g/	7.397	9.584	<0.00017	***
(time, V50pos) : A /g/	6.196	5.647	<0.00017	***

Note. The table displays the output of the model testing for consonantal differences within age cohort. The reference consonant is /b/. Significance codes '***': p < .00017; '**': p < .0017; '*': p < .008; '.': p < .017.

To summarize our main results, consonant-context-dependent differences in vocalic carryover coarticulation were found in every age cohort. Except for three-year-olds' /b/ and /g/ patterns, the three consonant contexts differed significantly for every age cohort. The across-cohort analysis testing for developmental differences revealed developmental trajectories of coarticulation degree to be consonant-specific, with a clear decrease in coarticulation degree with age for /g/ contexts, a slight decrease for /b/ contexts, and a special pattern for /d/ contexts indicating less coarticulation for children than adults in the domain of the consonant but slightly more coarticulation for younger participants again during the final schwa.

Figure 3.7

Relation Between the Tongue Position at V50 and at the Four Investigated Following Time Points per Consonant



Note. The figure displays the relation between the tongue dorsum position at V50 (x-axis) and that at each of the four investigated time points (y-axis, per row: V100, C50, C100, schwa50) per consonant context. Each age cohort is plotted separately.

3.5 Discussion

The present study is the first one using kinematic data to assess children's lingual carryover patterns as well as the first one ever investigating carryover coarticulation in German children. Because previous analyses within the same group of participants suggested a strong decrease of vocalic anticipation with increasing age, we asked whether vowel-related movements would also overlap more with following gestures in children's than in adults' speech. In a cross-sectional design we recorded speech movements via ultrasound tongue imaging to follow the development of coarticulatory processes in children from as young as three years of age until adulthood. The results confirm a developmental decrease in carryover coarticulation degree and therefore support our hypothesis of broader vocalic activation not only in the left (Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2018) but also in the right field of the utterance. The articulatory demands of the consonantal context were shown to impact the coarticulatory pattern within as well as across cohorts. The following sections discuss origins and implications of 1) the general developmental decrease of coarticulation, 2) the consonantal impact within age cohorts, and 3) the consonant-dependent developments across cohorts.

3.5.1 Carryover coarticulation decreases with age

The present study uncovered a developmental decrease of carryover coarticulation of stressed vowels in VCə sequences. In utterances with the consonants /b/ and /g/, this decrease was evident through the gradual shift from vowel-specific tongue dorsum positions towards a central position being significantly slower for young children than for older children and especially for adults (cf. Figure 3.5). Our interpretation within the coproduction framework (C. A. Fowler, 1980) is that the overlap of vocalic activation with following gestures is larger in younger than in older speakers. Although utterances containing the alveolar stop /d/ displayed a pattern differing tremendously from that of /b/ and /g/ contexts, we found evidence for a developmental decrease of vocalic activation here as well: While during the domain of the consonant the vocalic impact was lower in children than in adults, children produced the final schwa with tongue dorsum positions resembling more those of the vowel than adults'. The special coarticulatory pattern noted in alveolar contexts and the resulting discontinuous coarticulatory effects will be discussed in chapters 5.2.1 and 5.2.2.

The finding that not only young children's anticipatory but also their carryover coarticulation of stressed vowels is stronger than older participants', supports the hypothesis of a compression of activation curves with age that was implied by Nittrouer (1993) and revisited here. While our illustration of the hypothesized activation curves in Figure 3.1 is just a simplified sketch not integrating articulatory demands of the combined gestures, it depicts developmental differences in speech production strategies coherent with those identified in the present study. The symmetry of developments in both coarticulatory directions suggests anticipation and perseveration of a stressed vowel's articulatory gestures to have a common origin in the way gestures are phased to each other and overlap with their neighbors instead of being the result of two distinct processes.

A possible origin of broader vocalic activation in child than in adult speech may be related to the finding that a well-balanced degree of inhibition of temporarily irrelevant information in various cognitive domains only matures in the course of childhood (e.g., Bjorklund & Harnishfeger, 1990). Elements that are for some reason prominent or hyperactive would therefore be harder to inhibit for children than for adults. In chapter 3.2.3, various arguments for children to ascribe stressed vowels a special status in perception as well as production were summarized. C. A. Fowler (1980) proposed that exactly these stressed vowels serve as the basis for the organization of gestural activation and phasing. It therefore seems likely that a hyperactivation of vowels could in turn lead to broader gestural activation and execution, as for example suggested in Tilsen's (2016) selection-coordination theory of speech production. Accordingly, children's difficulty with inhibiting especially prominent parts of the speech plan would result in higher activation of vowels at a selection stage before execution leading to earlier initiation of the vocalic gesture and therefore to a higher degree of anticipatory coarticulation. It would, however, also delay the de-selection of the gesture after the vowel target was reached leading to broader overlap with following gestures, i.e., to a higher degree of carryover coarticulation. According to Tilsen (2018), the development of the speech production system across childhood could be driven by a change of the kind of feedback accessible to speakers as is known from other motor processes (e.g., Butz, Sigaud, & Gérard, 2003). While children would first rely on relatively slow external feedback from peripheral sensory organs, with experience, the faster internal feedback that works with predictions of sensory consequences of outgoing motor commands becomes accessible. It seems possible that the change in the use of feedback with increasing language experience and the maturation of inhibitory processes are closely related. Tilsen (2018) explicitly relates findings of a higher degree of anticipatory vowel-to-consonant coarticulation to an immaturity of the coordination of gestures. More precisely, he ascribes the CV hyper-coarticulation in child speech to an "asymmetry such that closure is more strongly coupled to V than release" (p. 33). At least in this form his reasoning does not provide a direct explanation either for children's greater degree of anticipatory long-distance coarticulation (Rubertus & Noiray, 2018), nor for the greater degree of carryover coarticulation found in the present study. We hope our findings stimulate a revival of empirical interest in carryover coarticulation to feed speech production (development) models.

While the maturation of inhibitory control possibly triggered by a change in the use of feedback can account for the anticipation and the perseveration of a vocalic gesture as well as their development across childhood, it would certainly be premature to exclude other scenarios. The decrease of coarticulation in the two directions might only resemble each other on the surface while being driven by different maturational processes: Can Flege's (1988) conclusion that inertial properties of the speech motor system do not change with age for example be replicated with articulatory measures and be transferred from velar to lingual motion? Only additional systematic investigations of carryover coarticulation in children's speech, for example via different speech rate conditions, as well as a direct comparison of anticipatory and carryover coarticulation can enlighten our understanding of the speech production mechanism and its development.

In order to investigate coarticulation development across childhood, the present study focused on age differences only, for reasons of simplicity. We would like to emphasize, however, that age itself should not be mistaken as the driving force for changes in spoken language. Instead, age is a mediating variable reflecting various cognitive and motor developments not illuminated here. In a recent study, it was for instance found that children with greater knowledge of the phonological structure of their language show more mature coarticulatory patterns than children with poor phonological awareness (Noiray, Popescu, et al., 2019). In addition, we have started examining possible influences of literacy acquisition on speech organization.

3.5.2 The impact of the consonant within cohorts

While we did find evidence for generally broader vocalic activation in children than in adults in all three consonant contexts, our results suggest that the consonant impacts on the coarticulatory trajectories within each cohort as well as on the development of vocalic carryover coarticulation across cohorts. Within each age cohort, there is a clear distinction of vowel-dependent tongue dorsum movement patterns as a function of consonants' degree of coarticulatory resistance. Utterances with low resistant consonants /b/ and /g/ are characterized by a gradual shift from vowel-specific towards central tongue dorsum positions while tongue dorsum movements in utterances with the high resistant consonant /d/ are clearly discontinuous, moving front in the domain of the consonant to then resume a rather central position during schwa. Except for the three-year-olds, a significant difference in coarticulation degree was found between the two low resistant consonants /b/ and /g/ with the bilabial allowing for more vocalic perseveration than the velar.

This set of findings is in line with C. A. Fowler and Brancazio's (2000) hypothesis of temporary resistance-dependent consonantal clamps of the tongue dorsum during continuous vowel productions. Because the tongue dorsum is not recruited for the bilabial plosive, it can follow its trajectory from vowel specific towards central positions without being disturbed while the lips form the closure for the consonant. Here, there is therefore no need for the tongue dorsum to reach a rather central position soon after the vocalic target because the bilabial plosive is intelligible independent of the tongue position. On the contrary, the velar plosive /g/ shares the primary articulator with the vowels. However, due to its low resistance, vocalic and consonantal movements are blended, resulting in vowel-dependent locations of the palatal constriction. Consequently, there is a gradual shift from the preceding vowel towards the center resembling that of the bilabial but reaching a central position earlier because of the relatively central position of the necessary palatal closure. Last, the highly resistant plosive /d/ needs a front movement of the tongue dorsum to support the tongue tip in forming the alveolar constriction which results in a strong temporal clamping of the tongue dorsum.

3.5.3 The impact of articulatory demands on the development of coarticulation

In terms of development across cohorts, we found that the decrease of coarticulation degree is strongest in /g/ contexts. The contour plots in Figure 3.5 suggest that the point of palatal contact for /g/ is more variable in younger than in older speakers. While tongue dorsum positions during the consonant's mid and end point are distributed widely from front to back positions in the young cohorts, the older the speakers are, the more central the tongue is during /g/. Again, this speaks for a strong blending of vocalic and consonantal gestures in young children resulting in very vowel-dependent points of palatal contact and therefore more variability in the constriction location of /g/. Older speakers on the other hand, display less vowel-dependency of the consonant which we interpret as evidence for less coproduction due to compressed vocalic activation curves compared to younger speakers. For utterances including the bilabial /b/, the decrease in coarticulation degree is significant as well although the difference between the youngest and the oldest cohort is not as strong as in the /g/ context. Presumably, this is because the strength of coproduction of vocalic and consonantal gestures does not result in changes of a point of lingual contact but only determines how long the vowel 'fades out.'

The developmental results from the alveolar context shed light on an aspect of the maturation of speech motor control across the age cohorts tested. This finding supplements existing research suggesting that the development of speech motor control is protracted and continues until adolescence at least in some aspects (Noiray et al., 2010, 2013; Smith & Zelaznik, 2004). To produce an alveolar stop, the tongue tip forms a constriction at the alveolar ridge. The strong forwards movement of the tongue dorsum that we see during the production of /d/ in young children but way less in adults (cf. Figure 3.5) indicates that children do not only move the tongue tip forward but the tongue dorsum as well. We interpret this as evidence for a tighter coupling or lack of independence between the tongue tip and the tongue dorsum in young children. Interestingly, results in Noiray, Wieling, et al. (2019, p. 3043) had indicated a more anterior tongue dorsum position for adults than for children during the production of /d/ as C_1 instead of C_2. A closer look at the data revealed that children's tongue dorsum positions during the production of the alveolar stop are approximately the same in C_1 and C_2 while adults' positions differ tremendously. This pattern is predicted by Tilsen's (2018) hypothesis of immature coordinative control: In C_1 = d contexts on the one hand, the vowel is coupled to the consonant's release for adults but to the consonant's closure for children. Vocalic and consonantal demands would therefore be blended more strongly in young participants, while adults' vocal tract is dominated by the alveolar constriction gesture. In C_2 = d contexts on the other hand, the vocal tract is in shape for the vowel when the consonantal gesture is initiated. Vocalic and consonantal gestures would therefore blend immediately from the beginning of C_2 in all cohorts. Regardless of the reasons behind the adults' pattern however, this observation shows that an independence of the tongue tip from the tongue dorsum creates the possibility for independent movements on the one hand and articulatory synergies on the other (cf. Noiray et al., 2013). Importantly, children in the younger cohorts have not mastered the independent functional control of tongue tip and tongue dorsum yet (cf. Nittrouer et al., 1996), but with increasing age and speech motor experience the independence of the two articulatory organs increases, approximating an adult-like pattern in five- and seven-year-olds. How these developmental changes in articulatory independence and synergies can be simulated in speech production models like the task dynamic model of inter-articulator speech coordination (TaDA) is the focus of another project currently run in our laboratory. While the tighter coupling of tongue tip and tongue dorsum prevents a higher activation of children's vocalic gestures from being measurable during the consonantal domain, it becomes apparent again during the schwa. Assuming C. A. Fowler and Brancazio's (2000) notion of consonants clamping the tongue dorsum temporarily during continuous vowel productions, this suggests that children's vocalic gestures are active until the final schwa but temporarily hidden by the consonantal requirements, while adults' vowel activation seems to decrease to a minimum before schwa production.

3.6 Conclusion

The present study provides first empirical evidence for vocalic carryover coarticulation to decrease with increasing age and therefore to develop similarly as anticipatory coarticulation. Although for now we cannot rule out other possible scenarios, this finding does not give rise to a discrepancy of underlying mechanisms between the two coarticulatory directions. Instead, we interpret our results as suggesting one common mechanism underlying anticipatory and carryover coarticulation: the coproduction of simultaneously active speech gestures that decreases across childhood in both directions because of a maturation of inhibitory control mechanisms responsible for accurate selection and de-selection of gestures. In addition to the width of gestural activation, our results support the notion that the degree of vocalic carryover coarticulation depends on the compatibility of articulatory demands of active speech gestures. Because of speech motor control maturation during childhood, this dependency is another source for developmental differences in coarticulatory patterns.

4 THE PROTRACTED DEVELOPMENT OF PHONEMIC BLENDING FLUENCY IS REFLECTED IN COARTICULATORY PATTERNS: EVIDENCE FROM BEGINNING AND PROFICIENT READERS[5]

4.1 Abstract

Learning to read is a crucial milestone in children's development that has a lasting impact on socio-economic integration and professional success. In transparent orthographies bearing consistent relations between graphemes and phonemes, beginning readers rely on sequential grapheme-to-phoneme decoding and blending. This leads to early accuracy but protracts word reading fluency. For the first time, the present study addresses this lack of fluency with kinematic measurements of anticipatory coarticulation, an essential feature of speech fluency. We employ state-of-the-art ultrasound imaging to track German beginning and proficient readers' tongue movements when reading aloud or repeating pseudowords. Using generalized additive mixed models, we show that unlike adults, children initiate lingual gestures for target vowels substantially later in the Reading compared to the Repetition condition. Their extent of anticipatory coarticulation is therefore lower in the former than in the latter condition. The present study provides first evidence that after one to two years of formal reading training, children lack phonemic blending fluency when reading aloud. Furthermore, it shows that with direct articulatory measures we can detect fine-grained differences in phonemic blending not revealed by standard measures of reading fluency. Implications for developmental changes in phonological representations and for the course of reading acquisition are discussed.

4.2 Introduction

Spoken language enables us to efficiently communicate with others. While it is ultimately composed of a finite set of speech sounds and their underlying articulatory speech gestures, the conveyed stream of information is fairly fast, continuous, and dynamic because the gestures are coproduced resulting in overlapping sounds (i.e., coarticulation) (C. A. Fowler, 1980; C. A. Fowler & Saltzman, 1993; Gafos & Goldstein, 2012). Most preliterate children are not yet aware that continuous speech is composed of a finite set of units corresponding to phonemes (Goswami & Bryant, 2016). When exposed to an alphabetic orthography, however, they see distinct letters black on white, one neatly following the other. The developmental path from print to fluent oral reading that requires children to learn to convert a sequence of letters into a

---

[5] Chapter 4 of this dissertation is currently under peer review:
Rubertus, E., Popescu, A., & Noiray, A. The protracted development of phonemic blending fluency is reflected in coarticulatory patterns: Evidence from beginning and proficient readers. *Developmental Science.* Adaptations were made regarding citation style, figure embedding, and cross references.

continuous speech stream is protracted. While standard assessments of reading proficiency infer fluency based on measures of accuracy and reading speed, the present study relies on direct kinematic measures of tongue motion to zoom in on phonemic decoding fluency - the identification and blending of the sounds corresponding to the letters of a word (Hudson, Pullen, Lane, & Torgesen, 2008). Phonemic decoding fluency is an important predictor of reading comprehension in beginning readers and therefore sets the basis for successful reading acquisition (Caravolas et al., 2019; Martens, Werder, Hier, & Koenig, 2013). Using measures of lingual coarticulation, we show that while German second- and third graders succeed in reading short utterances accurately, the single sounds are not blended as fluently as when the utterances are repeated after a model speaker. Ultimately, however, proficient (adult) readers do not exhibit differences in coarticulation based on whether they read aloud or repeat. The present study therefore provides first evidence for fine-grained fluency differences in the typical acquisition of reading that are not detected via standard measures like accuracy and speed to be reflected in measures of coarticulation.

4.2.1 The virus of literacy

By the time children learn to read, they have had experience with spoken language for about five to seven years. While most children develop spoken language fluency automatically with daily audio-visual input, learning to read requires instruction and extended practice (Boyer & Ehri, 2011; A. M. Liberman, 1989, 1991). The acquisition of literacy causes drastic changes in a person's brain (Dehaene, Cohen, Morais, & Kolinsky, 2015) and impacts speech production and perception in several ways: It affects how neighboring articulatory gestures are coordinated in the continuous speech stream (Popescu & Noiray, 2021). Relatedly, the presentation of the orthography in addition to acoustics leads to an increased phonetic production accuracy and governs participants' perception of ambiguous speech sounds (Bonte, Correia, Keetels, Vroomen, & Formisano, 2017; Saletta, 2019; Saletta, Goffman, & Brentari, 2016). Using transcranial magnetic stimulation, Pattamadilok, Knierim, Kawabata Duncan, and Devlin (2010) demonstrated that literacy acquisition changes existing phono-articulatory representations, instead of "only" coactivating phonological with newly built orthographic representations (Frith, 1998; J. C. Ziegler, Ferrand, & Montant, 2004). The idea that existing representations are specified by literacy acquisition, is also supported by the finding that categorical precision in phoneme perception is positively correlated with level of literacy, regardless of age (Kolinsky et al., 2021). Our orthographic knowledge may even mislead our perception of fluent speech: Brügelmann (1992) speaks of a filter of orthography through which literate adults perceive speech -

we “believe to hear what we actually see [Wir meinen zu hören, was wir eigentlich sehen”] (p. 80). Similarly, Frith (1998) compares the acquisition of an alphabetic orthography to a virus: “This virus infects all speech processing, as now whole word sounds are automatically broken up into sound constituents” (p. 1011).

4.2.2 From discrete letters to continuous speech

With the term “whole word sounds” Frith (1998) refers to coarticulated speech in which sounds overlap with each other because they are coproduced by smoothly overlapping articulatory gestures (e.g., concurrent tongue fronting for /i/ and lip closing movement for /b/ in the syllable /bi/; C. A. Fowler, 1980; C. A. Fowler & Saltzman, 1993; Gafos & Goldstein, 2012). This coarticulation is a fundamental feature of speech, the engine that enables efficient and fluent spoken communication. Over time, children naturally integrate the coarticulatory patterns of their native language into their speech (Noiray, Wieling, et al., 2019). When exposed to written words in an alphabetic orthography however, children see strings of single letters - a visual representation of language misleadingly highlighting discreteness and seriality. Fluent word reading requires learning to transpose these segments into a continuous stream of coarticulated speech.

The rules of this transposition vary across orthographies and hence the specific writing system a child is exposed to impacts the process of reading acquisition (Caravolas et al., 2019; J. C. Ziegler & Goswami, 2005). Alphabetic orthographies differ in the degree of transparency between graphemes and phonemes. Reading opaque orthographies (e.g., English, French) that bear inconsistent grapheme-phoneme relations requires processing units larger than the letter. In one of the most influential models of reading acquisition for English (Gathercole & Baddeley, 2014), Frith (1986) therefore, describes a relatively short phase of indirect access via alphabetic (letter-by-letter) decoding that is preceded by a direct logographic phase and followed by an orthographic phase in which whole words are directly accessed via the lexicon. In transparent orthographies instead, single graphemes reliably correspond to single phonemes (e.g., Finnish, Spanish). The present study investigates phonemic decoding and blending fluency in German, an orthography with transparent grapheme-phoneme relations (J. C. Ziegler, Perry, Jacobs, & Braun, 2001). In her model of reading acquisition for German, Scheerer-Neumann (1989, 2007, 2018) highlights that letter-by-letter decoding plays a more important role than it does in English. She points out that within the second year of formal literacy instruction, children learning to read German start using units larger than graphemes (i.e., syllables and morphemes). Crucially, however, she describes that (pseudo)words in this orthographic phase are, in contrast to English, still indirectly assembled via active decoding. Once beginning readers of German develop awareness about graphemes being placeholders for sounds and therefore being a way of representing the spoken language they know (i.e., the alphabetic principle), they quickly decode words accurately (Ehri, 1995; Landerl, Wimmer, & Frith, 1997; J. C. Ziegler, Perry, & Coltheart, 2000). What they struggle more with, is reading speed and fluency (Aro & Wimmer, 2003; Moll, Wallner, & Landerl, 2012), in other words, combining and blending the sounds that they decipher from left to right (Frith, 1986) to a fluently coarticulated speech stream.

While reading proficiency has traditionally been assessed via measures of speed and accuracy (Aro & Wimmer, 2003; Moll et al., 2012), the literature about the protracted reading fluency development in transparent orthographies lacks direct measures of articulatory fluency. Although speed and coarticulatory extent are closely connected, the two do not necessarily go hand in hand (see chapter 4.5). Our study aims to fill this gap by investigating the extent of children's and adults' anticipatory coarticulation in read aloud compared to repeated speech. In children, we focus on the transition between the alphabetic and the orthographic phase when they are expected to read mostly fluent as assessed with standard measures. To achieve this goal, we used the technique of ultrasound imaging that in the last decade has been proven successful to measure articulatory speech fluency in both adults and children. These measures offer the chance of detecting fine reading disfluencies otherwise easily overlooked and hence provide new insights about reading development as well as about factors impacting on fluent speech production.

4.2.3 Coarticulation as an index of decoding fluency

Measures of coarticulation have been widely used to track the typical development of fluent speech production across childhood (e.g., Canadian French: Barbier et al., 2020; American English: Cychosz, Munson, & Edwards, 2021; German: Noiray et al., 2018; Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2020; Scottish English: Zharkova et al., 2011) and to assess disfluencies in atypically developing children (e.g., apraxia of speech: Nijland et al., 2002; stuttering: Chang, Ohde, & Conture, 2002; speech sound disorders: Maas & Mailend, 2017).

When learning to read aloud in a language with a transparent orthography, even typically developing children are initially disfluent (Aro & Wimmer, 2003; Moll et al., 2012). As outlined above, oral word reading initially builds on sequential identification of sounds corresponding to letters and blending them (i.e., coarticulating), while fluent speech is characterized by dynamic patterns of segmental overlap. In that context, language-specific coarticulatory patterns in oral reading can only be produced when decoding is quick and efficient enough to anticipate upcoming segments (Kawamoto, Liu, & Kello, 2015; Rastle, Harrington, Coltheart, & Palethorpe, 2000; Wimmer, 1996). Early-stage readers, however, stumble through a word without necessarily capturing it as one larger unit. Measures of coarticulation i.e., of the amount and/or temporal extent of overlap between articulatory speech gestures, thus far limited to evaluations of spoken language fluency, may therefore provide a powerful method to assess the development of fluency during reading acquisition.

Coarticulation can be quantified either from acoustic analyses of formant frequencies or from measures of articulatory movements. In both acoustic and articulatory approaches, coarticulation amount is often measured as a regression between one segment, interpreted to undergo coarticulation and another segment interpreted to induce coarticulation, for example V-to-C coarticulation in CV syllables (acoustics: T. Gibson & Ohde, 2007; Nittrouer et al., 1996; Noiray et al., 2013, articulation: Butcher & Weiher, 1976; Noiray et al., 2018, 2013; Zharkova et al., 2011). When viewing coarticulation as coproduction of segments (C. A. Fowler, 1980; C. A. Fowler & Saltzman, 1993; Gafos & Goldstein, 2012) rather than the influence of one segment on the other, however, it is sensible to address the dynamic nature of speech looking at change over time. In the present study, we therefore opt for a tight temporal sampling of tongue position measures to provide a more accurate estimate of tongue dynamics. We examine vowel-dependent tongue trajectories in pseudowords including either the target front vowel /i/ or the back vowel /u/, representing the two endpoints of the front/back-dimension, in stressed position to detect the onset of vocalic anticipation in read aloud versus repeated speech (cf. Goffman et al., 2008).

4.2.4 Research questions and predictions

This study tests the hypothesis that the lack of word reading fluency reported for early-stage readers of German is reflected in their articulatory anticipation of vowels, as measured in tongue trajectories. To test this hypothesis, we ask the following two research questions:

1) Do early-stage readers anticipate vowels earlier in repeated than in read aloud speech? We predict that they do. Given their limited grapheme-to-phoneme decoding practice, German second- and third graders should initiate vocalic gestures earlier within utterances when asked to repeat pseudowords after a model speaker than when instructed to read them aloud.

2) Does proficient readers' vocalic anticipation differ between repeated speech and oral reading?

We suggest that it does not. Since decoding and blending efficiency is expected to speed up tremendously with experience, there should be no fluency difference between repeated and read aloud utterances for proficient adult readers.

4.3 Materials and Methods

4.3.1 Participants

32 children (16 female, 16 male) aged 7;03 (Y;MM) to 9;05 years (mean: 8;03), enrolled in the second (n = 13) or third (n = 19) grade of primary school, participated in the study. All children were tested within two months of the first half of the school year (mid-October until mid-December 2019) to have comparable duration of formal reading instruction within the cohorts. For the present analysis, second- and third graders were pooled together because they did not significantly differ in nonword reading proficiency as measured via a standard reading test (SLRT, see below), neither based on grade (Welch t-test: t(22.14) = -0.53, p = .6) nor on age (Pearson's correlation: r(30) = .11, p = .56). All children were raised in monolingual German homes, however, in school they had some input of English or French. An additional 11 children were recorded but were not included in the present analysis due to either technical issues during the recording (n=3), or because they read words silently before they read them out loud (n=7), or because they were not feeling well during the recording (n=1). In addition, 16 native German adults (9 female, 6 male, 1 other) between 19 and 55 years of age (mean: 29) were recorded. Due to technical issues during the recording, data from two further adults could not be used. No participant reported any hearing-related problem, and vision was normal or corrected. All participants (were) reported to have (had) a typical language acquisition process and had not received speech and/or language therapy in the two years before testing. The children and their parents as well as the adult participants gave written informed consent for their participation.

4.3.2 Stimulus material

Stimuli were disyllabic pseudowords consisting of the consonants /b, d, g/ and the vowels /i, u/ in the form consonant$_1$-vowel-consonant$_2$-schwa (C_1VC_2ə) where C_2 always differed from C_1. To examine vowel-induced differences in horizontal tongue trajectories across an utterance, the studied vowels represent the extreme ends of the horizontal dimension (i.e., the most front (/i/) and back (/u/) high vowels) of the German vowel space. Additional stimuli with the vowels /e, a, o/ were recorded but are not analyzed here. The fully crossed set of C_1VC_2 combinations resulted in 30 different target pseudowords, 12 of which are used here. Each pseudoword followed the German female indefinite article /aɪnə/ building short utterances such as /aɪnə biːdə/.

Stimuli were presented in two elicitation conditions: Reading and Repetition. In the Reading condition, they appeared on a screen in written form. Spelling of the article corresponded to German orthography (<eine>). Pseudowords were spelled with transparent orthography, not including any lengthening graphemes. Target words therefore always consisted of 4 letters corresponding to the 4 sounds, e.g., <bide>, <duge>, <gibe>, etc. No capital letters were used. For the Repetition condition, previously recorded utterances of a native German female adult were played acoustically. Pseudowords had a trochaic stress pattern as is most common for disyllables in German.

The Reading condition always preceded the Repetition condition. Per condition, there were 2 blocks, each including all 30 target words in random order. In total, participants were presented with 120 stimuli, 60 visually (Reading) and 60 acoustically (Repetition).

4.3.3 Experimental procedure

Participants were recorded with SOLLAR (Sonographic and Optical Linguo-Labial Articulation Recording system; Noiray et al., 2020), a child-friendly platform for recording acoustic and kinematic data, which we adapted for the purpose of presenting written stimuli by connecting an additional screen. This platform enables simultaneous recordings of tongue motion via ultrasound imaging (Sonosite Edge, sampling rate: 48Hz), head and labial movement via video recording (Logitech Webcam), and the acoustic signal via a microphone (Sennheiser, sampling rate: 48kHz). The ultrasound probe was fixed in a custom-made probe holder providing vertical flexibility via springs while being stable in the lateral and horizontal dimensions. Participants sat on a chair in front of the recording table and were instructed to put their lower jaw on the probe to record the tongue surface contour in the midsagittal plane.

The stimuli were presented using OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) which allows for both acoustic and visual output, a within-block randomized order of trials, as well as a flexible inter-stimulus-interval determined by the participant's speed. For children, the production task was embedded in a playful "junior scientist" scenario including obligatory distractor tasks as well as optional pauses to keep them engaged throughout the recording. This way, all invited participants completed all blocks.

After the recording, we assessed participants' reading proficiency using a subpart of the Salzburger Lese- und Rechtschreibtest II (SLRT II; Moll & Landerl, 2010) that asks participants to correctly read out loud as many nonwords from a list as possible. The proficiency score corresponds to the number of correctly read items within one minute, hence combining accuracy and speed. The list was the same for both children and adults.

4.3.4 Data processing

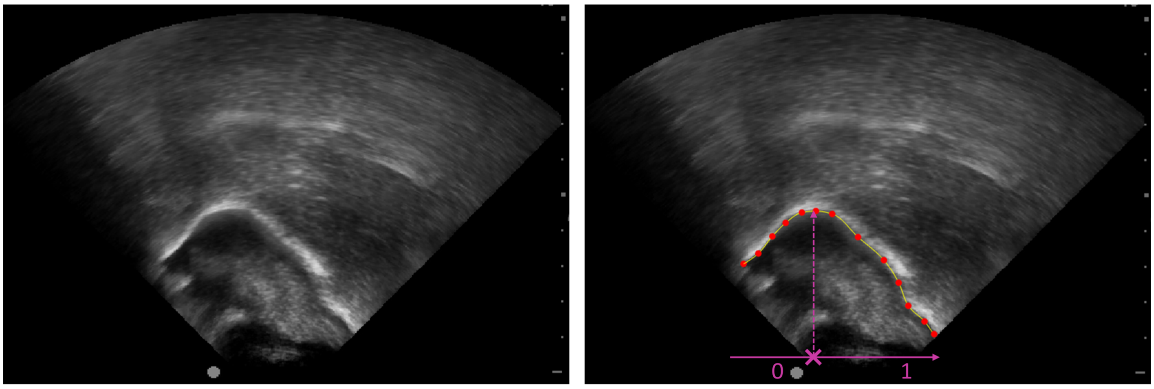
First, the acoustic signal was phonetically labelled using Praat (Boersma & Weenink, 2016). To identify all segment boundaries within a trial, a stable periodic cycle in the oscillogram as well as a stable formant pattern (especially a clearly detectable second formant (F2)) were used as indices for vowels. The first ascending zero-crossing in the oscillogram at the beginning of the periodicity was accordingly set as vocalic onset, the first ascending zero-crossing after the end of periodicity and disappearance of F2 as the beginning of the following consonant. Only trials without errors or hesitations are included in the present analysis. For this, two phonetically trained referees, minimally one of whom was a German native speaker, assessed each trial for audible hesitations as well as errors like metatheses, substitutions, insertions, and omissions.

The acoustically defined time stamps were used to identify corresponding frames in the ultrasound signal. For the present analysis, each of the segments preceding the vowel (/aɪ/, /n/, /ə/, and /C_1/) were assigned five keyframes: at the start of the segment, and at 20, 40, 60, and 80% of the temporal length of the segment. Within the vowel, the frames at the onset, at 20, 40, and 60% of the temporal length were considered. The tongue contours in the ultrasound signal were automatically detected by SLURP (Laporte & Ménard, 2018) and manually corrected when needed (Figure 4.1). The horizontal position of the highest point of the tongue dorsum surface contour (indicated in pink in Figure 4.1) was used as a representation of frontness of the tongue body. Frames displaying less than approximately 70% of the surface contour, as well as those for which the highest point of the tongue contour was not identifiable, were discarded. To prevent including tongue contours in which the highest point of the tongue is on the tongue tip instead of the tongue dorsum, frames with low horizontal values were visually inspected. These criteria along with the exclusion of erroneous trials as well as those with hesitations overall led to the exclusion of 28.8% of children's frames (39.9% of Reading frames, 17.7% of Repetition frames) and 13.73% of adults' frames (4.9% of Reading frames, 22.6 % of Repetition frames).

To ensure comparability across participants, the horizontal tongue dorsum positions were normalized. For this purpose, we set each speaker's most anterior tongue dorsum position at vowel midpoint to zero and his/her most posterior tongue dorsum position at vowel midpoint to one and scaled all his/her tongue dorsum positions in relation to this range.

Figure 4.1

Ultrasound Tongue Contour Detection



Note. Tongue contour detection exemplified on a third grader's /i/ at 40% within the temporal length of the segment. Left: raw ultrasound image, right: labelled tongue contour with indication of frontness measure. The front of the tongue is towards the left of the images.

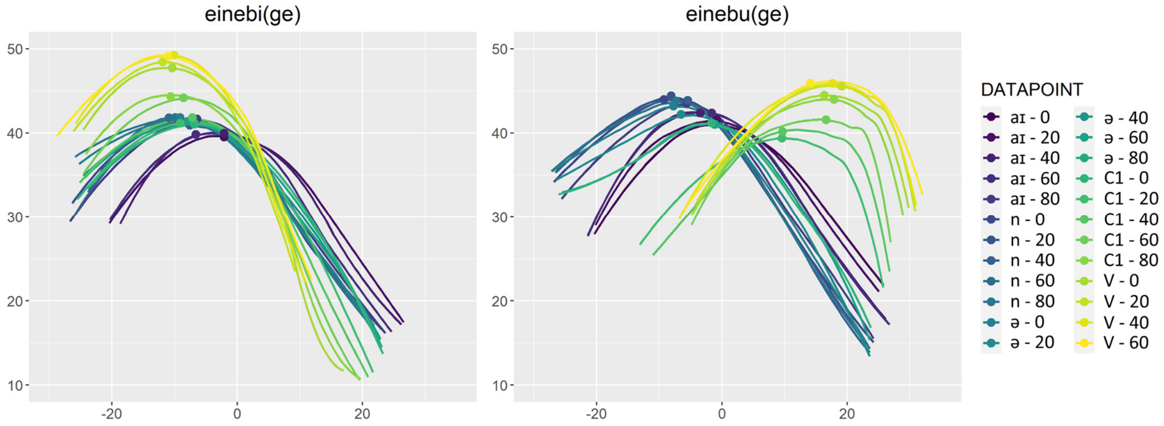
4.3.5 Data analysis

The temporal extent of participants' coarticulation was assessed by comparing horizontal tongue dorsum trajectories over the course of an utterance between stimuli with back and those with front vowels, similar to Goffman et al.'s (2008) approach to labial coarticulation. We determined the onset of vocalic anticipation as the time point of trajectory divergence for /i/- and /u/-stimuli, when in the former, the tongue moves front and, in the latter, backwards. A comparison of the contrast of /i/- and /u/-stimuli between the two experimental conditions Repetition and Reading allowed us to investigate whether target vowels are anticipated earlier in repeated as compared to read aloud speech.

Figure 4.2 shows the tongue configurations at the different time points of interest, as indicated by color shades. The left-hand side shows the key frames from /aɪnə biː/ of a third graders' 'einebige' trial, the right-hand side the same participant's /aɪnə buː/ from the trial 'einebuge'. Both trials are taken from the Reading condition and therefore only differ in the vowel. The highest point of the tongue dorsum is highlighted by a dot on every tongue shape. The horizontal position of this point neatly represents the frontness of the tongue body. The greatest difference between the two plots is in the tongue configuration for the vowels (yellow tongue shapes). As expected, /i/ is produced with a very front tongue position, while for /u/ the tongue is back. Crucially, the consonant (green shades) and in parts the schwa (petrol shades) differ between the two plots, tending towards the subsequent vowel.

Figure 4.2

Temporal Unfolding of /i/- Versus /u/ Stimuli



Note. Whole tongue surface contours of a female participant in third grade at the time points of interest. The left plot depicts one of her ‘einebige’ trials, the right plot ‘einebuge’. For every contour, a dot highlights the highest point. The front of the tongue is displayed towards the left side of each plot. The numbers in the labels correspond to the % of temporal length within the segment.

To assess coarticulation extent in the utterances, we fitted generalized additive mixed models (GAMMs), a statistical approach used to assess both linear and non-linear effects, that is becoming increasingly popular in studies on human speech production (Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2020; Strycharczuk & Scobbie, 2017; Tomaschek, Arnold, Bröker, & Baayen, 2018). Models were fitted using the function *bam* of the *mgcv* package in R (Wood, 2011, 2017). The function *gam.check* was used to examine the normality of residuals’ distribution, heteroscedasticity, and adequacy of the *k* parameter which specifies the maximal non-linearity of the fitted smooth. By-speaker variability over time in distinguishing between the vowels (/i/, /u/), the conditions (Reading, Repetition), and the consonants (/b/, /d/, /g/) was allowed for via random effects in every model. Because of a possible relation between coarticulatory extent and speech production speed (see chapter 4.5), speed was assessed and accounted for by inclusion as a fixed effect in each model. Following Wieling’s statistical approach (Wieling, 2018), we addressed vowel-induced trajectory differences within elicitation condition first, to then assess whether the vowel contrast in trajectories differed significantly between the two conditions. All in all, our research questions are addressed with two sets of models, three for children (those with extension .c), and three for adults (those with extension .a), all of which can be found in Appendix C. 1 and on OSF (https://osf.io/gn8q2) along with their respective outputs. The models m1.c and m1.a test whether the horizontal position of the highest point of the tongue dorsum depends on the target vowel and the elicitation condition. We included an interaction between the two as independent variable and allowed for non-linear patterns of this interaction over time. Possibly non-linear patterns over time for place of articulation of C1 (/b/, /d/, /g/) as well as speed (slow, fast) were also added as fixed factors. While m1.c and m1.a help us visualize and interpret the smooths of interest, they do neither directly inform us about the statistical difference between the two vowels within condition, nor about the statistical difference in vowel contrast between conditions. Therefore, two ordered factor models were fitted to address the former and two binary comparison models were fitted to address the latter question. Models m2.c and m2.a assess the vowel-induced trajectory difference within the two conditions directly, using ordered factors with contrast treatment for vowel and condition, one being TRUE for condition=Repetition and vowel=/u/, and the other being TRUE for condition=Reading and vowel=/u/. To finally address whether the Reading and the Repetition conditions differed significantly in the vowel contrast, we fitted m3.c and m3.a which include new smooths explicitly modelling the difference between the original smooths: One of these binary (i.e., dummy) variables distinguishes /i/ from /u/ without any requirement on condition by setting one level of the nominal variable to 0 (/i/ in our case) and the other level to 1 (/u/). The other binary variable distinguishes /i/ from /u/ within condition by being set to 1 whenever the vowel is /u/ and the condition is Repetition and to 0 otherwise. Therefore, if this second binary smooth is found to be significant, we can conclude that there is a considerable contrast in coarticulatory extent between the two conditions addressed. To reduce the risk of Type-1 error because of fitting two models for each question (i.e., one for children, one for adults), we Bonferroni-corrected our significance threshold to .025 to account for multiple comparisons. For further details on how to model articulatory data with GAMMs, we recommend Wieling's (2018) tutorial.

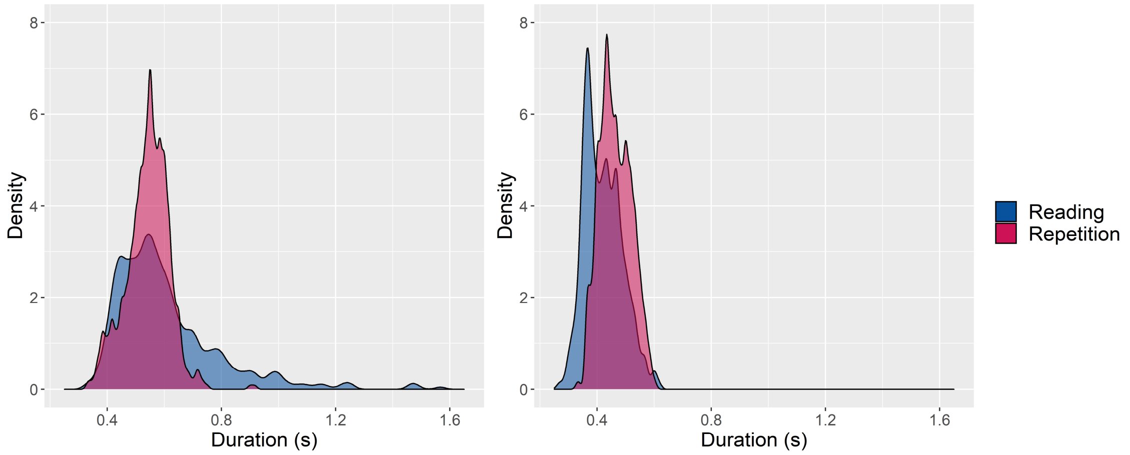
4.4 Results

4.4.1 Does speech rate differ between oral reading and repetition?

To rule out that differences in coarticulatory patterns between repeated speech and oral reading do not (only) result from speed differences, participants' speech rate was measured by taking the duration from frame aɪ_000 to frame V_060 in each trial. Figure 4.3 illustrates the distribution of durations per condition (Reading in blue, Repetition in pink) for children on the left-hand side and adults on the right-hand side. For both age cohorts, the conditions vary significantly in their trial duration, however, for children, Reading trials (mean = 0.61s) are longer than Repetition trials (0.54s; t(13506) = 33.781, p < 2.2e-16) while for adults, Repetition trials (0.46s) are longer than Reading ones (0.42s; t(15451) = -45.38, p < 2.2e-16). The density plots show that the distribution of children's Reading trials is very wide and includes few trials with very long durations while the majority of trials cluster with the Repetition trials. For adults on the other hand, there are clearly two different maxima for the two conditions. To account for speed-related differences, speech rate was included in the models as a fixed factor, categorized in fast and slow per cohort based on a split at the respective median.

Figure 4.3

Trial Duration per Condition



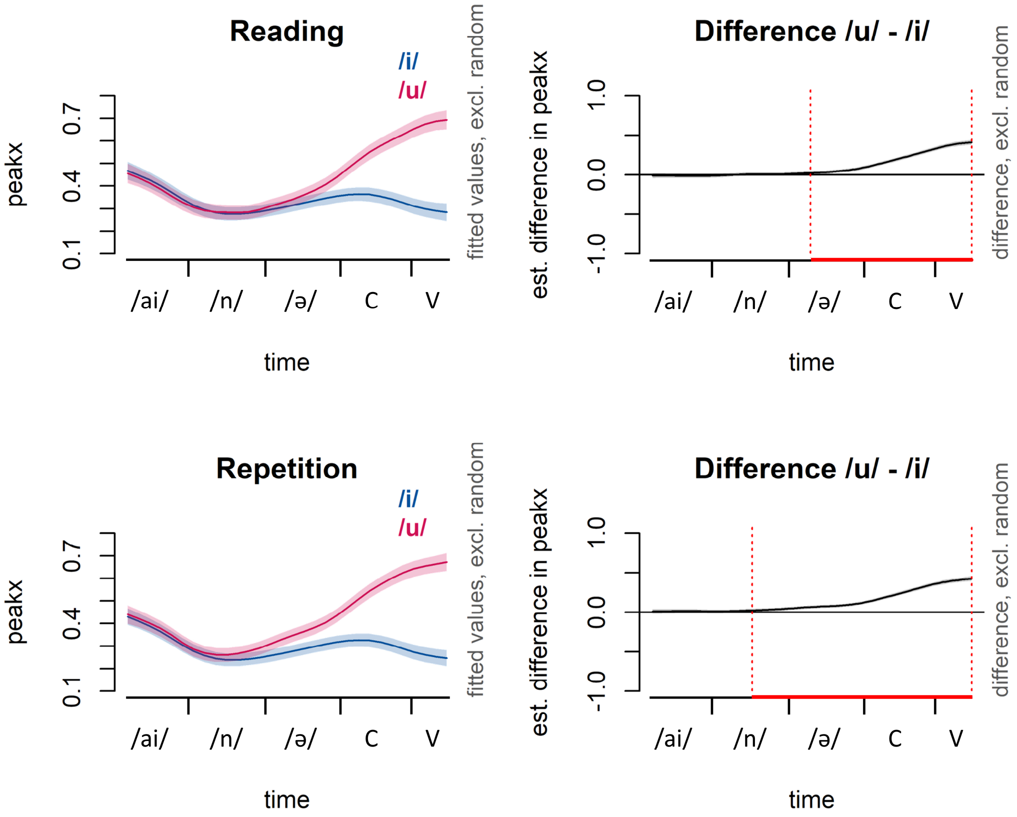
Note. Density plots of the duration of trials based on the Reading and Repetition condition. Left: Children, right: Adults.

4.4.2 Early-stage readers anticipate vowels earlier in repeated speech than in oral reading

The temporal unfolding of the 32 children's vocalic anticipation across conditions (as modelled in m1.c) is illustrated in Figure 4.4. On the left-hand side, the figure depicts the smooths for /i/- (in blue) and /u/-stimuli (in pink) in a coordinate system collapsing time and space: Normalized horizontal position of the highest point of the tongue dorsum (0 = front, 1 = back) is plotted over the temporal course of the utterance. The upper plot shows the Reading condition, the lower plot the Repetition condition. On the right-hand side, corresponding difference smooths are shown with intervals of significant difference highlighted in red. The width of this significance window depicts the temporal extent of significant vowel-related spatial differences in tongue positions, hence coarticulation. At the beginning of the utterances, both /i/- and /u/-stimuli display very similar tongue trajectories but start to diverge at some point. The tongue moves towards the front of the oral cavity for /i/, and towards the back for /u/. Importantly, the time point of trajectory divergence is not at the acoustically defined onset of the vowel but earlier, demonstrating anticipatory coarticulation in both conditions. As modelled in m2.c, the contrast between /i/- and /u/-stimuli is significant in both conditions (see Table 4.1 and Appendix C. 4 for the complete model output).

Figure 4.4

Children's Results



Note. Left: non-linear smooths for /i/- (blue) and /u/-stimuli (pink) displayed as the position (0 – front, 1 – back) of the highest point of the tongue dorsum (peakx) over time. The corresponding acoustic segment boundaries are indicated on the x-axes. Right: differences between the /i/- and /u/-smooths. Significant time windows are indicated in red. The Reading condition is displayed on the top, the Repetition condition on the bottom. Basis: m1.c (see Appendix C. 1 and Appendix C. 2).

Table 4.1

Partial Model Output m2.c

Effect	edf	Ref.df	F	p-value	
s(time): Reading /u/	11.956	14.514	40.827	< 2e-16	***
s(time): Repetition /u/	11.576	14.107	40.852	< 2e-16	***

Note. Results of the ordered factor smooths testing the vowel contrast within conditions for children. Significance codes (after Bonferroni-correction): '***': p < .0005; '**': p < .005; '*': p < .025; '.': p < 0.05.

Regarding the difference between oral reading and repetition, Figure 4.4 shows that vocalic anticipation (i.e., divergence of the two smooths) started later in read aloud (within /ə/) than in repeated (within /n/) stimuli. This is also visible on the right-hand side of the plots with a narrower window of significance in the former than in the latter condition. As tested in m3.c, the Reading condition differed significantly from the Repetition condition in the i/u contrast (see Table 4.2 and Appendix C. 6 for the complete model output).

Table 4.2

Partial Model Output m3.c

Effect	edf	Ref.df	F	p-value	
s(time): Repetition /u/	3.588	4.197	7.229	6.25e-06	***

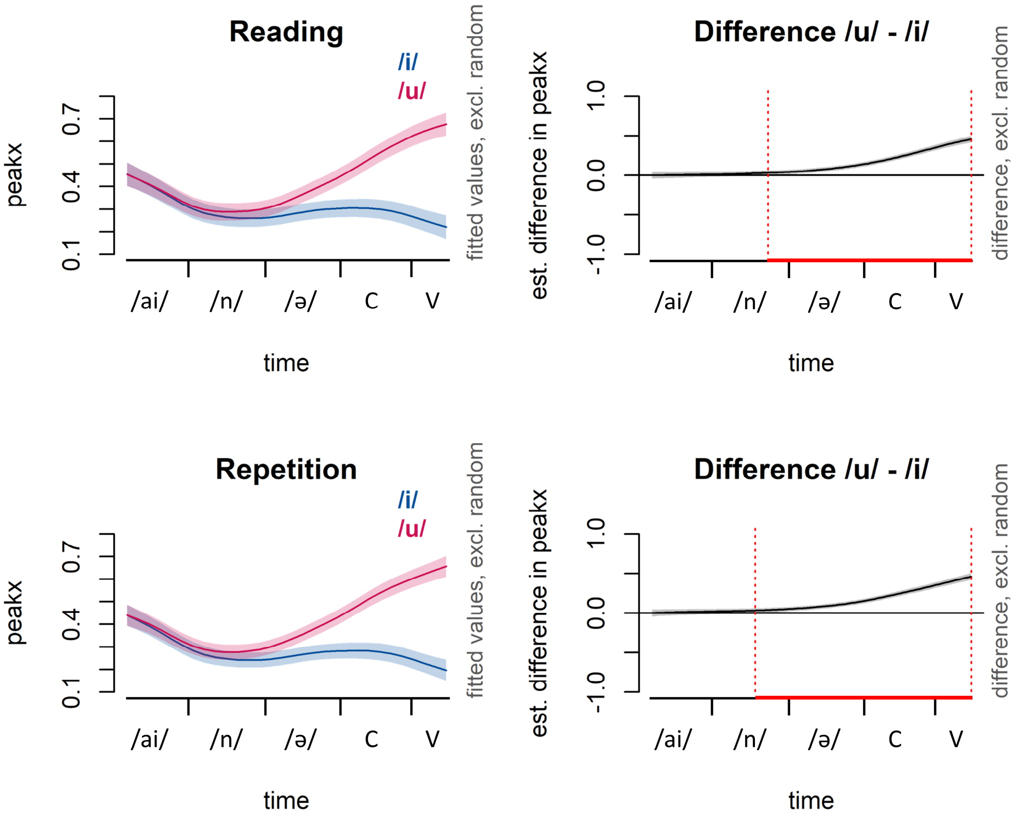
Note. Results of the binary difference smooth testing the difference in vowel contrast between conditions for children. Significance codes (after Bonferroni-correction): '***': p < .0005; '**': p < .005; '*': p < .025; '.': p < 0.05.

4.4.3 Proficient readers' extent of vocalic anticipation does not differ between oral reading and repetition

The results for the 16 adult speakers based on model m1.a are illustrated in Figure 4.5. Again, the ordered factor model m2.a showed that trajectories for /i/- and /u/-stimuli differed significantly within condition (see Table 4.3 and Appendix C. 10 for the complete model output.). In both conditions, they start out at approximately the same relatively central location but diverge soon with /i/-stimuli tending front and /u/-stimuli remaining central to then move backwards.

Figure 4.5

Adults' Results



Note. Left: non-linear smooths for /i/- (blue) and /u/-stimuli (pink) displayed as the position (0 – front, 1 – back) of the highest point of the tongue dorsum (peakx) over time. The corresponding acoustic segment boundaries are indicated on the x-axes. Right: differences between the /i/- and /u/-smooths. Significant time windows are indicated in red. The Reading condition is displayed on the top, the Repetition condition on the bottom. Basis: Model m1.a (see Appendix C. 7 and Appendix C. 8).

Table 4.3

Partial Model Output m2.a

Effect	edf	Ref.df	F	p-value	
s(time): Reading /u/	7.603	9.490	25.060	< 2e-16	***
s(time): Repetition /u/	5.926	7.310	29.896	< 2e-16	***

Note. Results of the ordered factor smooths testing the vowel contrast within conditions for adults. Significance codes (after Bonferroni-correction): '***': p < .0005; '**': p < .005; '*': p <. 025; '.': p < 0.05.

The difference between the two experimental conditions Reading and Repetition in the i/u contrast as addressed in model m3.a, however, is not significant for adult speakers (see Table 4.4 and Appendix C. 12 for the complete model output). The left part of Figure 4.5 illustrates that the point of divergence of smooths for /i/- and /u/-stimuli is very similar for both conditions, which is also illustrated by the significance windows on the right-hand side of the figure.

Table 4.4

Partial Model Output m3.a

Effect	edf	Ref.df	F	p-value
s(time): Repetition /u/	2.374	2.636	0.584	.476

Note. Results of the binary difference smooth testing the difference in vowel contrast between conditions for adults. Significance codes (after Bonferroni-correction): '***': $p < .0005$; '**': $p < .005$; '*': $p < .025$; '.': $p < 0.05$.

4.5 Discussion

The present study investigated vocalic anticipation in oral reading compared to repeated speech. We asked whether the often-reported protraction of reading fluency development in German early-stage readers (Aro & Wimmer, 2003; Moll et al., 2012) would be reflected in their extent of vocalic anticipation even when standard assessments suggest fluency. To this end, we employed ultrasound tongue imaging to capture differences in tongue movement trajectories in short utterances embedding high front versus high back vowels. With this measure we can specifically address phonemic decoding and blending fluency. Overall, we found that both proficient adult and early-stage readers anticipate target vowels well ahead of their acoustically defined onsets in both oral reading and repeated speech. However, in second- and third graders, movement trajectories of the tongue for /i/- and /u/-stimuli diverged substantially earlier for repeated than for read aloud speech, indicating they anticipate upcoming vocalic targets earlier when they repeat than when they read aloud. Crucially, all analyzed utterances were overtly fluent, without audible hesitations. In contrast, adults' extent of vocalic anticipation did not differ between the experimental conditions Reading and Repetition. Implications of these findings are discussed in the following sections.

Our main finding is that German children in second and third grade articulatorily anticipate vowels earlier in repeated speech compared to oral reading (Figure 4.4). Information about upcoming vowels is therefore integrated in their speech production earlier in repetition than in oral reading. This may be because the stronger serial left-to-right processing in oral reading decelerates the availability of information as compared to the repetition of heard speech. Based on a series of experiments comparing oral reading and object naming in adults, Roelofs (2004) provided evidence that both tasks involve similar serial processes. This suggests that effects of seriality may be located on the shared stage of phonological encoding. However, Mousikou, Rastle, Besner, and Coltheart (2015) highlighted that some seriality effects can be attributed to the reading-specific stage of orthography-to-phonology computation. In the present study, the clear seriality difference between conditions, as revealed by children's vocalic anticipation, suggests that it is the orthography-to-phonology conversion that lacks speed. Early-stage readers decode script in small portions from left to right and have not yet reached a decoding speed allowing for speech-like anticipatory coarticulation. The present data show that the recorded children surpassed the grapheme-by-grapheme decoding phase, producing coarticulated oral reading. However, the extent of vocalic coarticulation is more restricted in read aloud than in repeated speech. This active decoding of units broader than the letter neatly illustrates Scheerer-Neumann's (1989, 2007, 2018) *orthographic* phase – exactly that phase that she described as discrepant from reading development in more opaque languages like English. While the reading fluency protraction in transparent orthographies has been investigated with measures of speed and accuracy (Aro & Wimmer, 2003; Moll et al., 2012), to our knowledge this is the first study directly addressing phonemic blending with articulatory measures.

Thus far, most empirical investigations of coarticulatory processes in child speech provided evidence for a developmental decrease in their strength (American English: Nittrouer et al., 1989, 1996; Canadian French: Noiray et al., 2013; German: Noiray et al., 2018; Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2018, 2020; Scottish English: Zharkova et al., 2011). Rubertus and Noiray (2020) interpreted this finding as evidence for stressed vowels to serve as anchors during speech production, subdividing the speech stream and embracing close-by segments within their domain. Over time, the differentiation of individual speech segments becomes clearer and coarticulation decreases. Factors driving this decrease of coarticulation amount are numerous and may be found in different domains with varying strengths throughout development. One important factor is the changing articulatory system developing finer speech motor control (Maas & Mailend, 2017; Nittrouer et al., 1989) and more precise articulatory patterns (Abakarova, Fuchs, & Noiray, 2022). Other factors like vocabulary size (Cychosz et al., 2021; Noiray, Popescu, et al., 2019), phonological awareness (Noiray, Popescu, et al., 2019), production practice (Cychosz et al., 2021), and reading proficiency (Popescu & Noiray, 2021) have been shown to be negatively correlated with the amount of coarticulation. While production practice may also interact with motoric developments, these latter factors all relate to a change of the representations of speech segments from more holistic to more segmental entries ("phonological reorganization"; Macken, 1979; Vihman & Velleman, 1989). The present study adds up to this by providing evidence for the nature of transparent orthographies to decrease coarticulation extent in oral reading. Here, we show that for beginning readers, coarticulation in reading itself is influenced by the emphasis on individual segments in the transparent German orthography. The developing understanding of the correspondence between the discrete letters and the continuous speech stream is therefore likely to be one of several triggers for a growing differentiation of speech segments. These children, familiar with fluently coarticulated speech, who are now learning to decode and synthesize sequences of individual graphemes, are "infected" with the virus of literacy, as Frith (1998) put it, and are from now on tricked by their orthographic knowledge into believing they actually hear segmental speech (Brügelmann, 1992).

Adults did not exhibit a significant difference between Reading and Repetition in the i/u contrast. This means that the extent of vocalic anticipation in read aloud and in heard and repeated stimuli did not differ significantly. Because the stimuli were pseudowords, they could not be directly accessed lexically in the reading condition but had to be decoded by grapheme-to-phoneme-conversion. In various paradigms, effects of positional sensitivity have shown that even in proficient readers, this indirect route operates serially from left to right (e.g., masked priming: Forster & Davis, 1991; regularity effect: Rastle & Coltheart, 1999; phonological Stroop effect: Coltheart, Woollams, Kinoshita, & Perry, 1999). Therefore, the finding that coarticulation extent does not differ between the conditions in adults suggests that in contrast to early-stage readers, proficient readers are very quick in their grapheme-to-phoneme conversion, reaching a processing speed comparable to that in repeated speech. Any effects of seriality found here are the same in both conditions and may therefore well be part of the shared process of phonological encoding as suggested by Roelofs (2004). We conclude that the efficiency of indirect print-to-speech conversion is enhanced with practice – experience with the orthography of one's language seems to automatically improve decoding and blending fluency.

To make sure the Reading condition specifically addressed phonemic decoding and not holistic retrieval, we opted for the two Reading blocks to always precede the two Repetition blocks. This choice of experimental design could lead to a possible limitation: In addition to the switch of conditions, familiarity with the setup and practice over the course of the experiment could affect children's speech production. Effects of motor practice usually go hand in hand with an increase of movement speed and a decrease of duration (Schulz, Stein, & Micallef, 2001). The speed of speech movements is often related to coarticulation. However, the details about this relation seem to depend on various factors and are not fully understood. In his review, Berry (2011) points out that "rate-induced effects on phasing appear to be far from predictable. The literature includes reports that increasing speaking rate results in (a) increased overlap, (b) no change in overlap, or (c) decreased overlap" (p. 19) of articulatory movements. Shaiman and colleagues (Shaiman, 2001; Shaiman, Adams, & Kimelman, 1995) even provide evidence that strategies to increase speaking rate vary between manipulations of gestural overlap and velocity or a mixture of these across speakers and articulators. To make sure we measure an effect of elicitation condition (Reading versus Repetition) and not one related to speed, in the present analysis, we therefore included speed as a fixed factor in all our models. The outcome even under the consideration of speed differences that the condition effect was significant for children but not for adults, indicates that on top of possible speed-related differences in coarticulatory extent that could in turn hint at effects of familiarity or practice, reading proficiency is the driving force here. One reason for effects of practice and familiarity to be subordinate to the reading-proficiency-related effect of condition in the present data set lies in the specifics of our experimental design. The simple structure of our stimuli, the low number of repetitions, and the details of our procedure, specifically, the high number of playful interludes, do not promote effects of familiarization and motor practice.

4.6 Conclusion

This study is the first to implement the state-of-the-art technique of ultrasound tongue imaging to investigate the well-known protraction of word reading fluency in early-stage readers of transparent orthographies. Using fine-grained measures of lingual vocalic anticipation, we demonstrate that fine disfluencies in phonemic blending occur even in perceptually fluent utterances when comparing oral reading to repeated speech. To describe word reading fluency, coarticulatory extent as measured here, seems to surpass standard reading assessments of accuracy and speed as it detects fine disfluencies in phonemic blending in accurate productions independent of speed. This highlights the benefits of direct articulatory measures for assessments of reading fluency. Importantly, while early-stage readers articulatorily anticipated vowels later in oral reading than in repeated speech, the conditions did not differ in coarticulatory extent for adults. This leads us to conclude that the development of efficient grapheme-to-phoneme conversion and blending is not yet complete in German children in second and third grade.

5

GENERAL DISCUSSION

5.1 Summary of the empirical work

In this dissertation, I investigated developmental changes in lingual coarticulation across childhood. Speech consists of continuous movements whose context-dependency is not restricted to segmental, syllabic, or morphemic frames. Examining changes in speech fluency and specific coarticulatory patterns across childhood, therefore, calls for a close monitoring of longer distance coarticulatory processes. This way, we can identify developments of speech motor control as well as phonological representations across childhood, uncover potential causes thereof, and disentangle their implications for phonological theory and the human speech production mechanism. Before discussing the obtained results in light of different models, the following sections shortly summarize the empirical findings of this dissertation. In all three empirical studies of this dissertation participants produced pseudo-utterances of the form /aɪnə$_1$ C$_1$VC$_2$ə$_2$/ while their voice as well as their tongue motion was recorded. The setup was the same for all experiments (Noiray et al., 2020), except that the third study required an additional screen and a different presentation software to be able to use orthographic in addition to acoustic stimuli. In every ultrasound frame of interest (that varied between studies) we detected the horizontal position of the highest point of the tongue dorsum.

5.1.1 Anticipatory V-to-V coarticulation

In a cross-sectional investigation of anticipatory V-to-V coarticulation in 75 participants (chapter 2; Rubertus & Noiray, 2018), we recorded three- to seven-year-old children's and adults' tongue motion while they repeated pseudo-utterances after a model speaker. Using linear mixed effects models, we regressed their horizontal tongue position during the temporal midpoint of /ə$_1$/ on its position during the stressed vowel's (V) temporal midpoint, taking possible effects of the three C$_1$ contexts (/b, d, g/) and participants' age into account. Results provided evidence for substantial vocalic anticipation across an intervocalic consonant in all age cohorts. Importantly, children's tongue position during schwa resembled the vowel more closely than adults', which means their degree of vocalic anticipation was significantly greater compared to adults'. While we saw a trend of decreasing coarticulation degree with increasing age among the different age cohorts of children, significance was only reached in the comparison between the youngest (three-year-olds) and the oldest (seven-year-olds) children in /b/- and /d/-contexts. The intervocalic consonant did only affect adults' V-to-V coarticulation substantially in the direction that /g/ allowed for more vocalic anticipation in /ə₁/ than did both /b/ and /d/. Children's V-to-V coarticulation degree was not modulated by the intervocalic consonant.

5.1.2 *Carryover coarticulation*

In Rubertus & Noiray (2020; chapter 3 of this dissertation), we investigated the same data set as in Rubertus & Noiray (2018) for effects of vocalic coarticulation towards the right side of the utterance. While anticipatory effects are often attributed to planning, for carryover effects purely mechanic, inertial causes are highlighted. Comparing coarticulatory development between the two directions may, therefore, illuminate the question of underlying mechanisms. Instead of focusing on vocalic effects on the tongue position at one specific time point as in Rubertus and Noiray (2018), we applied a more dynamic analysis we had in the meantime developed for anticipation in Noiray, Wieling, et al. (2019) using generalized additive mixed models (GAMMs)[6] to model the tongue trajectory across four dependent time points (V100, C50, C100, schwa50). Like the degree of vocalic anticipation, the degree of carryover coarticulation decreased with increasing age. Substantial impacts of gestural demands of the intervocalic consonant on the tongue trajectory were found in every age cohort: While the low resistant consonants /b/ and /g/ allowed for stronger and more continuous vocalic effects, utterances with the highly resistant consonant /d/ indicated discontinuous vocalic effects. Additionally, developmental changes in coarticulation degree across childhood were shown to be strongest in utterances with /g/, whose place of articulation closely depended on the vowel in children but less so in adults. Together with previous findings (Noiray, Wieling, et al., 2019), utterances with the alveolar /d/ provided evidence for children to lack independent control of the tongue tip and the tongue dorsum.

5.1.3 *Anticipatory coarticulation in read versus repeated speech*

In chapter 4 of this dissertation, second and third graders as well as adults read out aloud written instances of the stimuli in addition to repeating acoustic presentations. Within the same setup as in Rubertus & Noiray (2018, 2020), we adopted a dynamic approach to coarticulation and sampled tongue position measures at tighter temporal intervals than in the two earlier studies to compare vowel-dependent tongue trajectories for the complete left part of the utterance (i.e.,

---

[6] While in chapter 3, the abbreviation GAM is used, please note, that both the second and the third study used generalized additive mixed models (GAMMs).

/aɪnə$_1$ C$_1$V/). Using GAMMs we compared movement patterns of the highest point on the tongue dorsum in utterances with target vowel /i/ to those in utterances with target vowel /u/. In contrast to the first two studies, here, the temporal extent of coarticulation is highlighted. In children, who are early-stage readers, the tongue trajectory became vowel-dependent at an earlier point within the utterance in the repetition condition than in the reading condition. Despite being produced correctly and without audible hesitations, children's read aloud stimuli, therefore, indicated weaker vocalic anticipation than their repeated stimuli - a sign of subtle disfluencies in phonemic blending. For adults, the extent of vocalic anticipation did not differ between the elicitation conditions.

5.2 Implications of the obtained results

The findings in the first two empirical studies of this dissertation (Rubertus & Noiray, 2018, 2020) as well as those in our related publications (Abakarova et al., 2022; Noiray et al., 2018; Noiray, Wieling, et al., 2019) provide compelling evidence for a substantially higher coarticulation degree in children than in adults. These findings in addition to earlier congruent findings in English (Nittrouer et al., 1989, 1996; Zharkova et al., 2011), actually lead to a broad agreement about a general decrease in coarticulatory degree across childhood (e.g., Cychosz, Munson, & Edwards, 2021; M. Gibson, Bunta, Johnson, & Huárriz, 2022; Zmarich et al., 2021). While for reasons of simplicity, we compared children's speech based on different age groups, the developmental change in coarticulatory behavior is not driven by age itself, but just like any other motor development is likely to result from a complex and dynamic interplay of multiple factors. As Esther Thelen (1995) emphasized, these causes blur the boundary between internal and external, between biology and experience. Regarding speech production, there is a direct interaction of motoric processes and higher order linguistic knowledge, both of which undergo drastic changes across childhood. While some of these changes may result from maturational processes, others may be caused by external input. The following sections discuss potential sources of the coarticulatory change across childhood and point out what our findings in turn teach us about the speech production mechanism and phonological representations.

5.2.1 Effects of speech motor control maturation

Considering the gestural hypothesis we posited in Noiray, Wieling, et al. (2019), we paid specific attention to the articulatory demands of the intervocalic consonants. The finding that the consonant's place of articulation impacts on the degree of coarticulation mostly within the consonant itself (Noiray et al., 2018; Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2020) and in parts as well within the preceding (Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2018) and the following vowels (Rubertus & Noiray, 2020) shows that the specific articulatory demands of neighboring speech segments may (temporarily) modulate the strength of vocalic coarticulation. While the whole tongue is relatively flexible during the production of the bilabial and velar consonants /b/ and /g/, respectively, the position of the tongue tip is very constrained during the production of the alveolar closure for /d/. In our data, young children indicated a lack of flexibility of the tongue dorsum during the production of /d/ compared to adults, uncovered by lower vocalic impacts in the domain of the consonant (Rubertus & Noiray, 2020). This finding is in line with further studies reporting a lack of independent control of individual articulators in children (Nijland et al., 2002; Nittrouer et al., 1989). Here, specifically, the tongue dorsum is closely coupled to the tongue tip and, therefore, needs to accompany the tip moving front to reach its goal at the alveolar ridge (Abakarova et al., 2022). While limiting coarticulation during /d/, a tighter coupling of tongue tip and tongue dorsum may be one reason for children to exhibit more vocalic coarticulation than adults in the context of /b/ and /g/: The movement of the tongue as a whole is governed by the stressed vowel which results in vocalic tongue positions during the labial closure of /b/ as well as vowel-dependent constriction locations of /g/. As children grow older, they develop a finer speech motor control and the subparts of the tongue become more independent. Among others, the coupling of the tongue's subparts, therefore, is one indication that some changes in coarticulation degree across childhood can be attributed to articulatory control and speech motor development (Maas & Mailend, 2017; Nijland et al., 2002; Nittrouer et al., 1989).

5.2.2 Representational changes: the vowel's prominence

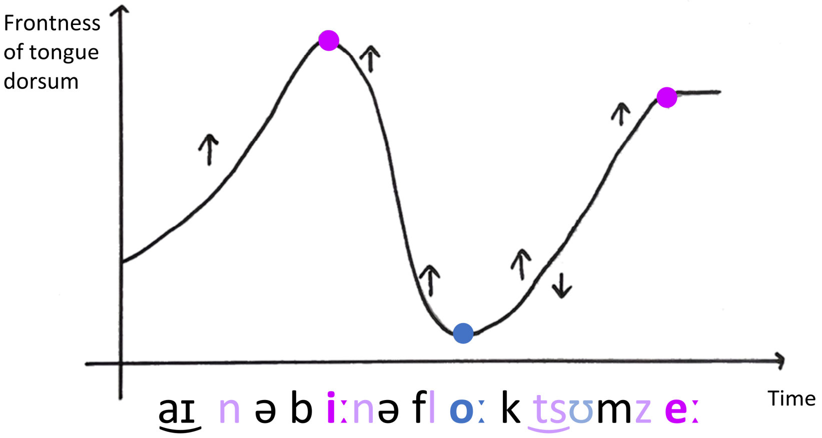
While developments in speech motor control like a decoupling of the tongue's subparts may explain changes in local coarticulatory processes, they cannot comprehensively explain the developmental decrease in longer distance coarticulation. In addition to motoric factors, it is likely, that there is a more fundamental, representational change in the building blocks underlying speech. And importantly, the developmental patterns revealed in our studies, provide further information about the nature of these underlying atoms of speech.

As Redford (2019) pointed out, the direction of the developmental change we found – i.e., a decrease of coarticulatory degree across childhood – cannot be modeled within a theory relying on complicated translation processes between an abstract, mental level and the physical events of speech production. In addition, as we have argued in Rubertus and Noiray (2020), the parallel findings in the two coarticulatory directions speak against different underlying mechanisms and, therefore, call early pre-planning for coarticulation into doubt. In contrast to the feature spreading model (Daniloff & Hammarberg, 1973) on the one hand and the window (Keating, 1988) and DIVA (Guenther, 1995) models on the other hand, the coproduction model (C. A. Fowler, 1980) does not require phonological or phonetic translation rules, respectively. It also does not predict a general dichotomy between anticipatory and carryover coarticulation effects. Instead, the coproduction framework unifies phonological representation and articulatory implementation by assuming articulatory gestures as phonological atoms which overlap upon execution in the vocal tract. Consequently, I take a closer look at how our findings could be explained within the coproduction framework.

As we have previously argued (Rubertus & Noiray, 2020), the finding of children's substantially stronger vocalic anticipation as well as perseveration compared to adults suggests that for children the stressed vowel is the most prominent element of the utterance, behaving like a strong anchor perceptually and impacting on vocal tract configurations for particularly long stretches of time. According to C. A. Fowler (1980), even in adult speech production stressed vowels govern the broad trajectory of the vocal tract shape while other articulatory gestures are subsumed within the general movement pattern of one stressed vowel to the next (cf. Öhman, 1966). To illustrate this line of thought, Figure 5.1 sketches a completely vowel-governed tongue dorsum trajectory for the German sentence "Eine Biene flog zum See" ('A bee flew to the lake') via its approximate horizontal position over time. Surely, articulatory gestures in between those of the stressed vowels (/iː/, /oː/, /eː/) must impact on the tongue dorsum trajectory for other speech segments to be intelligible, which is indicated by arrows pointing up (i.e., front) or down (i.e., back). The strength and temporal extent of this impact depends on how compatible the simultaneously active gestures are (C. A. Fowler & Brancazio, 2000; Recasens, 1984b, 1984a; Recasens et al., 1997). Importantly, our findings add that this relative impact also differs between children and adults: As summarized in chapter 3.2.3, for young children, stressed vowels seem to be exceptionally salient and may be hyperactive compared to adults. As children get older, the relative impact on the tongue dorsum trajectory of articulatory gestures constituting other segments increases, and, therefore, limits the dominance of the stressed vowels, that is, decreases vocalic coarticulation.

Figure 5.1

Sketch of the Tongue Dorsum's Horizontal Movement Over Time During the Production of the Sentence "Eine Biene flog zum See" ('A bee flew to the lake')



Note. The utterance is transcribed below the sketch to illustrate the element of speech dominating the signal. The sketched trajectory is governed by the stressed vowels /iː/, /oː/, and /eː/. Arrows indicate where other elements will impact the vocalic trajectory, pulling it up (front, in pink), or down (back, in blue).

This developmental change of stressed vowels' relative prominence within an utterance can be envisioned by the overlapping gestural prominence curves of the coproduction framework (C. A. Fowler, 1980; Nittrouer, 1993, see Figure 3.1). A lower vocalic coarticulation degree would be depicted by a later rise and earlier fall of vocalic activation (here the tongue dorsum gesture), i.e., narrower prominence curves. Since timing parameters are explicitly part of the specification of phonological atoms (C. A. Fowler, 1980), developmental changes in gestural timing can be considered as representational changes. The idea of a developmental compression of activation curves is not new – 30 years ago, Susan Nittrouer wrote:

> The curves sometimes used to depict the prominence of individual signal components (or, for present purposes, of articulatory gestures) across time might be envisioned as being flatter, broader, and having more shared areas in children's than in adults' productions. […] Clearly, the most appropriate method for studying the gestural organization of children's speech would be procedures recording articulatory movements directly. Unfortunately, such methods are often impractical to attempt with very young children, mainly due to the invasiveness of the procedures. (Nittrouer, 1993, p. 960f)

Owing to the implementation of noninvasive and, therefore, child-friendly ultrasound imaging to purposes of speech investigations, the large set of articulatory data acquired during the present research program can substantiate Nittrouer's hypothesis by providing more direct evidence that children's speech is characterized by larger gestural overlap, at least that of the stressed vowel, than adults'.

Two specific findings provide further support for the suggestion that children's vocalic gestures more strongly govern the vocal tract actions within a short utterance than adults': 1) We found evidence for discontinuous vocalic effects in carryover coarticulation across /d/. While children's tongue dorsum position during the production of the alveolar closure was always very front independent of the preceding stressed vowel, the position during the following schwa again oriented towards that of the stressed vowel. For adults, vocalic impacts on the schwa following /d/ were much weaker than for children. As C. A. Fowler and Brancazio (2000) and C. A. Fowler and Saltzman (1993) argue, discontinuous vocalic effects as in our child data are indicative of broad vocalic activation curves that are phased relatively invariantly to each other. Consonants are imposed on this vocalic trajectory as temporally limited events that briefly clamp the tongue, with a greater extent the more resistant the consonant is. Look-ahead models like the feature spreading account (Daniloff & Hammarberg, 1973) cannot explain these kinds of coarticulatory troughs in speech production data – why should a speaker revive the vowel after a consonant terminated planned carryover coarticulation (C. A. Fowler & Saltzman, 1993, p.181)? This finding also highlights the importance of investigating longer distance coarticulatory processes in addition to the more frequently studied local effects of coarticulation (e.g., CV-coarticulation). 2) For the anticipatory direction, we did not find robust effects of the consonant modulating the degree of V-to-V coarticulation in children, but we did in adults (Rubertus & Noiray, 2018). While the consonant's articulatory specifications, most importantly its coarticulatory resistance (Recasens et al., 1997), determine the degree of vocalic anticipation during the production of the consonant itself, both in children as well as in adults (Noiray et al., 2018; Noiray, Wieling, et al., 2019), they do not substantially influence the degree of V-to-V coarticulation in children's speech. Again, this speaks for children's tongue trajectories being guided by the stressed vowels with temporally very limited consonantal clamps of the tongue. Potential reasons for the developmental compression of vocalic activation curves are addressed in chapter 5.2.4.

5.2.3 Further support for coproduction: rhythm class differences

Interestingly, there is a striking discrepancy in recent findings about developmental changes in coarticulation that can well be explained when interpreted from a coproduction point of view: With regard to the claim that in English, articulatory trajectories are guided by stressed vowels while other gestures are rather integrated within this frame, C. A. Fowler (1981, p.128) raises the point that "this subsumption strategy may account for the linguists' (e.g., Abercrombie, 1964; Pike, 1945) and naïve listeners' (Donovan & Darwin, 1979; Lehiste, 1973) impression that English is stress-timed." Attempts to find isochrony in the signal to empirically substantiate the idea of different rhythm classes among languages failed (cf. Turk & Shattuck-Hufnagel, 2013). It was, therefore, concluded that it is not a predetermined rhythm triggering phonological specifics, but rather that language-specific phonological properties convey the impression of different speech rhythms (Nespor, Shukla, & Mehler, 2011). In line with C. A. Fowler (1981), Bertinetto & Bertini (2008) suggest that the rhythmic difference may be grounded in the specific degrees of segmental overlap a language allows for, hence in the way segments are blended or coproduced with each other. This is a very interesting claim given the discrepancy in findings between our study and the one of our colleagues Barbier et al. (2020) that we addressed in chapter 2.5.1 (Rubertus & Noiray, 2018). While we found striking evidence for vocalic coarticulation degree to decrease across childhood in German, their findings suggest that Canadian French children indicate a lower degree of anticipatory vocalic coarticulation than Canadian French adults. Surely, there are methodological differences between the two studies: In contrast to our stimuli with schwas in the dependent positions, their V_1 for example was a full vowel (one of the articulatorily flexible ones /ɛ/ and /a/, though), and they used the whole tongue surface contour instead of a point measure as in our investigations. A major reason for the contrasting findings, however, may be an underlying difference in the mechanisms of speech production between the two languages that is also responsible for the differing rhythmic impressions. German is traditionally classified as rhythmically stress-timed, like English, while French in contrast has a syllable-timed speech rhythm. If indeed the strong subsumption of segments under the trajectory from one stressed vowel to the next is responsible for the stress-timed speech rhythm of English and German (C. A. Fowler, 1981), the blending and coproduction of segments is guided by different mechanisms in syllable-timed languages like French. In French, we would, therefore, not expect the stressed/full vowel to be over-prominent in children's speech and hence vocalic coarticulation degree should not be higher than in adults.

A cross-linguistic comparison of coarticulation degree and its changes across childhood was not the focus of this dissertation. Therefore, the underlying difference in speech production between languages classified in different rhythm classes is only a post hoc argument driven by a comparison of our findings with those of our colleagues. The discrepancy calls for a more thorough investigation of cross-linguistic differences in coarticulatory degree and its changes across childhood. An especially interesting comparison could be drawn between two rhythmically different varieties of the same language, for example stress-timed American English versus syllable-timed Indian English (Crystal, 1994), because the phonetic material could be kept identical.

5.2.4 Alphabetic literacy as a driving force for linguistic restructuring

We provided evidence that one potential reason for the developmental decrease of vocalic coarticulation is speech motor control maturation (chapter 5.2.1) – a factor that comes into play only at implementation in the vocal tract. Another potential contributor we addressed, that may impact on the speech production process slightly earlier, is the development of more inhibitory control across childhood (chapters 2.5.3 and 3.5.1; cf. Bjorklund & Harnishfeger, 1990; Tilsen, 2013, 2016, 2018). However, most of the development of inhibitory control happens during the pre-school years (Best & Miller, 2010) while the largest change in coarticulation degree in chapters 2 and 3 (Rubertus & Noiray, 2018, 2020; as well as Noiray, Wieling, et al., 2019) was found between the oldest group of children (at the age of seven years) and the adults. This suggests further important developments within children's school years that cause changes in coarticulation degree. Modeling these changes via a developmental compression of vocalic gestures' activation width within the coproduction framework, suggests that changes occur within the building blocks of speech production, specifically in the timing component intrinsic to articulatory gestures. What may cause these phonological representations to change and via timing parameters delimit their relative prominence? One factor known to impact substantially on phonological representations of speech is literacy acquisition, potentially the most influential milestone in language development after the infant and toddler age, which we addressed in the third empirical study (chapter 4).

Our finding that early-stage readers of German exhibit significantly less vocalic anticipation in read aloud than in repeated stimuli indicates a lack of phonemic blending fluency. This provides insights into the process of literacy acquisition, especially in comparison to studies on literacy acquisition in the well-studied opaque orthography of English (cf. Frith, 1986; Scheerer-Neumann, 1989, 2007, 2018). Naturally, children's phonemic decoding and blending speeds up with experience and practice, as shown in the absence of a condition effect in adults. Importantly, this finding also shows how children's attention is shifted from the continuous stream of speech they are familiar with to the concatenation of separate discrete segments. As Goswami and colleagues (Goswami, 2000, 2001; Goswami & Bryant, 2016; J. C. Ziegler & Goswami, 2005) have argued, alphabetic literacy acquisition leads to an understanding of the composite nature of speech, which strengthens or maybe even first evokes phonemic awareness: "The representation of phoneme level information is thought to 'spurt' with the acquisition of literacy, because the feedback provided by graphemic information helps the child to represent segmental information at the phonemic level" (Goswami, 2000, p.146). The holistic articulatory templates are now broken up into a sequence of units; lexical entries likewise restructured from rather holistic to more specific, further subdivided representations.

In an alphabetic orthography like German, stressed vowels are not graphically emphasized compared to other segments. Rather, all graphemes are comparable and convey the idea of equivalence. While maturations in fine-grained phonological awareness due to literacy may not change the speech production mechanism at its core, experience with alphabetic scripts may subconsciously degrade the status of the stressed vowel to some extent and, therefore, set boundaries to the vocalic hyperactivation found in young children. The overlap between a vocalic gesture and its neighborhood is, therefore, assumed to decrease because the prominence curve of the vocalic gesture(s) gets narrower via a change in the vowel's timing parameters. While this is a systematic representational change, it does not necessarily need to happen in a purely mental entry, but as Fowler suggests may occur in representations that "have their primary home in the vocal tract, not in the mind" (C. A. Fowler, 2015, p.27). The "virus of literacy" that was introduced in chapter 4.2.1, may evoke awareness of the combinatorial nature of speech and may at the same time lead to a reduction in vocalic prominence. Both factors are likely to contribute to the decrease in vocalic overlap with neighboring speech segments.

5.2.5 Debating the size of articulatory units

Our findings support the hypothesis that young children's speech production does not imply a phonemically structured organization, but rather that segmental units get defined and tuned over time. Actually, the perceptual findings often cited as evidence for phoneme-sized units in child speech (see chapter 1.2), must be interpreted with caution: The ability to perceptually discriminate segmental contrasts does not necessarily imply that infants focus on segmental units as an entry into language acquisition. In fact, categorical perception along with particularly good discrimination ability at category boundaries have been shown in other mammals as well (e.g., chinchillas: Kuhl, 1981; macaques: Kuhl & Padden, 1983, Mongolian gerbils: Sinnott & Mosteller, 2001). This rather supports the view that the fundamentals of the mammalian auditory system determined how languages developed by making use of those sound contrasts that are well discriminated auditorily (e.g., Kuhl, 1981). Relatedly, perceptual narrowing towards information that is relevant in the child's input along with a loss of discrimination ability for irrelevant contrasts is not specific to language but was for example shown for face recognition (Krasotkina, Götz, Höhle, & Schwarzer, 2021; Pascalis, de Haan, & Nelson, 2002) and musical perception (Trehub & Hannon, 2006) as well. Outstanding phonetic discrimination abilities and perceptual narrowing, therefore, do not inevitably indicate linguistic knowledge and segmental representations.

Moreover, young children lack awareness of phoneme-sized units in speech as documented in various tasks like phoneme counting (I. Y. Liberman, Shankweiler, Fischer, & Carter, 1974), partial production (B. Fox & Routh, 1975), and grouping based on shared phonemes (Treiman & Zukowski, 1996). Holding on to the premise that children's speech is built upon underlying phonemic representations just as assumed for adult speech, the demonstrated failure to manipulate phonemes was ascribed to a difference between implicit and explicit knowledge (cf. A. E. Fowler, 1991). However, based on Ferguson's (1986) approach to view adult's phonology as growing out of the child's instead of children acquiring the adult system, A. E. Fowler (1991) and Studdert-Kennedy (1986) argue, lacking phonemic awareness could also be one of several indications that phoneme-sized units only emerge and undergo growth and change in the course of development: "The child does not build words with phonemes: phonemes emerge from words" (Studdert-Kennedy, 1986, p. 59).

Maybe, the classical phoneme is even merely an illusion, a by-product of literacy shaping linguists' as well as naïve speakers' intuitions about units of language (Brügelmann, 1992; Bybee & McClelland, 2005; Lotto & Holt, 2000; Port, 2010). While it has proven useful for linguistic descriptions, it may not necessarily correspond to an equally clear-cut representation in a speaker's mind. Instead of abstract phonemes, articulatory gestures may be the fundamental units, the atoms of speech perception as well as production. Within this Articulatory Phonology framework, it is debated whether phoneme-sized segments have a special status: Byrd (1996) provides evidence for higher phasing stability between gestures constituting one phonemic segment than between gestures belonging to different phonemes within a consonantal onset cluster. In contrast, Scobbie and Pouplier's (2010) data suggest that phonemic segments are simply coproductions of specific articulatory gestures, just like gestures in consonantal onset clusters are coproduced. Regarding developmental changes, Articulatory Phonology suggests that children do not start with units of a particular size relevant for phonological theory, but "initially master a few simple patterns of articulatory movement" (Nittrouer, 1993, p. 960). Over time, they continuously differentiate and tune the individual gestures they started to produce during babbling to reach a categorical system of contrastive gestures (Browman & Goldstein, 1989). Quantal articulatory-auditory relations (Stevens, 1989) as well as lexical pressure (Lindblom, 1986) may dynamically drive this development. Current network models and usage-based approaches agree that representations of segmental size develop along with lexical growth (for an overview see Vihman, 2017): The more words a child hears and produces, the more cross-connections are formed via ubiquitous phonological priming and similarity detection (Menn, Schmidt, & Nicholas, 2013). The close-meshed nets of connections then form new categories. Thus, building a categorical system of articulatory primitives used as atoms of spoken language is likely to be a lengthy process going hand in hand with lexical growth and literacy acquisition.

Even though our findings provide evidence for a developmental decrease in coarticulation degree, I hesitate to directly ascribe the presented broad coarticulatory effects exceeding syllable- and (pseudo-)word boundaries to the size of organizational articulatory units. Yes, there is compelling evidence that children do not start out with articulatory units of phonemic size, but more likely with broader vocal motor schemes resembling syllabic frames (McCune & Vihman, 1987). And our findings provide additional support for initial holistic patterns that are further defined and internally structured across childhood. In look-ahead models, it is obvious that the extent of coarticulation mirrors the size of a particular processing unit, as features can only be spread upon segments that are simultaneously active at some processing step. When coarticulation is interpreted as a coproduction process implemented only at the final stage of articulation, not pre-planned and processed at an early phonological or phonetic stage within the speech production process, however, we cannot directly infer articulatory unit size from the extent of coarticulatory effects. The coproduction framework, therefore, contests the frequently drawn inference from coarticulatory extent to developmental unit size changes and calls one of the most frequent motivations for investigating children's coarticulation patterns into question.

5.3 Perspectives

In order to draw conclusions about the human speech production mechanism or their phonology in general, it is important to consider structurally or prosodically different languages (see also Kidd & Garcia, 2022). As outlined in chapter 5.2.3, a direct comparison of coarticulatory patterns and their developments between children acquiring American English and those acquiring Indian English would for example provide interesting insights in the relation between gestural timing relations and the rhythmic class of a language. With regard to reading fluency development, a comparison of coarticulation between the relatively transparent alphabetic German orthography and an opaquer one like English, or even one based on syllables (e.g., Japanese), would help disentangle pure practice-based developments of reading fluency from those related to growing awareness of possible phoneme-sized units triggered by reliable grapheme-to-phoneme relations. This could also provide insightful results for the discussion about the existence of phoneme-sized linguistic units and whether they are only a by-product of alphabetic literacy.

The aim to identify potential causes of the found changes in coarticulatory patterns across childhood calls for an investigation of speech as varying with skills and developments in other domains, rather than with age. While age can serve as a first approach to sketch developments across childhood, it is merely a mediator of the actual causes (see chapters 3.5.1 and 5.2). Though substantially more complex, an age-independent analysis of the relation between different motoric and cognitive skills (e.g., independent control of the tongue's subparts, phonemic awareness, reading proficiency) and coarticulation degree would be a valuable approach reflecting the idea of speech production as a dynamically developing system (Thelen, 1995). In Noiray, Popescu, et al. (2019) we made important first steps in this direction. However, the children were only grouped according to the investigated skills post hoc; a study targeting participant samples of different levels of the respective skills would be more powerful.

The empirical data of this dissertation was acquired with ultrasound tongue imaging providing a high spatio-temporal resolution of articulatory material while being safe and relatively easy to acquire. However, as outlined in Hoole and Pouplier (2017), the noisiness of the images renders automatic tongue contour tracking complicated and leads to very high efforts necessary in data processing. To remain feasible, we drastically reduced our recorded material to single points in time and space for the earlier empirical investigations (Noiray et al., 2018; Noiray, Popescu, et al., 2019; Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2018, 2020). While this approach surely generates valuable insights into tongue motion, it, unfortunately, does not do justice to the dynamic nature of spoken language. In chapter 4 of this dissertation, we focused more on dynamic developments within an utterance by considering substantially more frames of each stimulus and looking at change over time. This theoretically grounded decision resulted in an enormous amount of ultrasound video frames to be manually corrected after automatic tracking. To neither reduce sample size nor drastically cut down the laboriously acquired articulatory material, future studies using ultrasound tongue imaging, especially those investigating children's speech, should opt for reliable automatic processing like the principle components analysis (PCA) approach suggested by Hoole and Pouplier (2017) to avoid massive labor-intensive manual processing. Only when a suitable mechanism for the processing of children's tongue data is found, it will be possible to efficiently implement the promising method of ultrasound tongue imaging to the field of clinical diagnostics with regard to atypical developments of speech fluency, speech motor control, and reading disorders.

6 CONCLUSION

The empirical work presented in this dissertation provides evidence for a developmental decrease of lingual vocalic coarticulation, as measured in anticipatory and perseveratory horizontal movements of the tongue dorsum. Reasons for this change in speech production are likely to be numerous and driving forces may dynamically interact with each other and vary in strength in the course of development (Thelen, 1995). Here, we addressed aspects of speech motor control maturation, a growing inhibition capacity possibly influencing the sequencing of gestures (Bjorklund & Harnishfeger, 1990; Tilsen, 2013), as well as representational changes, specifically the change in width of underlying gestural activation, possibly driven by literacy acquisition (Popescu & Noiray, 2021; empirical investigation in chapter 4). However, especially at the young age, physical growth of the vocal tract and associated relational changes between articulators (Vorperian et al., 2009) may play an important role. Similarly, vocabulary growth (Cychosz et al., 2021; Noiray, Popescu, et al., 2019), as well as effects of production frequency and speech motor practice (Tomaschek et al., 2021; Tomaschek, Tucker, Fasiolo, & Baayen, 2018) have been shown to affect coarticulation and may have a stake in decreasing articulatory overlap across childhood.

Importantly, the found developmental decrease of coarticulatory degree poses a problem to speech production models that rely on pre-planning and complex translation mechanisms from the underlying segments to their implemented form in the vocal tract (cf. Redford, 2019). Why should young children, who otherwise tend to strongly simplify speech, be better able to actively plan ahead? Instead, the coproduction framework (C. A. Fowler, 1980) offers an explanation of the found developmental change by ascribing context-effects to low-level interactions of temporally overlapping coordinative constraints during the implementation of underlying linguistic segments in the vocal tract. Here, the decrease of coarticulatory degree can be envisioned as a compression of vocalic activation curves. The hypothesis that stressed vowels have a special status in speech and serve as anchors in children's speech perception as well as production has been affirmed in a variety of studies (see chapter 3.2.3). It is likely that children start out using these anchors (or combinations of these stressed vowels with a preceding consonant in babbling-like CV syllables) as an entry to fluent speech production and only gradually tune articulatory movements for supplementing consonants and unstressed vowels. Within this perspective, the higher degree of coarticulation goes hand in hand with the tendency to simplify speech productions on the surface, as described via vocal motor schemes and phonological processes (chapter 1.1). The seesaw between efficiency and intelligibility in children often seems to tilt towards the efficiency side – specifically, articulatory effort is reduced both by assimilating places of consonantal articulation, as well as by subsuming and coproducing other segments within the broad vocal tract configuration governed by the stressed vowels.

The investigations presented in this dissertation imply that independent motoric control of the tongue's subparts is not given from early childhood but needs to be developed. They also provide evidence for human phonological representations to be intrinsically timed and overlapped during speech production. Processes of coarticulation, therefore, are one further example of how we can draw conclusions about human behavior by closely investigating ontogenetic developments. Advances in language acquisition that are eagerly awaited and trigger fascination in family members and friends (even outside the academic community), are an important source for linguistic theory and may help uncover some of the specifics of human speech and language.

7

REFERENCES

Abakarova, D., Fuchs, S., & Noiray, A. (2022). Developmental Changes in Coarticulation Degree Relate to Differences in Articulatory Patterns: An Empirically Grounded Modeling Approach. *Journal of Speech, Language, and Hearing Research*, *65*(9), 3276–3299. doi: 10.1044/2022_JSLHR-21-00212

Abakarova, D., Iskarous, K., & Noiray, A. (2017). Quantification of coarticulation resistance in German with ultrasound. *Stem-, Spraak- En Taalpathologie: 7th International Conference on Speech Motor Control Groningen*, 42.

Abercrombie, D. (1964). Syllable quantity and enclitics. In D. Abercrombie, D. B. Fry, P. A. MacCarthy, N. C. Scott, & J. L. Trim (Eds.), *In Honour of Daniel Jones* (Longman, pp. 216–222). London.

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*(4), 621–635. doi: 10.1017/S0142716403000316

Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M., & Perkell, J. (2013). Speech planning as an index of speech motor control maturity. *14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, 1278–1282.

Barbier, G., Perrier, P., Payan, Y., Tiede, M. K., Gerber, S., Perkell, J. S., & Ménard, L. (2020). What anticipatory coarticulation in children tells us about speech motor control maturity. *Plos One*, *15*(4), e0231484.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models; 2015. Preprint. Available from: arXiv:1506.04967. Cited 17 April 2018.*

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4. Preprint. Available from: arXiv:1406.5823. Cited 17 April 2018.*

Baum, S. R., & Waldstein, R. S. (1991). Perseveratory coarticulation in the speech of profoundly hearing-impaired and normally hearing children. *Journal of Speech and Hearing Research*, *34*, 1286–1292.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*(4), 591–627.

Berry, J. (2011). Speaking Rate Effects on Normal Aspects of Articulation: Outcomes and Issues. *Perspectives on Speech Science and Orofacial Disorders*, *21*(1), 15–26. doi: 10.1044/ssod21.1.15

Bertinetto, P. M., & Bertini, C. (2008). On modeling the rhythm of natural languages. *Proceedings of the 4th International Conference on Speech Prosody, SP 2008*, (Cc), 427–430.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641–1660.

Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review*, *10*(1), 48–71. doi: 10.1016/0273-2297(90)90004-N

Bladon, R. A. W., & Al-Bamerni, A. (1976). Coarticulation Resistance in English /l/. *Journal of Phonetics*, *4*(2), 137–150.

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer (Version 6.0.20) [Computer program]. Available from http://www.praat.org/.*

Bonte, M., Correia, J. M., Keetels, M., Vroomen, J., & Formisano, E. (2017). Reading-induced shifts of perceptual speech representations in auditory cortex. *Scientific Reports*, *7*(1), 1–11. doi: 10.1038/s41598-017-05356-3

Boucher, K. R. (2007). *Patterns of anticipatory coarticulation in adults and typically developing children.* Brigham Young University-Provo.

Boyer, N., & Ehri, L. C. (2011). Contribution of Phonemic Segmentation Instruction With Letters and Articulation Pictures to Word Reading and Spelling in Beginners. *Scientific Studies of Reading*, *15*(5), 440–470. doi: 10.1080/10888438.2010.520778

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*(1986), 219–252. doi: 10.1017/s0952675700000658

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(2), 201–251. doi: 10.1017/S0952675700001019

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

Browman, C. P., & Goldstein, L. (1992b). "Targetless" schwa: An articulatory analysis. *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, 26–56.

Brügelmann, H. (1992). *Kinder auf dem Weg zur Schrift* (4th ed.). Bottighofen: Libelle Verlag.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, *4*(1), 59–74. doi: 10.1016/S0095-4470(19)31222-7

Butz, M. V, Sigaud, O., & Gérard, P. (2003). Anticipatory behavior: Exploiting knowledge about the future to improve current behavior. In *Anticipatory behavior in adaptive learning systems* (pp. 1–10). Springer.

Bybee, J., & McClelland, J. L. (2005). *Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition.*

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209–244.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, *26*(2), 173–199.

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A Cross-Linguistic, Longitudinal Study of the Foundations of Decoding and Reading Comprehension Ability. *Scientific Studies of Reading*, *23*(5), 386–402. doi: 10.1080/10888438.2019.1580284

Chafcouloff, M., & Marchal, A. (1999). Velopharyngeal coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 69–79). Cambridge: Cambridge University Press.

Chang, S.-E., Ohde, R. N., & Conture, E. G. (2002). Coarticulation and formant transition rate in young children who stutter. *Journal of Speech, Language, and Hearing Research*, *45*, 676–688. Retrieved from http://dx.doi.org/10.1016/j.jfludis.2011.04.007

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English.* New York: Harper and Row.

Cole, J., Linebaugh, G., Munson, C., & McMurray, B. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics*, *38*(2), 167–184.

Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review*, *6*(3), 456–463.

Crystal, D. (1994). Documenting rhythmical change. In J. Windsor Lewis (Ed.), *Studies in General and English Phonetics* (pp. 174–179). London: Routledge.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, *21*, 101–108.

Cychosz, M., Munson, B., & Edwards, J. R. (2021). Practice and Experience Predict Coarticulation in Child Speech. *Language Learning and Development*, *17*(4), 366–396. doi: 10.1080/15475441.2021.1890080

Daniloff, R. G., & Hammarberg, R. E. (1973). On defining coarticulation. *Journal of Phonetics*, *1*(3), 239–248. doi: 10.1016/s0095-4470(19)31388-9

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, *16*(4), 234–244. doi: 10.1038/nrn3924

Donovan, A., & Darwin, C. J. (1979). The perceived rhythm of speech. *Proceedings of the Ninth International Congress of Phonetic Sciences*, *2*, 268–274. University of Copenhagen, Institute of Phonetics Copenhagen.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, *47*(2), 421–436.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, *18*(2), 116–125. doi: 10.1111/j.1467-9817.1995.tb00077.x

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [rl] distinction by young infants. *Perception & Psychophysics*, *18*(5), 341–347.

El Mogharbel, C., & Deutsch, W. (2007). 1.2 Von der Stimme zur Sprache: Die Ontogenese von Phonetik, Phonologie und Prosodie. *Sonderpädagogik Der Sprache*, 19.

Elsen, H. (1996). Two routes to language: stylistic variation in one child. *First Language*, *16*, 141–158.

Fant, G. (1970). *Acoustic theory of speech production*. Walter de Gruyter.

Farnetani, E. (1999). Labial coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 144–163). Cambridge: Cambridge University Press.

Farnetani, E., & Recasens, D. (1999). Coarticulation models in recent speech production theories. *Coarticulation: Theory, Data and Techniques*, 31–65.

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. In Joseph S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 36–57). Lawrence Erlbaum Associates.

Fikkert, P. (1994). *On the acquisition of prosodic structure*. Leiden University, Leiden, The Netherlands.

Flege, J. E. (1988). Anticipatory and carry-over nasal coarticulation in the speech of children and adults. *Journal of Speech and Hearing Research*, *31*(December), 525–536. doi: 10.1044/jshr.3104.525

Forster, K. I., & Davis, C. (1991). The Density Constraint on Form-Priming in the Naming Task: Interference Effects from a Masked Prime. *Journal of Memory and Language*, *30*, 1–25. doi: 10.1016/0749-596X(91)90008-8

Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. *Phonological Processes in Literacy: A Tribute to Isabelle Y. Liberman*, *106*, 97–117.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*(1), 113–133. doi: 10.1016/S0095-4470(19)31446-9

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech, Language, and Hearing Research*, *24*, 127–139. doi: 10.1044/jshr.2401.127

Fowler, C. A. (2015). The segment in articulatory phonology. In E. Raimy & C. E. Cairns (Eds.), *The segment in phonetics and phonology* (pp. 24–43). Wiley-Blackwell Chichester.

Fowler, C. A., & Brancazio, L. (2000). Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech*, *43*(1), 1–41. doi: 10.1177/00238309000430010101

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, *36*(2–3), 171–195. doi: 10.1177/002383099303600304

Fowler, C. A., Shankweiler, D., & Studdert-Kennedy, M. (2016). Perception of the Speech Code Revisited: Speech Is Alphabetic After All. *Psychological Review*, *123*(2), 125–150. doi: http://dx.doi.org/10.1037/rev0000013

Fowler, C. A., & Smith, M. R. (1986). Speech Perception as "Vector Analysis": An Approach to the Problems of lnvariance and Segmentation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123–139). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fox, B., & Routh, D. K. (1975). Analyzing spoken language into words, syllables, and phonomes: A developmental study. *Journal of Psycholinguistic Research*, *4*(4), 331–342.

Fox, A. V, & Dodd, B. J. (1999). Der Erwerb des phonologischen Systems in der deutschen Sprache [The phonological acquisition of German]. *Sprache-Stimme-Gehör*, *23*(4), 183.

Frith, U. (1986). A Developmental Framework for Developmental Dyslexia. *Annals of Dyslexia*, *36*, 69–81.

Frith, U. (1998). Literally changing the brain. *Brain*, *121*(6), 1011–1012. doi: 10.1093/brain/121.6.1011

Gafos, A., & Goldstein, L. (2012). Articulatory representation and organization. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 220–231). Oxford: Oxford University Press.

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language.* Psychology Press.

Gibson, M., Bunta, F., Johnson, C., & Huárriz, M. (2022). The production of English syllable-level timing patterns by bilingual English-and Spanish-speaking children with cochlear implants and their peers with normal hearing. *Journal of Phonetics*, *95*, 101194.

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. doi: 10.1044/1092-4388(2007/008)

Gleitman, L. R., & Wanner, E. (1982). The state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, UK: Cambridge University Press.

Goffman, L., Smith, A., Heisler, L., & Ho, M. (2008). The breadth of coarticulatory units in children and adults. *Journal of Speech, Language, and Hearing Research.* doi: 10.1044/1092-4388(2008/07-0020)

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in Phonetics and Speech Science*, 239–250.

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, Language, and Hearing Research*, *36*(4), 707–727. doi: 10.1044/jshr.3604.707

Goozée, J., Murdoch, B., Ozanne, A., Cheng, Y., Hill, A., & Gibbon, F. (2007). Lingual kinematics and coordination in speech-disordered children exhibiting differentiated versus undifferentiated lingual gestures. *International Journal of Language & Communication Disorders*, *42*(6), 703–724.

Goswami, U. (2000). Phonological representations, reading development and dyslexia: Towards a cross-linguistic theoretical framework. *Dyslexia*, *6*(2), 133–151.

Goswami, U. (2001). Early phonological development and the acquisition of literacy. *Handbook of Early Literacy Research*, *1*, 111–125.

Goswami, U., & Bryant, P. (2016). *Phonological skills and learning to read.* Taylor & Francis.

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.

Green, J. R., Nip, I. S. B., & Maassen, B. (2010). Some organization principles in early speech development. *Speech Motor Control: New Developments in Basic and Applied Research*, *10*, 171–188.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, *105*(4), 611.

Hardcastle, W. J., & Hewlett, N. (2006). *Coarticulation: Theory, data and techniques.* Cambridge: Cambridge University Press.

Hardcastle, W. J., Jones, W., Knight, C., Trudgeon, A., & Calder, G. (1989). New developments in electropalatography: A state-of-the-art report. *Clinical Linguistics & Phonetics*, *3*(1), 1–38.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation.* (Doctoral Dissertation). Massachusetts Institute of Technology Cambridge.

Hertrich, I., & Ackermann, H. (1995). Coarticulation in slow speech: Durational and spectral analysis. *Language and Speech*, *38*(2), 159–187. doi: 10.1177/002383099503800203

Hockett, C. F. (1955). *A manual of phonetics.* Bloomington, IN: Indiana University Press.

Hodge, M. M. (1989). *A comparison of spectral temporal measures across speaker age: Implications for an acoustic characterization of speech maturation.* (Doctoral dissertation). University of Wisconsin-Madison.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*(3), 262–274. doi: 10.1016/j.infbeh.2009.03.004

Hoole, P., Gobl, C., & Ní Chasaide, A. (1999). Laryngeal coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 105–143). Cambridge: Cambridge University Press.

Hoole, P., & Nguyen, N. (1999). Electromagnetic articulography in coarticulation research. *Coarticulation: Theory, Data and Techniques*, 260–269.

Hoole, P., & Pouplier, M. (2017). Öhman returns: New horizons in the collection and analysis of imaging data in speech production research. *Computer Speech & Language*, *45*, 253–277.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The Complex Nature of Reading Fluency: A Multidimensional View. *Reading & Writing Quarterly*, *25*(1), 4–32. doi: 10.1080/10573560802491208

Iskarous, K., Fowler, C. A., & Whalen, D. H. (2010). Locus equations are an acoustic expression of articulator synergy. *The Journal of the Acoustical Society of America*, *128*(4), 2021–2032. doi: 10.1121/1.3479538

Iuzzini-Seigel, J., Hogan, T. P., Rong, P., & Green, J. R. (2015). Longitudinal development of speech motor control: Motor and linguistic factors. *Journal of Motor Learning and Development*, *3*(1), 53–68.

Joos, M. (1948). Acoustic phonetics. *Language*, *24*(2), 5–136.

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, Language, and Hearing Research*, *34*(6), 1222–1232. doi: 10.1044/jshr.3406.1222

Kawamoto, A. H., Liu, Q., & Kello, C. T. (2015). The segment as the minimal planning unit in speech production and reading aloud: Evidence and implications. *Frontiers in Psychology*, *6*(September), 1–6. doi: 10.3389/fpsyg.2015.01457

Keating, P. A. (1988). The window model of coarticulation: articulatory evidence. *UCLA Working Papers in Phonetics*, *69*, 3–29.

Kelso, J. A. S., & Tuller, B. (1983). "Compensatory articulation" under conditions of reduced afferent information: A dynamic formulation. *Journal of Speech, Language, and Hearing Research*, *26*(2), 217–224.

Kent, R. D. (1983). The segmental organization of speech. In P. MacNeilage (Ed.), *The production of speech* (pp. 57–89). New York: Springer.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735.

Kisler, T., Schiel, F., & Sloetjes, H. (2012). *Signal processing via web services: The use case WebMAUS. Talk presented at Digital Humanities Conference.* Hamburg, Germany.

Kolinsky, R., Navas, A. L., Vidigal de Paula, F., Ribeiro de Brito, N., de Medeiros Botecchia, L., Bouton, S., & Serniclaes, W. (2021). The impact of alphabetic literacy on the perception of speech sounds. *Cognition*, *213*(March), 104687. doi: 10.1016/j.cognition.2021.104687

Krasotkina, A., Götz, A., Höhle, B., & Schwarzer, G. (2021). Perceptual narrowing in face-and speech-perception domains in infancy: A longitudinal approach. *Infant Behavior and Development*, *64*, 101607.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, *70*(2), 340–349. doi: 10.1121/1.386782

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. Retrieved from https://link.springer.com/content/pdf/10.3758/BF03212211.pdf

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, *73*(3), 1003–1010. doi: 10.1121/1.389148

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608. doi: 10.1126/science.1736364

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Nelson Education.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*, 315–334. doi: 10.1016/S0010-0277(97)00005-X

Laporte, C., & Ménard, L. (2018). Multi-hypothesis tracking of the tongue surface in ultrasound video recordings of normal and impaired speech. *Medical Image Analysis*, *44*, 98–114. doi: https://doi.org/10.1016/j.media.2017.12.003

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *The Journal of the Acoustical Society of America*, *54*(5), 1228–1234.

Liberman, A. M. (1989). Reading is hard just because listening is easy. *Brain and Reading*, 197–205.

Liberman, A. M. (1991). Observations from the sidelines. *Reading and Writing*, *3*(3–4), 429–433. doi: 10.1007/BF00354972

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*(2), 201–212.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, *35*(11), 1773–1781.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Orlando, FL: Academic Press.

Lleó, C., El Mogharbel, C., & Prinz, M. (1994). Babbling und Frühwort-Produktion im Deutschen und Spanischen. *Linguistische Berichte*, (151), 191–217.

Lotto, A. J., & Holt, L. (2000). The illusion of the phoneme. In S. J. Billings, J. P. Boyle, & A. M. Griffith (Eds.), *Chicago linguistic society* (pp. 191–204). Chicago.

Maas, E., & Mailend, M.-L. (2017). Fricative Contrast and Coarticulation in Children With and Without Speech Sound Disorders. *American Journal of Speech-Language Pathology*, *26*, 649–663. doi: 10.1044/2017_AJSLP-16-0110

Macken, M. A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua*, *49*, 11–49. doi: 10.1016/0024-3841(79)90073-1

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *25*(2), 187–205.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology*, *19*(23), 1994–1997. doi: 10.1016/j.cub.2009.09.064

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America*, *88*(3), 1286–1298.

Martens, B. K., Werder, C. S., Hier, B. O., & Koenig, E. A. (2013). Fluency Training in Phoneme Blending : A Preliminary Study of Generalized Effects. *Journal of Behavioral Education*, *22*(1), 16–36. doi: 10.1007/s10864-012-9159-8

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. doi: 10.3758/s13428-011-0168-7

MATLAB. (2016). *MATLAB and Statistics Toolbox Release, The MathWorks, Inc., Nattick, Massachusetts, United States.*

McCune, L., & Vihman, M. M. (1987). Vocal Motor Schemes. *Papers and Reports on Child Language Development*, *26*, 72–79.

McLeod, S., Van Doorn, J., & Reed, V. A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology*, *10*(2), 99–110.

Ménard, L., & Noiray, A. (2011). The development of lingual gestures in speech: Experimental approach to language development. *Faits de Langues*, *37*, 189–202.

Ménard, L., Schwartz, J.-L., & Boë, L.-J. (2004). Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1059–1080.

Menn, L., Schmidt, E., & Nicholas, B. (2013). Challenges to theories, charges to a model: The Linked-Attractor model of phonological development. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches and cross-linguistic evidence* (pp. 460–502). Cambridge, UK: Cambridge University Press.

Menn, L., & Vihman, M. M. (2011). Features in child phonology: Inherent, emergent, or artefacts of analysis? In N. Clements & R. Ridouane (Eds.), *Where do phonological features come from? The nature and sources of phonological primitives* (pp. 261–301). Amsterdam, The Netherlands: John Benjamins.

Modarresi, G., Sussman, H. M., Lindblom, B., & Burlingame, E. (2004). An acoustic analysis of the bidirectionality of coarticulation in VCV utterances. *Journal of Phonetics*, *32*(3), 291–312.

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese-und Rechtschreibtest; Weiterentwicklung des Salzburger Lese-und Rechtschreibtests (SLRT).* Huber.

Moll, K., Wallner, R., & Landerl, K. (2012). Kognitive Korrelate der Lese-, Leserechtschreib- und der Rechtschreibstörung. *Lernen Und Lernstörungen*, *1*(1), 7–19. doi: 10.1024/2235-0977/a000002

Mousikou, P., Rastle, K., Besner, D., & Coltheart, M. (2015). The locus of serial processing in reading aloud: Orthography-to-phonology computation or speech planning? *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(4), 1076–1099. doi: 10.1037/xlm0000090

Moyo, G. P. K., & Tetsiguia, J. R. M. (2020). Discussing the "First Cry" as an Initial Assessment for Neonates. *American Journal of Pediatrics*, *6*(2), 129–132.

Murdoch, B. E., & Goozée, J. (2003). EMA analysis of tongue function in children with dysarthria following traumatic brain injury. *Brain Injury*, *17*(1), 79–93. doi: 10.1080/0269905021000010203

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766. doi: 10.1037/0096-1523.24.3.756

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*(2), 203–230. doi: 10.1418/10879

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. *The Blackwell Companion to Phonology*, 1–13.

Nicholson, H., Munson, B., Reidy, P., & Edwards, J. (2015). Effects of age and vocabulary size on production accuracy and acoustic differentiation of young children's sibilant fricatives. *Proceedings of the International Congress on Phonetic Sciences.*

Nijland, L., Maassen, B., Van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, *16*(6), 461–483. doi: 10.1080/02699200210159103

Nip, I. S. B., Arias, C. R., Morita, K., & Richardson, H. (2017). Initial observations of lingual movement characteristics of children with cerebral palsy. *Journal of Speech, Language, and Hearing Research*, *60*(6S), 1780–1790.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech, Language, and Hearing Research*, *36*(5), 959–972. doi: 10.1044/jshr.3605.959

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language, and Hearing Research*, *32*(1), 120–132. doi: 10.1044/jshr.3201.120

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, Language, and Hearing Research*, *39*(2), 379–389. doi: 10.1044/jshr.3902.379

Nittrouer, S., & Whalen, D. H. (1989). The perceptual effects of child–adult differences in fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*, *86*(4), 1266–1276.

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. doi: 10.1044/2018_JSLHR-S-17-0148

Noiray, A., Cathiard, M.-A., Abry, C., & Ménard, L. (2010). Lip rounding anticipatory control: Crosslinguistically lawful and ontogenetically attuned. *Speech Motor Control: New Developments in Basic and Applied Research*, 153.

Noiray, A., Cathiard, M.-A., Abry, C., Ménard, L., & Savariaux, C. (2008). Emergence of a vowel gesture control: Attunement of the anticipatory rounding temporal pattern in French children. In S. Kern, F. Gayraud, & E. Marsico (Eds.), *Emergence of Language Abilities* (pp. 100–117). New Castle: Cambridge Scholars Publishing.

Noiray, A., Cathiard, M.-A., Ménard, L., & Abry, C. (2011). Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers.

The Journal of the Acoustical Society of America, *129*(1), 340–349. doi: 10.1121/1.3518452

Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *Journal of the Acoustical Society of America*, *133*(1), 444–452. doi: 10.1121/1.4763983

Noiray, A., Popescu, A., Killmer, H., Rubertus, E., Krüger, S., & Hintermeier, L. (2019). Spoken language development and the challenge of skill integration. *Frontiers in Psychology*, *10*, 2777. doi: https:/doi.org/10.3389/fpsyg.2019.02777

Noiray, A., Ries, J., & Tiede, M. (2015). Sonographic & Optical Linguo-Labial Articulation Recording system (SOLLAR). *Proceedings of Ultrafest VII, Hong Kong*, 8–9.

Noiray, A., Ries, J., Tiede, M., Rubertus, E., Laporte, C., & Ménard, L. (2020). Recording and analyzing kinematic data in children and adults with SOLLAR: Sonographic & Optical Linguo-Labial Articulation Recording system. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1), 14. doi: http://doi.org/10.5334/labphon.241

Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (2019). Back from the future: non-linear anticipation in adults' and children's speech. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 3033–3054. doi: 10.1044/2019_JSLHR-S-CSMC7-18-0208

Ohala, J. J. (1981). The listener as a source of sound change. In C. Masek, R. A. Hendrik, & M. F. Miller (Eds.), *Proceedings of the Chicago Linguistics Society 17* (pp. 178–203). Chicago.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, *39*(1), 151–168. doi: 10.1121/1.1909864

Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America*, *41*, 310–320.

Oller, D. K., & Eilers, R. E. (1992). Development of vocal signaling in human infants: Toward a methodology for cross-species vocalization comparisons. In H. Papusek, U. Jürgens, & M. Papousek (Eds.), *Nonverbal vocal communication* (pp. 174–191). Cambridge: Cambridge University Press New York.

Ostry, D. J., Gribble, P. L., & Gracco, V. L. (1996). Coarticulation of jaw movements in speech production: Is context sensitivity in speech kinematics centrally planned? *Journal of Neuroscience*, *16*(4), 1570–1579. doi: 10.1523/JNEUROSCI.16-04-01570.1996

Parush, A., Ostry, D. J., & Munhall, K. G. (1983). A kinematic study of lingual coarticulation in VCV sequences. *The Journal of the Acoustical Society of America*, *74*(4), 1115–1125.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*(5571), 1321–1323.

Pattamadilok, C., Knierim, I. N., Kawabata Duncan, K. J., & Devlin, J. T. (2010). How does learning to read affect speech perception? *Journal of Neuroscience*, *30*(25), 8435–8444. doi: 10.1523/JNEUROSCI.5791-09.2010

Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research*, *1275*, 73–80.

Perrier, P., & Fuchs, S. (2015). Motor equivalence in speech production. *The Handbook of Speech Production*, 223–247.

Piaget, J. (1952). *The origins of intelligence in children.* New York: Norton.

Pike, K. L. (1945). *The Intonation of American English* (U. of M. Press, Ed.). Ann Arbor.

Popescu, A., & Noiray, A. (2021). Learning to Read Interacts with Children's Spoken Language

Fluency. *Language Learning and Development*, *18*(2), 1–20. doi: 10.1080/15475441.2021.1941032

Port, R. F. (2010). Rich memory and distributed phonology. *Language Sciences*, *32*(1), 43–55.

Pouplier, M. (2011). The atoms of phonological representations. In M. van Oostendorp, K. Rice, B. Hume, & C. Ewen (Eds.), *The Blackwell companion to phonology*. Wiley-Blackwell.

R Core Team. (2015). R: A language and environment for statistical computing. R *Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org/.* (n.d.).

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 482–503. doi: 10.1037/0096-1523.25.2.482

Rastle, K., Harrington, J., Coltheart, M., & Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(3), 1178–1191. doi: 10.1037/0096-1523.26.3.1178

Recasens, D. (1984a). V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustical study. *Journal of Phonetics*, *12*(1), 61–73.

Recasens, D. (1984b). Vowel-to-vowel coarticulation in Catalan VCV sequences. *The Journal of the Acoustical Society of America*, *76*(6), 1624–1635. doi: 10.1121/1.391609

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech*, *28*(2), 97–114. doi: 10.1177/002383098502800201

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V: Coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics*, *15*, 299–312.

Recasens, D. (1999). Lingual coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (Vol. 4, pp. 80–104). Cambridge: Cambridge University Press.

Recasens, D. (2002). An EMA study of VCV coarticulatory direction. *The Journal of the Acoustical Society of America*, *111*(6), 2828–2841.

Recasens, D. (2018). Coarticulation. In *Oxford Research Encyclopedia of Linguistics*.

Recasens, D., Fontdevila, J., & Pallarès, M. D. (1995). Velarization degree and coarticulatory resistance for /l/ in Catalan and German. *Journal of Phonetics*, *23*(1–2), 37–52.

Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America*, *102*(1), 544–561.

Recasens, D., & Rodríguez, C. (2016). A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics*, *59*, 58–75. doi: 10.1016/j.wocn.2016.09.002

Redford, M. A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics*, *53*, 141–152. doi: 10.1016/j.wocn.2015.06.006

Redford, M. A. (2019). Speech production from a developmental perspective. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 2946–2962. doi: 10.1044/2019_JSLHR-S-CSMC7-18-0130

Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *The Journal of the Acoustical Society of America*, *79*(5), 1616–1619. doi: 10.1121/1.393298

Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition*, *32*(2), 212–222.

Rubertus, E., & Noiray, A. (2018). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. *PloS One*, *13*(9), e0203562. doi: 10.1371/journal.pone.0203562

Rubertus, E., & Noiray, A. (2020). Vocalic activation width decreases across childhood: Evidence from carryover coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1), 7. doi: http://doi.org/10.5334/labphon.228

Saito, M., Tomaschek, F., & Baayen, H. (2020). Relative functional load determines co-articulatory movements of the tongue tip. *12th International Seminar on Speech Production (Online Conference)*, 210–213.

Saletta, M. (2019). Orthography and speech production in children with good or poor reading skills. *Applied Psycholinguistics*, *40*, 905–931. doi: 10.1017/S0142716419000055

Saletta, M., Goffman, L., & Brentari, D. (2016). Reading skill and exposure to orthography influence speech production. *Applied Psycholinguistics*, *37*(2), 411–434. doi: 10.1017/S0142716415000053

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Series*, *15*, 129–144.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*(4), 499–526.

Sanguineti, V., Laboissiére, R., & Payan, Y. (1997). A control model of human tongue movements in speech. *Biological Cybernetics*, *77*(1), 11–22. doi: 10.1007/s004220050362

Scheerer-Neumann, G. (1989). Rechtschreibschwäche im Kontext der Entwicklung. In I. Naegele & R. Valtin (Eds.), *LRS in den Klassen 1-10* (pp. 25–35). Weinheim: Beltz.

Scheerer-Neumann, G. (2007). Rechtschreiben. In J. Walter & F. B. Wember (Eds.), *Handbuch Sonderpädagogik, Bd. 2, Sonderpädagogik des Lernens* (pp. 539–568). Göttingen: Hogrefe.

Scheerer-Neumann, G. (2018). *Lese-Rechtschreib-Schwäche und Legasthenie: Grundlagen, Diagnostik und Förderung*. Kohlhammer Verlag.

Schulz, G. M., Stein, L., & Micallef, R. (2001). Speech motor learning: Preliminary data. *Clinical Linguistics & Phonetics*, *15*(1–2), 157–161.

Scobbie, J. M., Lawson, E., & Stuart-Smith, J. (2012). Back to front: a socially-stratified ultrasound tongue imaging study of Scottish English/u. *Rivista Di Linguistica/Italian Journal of Linguistics, Special Issue:" Articulatory Techniques for Sociophonetic Research"*, *24*(1), 103–148.

Scobbie, J. M., & Pouplier, M. (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English/l. *Journal of Phonetics*, *38*(2), 240–259.

Shaiman, S. (2001). Kinematics of compensatory vowel shortening: the effect of speaking rate and coda composition on intra- and inter-articulatory timing. *Journal of Phonetics*, *29*, 89–107. doi: 10.1121/1.427640

Shaiman, S., Adams, S. G., & Kimelman, M. D. Z. (1995). Timing relationships of the upper lip and jawacross changes in speaking rate. *Journal of Phonetics*, *23*(1–2), 119–128. doi: 10.1016/S0095-4470(95)80036-0

Sinnott, J. M., & Mosteller, K. W. (2001). A comparative assessment of speech sound discrimination in the Mongolian gerbil. *The Journal of the Acoustical Society of America*, *110*(4), 1729–1732. doi: 10.1121/1.1398055

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*(4), 779–798.

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, *39*(5), 331–349.

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

Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., & Ménard, L. (2013). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *Journal of Phonetics*, *41*(3–4), 281–295. doi: 10.1016/j.wocn.2013.03.004

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1), 3–45.

Stoel-Gammon, C. (1998). Role of babbling and phonology in early linguistic development. *Transitions in Prelinguistic Communication*, *7*, 87–110.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, *19*(6–7), 455–501.

Story, B. H. (2005). A parametric model of the vocal tract area function for vowel and consonant simulation. *The Journal of the Acoustical Society of America*, *117*(5), 3231–3254.

Strycharczuk, P., & Scobbie, J. M. (2017). Fronting of Southern British English high-back vowels in articulation and acoustics. *The Journal of the Acoustical Society of America*, *142*(1), 322–331. doi: 10.1121/1.4991010

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. In Joseph S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 58–84). Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1987). The Phoneme as a Perceptuomotor Structure. *Haskins Laboratories: Status Report on Speech Research*, *91*, 45–57.

Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. *Studies in the Evolution of Language*, *3*, 235–254.

Sussman, H. M., Duder, C., Dalston, E., & Cacciatore, A. (1999). An acoustic analysis of the development of CV coarticulation: A case study. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1080–1096.

Sussman, H. M., Hoemeke, K. A., & McCaffrey, H. A. (1992). Locus equations as an index of coarticulation for place of articulation distinctions in children. *Journal of Speech, Language, and Hearing Research*, *35*(4), 769–781.

Sussman, H. M., Mccaffrey, H. A., Matthews, S. A., Sussman, H. M., Mccaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, *90*(3), 1309–1325. doi: 10.1121/1.401923

Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2009). Computational neural modeling of speech motor control in childhood apraxia of speech (CAS). *Journal of Speech, Language, and Hearing Research*, *52*(6), 1595–1609.

Terband, H., Van Brenk, F., Van Lieshout, P., Nijland, L., & Maassen, B. (2009). Stability and composition of functional synergies for speech movements in children and adults. *Interspeech 2009*, 788–791.

Thelen, E. (1995). Motor development: A new synthesis. *American Psychologist*, *50*(2), 79.

Tilsen, S. (2013). A dynamical model of hierarchical selection and coordination in speech planning. *PloS One*, *8*(4), e62800.

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, *55*, 53–77. doi: 10.1016/j.wocn.2015.11.005

Tilsen, S. (2018). *Three mechanisms for modeling articulation: selection, coordination, and intention*. Cornell Working Papers in Phonetics and Phonology. 1–50.

Tomaschek, F., Arnold, D., Bröker, F., & Baayen, R. H. (2018). Lexical frequency co-determines the speed-curvature relation in articulation. *Journal of Phonetics*, *68*, 103–116.

Tomaschek, F., Arnold, D., Sering, K., Tucker, B. V, van Rij, J., & Ramscar, M. (2021). Articulatory variability is reduced by repetition and predictability. *Language and Speech*, *64*(3), 654–680.

Tomaschek, F., Tucker, B. V, Fasiolo, M., & Baayen, R. H. (2018). Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard*, *4*(s2), 20170018.

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.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition*, *100*(1), 73–99.

Treiman, R., & Zukowski, A. (1996). Children's sensitivity to syllables, onsets, rimes, and phonemes. *Journal of Experimental Child Psychology*, *61*(3), 193–215.

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. doi: 10.1515/lp-2013-0005

Ushijima, T., & Sawashima, M. (1972). Fiberscopic observation of velar movements during speech. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics, University of Tokyo*, *6*, 25–38.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, *21*(1–2), 61–82. doi: 10.1016/s0095-4470(19)31321-x

Vihman, M. M. (2016). Prosodic structures and templates in bilingual phonological development. *Bilingualism*, *19*(1), 69–88. doi: 10.1017/S1366728914000790

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*(1), 1–27.

Vihman, M. M., & Croft, W. (2007). *Phonological development: Toward a "radical" templatic phonology.*

Vihman, M. M., & Velleman, S. L. (1989). Phonological reorganization: A case study. *Language and Speech*, *32*(2), 149–170. doi: 10.1177/002383098903200204

Volenec, V. (2015). Coarticulation. In J. Davis (Ed.), *Phonetics: Fundamentals, potential applications, and role in communicative disorders* (pp. 47–86). New York: Nova.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, *117*(1), 338–350.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., … Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*, *125*(3), 1666–

1678.

Waldstein, R. S., & Baum, S. R. (1991). Anticipatory coarticulation in the speech of profoundly hearing-impaired and normally hearing children. *Journal of Speech, Language, and Hearing Research*, *34*(6), 1276–1285. doi: 10.1044/jshr.3406.1276

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. doi: 10.1016/S0163-6383(84)80022-3

Werwach, A., Mürbe, D., Schaadt, G., & Männel, C. (2021). Infants' vocalizations at 6 months predict their productive vocabulary at one year. *Infant Behavior and Development*, *64*(May), 0–5. doi: 10.1016/j.infbeh.2021.101588

Westfall, P. H. (1997). Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association*, *92*(437), 299–306.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, *18*, 3–35.

Wickelgren, W. A. (1969). Context-Sensitive Coding in Speech Recognition, Articulation and Developments. In *Information processing in the nervous system* (pp. 85–96). doi: 10.1037/h0026823

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. doi: 10.1016/j.wocn.2018.03.002

Wimmer, H. (1996). The early manifestation of developmental dyslexia: Evidence from German children. *Reading and Writing*, *8*, 171–188.

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, *1*(1), 7–18.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 3–36. doi: 10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized additive models: an introduction with* R. Chapman and Hall/CRC.

Zharkova, N. (2017). Voiceless alveolar stop coarticulation in typically developing 5-year-olds and 13-year-olds. *Clinical Linguistics & Phonetics*, *31*(7–9), 503–513. doi: 10.1080/02699206.2016.1268209

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, *15*(1), 118–140. doi: 10.1123/mcj.15.1.118

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, *42*(2), 193–208. doi: 10.1017/S0025100312000060

Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory and Cognition*, *32*(5), 732–741. doi: 10.3758/BF03195863

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. doi: 10.1037/0033-2909.131.1.3

Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud : An extension to German. *European Journal of Cognitive Psychology*, *12*(3), 413–430. doi: 10.1080/09541440050114570

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*(5), 379–384. doi: 10.1111/1467-9280.00370

Ziegler, W., & Von Cramon, D. (1985). Anticipatory coarticulation in a patient with apraxia of speech. *Brain and Language*, *26*(1), 117–130.

Zmarich, C., Biondi, A. R., Bonifacio, S., Busà, M. G., Colavolpe, B., Gaiotto, M., & Olivucci, F. (2021). Coarticulation and VOT in an Italian child from 18 to 48 months of age. *Book Series Studi AISV*, *8*, 413–435.

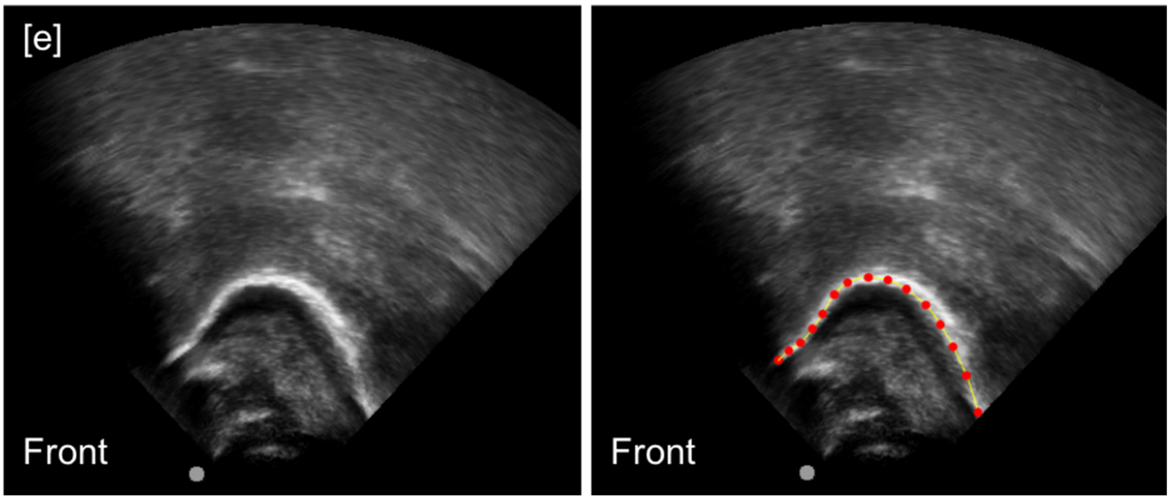
8 APPENDIX

Appendix A. 1. Summary of the number of analyzed trials per consonant context per age cohort in the study on anticipatory V-to-V coarticulation (chapter 2).

Cohort	Consonant Context	Number of trials
C3	əbV	534
	ədV	517
	əgV	485
C4	əbV	552
	ədV	566
	əgV	551
C5	əbV	522
	ədV	545
	əgV	509
C7	əbV	638
	ədV	655
	əgV	653
A	əbV	723
	ədV	723
	əgV	722

Cohort abbreviations are C3 – 3-year-old children, C4 – 4-year-old children, C5 – 5-year-old children, C7 – 7-year-old children, and A – adults.

Appendix A. 2. Example ultrasound image of the study on anticipatory V-to-V coarticulation (chapter 2).



Raw ultrasound image of a 5-year-old boy's tongue (CM5_005) at the temporal midpoint of the articulation of an [e] on the left and the semi-automatically labeled surface contour on top of the same frame on the right side. The tip of the tongue is to the left in both images.

Appendix A. 3. Model output for the vowel's effect on schwa in every consonant context for each cohort in the study on anticipatory V-to-V coarticulation (chapter 2).

Cohort	Consonant	β-coefficient	SE	t-value	*p*-value	
	b	0.654484	0.026293	24.892	<0.001	***
C3	d	0.615274	0.029026	21.197	<0.001	***
	g	0.621107	0.026886	23.101	<0.001	***
	b	0.587084	0.021988	26.700	<0.001	***
C4	d	0.560788	0.025054	22.384	<0.001	***
	g	0.54091	0.02448	22.094	<0.001	***
	b	0.584066	0.024605	23.738	<0.001	***
C5	d	0.537628	0.023936	22.461	<0.001	***
	g	0.588202	0.025618	22.961	<0.001	***
	b	0.530582	0.022336	23.755	<0.001	***
C7	d	0.501092	0.023329	21.480	<0.001	***
	g	0.57227	0.02453	23.332	<0.001	***
	b	0.16495	0.01577	10.462	<0.001	***
A	d	0.16034	0.01794	8.939	<0.001	***
	g	0.24961	0.01868	13.361	<0.001	***

Cohort abbreviations are C3 – 3-year-old children, C4 – 4-year-old children, C5 – 5-year-old children, C7 – 7-year-old children, and A – adults.

Appendix B. 1. Model codes and explanations for the study on carryover coarticulation (chapter 3).

Model m: Assessing coarticulatory patterns

```
m <- bam(peakX ~ te(time, VpeakX, k=c(4,10), by=cohort.consonant) + cohort.consonant +
             s(time, subject, by=consonant, bs='fs', m=1, k=4) +
             s(VpeakX, subject, by=consonant, bs='fs', m=1),
             data=datc, discrete=T, nthreads=32, rho=0.35, AR.start=datc$start.event)
```

Model m tests whether the horizontal position of the highest point on the tongue dorsum (*peakX*) depends on the horizontal position of the tongue dorsum during V50 (*VpeakX*) at the four target time points V100, C50, C100, and ə50. Within the tensor product (te) term, both predictors *time* and *VpeakX* as well as their interaction is included. The k parameter specifies the maximal non-linearity by setting the size of basis dimensions for both predictors. It is limited to the number of the predictors' unique points and therefore set to four for *time* and to the default value of 10 for *VpeakX*. The following by-parameter specifies the levels the non-linear patterns are fit for: Here, all 15 possible combinations of *age cohort* and *consonant* (i.e., three-year-olds-/b/, three-year-olds-/d/, three-year-olds-/g/, …, adults-/g/). Possible constant differences in the horizontal position of *peakX* between the age cohorts and consonants were considered by including the nominal variable *cohort.consonant* (the interaction between cohort and consonant). The random effect structure of the model, defined in the two factor smooth terms (s), included potentially non-linear patterns over *time* and for V*peakX* for each participant and consonant. In the final row of the model specification, the data set is defined (datc), a faster fitting method is employed, the number of processors used to run the model is specified, and autocorrelation (here at a level of about 0.35) in the data is accounted for.

Model mb7: Assessing differences of coarticulatory patterns between age cohorts

```
mb7 <- bam(peakX ~ te(time, VpeakX, k=c(4,10), by=consonant) + consonant +
              te(time, VpeakX, k=c(4,10), by=IsC3b) +
              te(time, VpeakX, k=c(4,10), by=IsC4b) +
              te(time, VpeakX, k=c(4,10), by=IsC5b) +
              te(time, VpeakX, k=c(4,10), by=IsAb) +
              te(time, VpeakX, k=c(4,10), by=IsC3d) +
              te(time, VpeakX, k=c(4,10), by=IsC4d) +
              te(time, VpeakX, k=c(4,10), by=IsC5d) +
              te(time, VpeakX, k=c(4,10), by=IsAd) +
              te(time, VpeakX, k=c(4,10), by=IsC3g) +
              te(time, VpeakX, k=c(4,10), by=IsC4g) +
              te(time, VpeakX, k=c(4,10), by=IsC5g) +
              te(time, VpeakX, k=c(4,10), by=IsAg) +
              s(time, subject, by=consonant, bs='fs', m=1, k=4) +
              s(VpeakX, subject, by=consonant, bs='fs', m=1),
              data=datc, discrete=T, nthreads=32, rho=0.35, AR.start=datc$start.event))
```

In contrast to Model m, model mb7 includes binary difference tensors to compare age cohorts to each other. Here, the variable *IsC3b* for example refers to a previously specified difference smooth for cohort C3 when the consonant is /b/. Since no difference smooth for cohort C7 is included in the model, this cohort is taken as reference.

Appendix C. 1. Model m1.c (generalized additive mixed model for children's data) in the study on coarticulation in reading (chapter 4).

m1.c: tonguepos ~ VCond + s(time, by=VCond, k=20) + s(time, by=C1) +
s(time, by=speed) +
s(time, subject, by=V, bs="fs", m=1, k=20) +
s(time, subject, by=Cond, bs="fs", m=1, k=20) +
s(time, subject, by=C1, bs="fs", m=1, k=20)

Appendix C. 2. Output of model m1c (study on coarticulation in reading (chapter 4)).

Parametric coefficients	**Estimate**	**Std. error**	**t-value**	**p-value**
(Intercept)	0.334993	0.013025	25.719	< 2e-16
Reading /u/	0.099716	0.008418	11.846	< 2e-16
Repetition /i/	-0.037553	0.016343	-2.298	0.0216
Repetition /u/	0.090198	0.017961	5.022	5.15e-07
Smooths	**Edf**	**Ref.df**	**F**	**p-value**
s(time): Reading /i/	0	0	0.012	0.99844
s(time): Reading /u/	9.964	11.785	60.301	< 2e-16
s(time): Repetition /i/	1.001	1.001	0	0.99229
s(time): Repetition /u/	9.97	11.791	39.245	< 2e-16
s(time): /b/	7.941	8.355	32.966	< 2e-16
s(time): /d/	8.178	8.515	28.888	< 2e-16
s(time): /g/	7.287	7.574	40.582	< 2e-16
s(time): fast	1	1	9.237	0.00237
s(time): slow	4.625	5.538	6.653	0.00000155
s(time, subject): /i/	142.451	639	0.346	< 2e-16
s(time, subject): /u/	224.878	639	0.857	< 2e-16
s(time, subject): Reading	217.884	559	1.039	< 2e-16
s(time, subject): Repetition	232.602	639	0.941	< 2e-16
s(time, subject): /b/	138.584	640	0.39	< 2e-16
s(time, subject): /d/	172.851	640	0.732	< 2e-16
s(time, subject): /g/	187.445	640	0.594	< 2e-16

Appendix C. 3. Model m2.c (generalized additive mixed model with ordered factors for children's data) in the study on coarticulation in reading (chapter 4).

m2.c: tonguepos ~ Cond + RuO + RepuO +
s(time, by=Cond, k=20) +
s(time, by=C1) +
s(time, by=speed)+
s(time, by=RuO, k=20) +
s(time, by=RepuO, k=20) +
s(time, subject, by=V, bs="fs", m=1) +
s(time, subject, by=Cond, bs="fs", m=1) +
s(time, subject, by=C1, bs="fs", m=1)

Appendix C. 4. Output of model m2.c (study on coarticulation in reading (chapter 4)).

Parametric coefficients	**Estimate**	**Std. error**	**t-value**	**p-value**
(Intercept)	0.334860	0.012986	25.787	< 2e-16
Repetition	-0.037395	0.016354	-2.287	0.0222
Reading /u/	0.099819	0.008382	11.909	< 2e-16
Repetition /u/	0.127628	0.008032	15.890	< 2e-16
Smooths	**edf**	**Ref.df**	**F**	**p-value**
s(time): Reading	1	1	16.96	0.0000384
s(time): Repetition	1	1	17.295	0.0000322
s(time): /b/	7.96	8.405	26.633	< 2e-16
s(time): /d/	8.191	8.551	23.622	< 2e-16
s(time): /g/	7.313	7.615	32.582	< 2e-16
s(time): fast	1	1	5.942	0.01478
s(time): slow	3.512	4.464	4.549	0.00102
s(time): Reading /u/	**11.956**	**14.514**	**40.827**	**< 2e-16**
s(time): Repetition /u/	**11.576**	**14.107**	**40.852**	**< 2e-16**
s(time, subject): /i/	101.929	287	0.582	< 2e-16
s(time, subject): /u/	154.529	287	1.472	< 2e-16
s(time, subject): Reading	154.394	251	2.005	< 2e-16
s(time, subject): Repetition	170.05	287	1.899	< 2e-16
s(time, subject): /b/	100.834	288	0.655	< 2e-16
s(time, subject): /d/	135.898	288	1.477	< 2e-16
s(time, subject): /g/	142.212	288	1.123	< 2e-16

Appendix C. 5. Model m3.c (generalized additive mixed model with binary smooths for children's data) in the study on coarticulation in reading (chapter 4).

m3.c: tonguepos ~ Cond + s(time, by=Cond, k=20) + s(time, by=C1) +
s(time,by=speed) +
s(time, by=Isu, k=20) +
s(time, by=IsRepu, k=20) +
s(time, subject, by=V, bs="fs", m=1) +
s(time, subject, by=Cond, bs="fs", m=1) +
s(time, subject, by=C1, bs="fs", m=1)

Appendix C. 6. Output of model m3.c (study on coarticulation in reading (chapter 4)).

Parametric coefficients	**Estimate**	**Std. error**	**t-value**	**p-value**
(Intercept)	0.33486	0.01299	25.776	< 2e-16
Repetition	-0.03742	0.01636	-2.287	0.0222
Smooths	**edf**	**Ref.df**	**F**	**p-value**
s(time): Reading	1	1	16.019	6.35e-05
s(time): Repetition	1	1	16.44	5.08e-05
s(time): /b/	7.966	8.416	26.714	< 2e-16
s(time): /d/	8.186	8.553	23.803	< 2e-16
s(time): /g/	7.315	7.621	32.695	< 2e-16
s(time): fast	1	1	6.002	0.014296
s(time): slow	3.537	4.493	4.623	0.000873
s(time): /u/	14.484	16.986	44.903	< 2e-16
s(time): Repetition /u/	**3.588**	**4.197**	**7.229**	**6.25e-06**
s(time, subject): /i/	101.480	287	0.578	< 2e-16
s(time, subject): /u/	154.69	287	1.477	< 2e-16
s(time, subject): Reading	154.676	251	2.003	< 2e-16
s(time, subject): Repetition	170.689	287	1.911	< 2e-16
s(time, subject): /b/	100.624	288	0.653	< 2e-16
s(time, subject): /d/	135.697	288	1.474	< 2e-16
s(time, subject): /g/	142.285	288	1.125	< 2e-16

Appendix C. 7. Model m1.a (generalized additive mixed model for adult's data) in the study on coarticulation in reading (chapter 4).

m1.a tonguepos ~ VCond + s(time, by=VCond, k=20) + s(time, by=C1) +
s(time, by=speed) +
s(time, subject, by=V, bs="fs", m=1, k=20) +
s(time, subject, by=Cond, bs="fs", m=1, k=20) +
s(time, subject, by=C1, bs="fs", m=1, k=20)

Appendix C. 8. Output of model m1.a (study on coarticulation in reading (chapter 4)).

Parametric coefficients	**Estimate**	**Std. error**	**t-value**	**p-value**
(Intercept)	0.30091	0.01737	17.326	< 2e-16
Reading /u/	0.12736	0.01174	10.847	< 2e-16
Repetition /i/	-0.01917	0.01834	-1.045	0.296
Repetition /u/	0.11565	0.02167	5.336	9.65e-08
Smooths	**edf**	**Ref.df**	**F**	**p-value**
s(time): Reading /i/	0	0	0.037	0.998258
s(time): Reading /u/	6.885	8.146	41.582	< 2e-16
s(time): Repetition /i/	1	1	0.133	0.715655
s(time): Repetition /u/	5.875	7.052	33.57	< 2e-16
s(time): /b/	6.845	7.364	11.051	< 2e-16
s(time): /d/	7.06	7.592	8.427	< 2e-16
s(time): /g/	6.516	6.915	14.613	< 2e-16
s(time): fast	1	1	11.128	0.000852
s(time): slow	3.939	4.729	4.407	0.000750
s(time, subject): /i/	130.405	319	0.999	< 2e-16
s(time, subject): /u/	120.780	319	0.905	< 2e-16
s(time, subject): Reading	116.277	319	1.06	< 2e-16
s(time, subject): Repetition	66.864	319	0.443	< 2e-16
s(time, subject): /b/	103.416	320	0.882	< 2e-16
s(time, subject): /b/	95.795	320	1.337	< 2e-16
s(time, subject): /b/	116.827	320	0.964	< 2e-16

Appendix C. 9. Model m2.a (generalized additive mixed model with ordered factors for adults' data) in the study on coarticulation in reading (chapter 4).

m2.a: tonguepos ~ Cond + RuO + RepuO +
s(time, by=Cond, k=20) +
s(time, by=C1) +
s(time, by=speed) +
s(time, by=RuO, k=20) +
s(time, by=RepuO, k=20) +
s(time, subject, by=V, bs="fs", m=1) +
s(time, subject, by=Cond, bs="fs", m=1) +
s(time, subject, by=C1, bs="fs", m=1)

Appendix C. 10. Output of model m2.a (study on coarticulation in reading (chapter 4)).

Parametric coefficients	Estimate	Std. error	t-value	p-value
(Intercept)	0.30083	0.01736	17.327	< 2e-16
Repetition	-0.01914	0.01831	-1.046	0.296
Reading /u/	0.12742	0.01181	10.793	< 2e-16
Repetition /u/	0.13492	0.01198	11.264	< 2e-16
Smooths	**edf**	**Ref.df**	**F**	**p-value**
s(time): Reading	1	1	0.114	0.7358
s(time): Repetition	0	0	0.068	0.9964
s(time): /b/	7.024	7.585	11.748	< 2e-16
s(time): /d/	6.753	7.363	12.119	< 2e-16
s(time): /g/	7.373	7.848	12.837	< 2e-16
s(time): fast	0	0	0.023	0.9993
s(time): slow	4.629	5.546	1.963	0.0813
s(time): Reading /u/	**7.603**	**9.49**	**25.06**	**< 2e-16**
s(time): Repetition /u/	**5.926**	**7.31**	**29.896**	**< 2e-16**
s(time, subject): /i/	90.02	143	1.969	< 2e-16
s(time, subject): /u/	75.383	143	1.4	< 2e-16
s(time, subject): Reading	84.583	143	2.15	< 2e-16
s(time, subject): Repetition	45.226	143	0.758	< 2e-16
s(time, subject): /b/	75.144	144	1.682	< 2e-16
s(time, subject): /d//	71.144	144	2.724	< 2e-16
s(time, subject): /g/	75.492	144	1.692	< 2e-16

Appendix C. 11. Model m3.a (generalized additive mixed model with binary smooths for adults' data) in the study on coarticulation in reading (chapter 4).

m3.a: tonguepos ~ Cond + s(time, by=Cond, k=20) + s(time, by=C1) +
s(time, by=speed) +
s(time, by=Isu, k=20) +
s(time, by=IsRepu, k=20) +
s(time, subject, by=V, bs="fs", m=1) +
s(time, subject, by=Cond, bs="fs", m=1) +
s(time, subject, by=C1, bs="fs", m=1)

Appendix C. 12. Output of model m3.a (study on coarticulation in reading (chapter 4)).

Parametric coefficients	**Estimate**	**Std. error**	**t-value**	**p-value**
(Intercept)	0.30060	0.01731	17.366	< 2e-16
Repetition	-0.01967	0.01827	-1.076	0.282
Smooths	**Edf**	**Ref.df**	**F**	**p-value**
s(time): Reading	1	1	0.135	0.714
s(time): Repetition	0	0	0.046	0.998
s(time): /b/	7.149	7.666	11.054	< 2e-16
s(time): /d/	6.9	7.464	11.616	< 2e-16
s(time): /g/	7.455	7.893	12.467	< 2e-16
s(time): fast	0	0	0	0.5
s(time): slow	3.974	4.821	1.707	0.12
s(time): /u/	11.081	13.487	26.436	< 2e-16
s(time): Repetition /u/	**2.374**	**2.636**	**0.584**	**0.476**
s(time, subject): /i/	89.605	143	1.967	< 2e-16
s(time, subject): /u/	75.451	143	1.414	< 2e-16
s(time, subject): Reading	83.544	143	1.987	< 2e-16
s(time, subject): Repetition	45.781	143	0.712	< 2e-16
s(time, subject): /b/	75.362	144	1.481	< 2e-16
s(time, subject): /d/	72.039	144	1.971	< 2e-16
s(time, subject): /g/	76.251	144	1.741	< 2e-16