Back from the future

Noiray, Aude; Wieling, Martijn; Abakarova, Dzhuma; Rubertus, Eline; Tiede, Mark

Back From the Future: Nonlinear Anticipation in Adults' and Children's Speech

Aude Noiray,[a,b] Martijn Wieling,[b,c] Dzhuma Abakarova,[a] Elina Rubertus,[a] and Mark Tiede[b]

Purpose: This study examines the temporal organization of vocalic anticipation in German children from 3 to 7 years of age and adults. The main objective was to test for nonlinear processes in vocalic anticipation, which may result from the interaction between lingual gestural goals for individual vowels and those for their neighbors over time.
Method: The technique of ultrasound imaging was employed to record tongue movement at 5 time points throughout short utterances of the form V1#CV2. Vocalic anticipation was examined with generalized additive modeling, an analytical approach allowing for the estimation of both linear and nonlinear influences on anticipatory processes.
Results: Both adults and children exhibit nonlinear patterns of vocalic anticipation over time with the degree and extent of vocalic anticipation varying as a function of the individual consonants and vowels assembled. However, noticeable developmental discrepancies were found with vocalic anticipation being present earlier in children's utterances at 3–5 years of age in comparison to adults and, to some extent, 7-year-old children.
Conclusions: A developmental transition towards more segmentally-specified coarticulatory organizations seems to occur from kindergarten to primary school to adulthood. In adults, nonlinear anticipatory patterns over time suggest a strong differentiation between the gestural goals for consecutive segments. In children, this differentiation is not yet mature: Vowels show greater prominence over time and seem activated more in phase with those of previous segments relative to adults.

Anticipation is a ubiquitous characteristic of motor programming (e.g., visual saccades: Zingale & Kowler, 1987; writing: Gentner, 1983; walking: Thelen & Smith, 1994), which plays a crucial role in movement dynamics (e.g., Lashley, 1951; Nadin, 2014). Given a motor goal (e.g., grasping a glass), anticipation expresses individuals' ability to use past experiences to predict (or anticipate) future events and build suitable motor responses (e.g., generating an appropriate hand trajectory for gripping

a full vs. empty glass vs. gripping a twig vs. a stone). Hence, in motor research, anticipation is taken to reflect the degree of adaptability and, importantly for the developmental field, of the way motor patterns can be learnt by individuals. As children gain more experience with a given goal in various contexts, the achievement of the goal-directed action is supposed to become more efficient and automatized (review in Butz, Sigaud, & Gérard, 2003).

In speech, anticipation is also a fundamental property of articulatory dynamics. It is commonly investigated via measures of the temporal binding between articulatory gestures, that is, through coarticulatory processes (Browman & Goldstein, 1992). As in other motor activities, speech anticipation reflects the interplay between planning processes (i.e., the selection of phonemic units together with their corresponding motor schemes) and their physical execution as coordinative structures that implement meaningful, syntactically structured utterances. The more practical experience with a given speech goal in various phonetic environments (e.g., a lingual constriction gesture for the vowel /i/ in different consonantal environments), the more proficient the anticipatory patterns are likely to be. For

[a]Laboratory for Oral Language Acquisition, Department of Linguistics, University of Potsdam, Germany
[b]Haskins Laboratories, New Haven, CT
[c]Center for Language and Cognition, University of Groningen, the Netherlands

Correspondence to Aude Noiray: anoiray@uni-potsdam.de

instance, in adults, frequent words have been associated with greater articulatory practice (Tomaschek, Arnold, Bröker, & Baayen, 2018) and pseudowords produced repeatedly were found to increase movement speed and decrease in variability (Tiede, Mooshammer, Goldstein, Shattuck-Hufnagel, & Perkell, 2011).

In this study, we were interested in the maturation of anticipatory mechanism in the speech of typically developing children. Given cumulative findings identifying deficiency in the temporal organization of speech gestures as a core symptom of certain developmental disorders (e.g., childhood apraxia of speech: Maas & Mailend, 2017; Maas, Robin, Wright, & Ballard, 2008; McNeil, Ballard, Duffy, & Wambaugh, 2017; Sussman, Marquardt, & Doyle, 2000; stuttering: Chang, Ohde, & Conture, 2002; Hardcastle & Tjaden, 2008; Lenoci, 2017; Walsh, Mettel, & Smith, 2015), understanding how anticipatory processes are implemented in the gestural organization of typically developing children's speech has become an increasingly significant research avenue for developmental theories of speech production and for clinical applications. Kinematic studies of anticipatory coarticulation in typically developing children have, for a long time, focused on examining labial anticipation because of the lips being more accessible for measurement than the tongue (e.g., Goffman, Smith, Heisler, & Ho, 2008; Noiray, Cathiard, Abry, Ménard, & Savariaux, 2008, 2011; Smith & Goffman, 1998; Smith & Zelaznik, 2004). It was found, for instance, that vocalic labial rounding can be initiated well ahead of the acoustically defined temporal domain of the vowel (Noiray, Cathiard, Abry, & Ménard, 2010). More recently, the optimization of ultrasound imaging to the developmental field has led to an explosion of lingual coarticulation studies in childhood (to cite only a few examples, in American English: Song, Demuth, Shattuck-Hufnagel, & Ménard, 2013; in Canadian French: Ménard & Noiray, 2011; Noiray, Ménard, & Iskarous, 2013; in German: Noiray, Abakarova, Rubertus, Krüger, & Tiede, 2018; in Scottish English: Zharkova, 2017; Zharkova, Hewlett, & Hardcastle, 2011, 2012).

In this study, we focused on the expression of anticipation over the course of short utterances to investigate two levels of gestural and linguistic organization: the intrasyllabic (or local) anticipatory coarticulation between a consonant and a vowel (CV) and the intersyllabic (or long-distance) coarticulation in vowel–consonant–vowel sequences across a word boundary (i.e., schwa#CV). A few important points are worth mentioning prior to reviewing the research relevant to this study. First, the expressions of local and long-distance lingual anticipation have mostly been examined separately in both adults and children, creating the artificial assumption they are two separate mechanisms. While both anticipatory processes may be related to different cognitive and gestural mechanisms (e.g., one is planned, and the other results from online gestural coproduction), they may not, at least, in young children. Unless local and long-distance anticipatory processes are examined together within the same population and with the same analytical approaches, the question of whether those are indeed two

fundamentally different processes or, on the contrary, must be considered within a single organizational scheme that is dynamically organized over time will remain unsolved. Second, knowledge about long-distance anticipatory organization remains relatively fragmented in comparison to local anticipation that has generated much more empirical interest. This discrepancy leaves many questions about organizational units open. Third, given the heterogeneity in empirical approaches and findings, various theoretical positions regarding the maturational trajectory of anticipatory process have flourished in the last decades. In the next section, we review the research that has specifically looked into developmental differences in coarticulatory organization and, when possible, relate them to similar findings at the representational level.

The Question of Units of Coarticulatory Organization

In the last half century, developmental psycholinguists, like archaeologists, have dissected children's early spoken forms in search of their primitive form. They have developed meticulous transcription procedures, speech error labeling, acoustic, and kinematic measurements of child speech to retrace the ontogenetic trajectory of coarticulatory organization. With recent technical advances, it has been possible to collect speech data in younger children and respond to the need of quantification and in-depth analyses of child language. However, whether children's organization of speech gestures corresponds to smaller or greater unit sizes compared to those of adults remains a difficult question to address, not only for practical reasons but also because of its theoretical complexity.

In fact, the question of the units of language organization is relevant across various domains pertaining to language in adults (see recent discussion in Caudrelier, Schwartz, Perrier, Gerber, & Rochet-Capellan, 2018) and its development in children, for instance, speech sound/ word processing production. Their maturation occurs during the same developmental window (albeit at different paces) and interacts over time in a nonlinear fashion (e.g., recognition stimulating production and vice versa between 10 and 12 months: DePaolis, Vihman, & Nakai, 2013). In a recent in-depth review of the question, Vihman describes the intricate relation between production and comprehension mechanisms as follows: "Do infants begin by learning speech sounds and then combine them to recognize and produce words? Or do they begin by producing word-like vocalizations and retaining bits of the speech signal that match their production? Or do these processes occur in parallel?" (Vihman, 2017, p. 1).

Based on previous empirical research including ours, three contrasting hypotheses emerge regarding the size and nature of the speech units employed by the young learner. Some studies support large units of spoken language organization (e.g., syllable, words, or prosodic phrases; hereinafter the holistic approach); some rather suggest an initially segmentally driven organization (the segmental approach), and finally, a body of research including ours argues that

both more segmental and more syllabic organizations may be found in children with gradients of coarticulation degree depending on the gestural demands associated with consecutive segments (the gestural approach). Note that this classification can only provide a simplified summary of a very rich and heterogenic literature.

The Holistic Approach

In favor of a holistic approach to coarticulatory organization is the finding of a greater vocalic influence on previous consonants resulting in greater coarticulation degree between consonants and vowels in children as compared to adults' productions (or local anticipation, e.g., Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer & Whalen, 1989). This result has been taken as evidence for an initially broad temporal organization of speech gestures in chunks from the size of the syllable with a gradual decrease in gestural overlap and of coarticulation degree with age. Similar findings were reported on the breadth of long-distance vowel-to-vowel anticipation (review in Rubertus & Noiray, 2018). For instance, Nijland et al. (2002) found a developmental decrease in long-distance vowel anticipation in six children aged 5–7 years. This trend was supported in a more quantitative investigation with 42 children aged 3, 4, and 5 years and 14 adults by Boucher (2007) as well as by Nittrouer, Studdert-Kennedy, and Neely (1996) in 30 American English children 3, 5, and 7 years old and adults. In the latter study, greater local CV anticipation was found in the same children tested than in adults. Interestingly, the view of large-sized units of language organization has been documented in research addressing infants' production of prosodic grouping in early word production (e.g., Snow, 1998), processing of prosodic units (Jusczyk, Cutler, & Redanz 1993; review in Speer & Ito, 2009), word learning (review in Vihman, 2017), and word-based production errors (review in Vihman & Croft, 2007), as well as in syllabic segmentation (Nazzi, Mersad, Sundara, Iakimova, & Polka, 2014). These findings (among others) suggest that lexical development is the backbone of phonological development (see discussions in Beckman & Edwards, 2000, and Edwards, Beckman, & Munson, 2004).

Turning to the implication of large-sized coarticulatory units for speech motor development, the holistic view suggests that children may exhibit interarticulator gestalts (e.g., Menn, 1983; Nittrouer, 1993) that are initially lexically driven (e.g., Keren-Portnoy, Majorano, & Vihman, 2009; Vihman & Velleman, 1989), that is, limited to segment combinations present in already acquired words. With the gradual expansion of the lexical repertoire, children may develop greater precision in existing articulatory coordination and greater independency of individual articulators for the coarticulation of new or less familiar segment combinations.

The Segmental Approach

The segmental approach to coarticulatory organization results from the opposite finding, that is, a relatively low coarticulation degree in children as compared to adults (e.g., Barbier et al., 2015; Kent, 1983; Whiteside & Hodgson,

2000). In this view, lingual gestures for consonants and vowels are produced rather independently from each other, and maturation of coarticulatory organization entails an increase in gestural cohesion for both segments. As regards long-distance vowel-to-vowel anticipation, a few studies employing formant frequency analyses of schwa#CV sequences have provided empirical evidence for a rather segmental organization of speech in the early years of life with an increase in segmental overlap with age (e.g., Repp's [1986] investigation of two American English daughters and their father as well as Hodge's 1989 investigation of 10 children and adults). This trend was later supported in Canadian French for some 4-year-old children whose lingual coarticulatory patterns were measured with the technique of ultrasound imaging (Barbier et al., 2015). However, for some other children of the same age, the opposite trend of greater vocalic anticipation was found with respect to adults. This result is important because it suggests that, at 4 years of age, anticipatory patterns are not uniform across children and that individual variability is a characteristic feature of developing spoken language fluency.

Regarding speech motor control, the segmental approach favors the view of a more incremental development of articulatory controls such that it is initially driven by segmental goals and the early support of the jaw as main achiever of speech goals (e.g., review in Green & Nip, 2010). Articulatory control later extends to broader phonological structures with the development of differentiated controls over other speech organs (e.g., the lips, the tongue) as well as their precise coordination over time (e.g., Green, Moore, & Reilly, 2002; Katz, Kripke, & Tallal, 1991; Kent, 1983). This view is congruent with a large body of research demonstrating infants' early segmental processing skills (e.g., categorical perception of consonants and vowels: Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984; sensibility to transitional probabilities: Saffran, Aslin, & Newport, 1996; see also the results of a meta-analysis: Bergmann, Tsuji, & Cristia, 2017; or in children's speech error patterns including segmental deletion or exchange: McLeod & Bleile, 2003).

The Gestural Hypothesis

A third body of research leads to suggest another approach to coarticulatory organization, which we call the gestural approach in reference to the principles of articulatory phonology (Browman & Goldstein, 1992). In this theoretical framework, gestural goals represent functional primitives of phonological organization conveying relevant information to the speech articulators (e.g., the tongue dorsum, the tongue tip) for units of various sizes to be assembled in speech (e.g., syllables and words). The developmental literature is replete with findings highlighting the role of articulatory gestures in language acquisition: in developmental psychology with research reporting early imitation of various language-related gestures in infants, with their capacity for self-correction (e.g., Meltzoff, 2007); in recent observations of a developmental increase in infants' attention to speakers' mouth when linguistically

relevant gestures are produced (e.g., babbling; de Boisferon, Tift, Minar, & Lewkowicz, 2018); in experimental phonetics with examples of between-/within-organ contrast distinctions (e.g., Goldstein, 2003; Studdert-Kennedy & Goldstein, 2003); and in perceptual studies with reports of poor discrimination of consonantal contrasts involving primary gestures from the tongue when movement from the tongue is restrained with a pacifier (Bruderer, Danielson, Kandhadai, & Werker, 2015).

Our recent research expands on existing evidence with insights on coarticulatory organization in the preschool age (Noiray et al., 2018; Rubertus & Noiray, 2018). Variations in how much consonants and vowels overlap within the time frame of a syllable (noted "coarticulation degree") were observed as a function of the identity of the onset consonant. While greater coarticulation degree was found in syllables involving a labial stop (e.g., with /b/), syllables including an alveolar onset (e.g., with /d/) exhibited lesser vocalic influence. This marked difference reflects the gestural (in)compatibility that affects the degree to which consecutive gestures can be coproduced with one another if they recruit the same speech organ (e.g., the tongue). The achievement of the labial consonantal gesture does not prevent the tongue dorsum gesture for the vowel to be coproduced during the temporal domain of the consonant, whereas the gestural goal for the alveolar stop /d/ requires a functional synergy between the tongue tip and the tongue dorsum to reach its target constriction in the alveolar region. This requirement prevents the tongue dorsum from setting in the position for the upcoming vowel early within the temporal domain of the consonant (e.g., Noiray et al., 2013). This phenomenon, coined *coarticulatory resistanc*e (Bladon & Al-Bamerni, 1976; Recasens, 1985), has been observed in numerous studies across languages in adults (in American English: Fowler, 1994; Fowler & Saltzman, 1993; Iskarous, Fowler, & Whalen, 2010; Australian languages: Graetzer, 2006; Canadian French: Noiray et al., 2013; Catalan: review in Recasens & Espinosa, 2009; German: Abakarova, Iskarous, & Noiray, 2018; Iskarous et al., 2013; Swedish: Lindblom & Sussman, 2012; Thai, Cairene Arabic, and Urdu: Sussman, Hoemeke, & Ahmed, 1993) as well as in children, albeit less extensively (e.g., in English: Gibson & Ohde, 2007; Katz & Bharadwaj, 2001; Munson, 2004; Smith & Goffman, 2004; Sussman, Duder, Dalston, & Cacciatore, 1999; Canadian French: Noiray et al., 2013; German: Noiray et al., 2018; Scottish: Zharkova, 2017; Zharkova et al., 2011).

Hence, our findings as well as those of others in the past suggest that vocalic anticipation in adults and children varies along a continuum, the magnitude of which is a function of whether articulatory gestures can be coproduced without affecting their respective perceptual intelligibility. Figure 1 provides an illustrative conceptualization of coarticulatory organization based on the findings reported in the literature. It represents coarticulation degree as a continuum along which various gradients of coarticulatory degree are simulated. Depending on the gestural compatibility between consecutive segments, coarticulatory organization can be viewed as more holistic (e.g., in CV sequences such as /bi/ allowing large coarticulatory overlap), or it can be more segmental when the physical organs recruited for adjacent consonantal and vocalic gestures compete with one another (e.g., /da/). In between, multiple gradients of coarticulatory overlap are also possible.

In summary, the gestural approach is not incompatible with current phonological perspectives on coarticulatory organization (as summarized in the holistic and segmental approaches). Instead, it reconciles various sets of findings that may a priori contradict each other but in fact characterize specific instances of coarticulatory organization among a variety of other possibilities. To our knowledge, developmental studies have not tested for differences in coarticulatory organization across an extended inventory of consonants and vowels because of children's limited ability to perform in long laboratory speech production tasks. Until quantitative investigations are conducted to determine whether children uniformly organize their speech in adult-based phonological categories (e.g., segments, syllables), the gestural approach provides a plausible scenario for explaining variation in coarticulatory degree with articulatory gestures being more flexible units of coarticulatory organization than phonological units. With this perspective, it is possible to explain a wider range of coarticulatory patterns across phonetic contexts, speech styles, or individuals. Importantly, it provides a unifying organizational scheme to relate adults' to children's patterns. How coarticulatory organization matures over time is then no longer solely a question of direction (toward a greater or lower coarticulatory degree) or categorical change in phonological organization (e.g., into segments or syllables) but a question of how a primitive gestural scheme shares similar tools (the articulators of speech), constraints, and principles (dynamic interarticulator coordination over time) with adults to instantiate complex phonetic combinations in line with the native language's phonological grammar. After all, before learning to read, children have had very little explicit knowledge of adults' units of phonological description such as segments and syllables. Yet, within a couple of years, they organize their speech in intelligible ways and display coarticulatory patterns in the direction of adults but not quite yet like adults. Intuitively, it seems counterproductive to learn to speak a language initiating one organizational scheme and move to a markedly different one rather than tuning an existing control system over time.

Why Another Study on Anticipatory Coarticulation

As highlighted in the previous section, well-defined relations between degree of gestural overlap and phonological organization have been hard to establish across developmental studies. Note that similar questions exist at the perception and representation level; however, those fall out of this study's scope (for a discussion of those, see for instance Hay, 2018). There are probably many reasons for the inconsistencies in these findings; some are obviously methodological, including large heterogeneity in experimental designs, stimuli, and analyses employed. Because

Figure 1. Illustration for gradients in coarticulatory degree between consecutive consonants (dotted circles) and vowels (crossed circles). Variations in coarticulation degree are represented along a continuum from large coarticulatory overlap between consonantal and vocalic gestures (i.e., more holistic organization) to instances involving coarticulatory resistance from the consonant (i.e., more segmental organization).
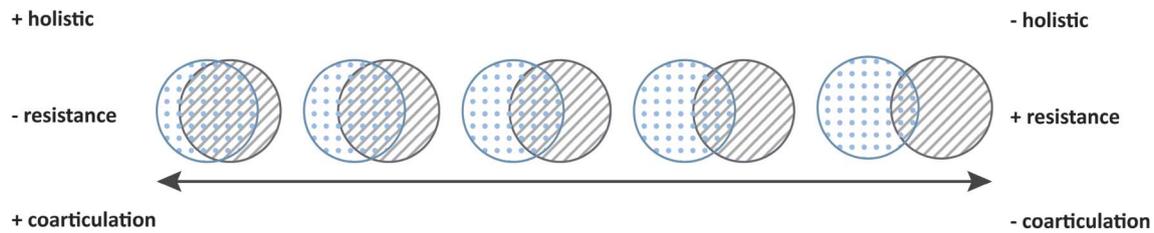
developmental research is often constrained in age span and sample size, it may be that studies extrapolate children's coarticulatory organization beyond the investigated age range. Given the nonmonotonic development of speech motor control (e.g., see Green, Nip, & Maassen, 2010, for a review), it may yet only characterize one of many developmental phases children undertake when learning to speak their language fluently. In addition to this confound, in the course of developing new skills, children may regress in performance for skills that have seemingly already been acquired. This phenomenon has been reported in the articulatory domain (e.g., temporary increase in variability for lip coupling during the lexical spurt at 2 years of age: Green et al., 2002; difference in lip–jaw coordination between 4 and 5 years of age: Smith & Zelaznik, 2004). It is not unique to language but pertains to other types of motor programming (e.g., walking: Thelen & Smith, 1994; hand coordination during the emergence of walking: Corbetta & Bojczyk, 2002; writing: Perret & Kandel, 2014).

This study responds to the necessity to examine the anticipatory process over time to elucidate possible nonlinearities in (a) how gestural goals are organized within the course of short utterances and (b) how this organization changes over developmental time. In two prior studies, we estimated the anticipatory imprint of a given vowel during the preceding consonant (Noiray et al., 2018) and schwa (Rubertus & Noiray, 2018) in short schwa#CV2 sequences uttered by German children (aged 3, 4, 5, and 7 years) and by adults. All groups of children exhibited both local and long-distance anticipation; however, we uncovered substantial developmental differences in spatiotemporal organization of tongue gestures with a greater degree of anticipatory coarticulation noted for the youngest cohorts in kindergarten (at 3–5 years of age) in comparison to school-aged children (at 7 years of age) and adults. One particularly intriguing pattern observed only in the examination of long-distance vocalic anticipation (i.e., in schwa#CV sequences) but not for local anticipation (i.e., CV sequences) motivated this study. While the degree of temporal overlap between an upcoming vowel and a preceding schwa varied significantly as a function of the medial consonant in adults, it did not at all in children: In their disyllables, target vowels were anticipated to the same degree regardless of the medial consonant. These separate results point at sharp differences

in children's organization of lingual gestures within as compared to beyond the syllabic frame. Whether the impact of consonantal gestures is restricted to the shorter temporal span of the syllable or modulates the degree of vocalic influence over more distant neighbors remains an open question in children and is not fully understood in adults. Importantly, these findings reaffirm that, while investigating how much children differ from adults at various ages is important for understanding the maturation of coarticulatory anticipation, examining why those differences occur has become even more imperative. Research addressing this question can tease apart contextual effects that are child independent (e.g., due to the [in]compatibility of vocalic and consonantal gestures) from maturational effects (e.g., control over tongue movement, differences in vocal tract anatomy or phonological representations; e.g., Ménard, Schwartz, & Boë, 2004) or highlight deviancy from typical trajectories (e.g., planning and phasing of speech gestures in childhood apraxia of speech; Nijland et al., 2002; Ziegler & Von Cramon, 1985).

Generalized Additive Measures to Account for Anticipation Over Time

A main conclusion in our previous investigation of intrasyllabic coarticulation degree in German (Noiray et al., 2018) was that the maturation of the coarticulatory mechanism may not consist in globally increasing or decreasing the magnitude of vocalic anticipation with age but in achieving fine-grained gradients of coarticulation degree depending on the gestural requirements associated with consecutive consonants. In that study, we had employed single time-point analyses; that is, we selected the midpoint of the consonant with respect to the vowel midpoint as a standard anchor representing its "steady" state. However, as colleagues in motor control research have commented: "Anticipation is an expression of change, i.e., of dynamics" (Nadin, 2014, p. 147; Bernstein, 2014). Reliably assessing the temporal organization of vocalic gestures over time requires accounting for time as a critical variable. Unfortunately, in many studies of coarticulation, including ours, the intrinsic dynamics of speech and of anticipation that expresses continuous change over time is estimated by single time-point analyses (e.g., simple linear regression or locus equation: Gibson & Ohde, 2007; Noiray et al., 2013;

Sussman et al., 1999; Sussman, Hoemeke, & McCaffrey, 1992) or linear mixed-effects models (e.g., Noiray et al., 2018; Rubertus & Noiray, 2018).

While research employing single time-point analyses has provided crucial insights on the maturation of coarticulatory processes, it may overlook complex features of movement patterns or paint a simplified picture that does not adequately reflect the reality of the underlying coarticulatory processes. In simple linear regression analyses, coarticulatory influences in CV syllables are measured via change in acoustic (e.g., F2) or articulatory (e.g., the tongue dorsum) parameters for a consonant across vocalic contexts. Linear relationships are therefore tested across syllables with the slope indicating the degree of coarticulation for a consonant across vocalic contexts and the correlation coefficient assessing the strength of the linear relationship observed. linear mixed model approaches are also useful in testing for significant differences in coarticulatory magnitude across given phonetic contexts but do not allow for analysis of dynamic (nonlinear) patterns over time(e.g., Wieling, 2018).

In this study, we expand on previous research by employing generalized additive modeling (GAM), a nonlinear regression method that is able to identify both linear and nonlinear patterns over time. In comparison to the methods mentioned above, GAM is hence more suitable to the fine-grained examinations of the speech dynamics, which is, by nature, continuous and variable. Importantly, this method also allows us to depart from standard measures of coarticulation expressing coarticulatory variation along a qualitative scaling (more/less *X* than *Y*) but instead look at interactions over time.

To assess the dynamics of anticipatory processes, we applied GAM with multiple time points. With this approach, we aimed to provide a finer-grained examination of how much the vocalic gesture impacts those of its neighbors and how long in advance it may be initiated in the speech stream.

Research Questions

The main objective of this study was to investigate variation in vowel anticipation over time in multiple age groups. We further examined whether the identity of the medial consonant impacts on the time course of the vocalic tongue gesture. This question was addressed within and between age groups. Given our previous findings in German (Noiray et al., 2018; Rubertus & Noiray, 2018), we predicted nonlinear trajectories of vocalic anticipation over time in adults to reflect the dynamical interaction between the lingual gestures for the target vowel and those of its consonantal neighbors. In children, especially in the kindergarten age, we did not expect such fine-grained interactions due to a lack of differentiation of tongue movement for consecutive gestural goals in comparison to adults or school-aged children.

Method
Participants

Seventy-four German native speakers all living in the Potsdam area (Brandenburg) were invited to take part in the study. We ensured none of the participants showed any regional influence on their speech. They were divided into four children age groups: nineteen 3-year-old children (10 girls, age range: 3;05–3;09 [years;months], *M* = 3;06), fourteen 4-year-old children (seven girls, age range: 4;04–4;08, *M* = 4;05), fourteen 5-year-old children (seven girls, age range: 5;04–5;07, *M* = 5;06), and fifteen 7-year-old children at the end of the first or beginning of the second grade in primary school (10 girls, age range: 7;00–7;06, *M* = 7;02). All children cohorts were selected from the large database of the Baby Lab at Potsdam University. They were enrolled in kindergarten and primary schools in Potsdam. For the purpose of this study, only participants with no known language-related, hearing-related, or visual problems were recruited.

The adult group of German speakers included 13 adults (seven women, age range: 19–28 years, *M* = 23 years). They were all living in the Potsdam and Berlin regions. We excluded participants with dialectal accent (e.g., from Bavaria). All participants, adults and children, were compensated for their participation in the study. Ethics approval was obtained from the ethics committee of the University of Potsdam.

Production Material

Trochaic pseudowords (i.e., conforming to German phonotactics) of the form schwa–consonant$_1$–vowel–consonant$_2$–schwa (C_1VC_2) were prerecorded by a native German female adult speaker and used as stimuli for a repetition task. Consonants used in both positions were /b/, /d/, and /g/. The vowel set consisted of the tense and long vowels /i/, /y/, /u/, /a/, /e/, and /o/. C_1Vs were designed as a fully crossed set of Cs and Vs. Target pseudowords were embedded in a carrier phrase with the article /aɪnə/ resulting in utterances such as /aɪnə biːdə/. In subsequent analyses, vocalic anticipation was estimated at four time points: midpoint and offset of the schwa in the article and midpoint and offset of the consonant prior to the full vowel of the pseudoword.

For all cohorts of children, trials were presented in six semirandomized blocks; for adults, nine blocks per participant were recorded. 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 the Appendix.

Experimental Procedure

The study took place at the Laboratory for Oral Language Acquisition at the University of Potsdam (Germany). Participants were recorded within the SOLLAR platform (Sonographic and Optical Linguo-Labial Articulation Recording system; Noiray, Ries, & Tiede, 2015). SOLLAR is a child-friendly custom-made platform for the recording and analysis of data from multiple sources (e.g., the tongue using ultrasound imaging with fps: 48 Hz, the lips using video camera with fps: 50 Hz, the audio speech signal via

microphone with fps: 48 KHz). It has been designed as a space rocket to be used with young children. To stimulate children's interest and motivation to complete the study, the production task was embedded in an interstellar journey. The ultrasound probe used for imaging the tongue is fixed in a custom-made probe holder that is integrated in the space rocket. It is flexible in the vertical dimension to follow natural speech-related vertical jaw movements but prevents lateral and horizontal motions. The probe is positioned below participants' chin between the maxillary bones to record the tongue surface contour in the midsagittal plane. In this study, additional head-to-probe stabilization was not employed to maximize the naturalness of speech and make the recording comfortable for young children. Trials during which participants moved were discarded subsequent to the recordings via visual inspection of the video data. All participants were recorded with the same equipment, except for the chair that differed between adults and children.

The production task was described to children as an interstellar journey during which children would repeat foreign words from the various planets they visited. For all participants, target words were arranged as randomized blocks and each block was associated with a mission. Upon completion of a block of target stimuli, children would complete a mission, get a reward, and travel to the next planet. With this experimental design, we stimulated children's curiosity and motivation for completing the study. For adults, the production task was presented as a repetition task without the child-friendly storyline.

Two experimenters were involved for each recording. The first one familiarized the participant with the SOLLAR platform and storyline for children. This experimenter maintained a face-to-face connection with the participant throughout the recording, controlled for head movement and correct pronunciation, and prompted the audio stimuli. The second experimenter operated SOLLAR's recording platform from a desk that was hidden from participants. The second experimenter also monitored both video and audio streams to control for the quality of the data collection. Both experimenters had experience with young children; they were also well trained with the equipment and the task. Prior to conducting the study, several pilot recordings were conducted to improve the setup and the storyline and to optimize the timing of the recording.

Data Processing

The acoustic signal was recorded together with the video from the ultrasound device and the video camera, enabling the generation of a common time code for subsequent data synchronization (via a cross-correlation function within MATLAB; cf. Noiray et al., 2013, 2015). First, the acoustic data were phonetically labeled using Praat (Boersma & Weenink, 1996). For adults, target words and segments were segmented semiautomatically using WebMAUSBasic (Kisler, Schiel, & Sloetjes, 2012) and manually corrected when necessary. For all children, native speakers of German manually labeled all target words and segments, using as

vocalic reference stable periodic cycles in the oscillogram and stable formant pattern, especially a clearly detectable second formant. In addition, the first ascending zero-crossing in the oscillogram at the beginning of the periodicity was used as schwa and vowel onset; the first ascending zero-crossing after the end of periodicity and disappearance of F2 was used as the beginning of the medial consonant. The output of the phonetic labeling was then used for the selection of the five relevant time points that provided measures for subsequent analyses (midpoint and offset of the schwa, midpoint and offset of the following consonant, midpoint of the target vowel).

Participants' productions that did not match the model speaker's word were discarded from further analysis, except for those of 3-year-old children. Given that kinematic data from young children are highly relevant for clinical outcomes but still scarce (five 2-year-olds: Song et al., 2013; seventeen 3-year-olds: Noiray et al., 2018, 2013), we opted for more flexibility in order to maximize quantification of anticipatory processes. We therefore used as many correctly produced CV syllables as possible, so words were kept as long as C_1V corresponded to the model speaker and C_2 did not differ in place of articulation from the model word (e.g., /aɪnə baːtə/ was kept for model /aɪnə baːdə/).

Ultrasound video frames corresponding to the five target time points (i.e., the midpoint and offset of the schwa, the midpoint and offset of the consonant, and the midpoint of the target vowel) were extracted automatically using the SOLLAR platform (Noiray et al., 2015). For each ultrasound frame, tongue contours were semiautomatically detected with scripts custom-made for MATLAB as part of the SOLLAR platform. For each ultrasound frame, a 100-point spline was automatically fit to the midsagittal tongue surface contour. x and y Coordinates for each of the 100 points of these splines were then automatically extracted. In this study, we used values for the highest point of the tongue dorsum surface contour in the x-coordinate reflecting the anterior–posterior position of the tongue dorsum.

Statistical Analyses

Preliminary Considerations

Before running statistical analyses, data were made comparable across participants. We set the most anterior position of the tongue dorsum position during all of the vowel pronunciations (at the midpoint of the vowel: V50) to 0 and the most posterior V50 position to 1. For all other relevant time points, tongue dorsum positions in the anterior–posterior dimension were scaled in this range (i.e., negative values or values greater than 1 are possible if there are more extreme positions or posterior positions of the tongue dorsum during the pronunciation of the consonant [or the schwa]). To assess potential nonlinear patterns over time, we used GAM. While this approach has been used to model the tongue's trajectories measured by electromagnetic articulography (Wieling, 2018; Winter & Wieling, 2016), to our knowledge, this is the first time GAM has been applied

to ultrasound tongue imaging data in the developmental field (but see Strycharczuk & Scobbie, 2017, in adults).

Testing for Consonantal and Age Differences in Vocalic Anticipation

The main goal in this study was to assess the influence of anticipatory coarticulation of the vowel on the preceding schwa and consonant. We predicted the anterior–posterior position of the tongue dorsum for each of the four time points (the midpoint of the schwa: schwa50, the offset of the schwa: schwa100, the midpoint of the consonant: C50, and the offset of the consonant: C100) on the basis of the anterior–posterior position of the tongue dorsum for the subsequent vowel (V50). Rather than analyzing the data for each of the preceding four time points separately, we explicitly looked for nonlinear patterns over these four time points. Of course, there is a limit to the amount of nonlinearity we are able to detect, given that there are only four time points, but the method will detect linear patterns if there is no support for a nonlinear pattern. We did not distinguish the vowel target in a categorical manner (i.e., /i, e, y, a, o, u/), but instead, we used the actual anterior–posterior position of the tongue dorsum during the pronunciation of the midpoint of the vowel as a numerical measure of the vowel target. Importantly, this allows us to investigate a nonlinear interaction between the two predictors, time and tongue dorsum position at V50. Because the pattern over time might be different depending on the target vowel (more specifically, the anterior–posterior position of the tongue dorsum during the midpoint of the vowel), we specifically test for a nonlinear interaction between time (i.e., the four time points preceding the vowel onset) and the anterior–posterior position of the tongue dorsum during the midpoint of the vowel (V50).

We were interested in two predictors: age group (3-, 4-, 5-, and 7-year-olds and adults) and the three consonants (/b, d, g/). For each combination of age group and consonant, we included a separate nonlinear interaction between time and V50 tongue dorsum position. While we might have included age as a numerical predictor, we decided against this, as there were large gaps between the age groups (especially between the 7-year-olds and the adults who had an average age of 23 years).

The specification of our first model was

m <- bam(PeakX ~ te(Time, VPeakX, k = c(4,10), by=Cohort.C) + Cohort.C + s(Time, Subject, by=C, bs="fs", m = 1, k = 4) + s(VPeakX, Subject, by=C, bs="fs", m = 1), data=dat, discrete=T, nthreads = 32, rho = 0.4, AR.start=dat$start.event).

To model the GAM, we used the function *bam* of the *mgcv* R package (Version 1.8-23; Wood, 2011; Wood & Wood, 2015). Our dependent variable was PeakX, which is the anterior–posterior position of the highest point on the tongue dorsum (peak) for each of the four time points (1: schwa50, 2: schwa100, 3: C50, and 4: C100). We predicted this value on the basis of a nonlinear interaction, which is modeled by a tensor product spline (te). A tensor product spline models both the (potentially) nonlinear effects across both predictors, Time and VPeakX, which is the anterior–posterior position of the peak at V50 (i.e., the target position of the tongue during the midpoint of the vowel, as well as their interaction; see Wieling, 2018, for a detailed explanation). The parameter k specifies the maximum nonlinearity in each of the two directions. It limits the nonlinearity as this specifies the maximum number of underlying functions (which are of increasing complexity; see Wieling, 2018), which may be combined to represent the complete nonlinear pattern. The value of k is limited by the number of unique points of each predictor and, for this reason, limited to 4 for the first predictor (Time) and set to the default value of 10 for the second predictor (VPeakX). The by-parameter allowed us to model different nonlinear interactions for each level of the nominal predictor (in this case, Cohort.C, which includes all 15 possible combinations of the age cohort and the consonant [i.e., 3-year-olds: /b/, 3-year-olds: /d/, 3-year-olds: /g/, …, adults: /g/]). Given that the nonlinear interactions were approximately centered (i.e., the mean value of each nonlinear interaction was approximately 0), we also included the nominal variable Cohort.C as a separate predictor to model potential constant differences in the anterior–posterior position of the peak for the different age groups and consonants. The final two s() blocks modeled the random-effects structure: For each individual subject, for each level of the consonant C, we allowed a nonlinear pattern over Time (the first block) and VPeakX using the so-called factor *smooths* (identifiable via bs="fs"). The k values were set equal to those in the general model specification, and the m parameter (set to 1) ensures that the random effects did not perfectly match the individual patterns but rather did account for shrinkage (i.e., the assumption that extreme observations are, in reality, a little bit less extreme: shrinkage toward the mean). The subsequent parameters of the function *bam* denote our data set (dat), a faster fitting method that employs discretization (i.e., binning of the numerical data to speed up the computation time; for this, the parameter discrete was set to TRUE), and the number of processors (nthreads) used to run the model, in our case 32, resulting in a time of about 80 s to fit the model. The final two parameters allowed us to correct for autocorrelation in the residuals: Measurements at subsequent time points are not necessarily independent. Given that these correlated at an average level of about 0.4, setting the rho parameter to 0.4, the model was able to correct for this autocorrelation. The parameter AR.start was used to delimit each individual sequence and was set to TRUE for the first time point in each series (i.e., Time Point 1: schwa50) and FALSE otherwise. The column start.event in our data set dat precisely contained these values. (Note that a requirement to adequately correct for autocorrelation is that the data are ordered, such that the time points belonging to an individual time series occupy subsequent rows in the data set.)

The above model specification only allowed us to assess whether the individual nonlinear interactions between time and the anterior–posterior position of the tongue were

significantly different from 0. In addition, we fitted four subsequent models using so-called binary difference tensors, allowing us to evaluate whether the nonlinear interactions differ significantly between the different sounds and/or age groups.

For example, the following model specification allowed us to assess whether different speaker groups differed significantly (by using the 3-year-olds as a reference):

```
m1 <- bam(PeakX ~ te(Time, VPeakX, k =
c(4,10), by=C) + C +
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=IsC7b) +
te(Time, VPeakX, k = c(4,10), by=IsAb) +
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=IsC7d) +
te(Time, VPeakX, k = c(4,10), by=IsAd) +
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=IsC7g) +
te(Time, VPeakX, k = c(4,10), by=IsAg) +
s(Time, Subject, by=C, bs="fs", m = 1, k = 4) +
s(VPeakX, Subject, by=C, bs="fs", m = 1),
data=dat, discrete=T, nthreads=32, rho = 0.4,
AR.start=dat$start.event)
```

In this case, the first tensor product spline models the nonlinear interaction between time and the anterior–posterior position of the peak at V50 for each of the three consonants. The next tensors all have by-variables that start with Is. These by-variables were constructed such that they are binary, that is, either 0 or 1. For example, IsC4b was set to be equal to 1 whenever the cohort equaled the 4-year-olds (i.e., dat$IsC4b <- (dat$Cohort == "C4" & dat$C == "b")*1) and the consonant equaled /b/; similarly, IsAg was set to be equal to 1 whenever the cohort was equal to the adults and the consonant equals /g/. Whenever a by-variable was not a nominal variable, but a binary variable, the interpretation of this tensor (i.e., nonlinear interaction) was as follows: Whenever the binary variable equals 0, the tensor was completely set to 0 (i.e., the interaction between Time and VPeakX is 0, and therefore the tensor did not contribute to the model fit). Whenever a by-variable equals 1, the tensor represents the difference compared to the reference level. However, what was the reference level? In this case, there were no binary by-variables associated with the 3-year-olds. Consequently, each time the cohort was equal to the 3-year-olds, all tensors with a by-variable starting with Is will be equal to 0. This means that the interaction surfaces for the 3-year-olds are represented by the first tensor (which models three interactions between time and position, one for each consonant). Suppose now we would like to know what the nonlinear interaction between time and position for the 4-year-olds for the /g/ consonant is. Given that the first tensor (i.e., the tensor for the 3-year-olds) is never 0, this tensor is included (for the sound /g/), and to this, we have to add the tensor where the by-variable equals

IsC4g. Given that the tensor for the 4-year-olds is thus constructed from two tensors (the one for the 3-year-olds and the one with IsC4g as a by-variable) and the first tensor is the interaction between time and position for the 3-year-olds, this must mean that the tensor with the by-variable IsC4g represents the difference between the 4-year-olds compared to the 3-year-olds for the consonant /g/. Analogously, we can argue that, for example, the tensor with the by-variable IsAb represents the difference between the adults compared to the 3-year-olds for the consonant /b/. By specifying the model in this way, we can then simply inspect the *p* values associated with these so-called difference tensors to assess if the differences between the 3-year-olds (i.e., the reference group) and the other groups are necessary.

In the following, we first use this approach to construct two models, one to test whether several age cohorts may be grouped (which corresponds to the model shown above) and one to examine whether consonants may be grouped. After potentially grouping consonants and/or age cohorts, we fit two final models, also using binary by-variables (similarly to that shown above) to assess which significant differences exist between the different age groups for the different consonants (the two models are similar, except that they use a different reference level for the age group). The total number of models therefore is 5, which is the reason we set our significance cutoff to $p = .01$. Indeed, an important shortcoming in running many models is that it increases the likelihood of falsely rejecting the null hypothesis and decreases researchers' trust in the obtained *p* values. Using a threshold of 0.05 with five models would lead to approximately 22% chance of falsely rejecting the null hypothesis, hence our decision for a more conservative cutoff to 0.01.

Results

General Trends

The output of GAM analyses is often represented with terrain plots or interaction plots, which visually represent interactions between target variables over time. Because this type of visualization is complex to interpret, we first provide an illustration of the interaction plot for the 3-year-olds in the context of the consonant /b/ together with the associated one-dimensional patterns (see Figure 2). The two figures directly to the right of the interaction plot are linked to the horizontal dashed lines in the interaction plot and show how the tongue dorsum position associated with the schwa and consonant evolves over time for two prespecified tongue dorsum positions associated with the target vowel (i.e., 0.3 and 0.7). The two figures on the second line are linked to the vertical dashed lines and show how the tongue dorsum position at the offset of the schwa (left) and the midpoint of the consonant (right) is related to the tongue dorsum position of the target vowel.

The terrain plot in the left panel of Figure 2 is a visual representation of changes in the tongue dorsum position over time with a color scaling starting from blue shades for low values (corresponding to more anterior tongue

Figure 2. Illustration of interaction plots visualizing tongue dorsum (TD) position over time dependent on the position of the TD during the midpoint of the vowel: schwa midpoint (@50%) and offset (@100%) and consonant midpoint (C50%) and offset (C100%). The dashed horizontal lines show the predicted position of TD over time (i.e., during the pronunciation of the schwa and consonant) dependent on a specific TD position for the vowel (i.e., 0.3 and 0.7). The associated graphs directly to the right of the interaction plot visualize these patterns in one dimension. Similarly, the dashed vertical lines show the predicted position of the TD depending on the TD position for the vowel for two time points (i.e., the offset of the schwa and the midpoint of the consonant). The associated graphs on the second line visualize these patterns in one dimension.
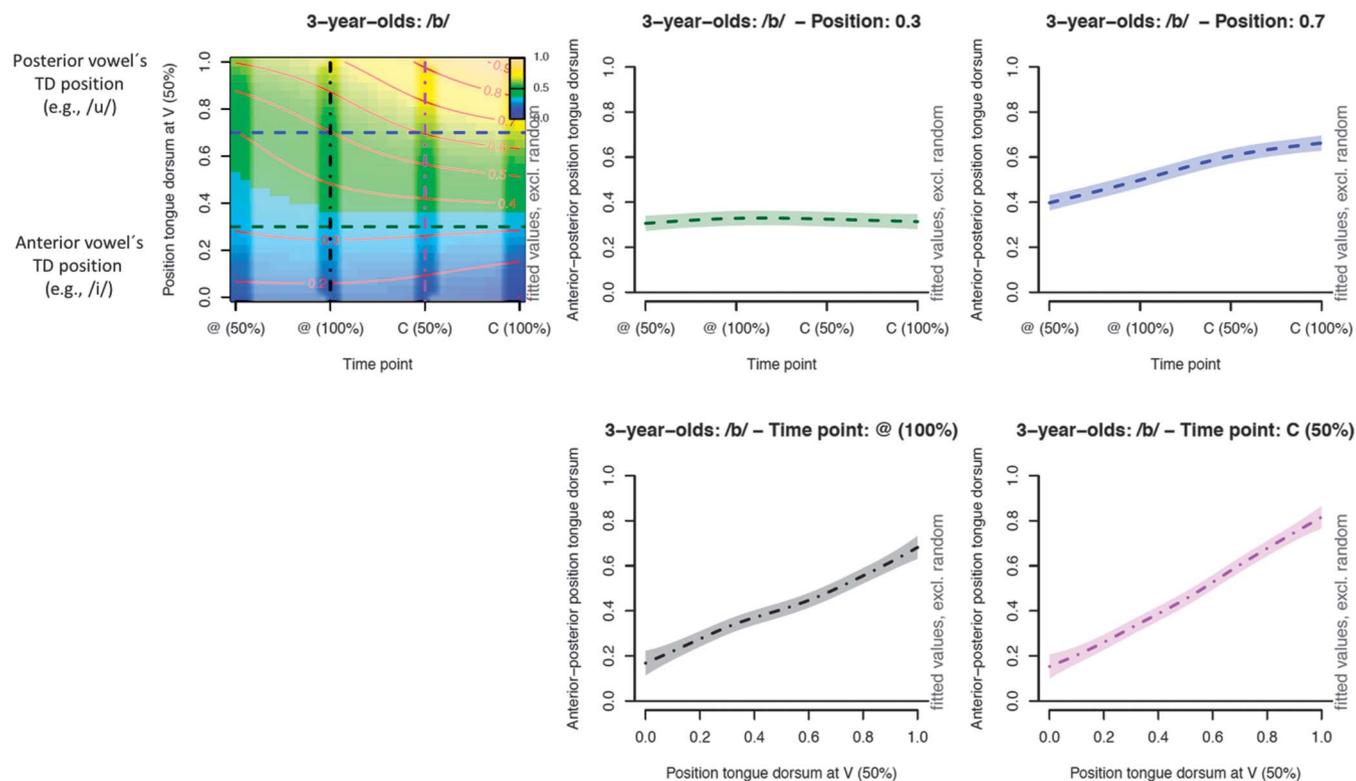
positions in the oral cavity, e.g., for /i/) to orange shades for higher values (corresponding to more posterior tongue positions, e.g., for /u/). In the same way that isolines are used in topographic maps to represent locations sharing the same altitude, the red contour lines connect points that have a similar (predicted, based on all trials) tongue dorsum position over time (i.e., during the pronunciation of the schwa and the consonant, on the *x*-axis) as a function of its vocalic environment (i.e., the tongue dorsum value during the pronunciation of the subsequent vowel, on the *y*-axis). The red contour lines also provide information regarding the direction of the change (i.e., increasing or decreasing; the values associated with each contour line are shown on the line) and whether the patterns are linear, that is, whether they increase or decrease steadily across the four time points (straight line) or nonlinear (curved lines) over time.

Figure 3 provides a general overview of the anticipatory patterns for each of the five age groups investigated (3-, 4-, 5-, and 7-year-olds and adults). Each plot depicts the time course of the vocalic tongue dorsum gesture over the four time points of interest (schwa midpoint: @50%,

schwa offset: @100%, consonant midpoint: C50%, and consonant offset: C100%) at the *x*-axis in interaction with the anterior–posterior position of the tongue dorsum at the vowel midpoint (V50%) on the *y*-axis as a function of consonant identity (/b, d, g/). All the patterns are significantly different from 0 ($p < .001$).

Based on these terrain plots, we can make the following observations. First, comparative observations for each age group show that the temporal organization of the vocalic tongue dorsum gesture varies as a function of consonantal context. This is illustrated by noticeable differences in the terrain plots between /b/, /d/, and /g/ for each cohort. Second, the position of the tongue dorsum at each of the four time points differs as a function of those for the subsequent vowel and its associated lingual gesture. This is evidenced by the vertical color change for a given time point. The predicted values for the tongue dorsum (dependent variable) are presented in the small referential color scaling in the upper right panels. While blue shades represent values for front vowels (e.g., /i, e, y/), orange shades characterize values for back vowels (e.g., /u, o/), and green shades characterize values for more central vowels.

Figure 3. Terrain maps illustrating the time course of the tongue dorsum gesture across three consonantal contexts (/b/: left column, /d/: middle column, and /g/: right column) and five age groups (3-year-olds: top row, 4-year-olds: second row, 5-year-olds: third row, 7-year-olds: fourth row, and adults: last row) and time points (positioned at the *x*-axis): midpoint of the schwa (@50%), schwa offset (@100%), consonant midpoint (C50%), and consonant offset (C100%). Finally, the interaction of time point with the position of the tongue dorsum at the midpoint of the vowel (*y*-axis) is shown. The bright vertical bands show that there are only four distinct time points across which the generalized additive model determines the nonlinear pattern (time points in between also have an associated position, but this is not linked to an actual measurement point).
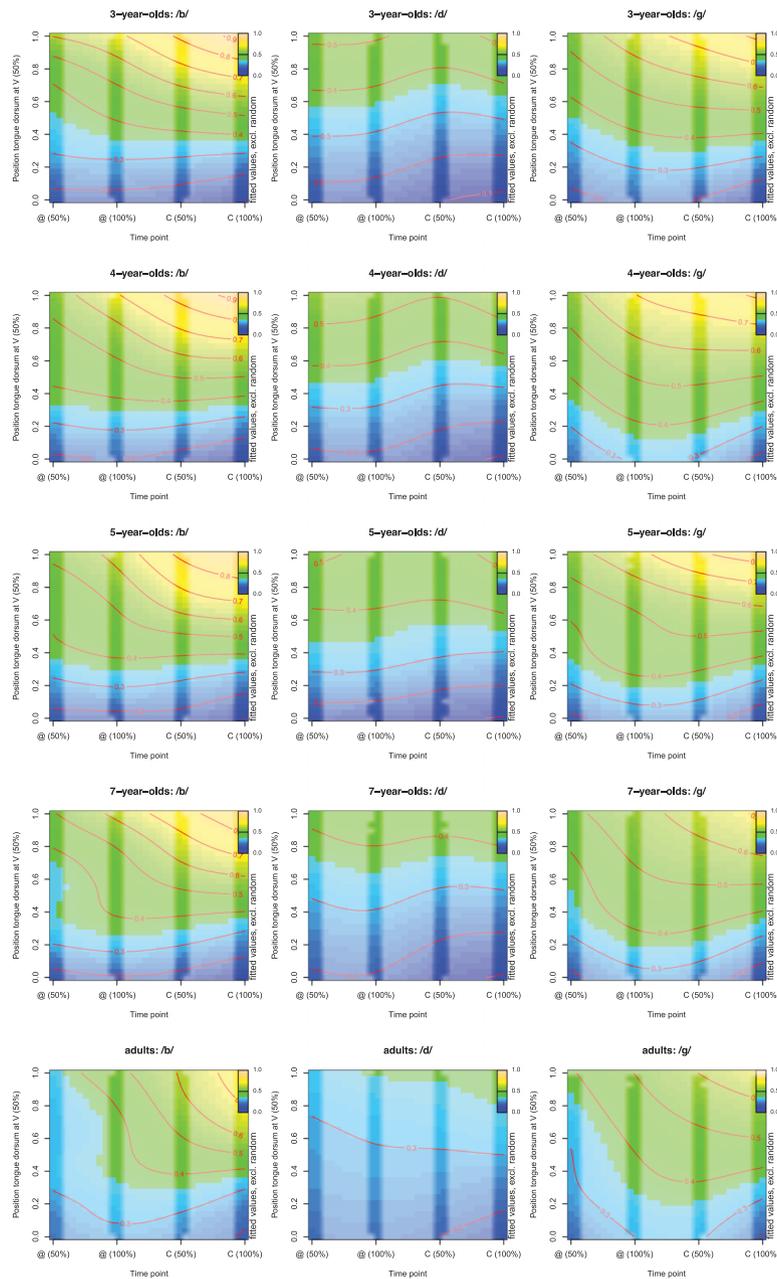
To contextualize this information with respect to vocalic anticipation, we may take as an example the tongue dorsum position at the midpoint of the schwa (@50%) in the context of /b/ for the 3-year-old group (upper left plot, in both Figures 2 and 3). If a single color would be observed across the vertical axis, it would mean that the position of the tongue dorsum at the midpoint of the schwa remained the same regardless of the upcoming vowel and therefore was insensitive to contextual influences. Here, on the contrary, the color contrast observed at @50% clearly evidences the influence of the individual vowels on the schwa. The strength of the vocalic impact is illustrated by the color gradients and the red contour lines. In this particular example, anticipation of vowels produced relatively in the front in the oral cavity (e.g., with a value of 0.3 on the y-axis) exerts greater influence on the tongue dorsum position at the midpoint of the preceding schwa (i.e., corresponding to a blue shade and contour line with a value close to 0.3) than anticipation of back vowels (e.g., with a value of 0.8 on the y-axis). For back vowels, the tongue dorsum position remains indeed more anterior during the schwa (i.e., green shade with a value between 0.4 and 0.5 as illustrated via the red contour lines). The closer we get to the temporal domain of the target back vowels (i.e., C100% at the x-axis), the more similar the tongue dorsum position is to those of the midpoint of the vowel (i.e., a value of 0.7 at C100% for a value of 0.8 at V50%, on the y-axis). The 4- and 5-year-old children overall exhibit a similar pattern as the youngest group, that is, an earlier vowel influence for more front vowels than for back vowels and an overall increase of vowel influence over time. Adults stand apart with tongue dorsum positions approaching those for subsequent vowels later than children, for both anterior and posterior vowels. The 7-year-old children stand in between the youngest cohorts and adults. Details of within–/across–age group differences are provided in the next sections.

The third and most important finding is that change in vocalic anticipation over time, that is, the interaction between the tongue dorsum position for the vowel and those for its neighbors, is nonlinear for all age groups. This is illustrated in the terrain plots by the red contour lines, which do not represent straight increasing or decreasing lines but curvatures. Interestingly, the nonlinearity of the anticipatory process as expressed by the different curvature shapes differs across consonantal contexts (comparing the three columns for a given row).

Since the patterns per consonant seem similar across the 3-, 4-, 5-, and perhaps 7-year-olds, we ran a second binary difference smooth model to assess whether data from the age cohorts could be grouped. Results indicate that nonlinear interaction surfaces for each of the three consonants separately did not significantly differ between the 4- and 5-year-olds to the 3-year-old children. However, it did show differences comparing the 7-year-olds (and the adults) to the 3-year-olds (most strongly for the /g/). Hence, we grouped the 3/4/5-year-old children in subsequent analyses.

Within–Age Group Comparisons of Vocalic Anticipation

Figure 4 illustrates the patterns of vocalic anticipation for each consecutive time point separately. The four rows correspond to the four time points examined with respect to the vowel (@50%, @100%, C50%, and C100% from top to bottom) for the three age cohorts (3/4/5-year-olds, 7-year-olds, and adults) shown in the three columns. In each graph, there are three patterns shown in different colors, one for each of the three consonants. In each of these graphs, the x-axis shows the anterior–posterior position of the tongue dorsum associated with the subsequent vowel, whereas the dependent variable (i.e., the anterior–posterior position of the tongue dorsum associated with the four time points spread out over the preceding schwa and consonant) is represented by the value of the y-axis.

The interpretation of these graphs can again be illustrated using an example. Consider the top-left plot of Figure 4, which shows the amount of anticipatory coarticulation for the 3/4/5-year-olds. Recall that the x-axis shows the tongue dorsum position associated with the upcoming vowel, whereas the y-axis shows the tongue dorsum position associated with the midpoint of the schwa (i.e., the first time point). If there was no vocalic anticipation, one would not expect any influence of the vowel tongue dorsum position on its position during the previous schwa's pronunciation. However, there is clear vowel anticipation across time points. For the youngest kids, the lines seem to have the steepest angle, showing the greatest amount of overlap between the tongue dorsum position for individual vowels and those during the schwa or consonant as compared to the other two groups (i.e., 7-year-olds and adults).

Overall, regardless of the consonantal context, anticipation of the upcoming vowel is already present within the schwa (first row of plots). Second, greater vocalic anticipation is found with labial and velar stops compared to the alveolar stop /d/ for all age groups. Third, both the magnitude of vowels' influence over time and the effect of medial consonants vary for each age group. For the younger cohorts (at ages 3, 4, and 5 years), we note differences in vowels' influence over the anteroposterior position of the tongue dorsum as a function of consonant emerging at the vicinity of the acoustically defined temporal domain for the consonant (at the offset of the schwa). This is illustrated in Figure 4 by the growing separation between the consonant-specific slopes across consonantal contexts. Third, while the influence of individual vowels increases rather steadily over time and becomes more linear in the labial (as soon as schwa offset) and velar (consonant offset) contexts, this is not the case for the resistant alveolar stop /d/. In that last case, the tongue position remains relatively anterior (even in the context of upcoming back vowels), which indicates a lower magnitude of vocalic influence over the tongue dorsum position during the consonant (as noted in the terrain plot, see Figure 3). Reasons for such patterns are suggested in the Discussion section.

Figure 4. Relation between the position of the tongue dorsum at four time points (per row): schwa 50%, schwa 100%, C50%, and C100% as a function of medial consonants /b, d, g/. Results are presented for each age cohort (per column: 3/4/5-year-olds, 7-year-olds, and adults).
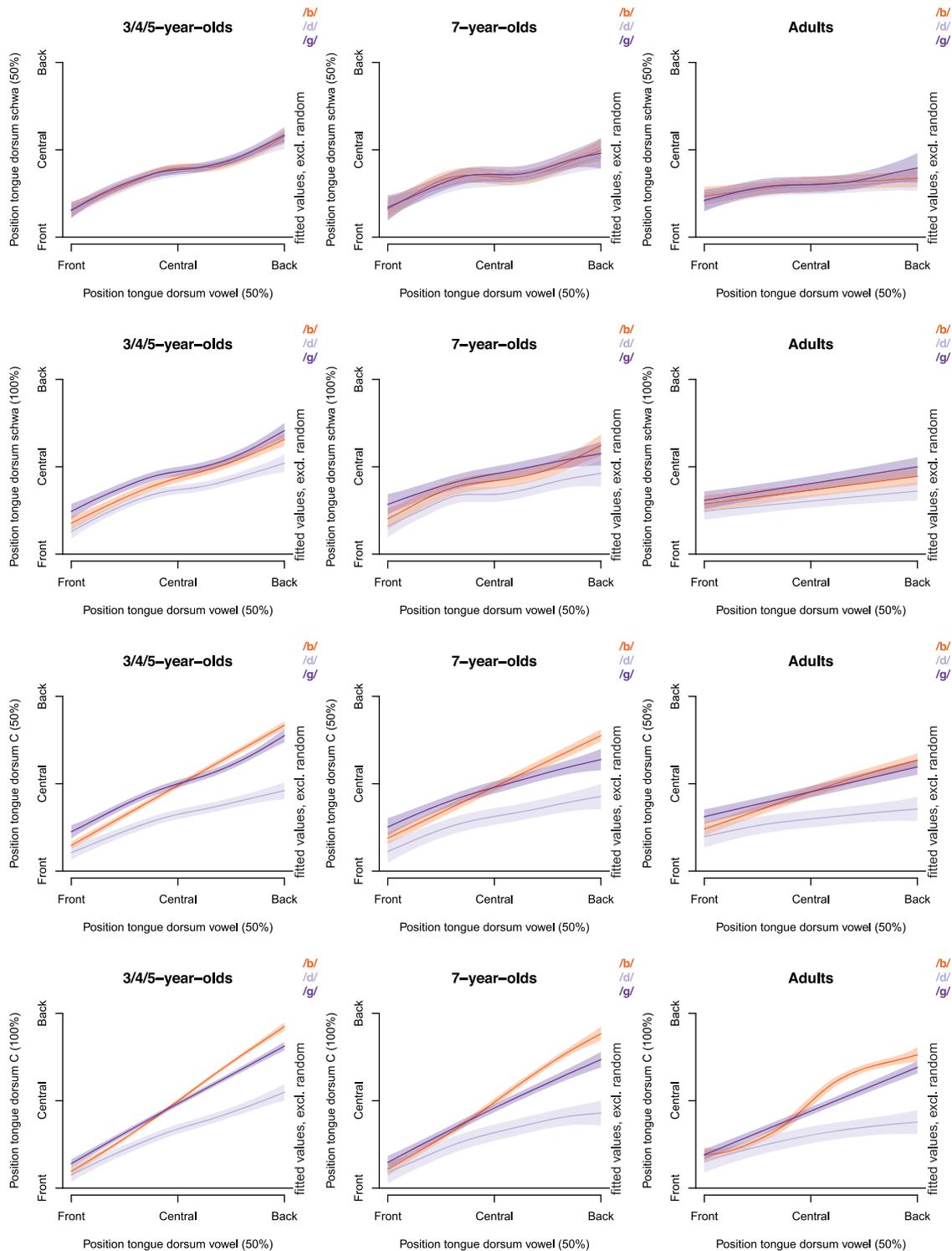
The overall trajectory in anticipatory patterns for older children at the age of 7 years also shows large overlap in slopes across consonants during the schwa (@50%) and an increasing differentiation of anticipatory patterns across consonants over time (i.e., subsequent time points). Hence, there is no any specific effect of consonant identity on children's anticipatory patterns at an early stage of the utterance but only closer to the temporal domain of the consonant. Furthermore, it can be noted that the influence of vowels' tongue dorsum position becomes more linear in labial and velar contexts from the midpoint of the consonant (C50%), while it does not in the alveolar context.

In adults, the magnitude of vocalic anticipation is overall lower over time than in all children. In the context of /b/, the tongue dorsum position during the schwa (e.g., @50%) has a front to central position regardless of the upcoming vowels (i.e., font, central, or back; seen as well in the terrain plot, see Figure 3). This suggests the tongue dorsum position is unaffected by the upcoming vowel but instead reflects the lingual posture for the schwa. The influence of individual vowels becomes more prominent during the temporal domain of the labial stop (e.g., back vowels are associated with more posterior position of the tongue dorsum at C100%). The anticipatory trajectory for sequences involving the stop /b/ exhibits a nonlinear relationship between the tongue position for target vowels and those at the labial stop offset. The pattern for the velar /g/ shows a roughly similar progression as for /b/, but we note the relation between the tongue dorsum position at C100% with respect to upcoming target vowels is linear. Furthermore, the vowels that are associated seem to affect tongue dorsum position for the velar to a lesser extent with respect to /b/ context. Finally, in the context of the alveolar stop /d/, the position of the tongue dorsum remains relatively front to central during the schwa and more anterior at C50% and C100% that correspond to the temporal domain of the consonant.

Across–Age Group Comparisons of Vocalic Anticipation

To compare developmental differences in anticipation, it is most useful to refer to Figure 5, which allows for a direct comparison of the age cohorts per consonant. Tables 1 and 2 summarize the results for the age comparisons made. Comparisons across age groups and consonants using two binary-difference smooth models (one with the adults as the reference level and another one with the 3/4/5-year-olds as the reference level). Our first binary difference smooth model showed that all consonantal contexts are associated with significantly greater vocalic anticipation in all children groups than in adults ($p < .001$), except between the adults and the 7-year-olds for the velar stop /g/ ($p = .08$). The second binary difference smooth model revealed that the youngest children (i.e., the 3/4/5-year-olds) did not show significantly greater anticipation than the 7-year-old for the alveolar /d/ ($p = .02$; note that our significance threshold was set to .01) or the

velar /g/ ($p = .03$) but significantly greater vocalic anticipation for the /b/ ($p = .0095$).

Discussion

Speech is a complex dynamical system encompassing various processes in the cognitive, perceptual, and motor domains. In the past decades, tremendous effort has been devoted to the understanding of the temporal organization of articulatory gestures supporting fluent speech. In this study, we examined the dynamics of vocalic anticipation from the age of 3 years to adulthood. We utilized the technique of ultrasound imaging, which allows for the continuous recording of the tongue movement during speech while being suitable with young children. We then used GAM to estimate both linear and nonlinear influences on coarticulatory processes. In the next sections, we discuss our findings with respect to the temporal organization coarticulatory across consonants and vowels and its change over development.

Nonlinear Patterns of Anticipation: Role of Consonantal and Vocalic Gesture

A main objective was to test for nonlinear patterns of vocalic anticipation, which may result from the interaction between tongue gestures for individual vowels and those for their neighbors over time. Results indicate nonlinearities in vowel anticipation over time in all cohorts, albeit to a lesser extent in children than in adults. This is a new finding relative to our previous research that has tested for linear relationships between consecutive gestures. The present results show that vocalic anticipation is a more complex process with a rate of change that differs over time. We discuss two sources for the nonlinearities observed. First, the magnitude of the anticipation over time changes as a function of the identity of the medial consonant between the schwa and the target vowel. This is most salient in the terrain plots (see Figure 3) and in Figure 4 (third and fourth rows illustrating the temporal domain of the consonant). When the organs involved in the achievement of neighboring gestural goals are anatomically relatively independent from each other (lips/jaw and tongue in the syllable /bi/), vocalic anticipation was greater in the temporal domain of the stop than when articulators are mechanically coupled (e.g., the tongue tip and tongue dorsum for /da/). In this case, vocalic anticipation is reduced due to the gestural demand for the alveolar stop in its temporal domain. To achieve a target constriction gesture in the alveolar region (e.g., for the alveolar stop /d/ or for the vowel /i/), the tongue body needs to move front (e.g., review in Buchaillard, Perrier, & Payan, 2009) for the tongue tip to then raise to its target position. Can we conclude that vocalic anticipation is solely modulated by the gestural demands for the medial consonant? Not really. A second important factor for the observed nonlinearity in anticipatory patterns comes from the identity of the target vowel and its associated tongue dorsum position in the antero-posterior dimension (see Figure 3). This result expands on our research with German adults (Abakarova et al., 2018)

Figure 5. Relation between the position of the tongue dorsum at four time points (per row: schwa 50%, schwa 100%, C50%, and C100%) as a function of consonant (per column: /b, d, g/) for each age group: 3/4/5-year-olds, 7-year-olds, and adults.
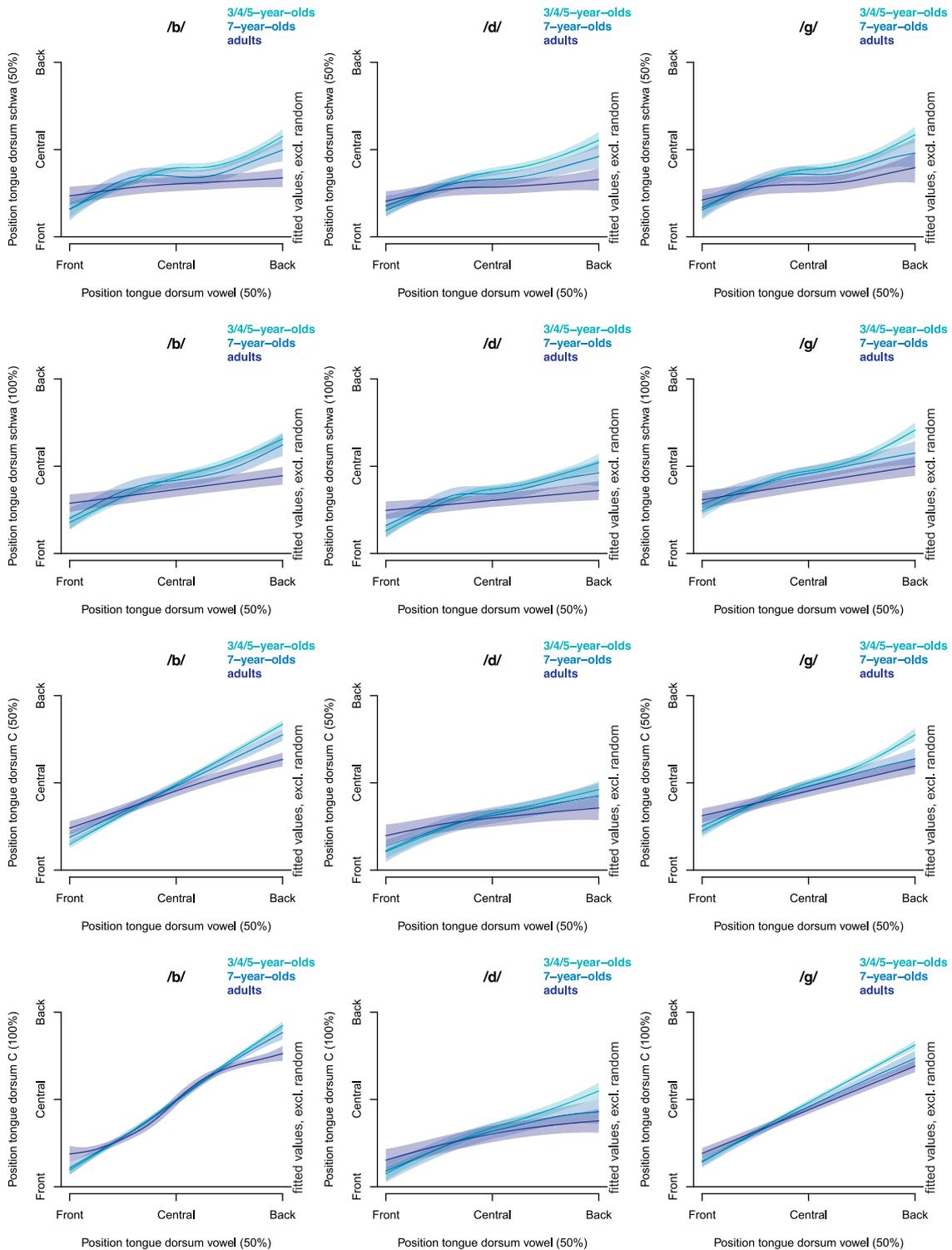
Table 1. Smooth function terms of the generalized additive model testing vowel anticipation over time across all age groups.

Smooth functions	edf	F	p
S(Time): children, consonant /b/	23.948	62.029	< .001
S(Time): children, consonant /d/	19.246	23.680	< .001
S(Time): children, consonant /g/	21.563	35.234	< .001
S(Time): children 3/4/5, consonant /b/	15.262	8.063	< .001
S(Time): children 7, consonant /b/	6.234	5.165	< .001
S(Time): children 3/4/5, consonant /d/	9.560	2.596	< .001
S(Time): children 7, consonant /d/	7.066	2.596	< .01
S(Time): children 3/4/5, consonant /g/	4.001	9.814	< .001
S(Time): children 7, consonant /g/	4.001	2.036	.087

Note. Both children cohorts (3/4/5-year-olds and 7-year-olds) were compared to adults.

and on findings made in 6- and 9-year-old Scottish children with /a-i-u/ pairs (Zharkova et al., 2012). Our findings further suggest that the time course of vocalic anticipation reflects the compatibility between the gestural goal for individual vowels and those of their neighbors and it is the interaction of those goals over time that determines the linearity of the anticipatory process or lack thereof. Note that this is the intuition that stimulated us in using GAM to investigate anticipation over time: The method allows for revealing complex gestural interactions over time, which may result in linear or nonlinear patterns.

With respect to the three general approaches to coarticulatory organization laid out in the introduction, we interpret our findings as supportive of a gestural approach to speech production (e.g., articulatory phonology: Browman & Goldstein, 1992). While gestural goals are discrete and language specific, they can be achieved via different coordinative strategies, especially in the developing language and motor systems of children. This leads us to the discussion of the developmental differences noted in our study.

Developmental Differences in Movement Dynamics

Results from this study suggest that the developmental differences in anticipatory organization observed in our study are related to differences in movement dynamics. As already mentioned, the speech articulators from which

Table 2. Smooth function terms of the generalized additive model testing vowel anticipation comparing all 7-year-olds and adults to the younger 3/4/5-year-old cohort.

Smooth functions	edf	F	p
S(Time): children, consonant /b/	22.467	108.285	< .001
S(Time): children, consonant /d/	19.572	19.047	< .001
S(Time): children, consonant /g/	21.548	67.461	< .001
S(Time): children 7, consonant /b/	10.477	2.126	< .01
S(Time): adults, consonant /b/	16.858	7.738	< .001
S(Time): children 7, consonant /d/	6.480	2.480	< .05
S(Time): adults, consonant /d/	9.459	4.895	< .001
S(Time): children 7, consonant /g/	4.003	2.653	< .05
S(Time): adults, consonant /g/	5.221	6.744	< .001

movements emerge are mechanically coupled, and their movements do not start and end abruptly as their phonetically defined boundaries in acoustic transcriptions. Instead, speech movements may be conceptualized like hysteresis curves; they gradually increase and decrease in prominence and have their own intrinsic timing (Fowler, 1980), which leads to gradients in coarticulatory overlap. This phenomenon has been described with respect to labial anticipation (e.g., in adults: Fowler & Saltzman, 1993; Noiray, Cathiard, Ménard, & Abry, 2011; in children: Noiray et al., 2010) and lingual anticipation (e.g., Fowler & Brancazio, 2000). For a given gestural goal (e.g., for /u/), articulators gradually move toward their target increasing in velocity and decelerating upon reaching the vowel "steady" state (zero velocity). Depending on the next gestural goal, each organ may then move toward the next gestural target or reset to a more neutral position if not involved in the next gestural goal. Fowler and Brancazio (2000, p. 37) explicate this phenomenon in American English speakers as follows: "One can think of the gestures of a consonant or vowel first strengthening then weakening over time. The strength of the consonant's clamping of the tongue dorsum then would be strongest in the time interval identified as the temporal domain of the consonant (perhaps strongest of all during consonant closure) and weaker earlier than that and later than that time." Our findings support the view of gestural clamping on the tongue dorsum and further point at developmental differences in the phasing between individual gestures over time. Overall, preschoolers' anticipation is organized along a broader temporal span compared to adults and, to some extent, to 7-year-old children too (e.g., Goodell & Studdert-Kennedy, 1993; Nittrouer, 1993; Nittrouer et al., 1989). However, our results also indicate an effect from individual consonants and vowels (as exemplified in some previous research). How can we reconcile the fact that children do exhibit context-specific anticipatory patterns but also anticipate upcoming vowel targets to a globally greater extent than adults?

Those differences may be explained by an interplay between several factors. First, greater anticipation in children may partly result from differences in the anatomy of children's vocal tract as compared to adults. While children rather effortlessly learn to speak their language fluently, the geometry of their vocal tract (e.g., descent of the hyoid bone at around 4 years of age leading to a more posterior position of the tongue: Buhr, 1980; Vorperian & Kent, 2007) changes nonlinearly over time. This means that children have to regularly readjust their gestural organization to achieve adultlike vocalic targets, which results in long-lasting articulatory and acoustic variability until children reach adultlike vocal tract anatomy (e.g., Vorperian, Kent, Gentry, & Yandell, 1999). Anatomical influences may have well impacted children's anticipatory patterns recorded in our study (e.g., via an overall more anterior position of the tongue dorsum irrespective of the target utterances in the youngest cohorts); unfortunately, in this study, anatomical differences could not be quantified. While measuring the direct impact of anatomical development onto children's

speech has remained methodologically challenging, promising models have been developed to address these aspects (see, for instance, Story, Vorperian, Bunton, & Durtschi, 2018).

Second, developmental differences in the temporal organization of vocalic anticipation also result from discrepancies in control of the speech motor system and lack of differentiation of gestural goals for consecutive segments. This is unsurprising given it takes over a decade for children to achieve mature coordinated patterns in their native language (e.g., Kent, 1976; Smith & Zelaznik, 2004; Walsh & Smith, 2002). While vowels are usually said acquired by the age of 3 years, accuracy in words varying in phonological complexity takes at least another 3 years (e.g., James, Van Doorn, and McLeod's [2001] study of 354 children from 3 to 7 years of age). The children tested in our study fall within that age range and are hence still in the process of learning to control the speech machinery to create precise coordinative structures for producing vocalic and consonantal gestures over time. In that context, the tongue gesture for the vocalic target may be integrated to those for neighboring segments, following the principle of "all move at once" (Kent, 1983, p. 70; Nittrouer, 1993). Instead, adults may behave more along a principle of economy of energy (e.g., Lindblom, 1990; Nelson, 1983; Sporns & Edelman, 1998), achieving the vowel gesture later in the utterance only when necessary.

Last, the discrepancies in vocalic anticipatory patterns may also reflect developmental differences in gestural planning. Vowels are, in general, perceptually very salient due to their long duration, loudness, and formant patterns (e.g., Cutler & Mehler, 1993). They are also acquired developmentally earlier compared to consonants (e.g., Kuhl et al., 2006; Polka & Werker, 1994) and associated with a greater focus in stressed syllables than in unstressed ones (e.g., in German: Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2013). Hence, in early childhood, vowels may function as attractors in utterances and be initiated earlier than adults, leading to broader temporal overlap (i.e., in the schwa; see Figure 3). Adults instead show greater differentiation between the gestural goals for consecutive segments than children. For instance, the lingual gesture for the target vowel is not activated early in the schwa, even in sequences including a labial stop, which does not recruit the same organ and offers an opportunity for maximal vocalic anticipation (see Figure 3). Instead, the vocalic gesture seems more active later toward the end of the acoustically defined temporal domain of the consonant. Hence, if adults plan their speech from one vowel to a subsequent vowel, our results suggest they have optimized their anticipatory patterns compared to children in that the speech plan takes the gestural constraints for the upcoming segments into account (Fowler & Saltzman, 1993). In children, the timing of the vocalic anticipatory trajectory is not as finely adjusted to accommodate these gestural constraints.

Figure 6 provides a hypothetical depiction of the differences in lingual organization and prominence over time between adults (left panel) and children (right panel). In the figure, greater prominence of a gesture is illustrated by

a higher activation curve than for gestures with lower prominence (in our study: for the stressed vowel in comparison to the schwa). As seen in the terrain plots (see Figure 3), the interaction between individual vowels and consonants is clearly more complex than the simplistic depiction provided here. However, it can be noted that children's curves characterizing gestural prominence are overall temporally broader than those of adults, following the conclusion drawn by Nittrouer (1993). Tilsen (2016) explains the phenomenon of broader temporal activation curves for stressed vowels in children in comparison to adults (and greater coarticulation) as resulting from a general lack of inhibitory control observed in childhood. With increased experience, children's lingual gestures should become more precisely controlled over time. Tilsen proposes that the internalization of feedback collected through repetitive experiences with a gestural goal would allow children to build "anticipated sensory consequences of motor commands" or efferent copies and inhibit motor plans that are not suitable to be deployed in speech (Tilsen, 2016, p. 57). Benefiting from greater exposure to the native language diversity and greater experience speaking the language, our older group of children at 7 years old differs from younger children. The influence of the vocalic gesture is less prominent and more nonlinear than for the younger group. Note, however, that their anticipatory patterns still differ from adults.

Tilsen's proposition aligns well with the view in the motor control domain that anticipatory behavior is tightly related to knowledge about the future (e.g., Butz et al., 2003; Nadin, 2015). "The fact that the sequential model (serial order) is only an approximation becomes evident when a certain action (hammering, hitting the golf ball) involves parallel components. The action depends on the perception. The hand seems to 'know what resistance it will meet'" (Nadin, 2015, p. 331). Hence, experience (with its internalized sensory feedback) plays a crucial role on the efficiency of motor coordination. While adults can anticipate the force to apply to the grip, the resistance due to the weight, and the trajectory to employ because of past experiences with similar goals and contexts, children do not benefit from such rich experience yet and lack of feedback to construct skilled motor patterns (for a similar discussion with respect to word comprehension and production, see review in Hall, Hume, Jaeger, & Wedel, 2018). Furthermore, before entering school, children are often exposed to child-directed speech that consists of rather simplistic (and often hyperarticulated) utterances. It may take several years for children to benefit from the rich input provided by their social environment and from the practical experience gained in speaking the language to display skilled anticipatory patterns. A lot more work is needed to disentangle maturational processes from social and environmental aspects, all of which interact in fundamental ways to shape language acquisition. In the last decade, assessing the role of experience in social interaction and, more specifically, its contribution to shaping production and perception mechanisms have been two major foci in sociophonetics (e.g., reviews in Foulkes & Hay, 2015; Hay, 2018). Because anticipation is largely

Figure 6. Schematic representation of vowel (V) and consonant (C) prominence over time. The schwa is represented by a plain line; the consonant, by a mixed dashed line; and the target vowel, by a small dashed line. The vertical lines represent hypothetical segments' acoustic onsets and offsets.
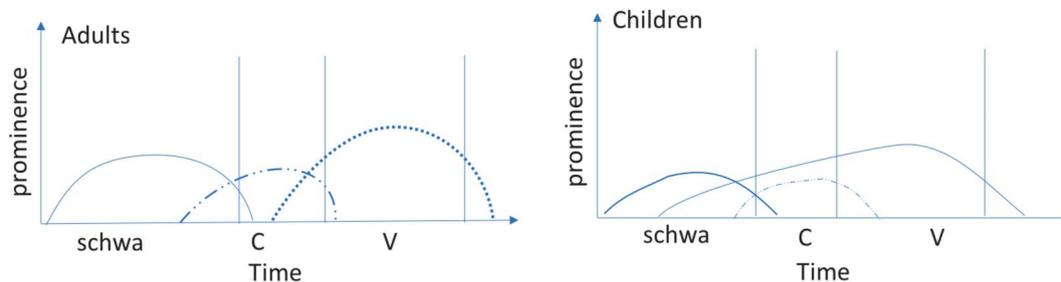
related to feedforward representations, which in turn are driven by the (sensory) information drawn from past experiences, future studies looking at its maturation in childhood should greatly benefit from research, the primary interest of which is to describe speech in its natural communicative context.

Limitations of the Study

While learning to speak their native language fluently, children develop various cognitive skills (e.g., lexicon, phonemic awareness, reading) in parallel to gaining greater control over their speech motor system. While it was not possible in this study to estimate how the interaction between cognitive and motor processes directly affects the maturation of vocalic anticipation over time, the question is crucial for advancing our understanding of the factors responsible for variation in anticipatory processes. This also means we may need to depart from age-related descriptions, which are of practical convenience but do not accurately reflect well children's developmental stages with respect to specific skills. For instance, in our study, it was found that 3-, 4-, and 5-year-olds do not fundamentally differ in their anticipatory patterns but only with respect to older children at 7 years old. Inversely, Barbier et al. (2015) reported individual profiles in their 4-year-old children; some exhibited patterns in the direction of adults, while others showed great lingual coarticulation than their peers. Taken together, these results should probe us into carefully examining individual variability (vs. focus on age group analyses, as was the case in our study). The sources of the developmental differences seem indeed to result from complex interactions between diverse maturational trajectories, with some being intrinsic to the speech system (e.g., anatomical development), some external (e.g., degree of exposure to the language), and some the product of both external and internal factors (e.g., speech motor control) rather than purely age dependent. In a recently funded project, we have made a first step in that direction and hope soon to provide new insights on how those multifaceted developments shape the maturation of anticipatory processes in speech. Another important shortcoming in our study stands in its limitation to a description of nonlinearities in anticipatory processes in childhood without providing any

prescriptive outcome. Increased focus on investigating the sources of the differences observed in typical development of anticipatory patterns (e.g., via modeling that allows for greater flexibility in hypothesis testing compared to time-consuming recording of yet small samples of children) should in turn help researchers predict the challenges some children may encounter when learning to speak their native language fluently and determine whether idiosyncrasies may be viewed as typical for an age range or a feature of disordered language (e.g., apraxia, stuttering).

Conclusion

The main objective of this study was to investigate the expression of anticipation, a fundamental property of motor programming in the speech of German children and adults. Using ultrasound imaging, we recorded the movement of the tongue in short utterances and examined the pervasiveness of vocalic gestures on gestures for preceding segments. Results support the hypothesis of a developmental transition of coarticulatory organization toward more segmentally differentiated and contextually specified organizations in primary school and adulthood. Expanding on previous research, we provide evidence for nonlinear interactions between vocalic and consonantal gestures over time in adults and, to some extent, in children. This suggests that the time course of vocalic anticipation is a function of the compatibility between the gestural goals for individual vowels and those of their neighbors and it is the interaction of those goals over time that determines the linearity of the anticipatory process or lack thereof. Substantial differences were found between children and adults and, to some extent, between school-aged children and younger children in kindergarten. While in adults, nonlinear anticipatory patterns over time suggest a strong differentiation between the gestural goals for consecutive segments, in children, maturation toward more individuated lingual gestures and greater precision is protracted.

Acknowledgments

References

Abakarova, D., Iskarous, K., & Noiray, A. (2018). Quantifying lingual coarticulation in German using mutual information: An ultrasound study. *The Journal of the Acoustical Society of America, 144*(2), 897–907.

Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M., & Perkell, J. (2015). *Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses.* Paper presented at the 16th Annual Conference of the International Speech Communication Association, Dresden, Germany.

Beckman, M. E., & Edwards, J. (2000). Lexical frequency effects on young children's imitative productions. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon* (pp. 208–218). Cambridge, England: Cambridge University Press.

Bergmann, C., Tsuji, S., & Cristia, A. (2017). Top-down versus bottom-up theories of phonological acquisition: A big data approach. In *Interspeech 2017* (pp. 2103–2107).

Bernstein, N. A. (2014). *Dexterity and its development.* New York, NY: Psychology Press.

Bladon, R. A., & Al-Bamerni, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics, 4,* 135–150.

Boersma, P., & Weenink, D. (1996). *Praat, a system for doing phonetics by computer, version 3.4, report* (Vol. 132, pp. 1–182). Amsterdam, Netherlands: Institute of Phonetic Sciences of the University of Amsterdam.

Boucher, K. R. (2007). Patterns of anticipatory coarticulation in adults and typically developing children.

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

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 112*(44), 13531–13536.

Buchaillard, S., Perrier, P., & Payan, Y. (2009). A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *The Journal of the Acoustical Society of America, 126*(4), 2033–2051.

Buhr, R. D. (1980). The emergence of vowels in an infant. *Journal of Speech and Hearing Research, 23*(1), 73–94.

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

Caudrelier, T., Schwartz, J. L., Perrier, P., Gerber, S., & Rochet-Capellan, A. (2018). Transfer of learning: What does

it tell us about speech production units? *Journal of Speech, Language, and Hearing Research, 61,* 1613–1625.

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*(4), 676–688.

Corbetta, D., & Bojczyk, K. E. (2002). Infants return to two-handed reaching when they are learning to walk. *Journal of Motor Behavior, 34*(1), 83–95.

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

de Boisferon, A. H., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2018). The redeployment of attention to the mouth of a talking face during the second year of life. *Journal of Experimental Child Psychology, 172,* 189–200.

DePaolis, R. A., Vihman, M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development, 36*(4), 642–649.

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.

Foulkes, P., & Hay, J. B. (2015). 13 The Emergence of Sociophonetic Structure. *The Handbook of Language Emergence, 87,* 292.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*(1), 113–133.

Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & Psychophysics, 55*(6), 597–610.

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.

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

Gentner, D. R. (1983). The acquisition of typewriting skill. *Acta Psychologica, 54*(1–3), 233–248.

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.

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, 51*(6), 1424–1437.

Goldstein, L. (2003). Emergence of discrete gestures. *Proceedings of the 15th International Congress of Phonetic Sciences, 2003,* 85–88.

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

Graetzer, S. (2006). Consonantal coarticulation resistance in vowel–consonant–vowel sequences in two Australian languages. *Proceedings of the 11th Australasian International Conference on Speech Science and Technology, 1*(1), 270–275.

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. (2010). Some organization principles in early speech development. In B. Maaseen & P. H. H. M. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 171–188). New York, NY: Oxford University Press.

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

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard, 4*. https://doi.org/10.1515/lingvan-2017-0027

Hardcastle, B., & Tjaden, K. (2008). Coarticulation and speech impairment. In M. J. Ball, M. R. Perkins, N. Müller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 506–524). Cambridge, England: Blackwell.

Hay, J. (2018). Sociophonetics: The role of words, the role of context, and the role of words in context. *Topics in Cognitive Science, 10*(4), 696–706.

Hodge, M. M. (1989). A comparison of spectral temporal measures across speaker age: Implications for an acoustic characterization of speech maturation. Ann Arbor, MI: University Microfilms International.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2013). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior Development, 32*(3), 262–274.

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.

Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C. H., Saltzman, E., & Whalen, D. H. (2013). The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance. *The Journal of the Acoustical Society of America, 134*(2), 1271–1282.

James, D., Van Doorn, J., & McLeod, S. (2001). Vowel production in mono-, di-and poly-syllabi words in children 3; 0 to 7; 11 years. In Speech Pathology Australia National Conference: Evidence and Innovation, (pp. 127–136). Speech Pathology Australia.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687.

Katz, W. F., & Bharadwaj, S. (2001). Coarticulation in fricative-vowel syllables produced by children and adults: A preliminary report. *Clinical Linguistics & Phonetics, 15*(1–2), 139–143.

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

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research, 19*(3), 421–447.

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

Keren-Portnoy, T., Majorano, M., & Vihman, M. M. (2009). From phonetics to phonology: The emergence of first words in Italian. *Journal of Child Language, 36*(2), 235–267.

Kisler, T., Schiel, F., & Sloetjes, H. (2012). Signal processing via web services: The use case WebMAUS. *Digital Humanities Conference 2012, 2012*, 30–34.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21.

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.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior; the Hixon Symposium* (pp. 112–146). Oxford, England: Wiley.

Lenoci, G. (2017) Anticipatory coarticulation in the speech of people who stutter. In C. Bertini, C. Celata, G. Lenoci, C. Meluzzi, & I. Ricci (Eds.), *Fattori sociali e biologici nella variazione fonetica* [Social and biological factors in speech variation] (pp. 397–415). Milano, Italy: Officinaventuno.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht, the Netherlands: Springer.

Lindblom, B., & Sussman, H. M. (2012). Dissecting coarticulation: How locus equations happen. *Journal of Phonetics, 40*(1), 1–19.

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*(2S), 649–663.

Maas, E., Robin, D. A., Wright, D. L., & Ballard, K. J. (2008). Motor programming in apraxia of speech. *Brain and Language, 106*(2), 107–118.

McLeod, S., & Bleile, K. (2003, November). Neurological and developmental foundations of speech acquisition. In *American Speech-Language-Hearing Association Convention*, Chicago, IL: ASHA.

McNeil, M. R., Ballard, K. J., Duffy, J. R., & Wambaugh, J. (2017). Apraxia of speech theory, assessment, differential diagnosis, and treatment: Past, present, and future. In P. van Lieshout, B. Maassen, & H. Terband (Eds.), *Speech motor control in normal and disordered speech: Future developments in theory and methodology* (pp. 195–221). Rockville, MD: ASHA Press.

Meltzoff, A. N. (2007). Like me': A foundation for social cognition. *Developmental Science, 10*(1), 126–134.

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. *Language Production, 2*, 3–50.

Ménard, L., & Noiray, A. (2011). The development of lingual gestures in speech: Comparing synthesized vocal tracts with natural vowels. *Faits de Langue, 37*, 189–202.

Ménard, L., Schwartz, J. L., & Boë, L. J. (2004). Role of vocal tract morphology in speech development. *Journal of Speech, Language, and Hearing Research, 47*(5), 1059–1080.

Munson, B. (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research, 47*(1), 58–69.

Nadin, M. (2014). Quantifying anticipatory characteristics. The AnticipationScope and the AnticipatoryProfile. In B. Iantovics & R. Kountchev (Eds.), *Advanced intelligent computational technologies and decision support systems* (pp. 143–160). Cham, Switzerland: Springer.

Nadin, M. (2015). *Anticipation: Learning from the past. Cognitive systems monographs*. Cham, Switzerland: Springer.

Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: Task-dependent and dialect-specific aspects. *Journal of Child Language, 41*(3), 600–633.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics, 46*(2), 135–147.

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

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

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 and Hearing Research, 32*(1), 120–132.

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

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,* 1355–1368.

Noiray, A., Cathiard, M. A., Abry, C., & Ménard, L. (2010). Lip rounding anticipatory control: Crosslinguistically lawful and ontogenetically attuned. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 153–171). Oxford, United Kingdom: Oxford University Press.

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 linguistic abilities* (pp. 100–116). Newcastle, England: 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.

Noiray, A., Iskarous, K., & Whalen, D. H. (2014). Variability in English vowels is comparable in articulation and acoustics. *Laboratory Phonology, 5*(2), 271–288.

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

Noiray, A., Ries, J., & Tiede, M. (2015). Sonographic & optical linguo-labial articulation recording system (SOLLAR). Oral presentation at Ultrafest VII, Hong Kong.

Perret, C., & Kandel, S. (2014). Taking advantage of between- and within-participant variability? *Frontiers in Psychology, 5,* 1235.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421.

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

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America, 125*(4), 2288–2298.

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

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.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research, 41*(1), 18–30.

Smith, A., & Goffman, L. (2004). Interaction of motor and language factors in the development of speech production. In B. Maassen,

R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 227–252). Oxford, United Kingdom: Oxford University Press.

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

Snow, D. (1998). A prominence account of syllable reduction in early speech development: The child's prosodic phonology of tiger and giraffe. *Journal of Speech, Language, and Hearing Research, 41*(5), 1171–1184.

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), 281–295.

Speer, S. R., & Ito, K. (2009). Prosody in first language acquisition —Acquiring intonation as a tool to organize information in conversation. *Language and Linguistics Compass, 3*(1), 90–110.

Sporns, O., & Edelman, G. M. (1998). Bernstein's dynamic view of the brain: The current problems of modern neurophysiology (1945). *Motor Control, 2*(4), 283–305.

Story, B. H., Vorperian, H. K., Bunton, K., & Durtschi, R. B. (2018). An age-dependent vocal tract model for males and females based on anatomic measurements. *The Journal of the Acoustical Society of America, 143*(5), 3079–3102.

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.

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., & Ahmed, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *The Journal of the Acoustical Society of America, 94*(3), 1256–1268.

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 and Hearing Research, 35*(4), 769–781.

Sussman, H. M., Marquardt, T. P., & Doyle, J. (2000). An acoustic analysis of phonemic integrity and contrastiveness in developmental apraxia of speech. *Journal of Medical Speech-Language Pathology, 8*(4), 301–313.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of perception and action.* Cambridge, MA: MIT Press.

Tiede, M., Mooshammer, C., Goldstein, L., Shattuck-Hufnagel, S., & Perkell, J. S. (2011). Motor learning of articulator trajectories in the production of novel utterances. In *ICPhS* (pp. 1986–1989).

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

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.

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. *Linguistics, 45*(4), 683–725.

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

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research, 50*(6), 1510–1545.

Vorperian, H. K., Kent, R. D., Gentry, L. R., & Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: Preliminary results. *International Journal of Pediatric Otorhinolaryngology, 49*(3), 197–206.

Walsh, B., Mettel, K. M., & Smith, A. (2015). Speech motor planning and execution deficits in early childhood stuttering. *Journal of Neurodevelopmental Disorders, 7*(1), 27.

Walsh, B., & Smith, A. (2002). Articulatory movements in adolescents. *Journal of Speech, Language, and Hearing Research, 45*(6), 1119–1133.

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.

Whiteside, S. P., & Hodgson, C. (2000). Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. *Speech Communication, 32*(4), 267–285.

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.

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.

Wood, S. N., & Wood, M. S. (2015). Package 'mgcv.' *R package version,* 1–7.

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.

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.

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.

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

Zingale, C. M., & Kowler, E. (1987). Planning sequences of saccades. *Vision Research, 27*(8), 1327–1341.

Appendix

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

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