



HAL
open science

Developmental Changes in Coarticulation Degree Relate to Differences in Articulatory Patterns: An Empirically Grounded Modeling Approach

Dzhuma Abakarova, Susanne Fuchs, Aude Noiray

► To cite this version:

Dzhuma Abakarova, Susanne Fuchs, Aude Noiray. Developmental Changes in Coarticulation Degree Relate to Differences in Articulatory Patterns: An Empirically Grounded Modeling Approach. *Journal of Speech, Language, and Hearing Research*, 2022, 65 (9), pp.3276-3299. 10.1044/2022_JSLHR-21-00212 . hal-04086083

HAL Id: hal-04086083

<https://hal.science/hal-04086083>

Submitted on 1 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Research Article

Developmental Changes in Coarticulation Degree Relate to Differences in Articulatory Patterns: An Empirically Grounded Modeling Approach

Dzhuma Abakarova,^{a,b}  Susanne Fuchs,^b and Aude Noiray^{c,d}^aLaboratory for Oral Language Acquisition, Department of Linguistics, University of Potsdam, Germany ^bLeibniz-Centre General Linguistics (ZAS), Berlin, Germany ^cLaboratoire Dynamique du Langage, Lyon, France ^dHaskins Laboratories, New Haven, CT

ARTICLE INFO

Article History:

Received April 13, 2021

Revision received September 22, 2021

Accepted May 9, 2022

Editor-in-Chief: Bharath Chandrasekaran

Editor: Antje Sabine Mefferd

https://doi.org/10.1044/2022_JSLHR-21-00212

ABSTRACT

Purpose: Coarticulatory effects in speech vary across development, but the sources of this variation remain unclear. This study investigated whether developmental differences in intrasyllabic coarticulation degree could be explained by differences in children's articulatory patterns compared to adults.

Method: To address this question, we first compared the tongue configurations of 3- to 7-year-old German children to those of adults. The observed developmental differences were then examined through simulations with Task Dynamics Application, a Task Dynamics simulation system, to establish which articulatory modifications could best reproduce the empirical results. To generate syllables simulating the lack of tongue gesture differentiation, we tested three simulation scenarios.

Results: We found that younger speakers use less differentiated articulatory patterns to achieve alveolar constrictions than adults. The simulations corresponding to undifferentiated control of tongue tip and tongue body resulted in (a) tongue shapes similar to those observed in natural speech and (b) higher degrees of intrasyllabic coarticulation in children when compared to adults.

Conclusions: Results provide evidence that differences in articulatory patterns contribute to developmental differences in coarticulation degree. This study further shows that empirically informed modeling can advance our understanding of changes in coarticulatory patterns across age.

Coarticulation, defined here as the articulatory and acoustic overlap between consecutive speech segments, is a ubiquitous characteristic of spoken language. The amount of coarticulatory overlap reflects a balance that is essential for effective communication—the balance of efficiency for the speaker and comprehensibility for the listener (Lindblom, 1963, 1983). Consequently, the study of coarticulatory patterns in developing speech has long been considered to provide a window into speech maturation. Previous research demonstrated a higher degree of intrasyllabic lingual coarticulation in children compared to adults (e.g., Nittrouer et al., 1989; Noiray et al., 2018; Noiray, Wieling, et al.,

2019). This means that in consonant–vowel (CV) syllables produced by children, the tongue position during the consonant overlaps more with the tongue position during the vowel than in adult productions. For example, in Noiray et al. (2018), the tongue position at /d/ constriction in the front–back dimension was more affected by the vowel context in children than in adults. Many have attributed these findings to developmental differences in independent control of speech articulators (e.g., Gibson & Ohde, 2007; Nijland et al., 2002; Nittrouer et al., 1989; Noiray et al., 2013, 2018). However, it is still not fully understood if differences in coarticulatory patterns are related to differences in control and coordination of individual speech articulators.

In this study, we tested whether the relationship between the developmental differences in CD reported in previous studies (Noiray et al., 2018; Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2018) relates to developmental

Correspondence to Dzhuma Abakarova: dzhuma.abakarova@uni-potsdam.de. **Disclosure:** The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.

differences in articulatory patterns. An articulatory pattern is understood here as the pattern of displacement and deformation of the speech articulators for achieving an articulatory–acoustic goal (Serrurier et al., 2019). A motor goal can be achieved in various ways. This phenomenon is known as motor equivalence (e.g., Kelso & Tuller, 1983; Perrier & Fuchs, 2015). For example, an alveolar constriction for /d/ can be produced in a variety of ways (see Figure 1). The tongue tip (TT) and tongue body (TB) may both be actively, synergistically involved in contributing to the alveolar constriction. This scenario is defined as *differentiated control* (for further studies using this term, see The Ontogenetic Maturation of Speech Motor Control section). Alternatively, when only the TT or the TB is actively engaged in the alveolar constriction, we consider the other articulator “comes along for the ride.”

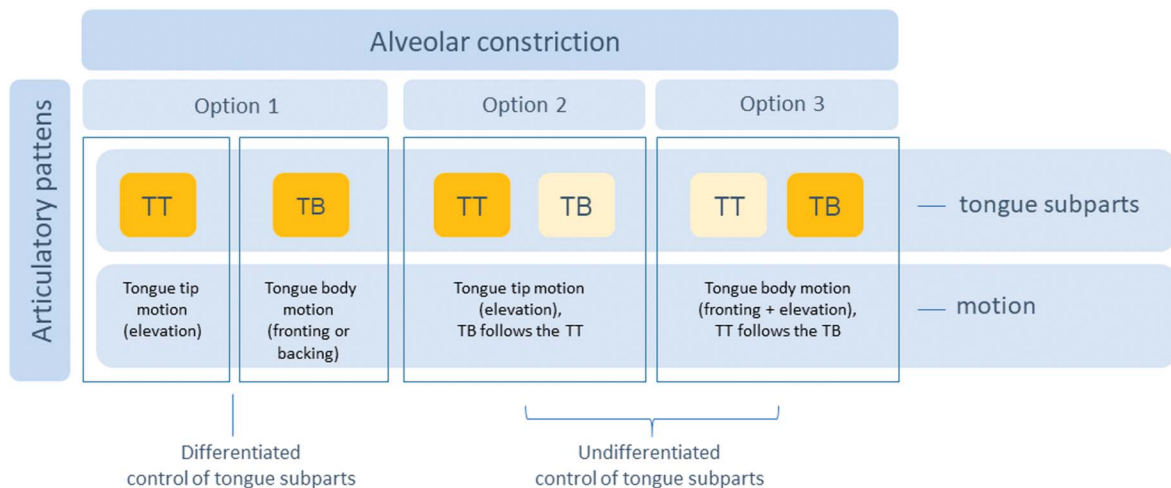
To test the relationship between the developmental differences in CD and developmental differences in articulatory patterns, this study combined analyses of ultrasound tongue data and task dynamical computational simulations. The ultrasound data were collected in 3-, 4-, 5-, and 7-year-old German-speaking children and adults (Noiray et al., 2018; Noiray, Wieling, et al., 2019; Rubertus & Noiray, 2018). For this data set, we have reported higher degrees of vocalic coarticulation onto previous segments in children compared to adults, both onto consonants within CV syllables and in longer sequences (e.g., V-to-V coarticulation in /amə #bidə/ where /i/ is the target vowel) in four consonant contexts. Here, we first reexamined the data for age-related differences in articulatory patterns, specifically, in tongue position at consonant constriction (temporal midpoint). Next, for each utterance, we generated a set of simulations with varying articulatory patterns to test whether they would exhibit the

same pattern of developmental differences in CD as observed experimentally.

The corpus of simulated utterances was created with the Task Dynamics Application (TaDA; Nam et al., 2004), a computer implementation of the Articulatory Phonology framework (Browman & Goldstein, 1989) that combines the Task Dynamics model of interarticulator speech coordination (Saltzman & Munhall, 1989) with a coupled-oscillator model of intergestural planning, and a gestural-coupling model. Task Dynamics is a dynamical theory of speech production that attempts to model how articulators move during speech, based on contrastive specifications of gestural parameters. This theoretical framework, along with the principles from articulatory phonology, helps us conceptualize why and how the differences in articulatory patterns may be relevant for coarticulation. According to the Task Dynamics and Articulatory Phonology perspectives, coarticulation results from the coproduction of speech gestures. Speech gestures are synergistic movements of the individual speech articulators (independently movable parts) that work together to achieve a gestural goal within a vocal tract. As exemplified in Figure 1, the gestural goal (in this case, a constriction for the alveolar stop /d/) can be achieved through several articulatory patterns. These patterns, in turn, affect the way gestures for consecutive vowels and consonants interact with one another and, therefore, their degree of coarticulation.

TaDA allows manipulating various parameters (e.g., set of articulators associated with a gesture) that may underlie the developmental differences in CD observed in our experimental data. While the Task Dynamics approach has mainly been applied to adult speech, it is a rich enough theory that different hypotheses about development can be

Figure 1. Schematic depiction of three hypothetical lingual patterns for the production of an alveolar constriction goal (e.g., /d/). In each scenario, the actively controlled articulators are represented in dark yellow. TT = tongue tip; TB = tongue body.



framed within it. In this work, we demonstrate the value of an explicit model for testing specific predictions about speech development. Before describing our experimental approach in more detail, we first summarize previous findings on speech motor control development and outline the relationship between coarticulation and articulatory patterns as viewed from the Task Dynamics perspective. Last, we lay out our hypotheses about the relationship between articulatory patterns and coarticulation across development.

The Ontogenetic Maturation of Speech Motor Control

Many studies on speech development have viewed coarticulatory patterns as an index of maturity of speech motor control (Maas & Mailend, 2017; Nijland et al., 2002; Nittrouer et al., 1989). Specifically, the finding of greater CD in children as compared to adults has been interpreted as evidence for lack of independent control over individual speech articulators in young children (Nijland et al., 2002; Nittrouer et al., 1989). Indeed, CD and control over individual speech articulators appear related. As mentioned above, from the articulatory phonology perspective, coarticulation arises from the coproduction of speech gestures. Speech gestures are coordinated movements of individual speech articulators for achieving a speech goal (Browman & Goldstein, 1989). Individual articulators can be coordinated in many ways, and the task of a developing speech motor system is to find the optimal way to achieve a speech goal. Speech movements should be invariant enough to keep the resulting acoustic productions immediately recognizable to listeners but flexible enough to achieve gestural goals in a variety of situations (e.g., different speech rates, task modification; Latash, 2012). A prerequisite for developing such coordinative structures is the ability to independently control speech articulators, including the subparts of the tongue (e.g., TT, TB). However, developing motor systems often show limited independent control of distinct anatomical units across a range of behaviors, including reaching, intermanual coordination, and locomotion (D'Souza et al., 2017; Shim et al., 2008; Sweeney et al., 2018). This is the case for speech articulators too (Green et al., 2000; Smith & Zelaznik, 2004).

The development of *lingual* control has not been studied extensively in children *younger than 6 years*, likely due to methodological challenges associated with collecting and analyzing tongue movement data in young children. A recent exception is the work of Kabakoff et al. (2021), which is described below.

However, there are studies on motor control development in young children for speech articulators other than tongue. For example, Green et al. (2002) investigated

the development of jaw (JA) and lip control in 1-, 2-, and 6-year-old children and young adults during productions of *baba*, *papa*, and *mama* reduplicative sequences. They reported a sequential development of articulatory control: while the JA exhibited stable adultlike patterns already in 1-year-olds, labial control approximated adults' patterns only by the age of 6 years.

Children are likely to need even more time to acquire adultlike control of the tongue subparts. The tongue, the object of investigation here, is an incompressible hydrostat, highly mobile and deformable. It has a nearly infinite number of mechanical degrees of freedom and no clearly identifiable anatomic landmarks (Sanguineti et al., 1997) as most muscles extend throughout its volume. It is, therefore, challenging to learn how to use the tongue subparts independently to achieve very precise goals in specified locations in the vocal tract.

Gibbon (1999) reviewed nine studies that used electropalatography (EPG) to study lingual articulation in typically developing children and children with speech sound disorders (SSDs). EPG data collected from 24 typically developing children, ages 6–14 years, showed that they produced alveolar stops such as /t/ and /d/ by a combination of lateral bracing and an upward movement of the TT/blade to the alveolar ridge, same as adults. In a sample of 17 children with SSDs, 4–12 years of age, 71% showed *undifferentiated lingual gestures* that involved unusually large areas of tongue-to-palate contact exceeding what is typical in adult production. Specifically, most of these children produce alveolar stop targets with lingual contact that spanned both alveolar and palatal/velar regions instead of localized contact on the alveolar ridge. The author suggested that undifferentiated lingual gestures point to a lack of independent control of the TT/blade, body, and lateral margins of the tongue, and hypothesized that undifferentiated gestures (UG) might be common in the speech of typically developing children younger than 6 years old. Cheng et al. (2007) also used EPG to study the development of lingual control in 6-year-old typically developing children and observed a higher amount of palatal contact in younger than older speakers. A more recent study (Cleland & Scobbie, 2021) investigated the lingual differentiation in alveolar and velar stops in 30 English-speaking typically developing children using ultrasound tongue imaging. They found no correlation between age and the degree of lingual differentiation across consonants. However, the authors indicated that the data used in the study had been collected in children aged 5;8 (years; months) to 12;10, with only seven children under the age of 8 years. The age effect on the degree of lingual differentiation may, therefore, be observed with a larger sample of young children.

Another recent study by Kabakoff et al. (2021) explored the relationship between tongue shape complexity

and phonemic development in 24 children (aged between 4;0 and 6;3 with seven children) and without SSD (17 children) producing various phonemes. They measured tongue shape complexity from ultrasound images and found that it was not significantly affected by age in typically developing children, except for the alveolar stop /t/. Interestingly, the correlation was negative, meaning that tongue shape complexity during /t/ production became lower with age. However, as the authors remarked, this result should be interpreted with caution. They also suggested that “the absence of a general association between age and tongue complexity may be attributed to the small age range of the children in the data set” (Kabakoff et al., 2021, p. 2571). Children’s patterns were not compared with adults in this study.

Previous studies have also demonstrated continued developmental differences in interarticulator coordination even after the independent control of speech articulators has been acquired. For example, Green et al. (2000) investigated the development of JA and lips coordination in 1-, 2-, and 6-year-old children and young adults and found significant changes in the relative contribution of the upper lip (UP), lower lip (LL), and JA into syllable production up to 6 years of age with the contribution of JA decreasing and the contribution of LL increasing significantly with age. Smith and Zelaznik (2004) further investigated the development of lip–JA coordination across the age span of 4–21 years and reported a decrease in the variability of interarticulator coupling and an increase in movement synchrony for children with increasing age.

Thus, it may be inferred from previous research that children differ from adults in their speech production patterns. One possible explanation for this finding is the non-uniform development of control over distinct speech articulators (e.g., Nittrouer, 1993): Children under 6 years of age may not have yet acquired adultlike independent control of the TT and tongue. However, it is essential to remember that motor development occurs in the context of significant anatomical changes from the first year of life (e.g., Boë et al., 2019; Ménard et al., 2004; Vorperian & Kent, 2007). On one hand, growing oral space (Bosma & Showacre, 1975) and increasing movement potential (Bosma, 1963) facilitate motor control maturation. On the other hand, this changing environment (Ménard et al., 2004, 2007) requires children to regularly adjust their articulatory patterns and thus presents a challenge that necessitates the development of fine motor control. It is possible that the child-like vocal tract morphology is not a limiting factor (Ménard et al., 2009) but rather a feature that allows for the use of simpler articulatory patterns than those produced by adults. From this perspective, children may use simplified patterns because they are just enough. Although originally undifferentiated lingual patterns were viewed as indicative of lack of ability, if found in young, typically developing children, they may also

indicate lack of necessity. It is not the goal of this study to reveal the origins of the potential absence of articulators’ differentiation in child articulations, whether caused by lack of motor control abilities or lack of intention. Here, we merely suggest that undifferentiated articulatory patterns should not be automatically associated with immature motor control. The aim of this article is first to test whether developmental changes in lingual CD are related to potential age-related differences in articulatory patterns.

Task Dynamics View of the Relationship Between CD and Articulatory Patterns

Measures of CD (e.g., Abakarova et al., 2018; Lindblom, 1963; Recasens, 1987) have been widely used to understand linguistic and motor development (Cychosz et al., 2021; Fricke & Johnson, 2012; Nijland et al., 2002; Noiray et al., 2013, 2018; Rubertus & Noiray, 2018; Zharkova et al., 2014). These measures determine the amount of articulatory overlap between two speech segments by using direct (kinematic) or indirect (acoustic) measurements to describe the shape of the vocal tract during their production and then quantifying the similarity between the two shapes. However, it is not clear what underlying phenomena are captured by such measures of CD.

The articulatory implementations of coproduced gestures can vary depending on their relative timing and mechanical interactions (Fowler & Saltzman, 1993; Iskarous et al., 2013; Recasens & Espinosa, 2009). According to Fowler and Saltzman (1993), the variations in articulatory implementations induced by coproduction (temporal overlap) of gestures depend on the degree to which the gestures share articulators, that is, on the degree of spatial overlap. Temporally overlapping gestures combine (blend) their influences on vocal tract configurations. When consecutive gestures share a single articulator (e.g., only the JA is shared by /b/ and V in /VbV/ sequences), the effects of gestural interference will be small, and temporal overlap between vocalic and consonantal gestures will take place with minimal spatial perturbations. The highest degree of spatial perturbation occurs when two overlapping gestures share the articulators and impose competing demands on them. For example, in the production of alveolar stop consonants in VCV sequences, the TT is employed by the consonantal gesture alone. However, the JA and the TB are employed by both the consonantal and the overlapping vocalic gestures that compete for the control of the two articulators. Then, the observed coarticulatory V-to-C effects on the TB and the JA reflect a blending of the influences of the overlapping V and C gestures.

Thus, systematic differences in CD may be related not only to differences in motor planning, that is, temporal coordination of gestures but also to differences in the set of independent articulators involved in producing a target gesture. As suggested in the previous section, children may use a different set of articulators for producing a gesture compared to adults. For example, although different parts of the tongue, like the tip and the body, are connected, they may be independently controlled as two separate units by adults or as a single unit by children. The differential use of different subparts of the tongue may result in differences in CDs.

The use of different articulatory patterns can directly affect the amount of variability allowed for an articulator. In the case of alveolar stops, according to Mermelstein (1973), the adult pattern for alveolar constriction has a consonantal TB target to avoid undue extension of the tongue blade during TT raising toward the alveolar ridge. Because TB position needs to prevent such extension, it cannot go backward for back vowels. This would explain the observation from our previous study (Noiray, Wieling, et al., 2019): We found that in adult utterances in the context of the alveolar stop /d/, the position of the TB remained in a front to central position during the schwa, at consonant onset, and even at consonant midpoint regardless of the vowel context. If children only use the TT to create an alveolar constriction, the TB may be freer to move in correlation with both front and back vowels.

Research Questions, Hypotheses, and Predictions

This study investigated whether developmental differences in lingual CD are related to differences in articulatory patterns. For that, we needed to answer two questions: (a) Are there differences in articulatory patterns for the production of stop consonants between children and adults? (b) If yes, do these differences result in the previously reported related differences in CD?

Regarding our first research question, we started with the assumption that different articulatory patterns (e.g., different involvement of the tongue subparts) should be reflected in tongue shape and positioning during consonant production. Based on the previous findings on the development of motor control and vocal tract anatomy, we hypothesized that children would use different lingual articulatory patterns for stop production than adults.

To test this hypothesis, we investigated TB positions for stop production in 3-, 4-, 5-, and 7-year-old children and adults producing /b, d, g/ stops in nonwords preceded by a carrier word “eine” (/a ɪ n ə/). Since the development of speech motor control is often nonlinear and changes can occur rapidly (e.g., Nittrouer, 1993), we examined

each age group separately to identify precisely at which age the transition to more differentiated articulatory patterns may occur. We also looked at different consonant contexts separately, as different mechanisms may explain age differences in CD, depending on the consonant. For example, for consonants that do not involve differentiated actions from the tongue subparts (e.g., /b/), variation in CD may result from differences in the temporal coordination of articulatory gestures within CV syllables, not from differences in their articulatory implementations. To quantify differences in tongue position, both in terms of tongue front (TF)-backness and raising, we used measures of curvature position and curvature degree, respectively.

Since /b/ is not a lingual consonant, it may not be intuitive why independent actions of the tongue subparts should play a role in its coarticulatory behavior. However, previous analyses of our data (Noiray, Wieling, et al., 2019) have demonstrated age differences in CD as early as the first schwa midpoint in the utterances of the form /amə1#C1V1ə2/. Therefore, we hypothesized that this difference might be due to the tongue position for /n/ production.

Regarding our second research question, we hypothesized that age differences in articulatory patterns result in the observed age differences in CD. To test this hypothesis, we manipulated the articulatory patterns in TaDA to simulate the experimental productions and then compared CDs in child-like and adultlike simulated utterances. We expected our simulations to result in similar developmental differences as the experimental data, that is, greater vocalic CDs for all stops in children compared to adults.

Experiment 1

Method

Data Set

The experimental data used as a reference for modeling come from ultrasound recordings collected in 65 German-speaking 3-, 4-, and 5-year-old preschoolers as well as seventeen 7-year-olds and 11 adults with no reported history of speech, language, or hearing difficulties. Table 1 summarizes the information about participants and observations.

Participants were all recorded at the Laboratory for Oral Language Acquisition (LOLA) at the University of Potsdam. The productions were collected in a nonword repetition task. Participants were instructed to repeat pre-recorded and auditorily presented stimuli produced by an adult female speaker of German. The stimuli consisted of disyllabic /C1VC2ə/ nonwords preceded by the carrier word “eine” (/a ɪ n ə/). The first target syllable consisted of one of the three consonants /b/, /d/, or /g/ and, to get a

Table 1. The number of participants (sex distribution is given in parentheses) and experimental observations (obs) for five age groups (3-year-olds [C3], 4-year-olds [C4], 5-year-olds [C5], 7-year olds [C7], and adults [A], and three consonants [b, d, g]).

Age group		C3	C4	C5	C7	A
<i>n</i> subjects		18 (9f)	14 (9f)	16 (7f)	17 (11f)	11 (6f)
Age range		3,5–3,7	4,4–4,7	5,4–5,7	6,11–7,5	19–28
<i>n</i> obs	/b/	6639	7201	7701	8827	6226
	/d/	6410	7234	7859	9018	6226
	/g/	6396	7094	7534	8902	6125
	Total	19,445	21,529	23,094	26,747	18,577

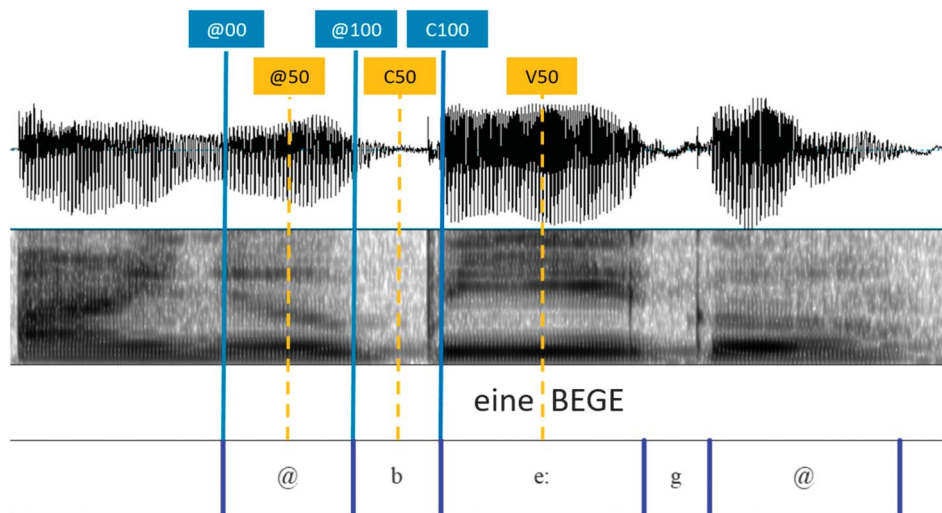
large degree of potential context effects, one of the six long vowels /i/, /y/, /u/, /a/, /e/, or /o/. The second syllable also included one of the consonants above and the vowel schwa (/ə/). The first and the second syllables were combined in such a way that C1 was never the same as C2. This resulted in 24 target utterances. Stress was always on the first syllable. Target consonants were selected such that children as young as 3 years of age could reliably produce them. Target vowels were chosen to represent the vowel space maximally, that is, by including the opposition of front–back vowels, high–low vowels, and rounded–unrounded vowels while keeping the number of stimuli minimal to accommodate for young children’s attention span. The carrier word “eine” (/a i n ə/), which ends with the neutral vowel /ə/, was selected to approximate a neutral tongue position and, thus, prevent strong coarticulatory effects on the target syllable from the unstressed syllable of the preceding word.

All participants were recorded within the SOLLAR platform created in our laboratory (Sonographic and Optical Linguo-Labial Articulation Recording system; Noiray et al., 2020). The platform allows for the recording of tongue movement via ultrasound imaging (Sonosite Edge portable ultrasound imaging device, sampling rate: 48 Hz, ultrasound zoom depth: 10 mm), the audio speech signal (Sennheiser microphone, sampling rate: 48 kHz), and labial-shape tracking via video recording (Sony camera, 60 fps). In SOLLAR, the ultrasound probe is positioned below the participants’ chin to record the tongue surface contour on the midsagittal plane. It is placed in a custom-made probe holder constructed with a system of light springs and ball bearings to allow the probe to move smoothly down with the JA while the participant speaks. The probe holder is mounted in a custom-made pedestal and on an electrical table adjustable to the participant’s height. The participant sits perpendicular to the probe holder with the small probe positioned below his chin between the maxillary bones. In this study, we did not use a fixed headset (a) to avoid constraining children’s JA movements and hence maximize the naturalness of the speech recorded and (b) to maximize the probability of a successful recording. For a complete description of the data collection procedure, see Noiray et al. (2018) and Noiray, Wieling, et al. (2019).

All acoustic and ultrasound data processing was performed by the first and the third author, as well as university students with project-specific training. First, acoustic data were phonetically transcribed using Praat (Boersma & Weenink, 2021). Repetitions that did not match the model speaker’s production were discarded from further analysis, except for 3-year-olds. Here, productions were kept in the data set as long as ə#C1V productions corresponded to the model word and C2 did not differ in the place of articulation from C2 in the model word (e.g., /amə ba:tə/ was kept for /amə ba:də/). This resulted in two instances of words with C2 = /k/ instead of /g/, 17 with C2 = /t/ instead of /d/, and 10 with C2 = /v/ instead of /b/. An example of segmentation is provided in Figure 2. Segments were delimited based on the stability of both formants and the periodic cycle in the oscillogram. The onset of stop consonants was identified as the beginning of the closure phase. The vocalic onset was identified as the first ascending zero-crossing in the oscillogram at the beginning of the periodicity (see Figure 2). From the resulting intervals, the timestamps for the eight target time points (i.e., the onset [@00], the midpoint [@50], and offset [@100] of the schwa; the temporal midpoint of the consonant [C50] and offset of the consonant [C100]; and the temporal midpoint of the vowel [V50]) were automatically extracted. These time stamps were subsequently used to find the corresponding frames in the ultrasound video signal.

For each relevant frame, tongue contours were then detected semi-automatically with MATLAB (The MathWorks, Inc., 2020) custom-made scripts that were a part of the SOLLAR Suite (see Figure 3; Noiray et al., 2020). Detection was done by placing 10 anchors along the lower boundary of the white line visible on the ultrasound image (see Figure 3). Any frame that was judged to be off-center or unclear was discarded. In most included contours, the tongue’s surface extended from the hyoid shadow to the mandibular shadow. In cases where the whole tongue was not visible, the following procedure was used: The preceding and following frames were consulted to see if the contour could be reconstructed. In the case of /o/ and /g/, the frames with hyoid shadow were not discarded as all visible tongue contour in the production of

Figure 2. Example of phonetic segmentation of /ainə #be:ga/ with Praat. C50 corresponds to the temporal midpoint of the consonant; V50 corresponds to the temporal midpoint of the vowel. The “@” label refers to schwa. 00, 50, and 00 labels refer to temporal onset, midpoint, and offset of the segment, respectively.



these segments is located in the mid-to-back area. Consequently, even if the hyoid shadow may have obstructed the most anterior part of the tongue during the production of those segments, it did not prevent us from tracking the mid and posterior portion of the tongue contour. All tongue contours were checked by a second labeler and manually corrected if necessary. For each relevant contour, the coordinates of the highest point of the TB were automatically extracted and used for subsequent analyses.

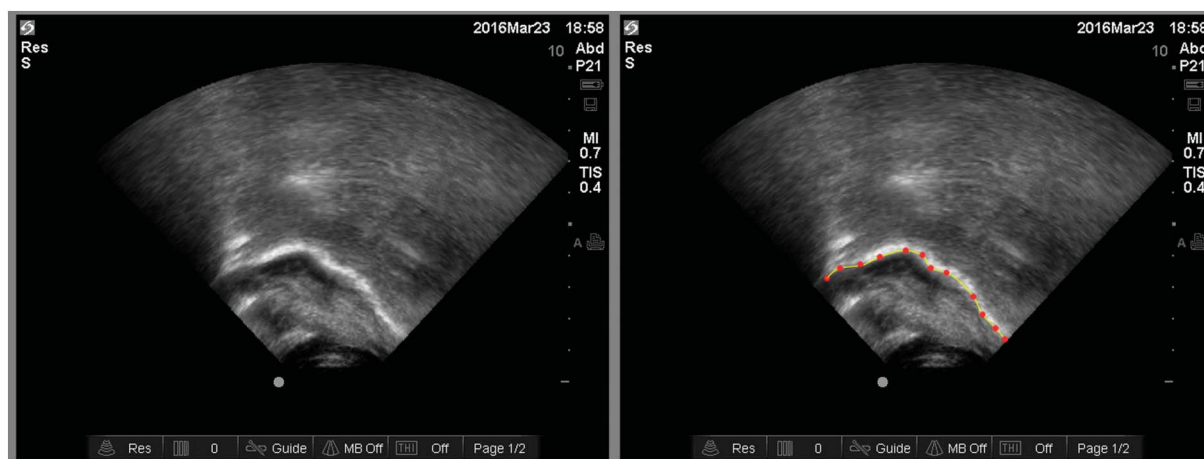
Finally, we checked for outliers that could be attributed to a difference in view range where the most anterior or the most posterior regions of the contour were not captured. For this, the outliers in labeled contours were

identified within speaker and target segment. Outliers were defined as contours for which the horizontal position of the highest point on the tongue was more than 2 *SDs* away from the average horizontal position of the highest point on the tongue contour. In this manner, a total of 102 observations were discarded.

Quantifying the Differences in Articulatory Patterns

In previous studies (Noiray et al., 2018; Noiray, Wieling, et al., 2019), we found a greater degree of anticipatory V-to-C coarticulation in the child than in adult speech productions.

Figure 3. Example of an ultrasound image capturing the midsagittal tongue contour at the temporal midpoint of the alveolar stop /d/ produced by a 4-year-old female. The left panel presents the raw ultrasound image; the right panel shows the highlighted tongue contour resulting from SOLLAR’s semi-automatic tracking. The gray circle indicates the anterior part of the tongue.



Specifically, in adults, the horizontal position of the TB highest point at C50 was relatively fronted in all vowel contexts for the stops (/b, g, d/). In children, however, the position of TB's highest point varied as a function of the upcoming vowel position. To answer our first research question, whether children and adults employ different articulatory patterns for producing the target consonants, we examined the articulatory data. The assumption here was that different articulatory patterns (e.g., different involvement of tongue subparts) should be reflected in tongue shape and positioning during consonant production.

Figure 4 provides examples of midsagittal tongue contours captured at two time points, C50 (solid lines) and V50 (dotted lines), during the production of CV syllables produced by a 4-year-old girl and an adult woman. CV syllables consisted of C /b, d, g/ produced in /i:/ and /o:/ vocalic contexts (color online: shown in red and blue, respectively). For example, in the upper left panel, red contours show the tongue shape at two time points during the production of /bi:/, at C50 (solid line) and V50 (dotted line). Blue contours show the tongue shape at two time points during the production of /bo:/ at C50 (solid line) and V50 (dotted line). For visualization purposes, we only show some instances. Vowel contexts are chosen here to represent front and back vowels; displaying all vowel contexts would result in unreadable plots due to the high number of observations. This figure, therefore, illustrates both tongue positions for different stops and the extent of similarity between the tongue shapes for the stop and the vowel in the same CV syllable.

Regarding tongue position in the front-back dimension, in the child compared to the adult speaker, the highest point of the tongue appears more fronted relative to the whole tongue in /b/ and /d/ examples. In adults, for all consonants except for /g/, the vertical position of the tongue during vowel production (dotted lines) is higher than that during consonant production (solid lines). In child productions, the tongue shape at the consonant midpoint (solid line) seems to be closer to the tongue shape at

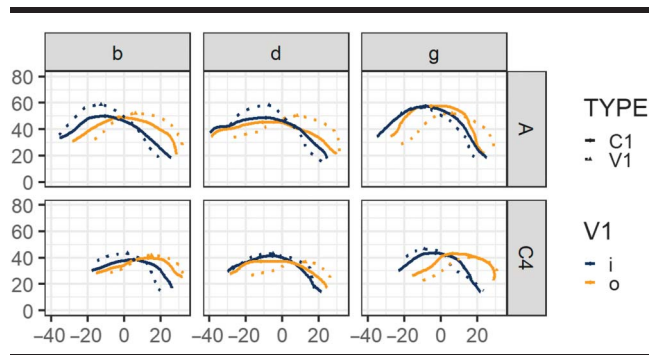
the respective vowel midpoint (dotted line) for all stops. There is hardly a difference in the vertical position of the highest point of the tongue between C50 and V50.

To check whether these individual examples represent general patterns, we quantified the observed differences in consonantal tongue contours between child and adult participants. Various measures may be used for this purpose, such as the measures of tongue shape complexity and relative horizontal and vertical position of the tongue (Dawson et al., 2016; Kabakoff et al., 2021; Ménard et al., 2012). Measures of tongue shape complexity are problematic for our purposes, especially in the case of alveolar consonants. A more complex tongue shape for alveolar production implies differentiated action of the TT and TB. If such action were present, we would see TT raised whereas TB not necessarily fronted or raised. Thus, if adult tongue shapes for alveolar consonants were more complex, the difference would be found in the very front region, at the TT. Yet, with ultrasound imaging, the TT is not always captured in adult speakers; it may be hidden by the mandible bone shadow. Consequently, we turn to positional measures for tongue shape quantification. Specifically, we chose the positional measures of the TB.

When choosing quantification measures, another point to keep in mind was that our data are not corrected for head movement. This kind of setting is common for studies with young children. Typically, the experimenter holds the probe under the subject's chin, or a spring-loaded probe stand is used to allow the probe to follow JA motion. The latter was the case in our experiment. The advantage of such a setting is that the ultrasound transducer follows JA movement, which limits tissue compression under the tongue. In such an unconstrained setting, the displacements of the head and/or probe can be tracked using flesh-point tracking methods such as HOCUS (Whalen et al., 2005). In our data set, movement correction was not applied because we did not have a movement correction system in place at the moment of data collection. It was also not strictly necessary because the data set was originally meant for investigating coarticulation within CV syllables. The measures of coarticulation were always relative and compared the tongue position at consonant midpoint with the tongue position at vowel midpoint within a CV syllable. These two points are so close in time that participants' head position relative to the probe could not change substantially between C and V midpoints.

However, unlike with coarticulation measures, measures based on single curves were required for this study. In a constrained head-probe setting, the reference coordinates are constant, so a variety of parameters can be considered. In an unconstrained setting like in our data, it is more appropriate to employ methods specifically designed for analyzing data that do not rely on a fixed reference plane. For this reason, we calculated tongue curvature

Figure 4. Examples of midsagittal tongue contours for three consonants /b, d, g/ (solid line), and two vowels (dotted line) at their temporal midpoint, for a female 4-year-old (C4) and an adult female (A). The left side of a panel: anterior part of the tongue.



position and degree using LINGUA (Ménard et al., 2012). Curvature position is a measure of the position of the tongue's mass relative to the whole tongue and is related to front-backness. Tongue curvature degree is a correlate of tongue bunching-flatness. These measurements have been shown (Ménard et al., 2012; Zharkova et al., 2015) to be interpretable in the presence of JA or head displacement (unlike raw position parameters), thus making them particularly suitable for child studies.

When comparing tongue shapes across age groups, it is important to consider the potential age differences in the anatomical shape of the tongue. For example, tongue curvature position could be naturally more fronted in children. To address this, we normalized both curvature position and curvature degree values for consonant midpoint by relating them to curvature positions and curvature degree at vowel midpoint, respectively, for each participant.

While curvature degree gives insight into tongue shaping, it does not quantify tongue elevation relative to the hard palate because it does not measure the distances between selected points along the tongue surface and points at the palate, perpendicular to the tongue. For example, if we find a flatter tongue shape in children, there are different possible explanations for this result. First, the flatter tongue shape may be a consequence of the flatter hard palate in children. Second, it is possible that the TT and TB are not controlled independently so they may be raised, but to the same level, resulting in a flat tongue shape. If the TB was raised alone, TT would be lowered down; if it was TT alone, TB would not be so high up at the same level. To estimate the relative height of the TB, we used a relative measure comparing the curvature degree difference at C50 and V50. This was calculated by subtracting the curvature degree for the consonant from the curvature degree for the high vowels. Curvature degree values for consonant and vowel were normalized to the 0:5 range.

Then, to test whether curvature position and curvature degree differed significantly between age groups, we fitted linear mixed models in R (lme4 package, Bates et al., 2015) with curvature position or curvature degree as the dependent variable, and the interaction of age and consonant as predictors. The structure of random effects for these models and all the other linear mixed models presented later in the paper was determined following the approach suggested by Bates et al. (2018). This approach combines principal components analysis (PCA) to determine the maximal number of dimensions supported by the data with likelihood ratio tests to assess the goodness of fit. We began by testing the full random effects structure for subject and word. If the maximal model failed, we dropped variance components until the identification was achieved. Once the model converged, we used PCA to check whether the data supported this number of

dimensions. If the number of dimensions was not supported (i.e., if the number of principal components that cumulatively accounted for 100% of the variance in the data was lower than the number of dimensions in the model), we dropped the smallest variance component. As a result, random intercepts for subjects were included as random effects in both models. The model assumptions were checked by visual inspection of the residual plots. Outliers were checked individually and either removed (in case of experimental errors) or corrected (in case of processing errors). Removing outliers did not result in any changes in the outcome pattern. In both models, the curvature values were power-transformed to better approximate normality using the *BoxCox* function from the R package “forecast” (Hyndman & Khandakar, 2008). The pairwise comparisons between age groups were performed with the help of the *emmeans* function (Lenth et al., 2018). In this function, the *p* values are automatically corrected for multiple comparisons using Tukey's honestly significant difference (HSD) method (Bretz et al., 2008).

Results

Table 2 provides mean normalized values for curvature positions, curvature degrees, and differences between curvature degrees for C and V (Ccurvdeg-Vcurvdeg) for the three consonants (/b, d, g/) across five age groups. The results of the linear mixed model analysis comparing the average tongue curvature position indices across age are presented in Table 3. When the value for curvature position index is low, the most elevated part of the tongue

Table 2. Means (*M*) and standard deviations (*SD*) for curvature position, curvature degree, and difference in curvature degree between consonant and vowel (Ccurvdeg-Vcurvdeg) for three consonants (/b, d, g/) across five age groups.

C	Age	Curvature position		Curvature degree		Ccurvdeg-Vcurvdeg	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
/b/	C3	1.01	1.18	2.89	0.66	0.08	0.21
	C4	1.14	3.77	3	0.83	0.13	0.18
	C5	1.22	4.34	2.97	0.72	0.06	0.21
	C7	1.09	1.21	2.99	0.78	0.05	0.19
	A	1.09	0.4	2.87	0.81	-0.03	0.17
/d/	C3	0.84	0.34	3.56	0.81	-0.08	0.25
	C4	1.23	9.41	3.82	1.13	0.02	0.23
	C5	0.95	0.41	3.54	0.75	-0.06	0.26
	C7	0.91	0.51	3.72	0.88	-0.07	0.24
	A	1.09	0.39	3.64	1.26	-0.16	0.2
/g/	C3	1.07	1.53	2.39	0.5	0.18	0.22
	C4	1.46	8.25	2.27	0.48	0.22	0.2
	C5	2.12	9.49	2.32	0.59	0.19	0.22
	C7	1.76	12.6	2.14	0.48	0.18	0.2
	A	1.11	2.11	1.83	0.31	0.16	0.2

Table 3. The output of the linear mixed models comparing the curvature position at C50 for three stop contexts (/b, d, g/) across five age groups.

C1	Contrast	Estimate	SE	Z ratio	p value
/b/	A-C3	0.232	0.045	5.181	< .001
	A-C4	0.163	0.045	3.61	.003
	A-C5	0.076	0.044	1.73	.415
	A-C7	0.103	0.043	2.389	.118
	C3-C4	-0.070	0.043	-1.64	.472
	C3-C5	-0.157	0.041	-3.792	.001
	C3-C7	-0.130	0.041	-3.196	.012
	C4-C5	-0.087	0.042	-2.091	.224
	C4-C7	-0.060	0.041	-1.465	.585
	C5-C7	0.027	0.039	0.689	.959
/d/	A-C3	0.246	0.044	5.576	< .001
	A-C4	0.303	0.045	6.724	< .001
	A-C5	0.142	0.044	3.232	.011
	A-C7	0.167	0.043	3.883	.001
	C3-C4	0.056	0.042	1.344	.664
	C3-C5	-0.105	0.041	-2.57	.076
	C3-C7	-0.079	0.040	-1.999	.267
	C4-C5	-0.161	0.047	-3.868	.001
	C4-C7	-0.136	0.041	-3.339	.008
	C5-C7	0.025	0.039	0.638	.969
/g/	A-C3	0.067	0.045	1.501	.562
	A-C4	-0.026	0.045	-0.568	.980
	A-C5	-0.043	0.044	-0.954	.876
	A-C7	0.057	0.043	1.325	.676
	C3-C4	-0.092	0.042	-2.191	.183
	C3-C5	-0.109	0.041	-2.635	.064
	C3-C7	-0.010	0.040	-0.239	.999
	C4-C5	-0.017	0.042	-0.397	.995
	C4-C7	0.083	0.041	2.031	.251
	C5-C7	0.099	0.040	2.49	.093

Note. Significant contrasts ($p < .05$) are shown in bold.

curve is further front relative to the whole tongue. A higher curvature position index indicates that the most bunched part of the tongue is positioned relatively back with respect to the whole tongue. The direction of the comparison can be seen in the “contrast” column. For example, in the first line, (“A-C3”), C3 age group is compared to the A age group. When the sign is positive, this means that the curvature position value for adults is higher than that for children.

First, all child age groups show significantly more fronted tongue curvature position compared to adults at /d/ temporal midpoint ($p < .001$ for 3- and 4-year-olds, $p = .01$ in 5-year-olds, and $p = .001$ in 7-year-olds). In 4-year-olds, the curvature position relative to the whole tongue was significantly less fronted than in 5-year-old and 7-year-old children ($p = .001$ and $p = .008$, respectively). There were no differences between other child groups. For /b/, the curvature position relative to the whole tongue was more fronted in 3- and 4-year-old children when compared to adults ($p < .001$ and $p = .003$, respectively), but not in 5- and 7-year-old children. The curvature position also differed significantly in C3 when compared to C5 and C7 ($p = .001$ and $p = .012$,

respectively). Last, for /g/, the tongue curvature position did not differ between adult and child cohorts.

Table 4 presents the results of the linear mixed model analysis comparing the average curvature degree values at C50 for three consonants (/b, d, g/) across ages. As a reminder, a low curvature degree index means the tongue is more bunched, whereas a high curvature degree index indicates that the tongue is flatter. In the /b/ context, 3-year-old children showed significantly less bunching than all other age groups (with $p < .001$ when compared to 4-, 5-year-olds, and adults and with $p = .002$ when compared to 7-year-olds). In /d/ contexts, the tongue was significantly flatter in 3-year-olds than in 4-, 5-, and 7-year olds (with $p < .001$, $p = .002$, and $p = .03$, respectively). The difference between 3-year-olds and adults was only marginally significant ($p = .07$). Curvature degrees were also significantly lower in 5-year-olds when compared to 7-year-olds ($p = .03$) and adults ($p = .01$). No age differences in curvature degree were found in the /g/ context.

Finally, we tested for age difference in curvature degree for the consonant relative to the vowel. For that, we compared the difference between the curvature degree at C50 and V50 for each CV syllable. Results are

Table 4. The linear mixed models-based pairwise comparisons of the curvature degree at C50 for five age groups in three stop contexts (/b, d, g/).

C1	Contrast	Estimate	SE	Z ratio	p value
b	A-C7	-0.017	0.025	-0.701	.956
	A-C5	-0.011	0.024	-0.445	.992
	A-C4	-0.027	0.024	-1.149	.780
	A-C3	-0.116	0.026	-4.452	< .001
	C7-C5	0.007	0.025	0.27	.999
	C7-C4	-0.01	0.024	-0.405	.994
	C7-C3	-0.099	0.026	-3.728	.002
	C5-C4	-0.016	0.023	-0.701	.956
	C5-C3	-0.105	0.026	-4.074	< .001
	C4-C3	-0.089	0.025	-3.51	< .001
d	A-C7	0.006	0.025	0.241	.999
	A-C5	0.078	0.024	3.213	.011
	A-C4	0.026	0.025	1.089	.812
	A-C3	-0.068	0.026	-2.606	.069
	C7-C5	0.072	0.025	2.913	.029
	C7-C4	0.02	0.024	0.82	.924
	C7-C3	-0.074	0.026	-2.794	.042
	C5-C4	-0.052	0.023	-2.231	.168
	C5-C3	-0.146	0.026	-5.647	1.63E-07
	C4-C3	-0.094	0.025	-3.71	.002
g	A-C7	-0.008	0.025	-0.333	.997
	A-C5	-0.014	0.024	-0.562	.980
	A-C4	-0.006	0.024	-0.247	.999
	A-C3	-0.021	0.026	-0.807	.929
	C7-C5	-0.005	0.025	-0.216	.999
	C7-C4	0.003	0.024	0.104	.999
	C7-C3	-0.012	0.027	-0.479	.989
	C5-C4	0.008	0.023	0.336	.997
	C5-C3	-0.007	0.026	-0.284	.999
	C4-C3	-0.015	0.025	-0.603	.975

Note. Significant contrasts ($p < .05$) are shown in bold.

presented in Table 5. In /b/ context, differences were significantly larger in adults as compared to all child groups ($p < .01$). Similarly, in the /d/ context, there were larger differences between curvature degree for C and V for adult than child groups (with $p = .003$, $p < .0001$, $p = .04$, and $p = .049$ for comparisons with 3-, 4-, 5-, and 7-year-olds, respectively). In /d/ context, the difference in tongue bunching for C and V was also significantly larger for 5- and 7-year-olds when compared to 4-year-olds ($p = .04$ and $p = .02$, respectively). The relative curvature degree for /g/ did not differ across age groups.

Discussion of the Experimental Results

To evaluate developmental differences in articulatory patterns, we compared two measures of tongue shape—curvature position and curvature degree—between adults and children. From the results of these measurements, several conclusions can be drawn. First, TB positions for the production of the stop consonants /b/ and /d/ were more fronted in all child groups compared to adults. It was, however, not the case for the velar stop /g/. This

result suggests that adults produce alveolar stops without moving the TB as far forward as children do. As for curvature degree, in /b/ context, the only significant difference involved it being significantly lower in 3-year-old children compared to all other age groups. In /d/ context, the results were rather inconsistent. There was no significant difference for /g/ among all age groups. However, the relative curvature degree (the difference in curvature degree for the consonant relative to the curvature degree for the vowel) was smaller for all child age groups as compared to adults for /b/ and /d/. There was again no significant difference for /g/. Based on the experimental data, we conclude that the tongue shapes employed by children differ from adults for /b/ and /d/, but not for /g/ production.

A general challenge in interpreting speech acquisition data is that they can be driven by a variety of underlying factors. The only way to reliably test and understand the driving forces of speech acquisition is to model the data. Modeling has the advantage of controlling the input and allowing to compare it with the experimental data. In this study, we use the TaDA application to compare whether the previously observed age differences in CDs result from differences in articulatory patterns.

Table 5. The output of the linear mixed models comparing the curvature degree difference at C50 and V50 across five age groups for the three target consonants.

C1	Contrast	Estimate	SE	Z ratio	p value
/b/	A-C7	0.12	0.026	4.551	< .001
	A-C5	0.101	0.027	3.791	.001
	A-C4	0.102	0.026	3.912	< .001
	A-C3	0.089	0.026	3.47	.005
	C7-C5	-0.019	0.025	-0.749	.945
	C7-C4	-0.018	0.025	-0.723	.951
	C7-C3	-0.031	0.024	-1.308	.686
	C5-C4	0.001	0.025	0.043	.999
	C5-C3	-0.013	0.024	-0.514	.986
	C4-C3	-0.014	0.024	-0.572	.979
/d/	A-C7	0.072	0.026	2.734	.049
	A-C5	0.074	0.027	2.778	.043
	A-C4	0.146	0.026	5.582	< .001
	A-C3	0.093	0.025	3.634	.003
	C7-C5	0.002	0.025	0.092	.999
	C7-C4	0.074	0.025	3.004	.022
	C7-C3	0.021	0.024	0.863	.910
	C5-C4	0.072	0.025	2.858	.035
	C5-C3	0.018	0.024	0.752	.944
	C4-C3	-0.053	0.024	-2.248	.162
/g/	A-C7	0.017	0.026	0.649	.967
	A-C5	0.011	0.027	0.418	.993
	A-C4	0.008	0.026	0.292	.998
	A-C3	0.016	0.026	0.616	.973
	C7-C5	-0.006	0.025	-0.233	.999
	C7-C4	-0.01	0.025	-0.384	.995
	C7-C3	-0.001	0.024	-0.059	.999
	C5-C4	-0.004	0.025	-0.142	.999
	C5-C3	0.005	0.024	0.184	.999
	C4-C3	0.008	0.024	0.34	.997

Note. Significant contrasts ($p < .05$) are shown in bold.

Experiment 2

Method

Task Dynamics Application

The simulations were produced using TaDA (Nam et al., 2004), a computer implementation of Articulatory Phonology framework (Browman & Goldstein, 1992) that combines the Task Dynamic model (Saltzman & Munhall, 1989), a coupled-oscillator model of intergestural planning (Goldstein et al., 2009), and a gestural-coupling model. The flow of information through TaDA is schematically represented in Figure 5. TaDA works in the following way: An utterance in the form of either an orthographic or ARPABET text grouped into syllables is used as input to the gestural coupling model. For each phonemic unit in the utterance, the corresponding entry is found in the segment-to-gesture dictionary that specifies the gestures and dynamical control parameters associated with the unit. Each gesture controls one of the organs of the vocal tract: LIPS, TT, TB, velum (VEL), and glottis (GLO). The gestural targets for the constriction of these vocal tract organs are defined along a pair of tract variables (TVs): lip protrusion (LP), lip aperture (LA), tongue tip constriction location (TTCL), tongue tip constriction degree (TTCD), tongue body constriction location (TBCL), tongue body constriction degree (TBCD), VEL, and GLO (see

Figure 5. The flow of information through Task Dynamics Application. Modules are presented by squares, and output is represented by circles.

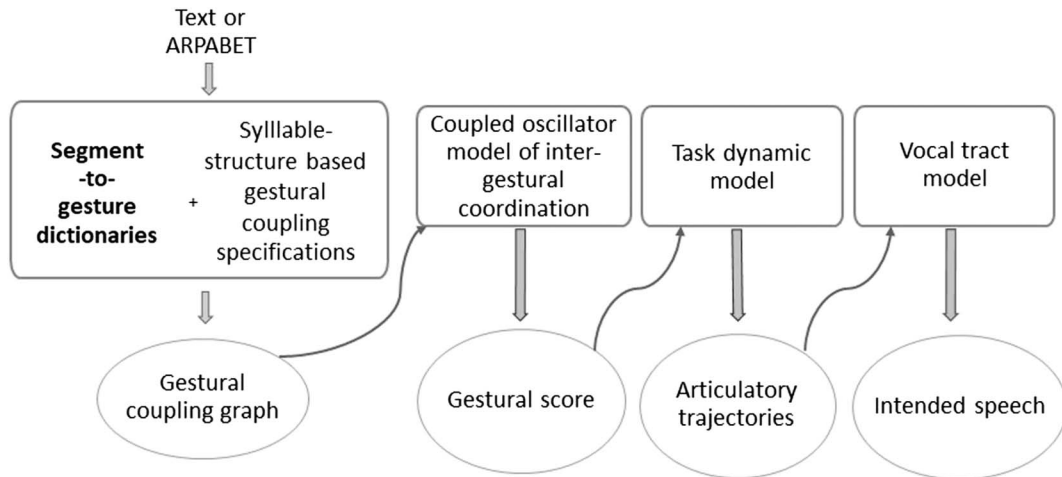
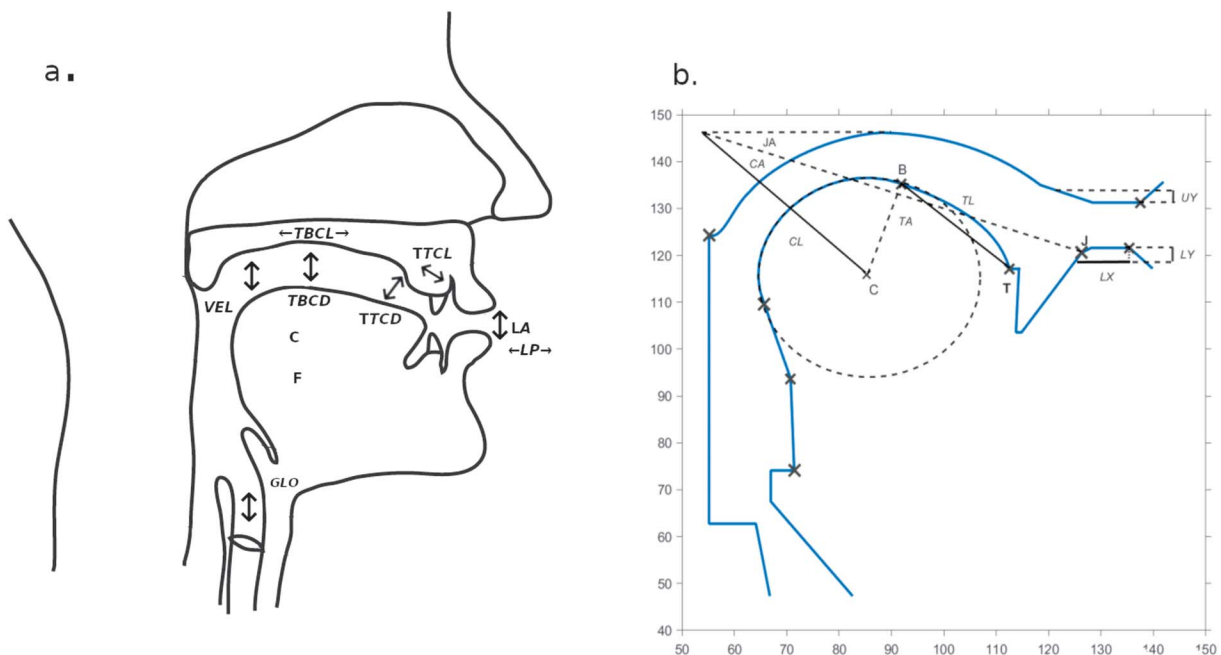


Figure 6). Constriction location represents the placement of the gestural constriction along the vocal tract (i.e., further front or back) and constriction degree represents the size of the constriction at the gestural target (i.e., wider or narrower constriction). The targets of LP, LA, TTCD, and TBCL gestures are defined in millimeters and those of

VEL and GLO gestures in arbitrary units. The targets of LP are the horizontal location of the lips, and those of LA, TTCD, and TBCL are the constriction degree of the constriction organs. The targets of TTCL and TBCL are defined using a polar grid, ranging from 0° to 180° as shown in Figure 6, in which 0° is in front of the chin, 90°

Figure 6. Figure 6. (a) Midsagittal section of the vocal TVs at five distinct constriction organs. (b) A visualization of the Configurable Articulatory Synthesizer (CASy) in a neutral position. The vocal tract outline is shown in gray. The model articulators are shown as crosses and labeled in bold. Dotted lines are geometric reference lines used to define the articulatory parameters (black lines and angles, labeled in italics). TBCL = tongue body constriction location; TTCL = tongue tip location constriction; LP = lip protrusion; LA = lip aperture; TTCD = tongue tip constriction degree; TBCL = tongue body constriction degree; VEL = velum; GLO = glottis. Adapted from Lammert, A., Goldstein, L., Narayanan, S., & Iskarous, K. (2013). Statistical methods for estimation of direct and differential kinematics of the vocal tract. *Speech Communication*, 55(1), 147–161. <https://doi.org/10.1016/j.specom.2012.08.001>



at the center of the hard palate, and 180° is in the center of the pharynx.

From the entries found in the segment-to-gesture dictionary, the gestural coupling model creates a syllable-based gestural coupling graph with coupling links and activation phases (see Figure 5). Given a coupling graph, the coupled-oscillator model of intergestural planning produces a gestural score that specifies gestures' constriction goals and their relative timing. Based on a selected gestural score, the task-dynamic model calculates the tract variable and articulator trajectories. These trajectories serve as input to the Haskins Configurable Articulatory Synthesizer (CASy; Iskarous et al., 2003), a vocal tract model that models constrictions and articulator movements in a two-dimensional sagittal plane. CASy generates time-varying midsagittal vocal tract shapes, area functions, and formant frequencies.

The full vocal tract shapes are computed, as well as the 7 points on the controlled articulators that correspond to EMA pellet locations: UL, LL, JA, TT, TF (referred to as "TB" henceforth), tongue dorsum (TD), and tongue rear (TR); see Figure 7. Finally, based on the CASy parameters, Hlsyn (Hanson & Stevens, 2002) generates acoustic output that consists of the fundamental frequency and the first four formants.

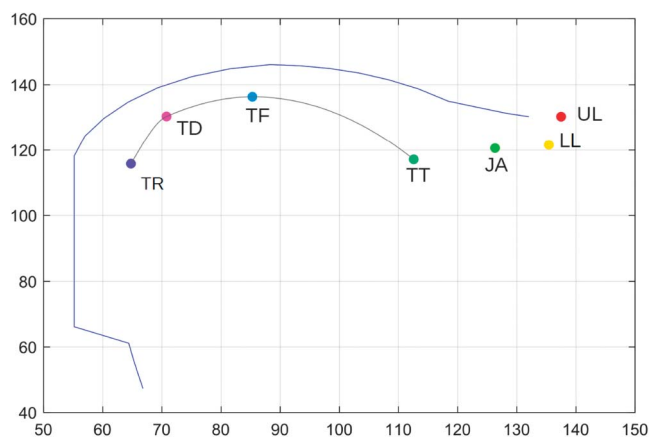
TaDA allows manually manipulating gestural parameters and thus testing specific hypotheses concerning developmental differences in speech production. Each TaDA module (e.g., model of intergestural coordination, task dynamic model) takes explicit input, and this input can be modified for each module independently of other modules. For example, one can specify different coupling relationships in the gestural coupling graph without changing the relative timing in the gestural score. In this

study, we simulated child-like productions to test the hypothesis that differences in articulatory patterns relate to differences in coarticulation patterns. To do so, we modified the default gestural specifications in the gestural coupling graphs.

The kinematic measure for our simulations was the horizontal position of the highest points on the tongue at the following time points: schwa onset (@00), schwa midpoint (@50), consonant onset (C00), consonant midpoint (C50), vowel onset (V00), and vowel midpoint (V50). We simulated the tongue position at multiple time points across the target utterances because we hypothesized that the tongue position for /n/ in the carrier word would affect the subsequent C. Because the tongue position during /n/ was not labeled in the experimental data, the tongue position at the schwa onset was taken to indicate the patterns in /n/ production. In this study, the simulated TB position was represented by the tongue front point (TF, depicted in blue in Figure 7) in TaDA output for the simulated utterances. From this point on, we will refer to this point as TB for congruency with experimental data. The time points for measuring the TB position at C and V constriction were determined in the same way as for the experimental data, that is, from the simulated acoustic speech signal.

German alveolar consonants are modeled in TaDA as a TT closure gesture coupled in-phase to a TB gesture. TT gestures are specified according to TTCL and TTCD tract variables that are associated with the following articulatory parameters, shown in Figure 6b: jaw (JA), circle location (CL), and circle angle (CA) with respect to the jaw (J), tip location (TL), and tip angle (TA) with respect to TB (B). TB gestures are specified along with TBCL and TB CD tract variables that are associated with CL, CA, and JA articulatory parameters. These specifications were devised by Pastätter and Pouplier (2014) for the modeling study of German coronals. In the American English version of TaDA, coronals are generally implemented by a TT constriction gesture. However, when modeling CV coarticulation for the German stops /t, d, n/, Pastätter and Pouplier (2014) found that in the absence of a TB target, the tongue shape posterior to the TT region was entirely dominated by the preceding or the following vowel, making the consonant acoustically unidentifiable across vowel contexts. Based on the articulatory and auditory evaluations of the simulated utterances, the authors suggested that to adequately model coarticulation between German vowels and alveolar stop consonants, gestural specifications need to include a TB target along with a TT target. To ensure the TB configuration is actively controlled and maintained, the authors introduced the TB target for the coronal stops and nasal and adjusted the blending values (alpha) such that the consonantal TB target weights more than the vowel target. Alpha values for consonant TB gestures were adjusted to 10 and one for

Figure 7. Task Dynamics Application spatial display of articulator positions for "dube" at the temporal midpoint of the consonant: upper lip (UL), lower lip (LL), jaw (JA), tongue tip (TT), tongue front (TF), tongue dorsum (TD), and tongue rear (TR).



TBCL and TBCD, respectively. For vowels, alpha was set to one for both parameters, following the specifications by Pastätter and Pouplier (2014).

Simulating Adult Productions

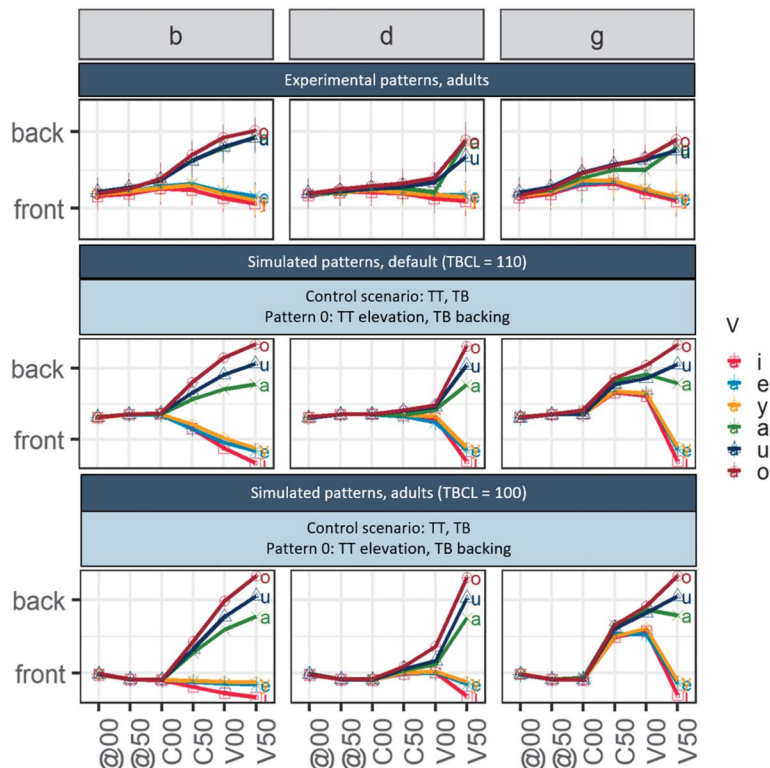
First, adult productions of the 24 utterances collected in our past studies were simulated based on the parameters set by Pastätter and Pouplier (2014) for German TaDA. Gestural scores for each utterance were then generated. For each consonant, the activation intervals (the time interval over which gestures are active) were identical across six vowel contexts. The scores differed only in the TBCL and TBCD target parameters for the different vowels. The default values from the segment-to-gesture dictionary were used for the vowels and were the same in all cases. The gestural scores were used as input to the Task Dynamic model (Saltzman & Munhall, 1989) that produced models' articulator motions (see Figure 6).

Figure 8 compares the normalized horizontal position of the highest points on the tongue across six time points as observed in experimental data (upper panel, labeled “exp”) with the normalized horizontal position of the TB across six time points as in simulated tongue contours (second panel, labeled “def”). Standard deviation

(SD) bars in the upper panel of Figure 8 show the variability in experimental data resulting from multiple repetitions by 11 adult speakers. The bottom plots do not visualize SD bars because each simulated trajectory is based on only two simulated shapes (one for each CVC). In experimental adult productions of “eine#CV” (upper panel, labeled “exp”), the horizontal position of the highest point of the TB at schwa onset (@00), schwa midpoint (@50), and consonant onset (C00) correlates with the tongue position for front vowels (color online: shown in red, mustard, and green) at V50, regardless of the vowel context and consonant. However, simulations based on the specifications from Pastätter and Pouplier (2014) for German TaDA resulted in the highest horizontal tongue position being closer to the tongue position for the central vowel from schwa onset through consonant onset for /b/ and /g/ and though vowel onset for /d/, whereas in experimental data, it was more correlated with the front vowels at these time points. We suggested that this discrepancy had to do with the specifications for alveolar consonant production. The TB position at schwa onset for /b/ and /g/ must have been affected by immediately preceding alveolar /n/.

To simulate TB trajectories that better correspond to the experimental observations, we first changed the

Figure 8. The normalized horizontal position of the highest points on the tongue body at six time points as observed in experimental data and tongue contours simulated for adults. TBCD = tongue body constriction degree.



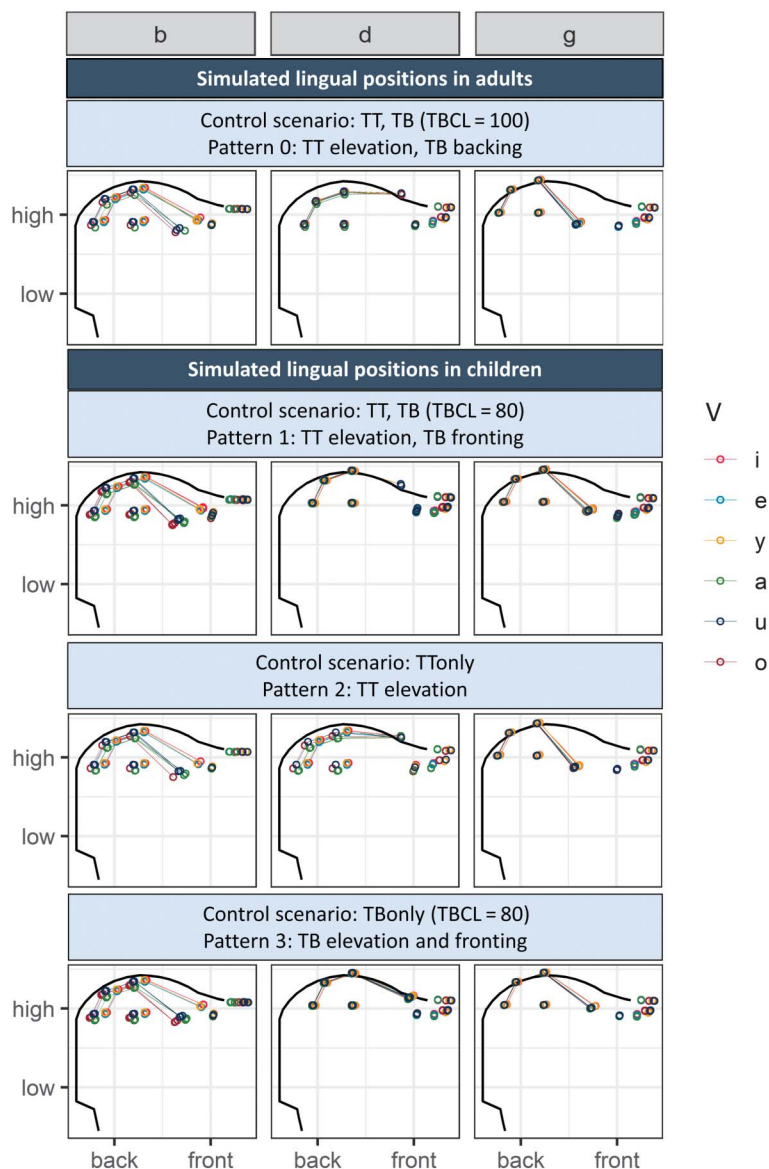
TBCL target for all alveolar gestures to a fronter position, from 110° to 100 . The schwa TBCL target was also changed to depend on the surrounding context, in this case, the preceding alveolar /n/. Results of the modifications are presented in the lowest panel of Figure 8 (labeled “sim. adults”). To test whether simulations better approximated the experimental TB trajectories based on the default or the modified gestural specifications, we compared each simulated trajectory to the average experimental trajectory using the root-mean-square deviation (RMSD), commonly used to measure the difference between observed and

estimated data. According to RMSD analysis, TB trajectory in the simulations with modifications approximates the experimental trajectories better than the simulations based on German TaDA, with RMSD of 0.14 compared to RMSD of 0.21.

Simulating Child-Like Productions

To investigate the relationship between children’s articulatory patterns and the degree of articulatory overlap observed in their speech, we first conducted a series of

Figure 9. Simulated tongue shapes at the consonant midpoint for three consonants (/b, d, g/) in six vowel contexts (/i, e, y, a, u, o/) for adults (“TT, TB” scenario with the TBCL target for alveolar consonants was set to 100), and in three scenarios for children: The scenario with the TBCL target for alveolar consonants was set to 80 (“TT, TB”), the scenario without the TB gesture (“TTonly”), and the scenario without the TT gesture (“TBonly”).



simulations to capture the tongue shapes observed in the experimental data. We found evidence of developmental differences in tongue shapes during alveolar and bilabial consonant production: in children, TB location was more fronted and elevated relative to the whole tongue than in adults. Based on previous findings on speech motor control development as well as on muscle development, we hypothesized that developmental differences in CD resulted from developmental differences in articulatory patterns for alveolar production (/d/ for dV syllables and /n/ in all other cases). This difference may be related to differences in articulatory patterns in several ways.

First, it is possible that TB position was more fronted in children during alveolar production because they move TB forward to help TT to achieve the alveolar constriction goal, but TT would still make the constriction. “Control scenario: TT, TB” in Figure 9 illustrates the simulated tongue shapes for the scenario in which the alveolar production involves both TT and TB gestures but the TB target for children is fronter than for adults. Recall that in the adult articulatory settings for German, two gestures are specified for /d/ and /n/: TT and TB. It is then possible that children have independent control over TT and TB and need to move TB forward to allow TT to make a constriction. To simulate the hypothetical tongue shape for this scenario, we changed the gestural specifications for alveolar stops such that they still contained a target for both TT and TB, same as for adults, but the target for TB is changed to a fronter and higher position compared to the adult setting to reflect the difference observed in the experimental data.

Second, children’s alveolar gestures could be specified only for the TT target (control scenario: “TTonly” in Figure 9). In this case (TT gesture only), TB would follow TT because of anatomical coupling but would not have its own goal. Still, in back vowel context, TB would not be able to go too far back for the upcoming vowel because then TT would not be able to reach the alveolar ridge. The difference between adults and children would thus be the following: During alveolar production in adults, TB has its own target independent of the TT, and is constrained so that the vowel cannot influence its spatial position strongly (see Option 1 in Figure 1). Children, however, may use subparts of the tongue in an undifferentiated manner (using TT and TB as a single anatomical and/or functional unit) and therefore not have two independent goals for the production of alveolar stops. To simulate the undifferentiated control of the tongue’s subparts in children, we changed the gestural specification for alveolar consonants such that it does not contain a separate target for TB (see Option 2 in Figure 1).

Third, TB may alone achieve the alveolar constriction goal, without independent action from TT (Control scenario: “TBonly” in Figure 9). The TT would then be

dragged forward and upward because of its mechanical coupling with the TB. To simulate the hypothetical tongue shape for this scenario, we changed the gestural specifications for alveolar stops such that they do not contain a target for TT (see Option 3 in Figure 1). The target for TB was changed to a fronter position (“80°” vs. “100°” in adults) compared to adults (bottom row, TBonly) to reflect the difference observed in the experimental data. The alpha value for the consonant was changed from 1 to 10.

In the case of the bilabial stop /b/, although its production does not involve independent control of tongue subparts, in our data set, the amount of constraint on TB during bilabial production may be affected by the nasal stop /n/ produced right before the target CV syllable in the carrier word /amə/. The only segment between /n/ and /b/ in our target nonwords (/amə1#C1V1ə2/) is /ə1/ that does not strongly affect the tongue shape. Thus, by the time of /b/ onset, the tongue shape may still retain the shape for /n/. For example, if children do not specify both TT and TB targets, but rather only TT target for alveolar /n/ production, same as for /d/, TB would be less constrained during /b/ production. That would allow the TB to reach the position for the upcoming vowel with less interference and, consequently, more coarticulation. If developmental differences in alveolar /n/ production do affect the target CV production, we assume that then this effect should also be observable in tongue positioning and shaping during schwa.

Regarding the velar stop /g/, we did not find any difference in tongue shapes across age groups. Both for children and adults, the position of the highest point of the tongue was fairly central relative to the whole tongue. However, it is still possible that the difference in CD for /g/ that we observed across age in our previous studies was related to differences in tongue shapes for /n/ production in /amə1#C1V1ə2/. Specifically, in adults, the TB position during /g/ can still correlate more with front vowels because of the /n/ production, as is the case for alveolar contexts. Instead, in children, the TB position at the consonant midpoint is more correlated with the mid vowel position and therefore quicker to adapt to back vowels, which results in the higher overlap between tongue positions for the consonant and vowel.

We tested these three scenarios by simulating utterances identical to those collected in the experimental data and manipulating the gestural specifications for alveolar consonants. We run a series of simulations for each scenario, varying the TBCL target (see Figure 6), the blending, and stiffness (the relative speed with which a gesture reaches or moves away from its target) values. The TBCL value for adults was set 100° (see Simulating Adult Productions section). To simulate a more fronted position for TB in children, we set the TBCL value to 80° which slightly more fronted than TBCL value for /i/ (88°). The

constriction degree target for TB was changed from 10 to 3.5, using the same value as for /i/. This value allowed for creating a closure with TB.

To decide which simulation best approximated the experimental data, we used the following criteria: TB in children should be more fronted to reflect the age differences in tongue positioning, and the positioning of TB in the simulated utterance at each time point of interest should correlate with the position of the TB at the corresponding time point in the observed data.

The simulated shapes for each condition are reported in Figure 9. The upper panel displays the simulated adult productions. In all scenarios for child production simulation (three lower panels), the alveolar constriction is achieved. However, it is not the case that all of the resulting articulatory patterns correspond to the experimental finding that the average TB position is more fronted in children than in adult simulations. Figure 10 shows the horizontal position of the TB for all scenarios. Two of the three tested scenarios, “TT, TB” and “TBoNly,” appear to simulate the difference between adult and child tongue shape, one without raising the TT and the other with raising the TT. In the TToNly condition, the TB position remains adultlike.

To examine the effect of these changes on CD across utterance, we analyzed the changes in the tongue’s horizontal position over time during the production of target utterances for three C1 /b, d, g/ and six V /i, e, y, a, o, u/ (see Figure 11). The experimental data (the panel named “C” for “child”) are visualized as the normalized average position of the highest point of the tongue (*SD*

shown by bars) between the acoustically defined schwa onset (@00) and vowel midpoint (V50). For the simulated utterances (the lower three panels), it is the horizontal TB position at the same time points as defined acoustically from the simulated acoustic wave. *SD* bars in the “C” panel show the variability in experimental data that results from multiple repetitions by multiple speakers. The simulated trajectories result from only two simulated shapes (one for each CV) and, thus, have almost no *SD* bars. Here, we can see that in the TToNly condition, TB starts to move toward the vowel targets earlier when compared to the experimental data. “TT, TB” and “TBoNly” conditions both provide similarly good TB trajectory approximations. Thus, the two scenarios representing the greater anatomical coupling between TB and TT best reproduce the tongue shapes at consonant constriction as well as TB trajectory over time.

We then compared each simulated trajectory to the trajectory in the experimental data using the root-mean-square deviation (RMSE) measure (see Table 6). For all consonants, the best approximation of the experimental trajectories for TB is achieved with “TBoNly” and “TT, TB” scenarios that were equally successful.

CD in Simulated Data

Finally, to determine whether the differences in CD between the simulated adultlike and child-like utterances align with the experimental findings, we analyzed the simulated data in the same manner as in our previous empirical studies. However, to be able to do statistical

Figure 10. The average horizontal position of the highest points on the tongue body at C50 in simulations for adults (“sim.adults”) and three child-like scenarios (“TT, TB”, “TToNly, and “TBoNly”).

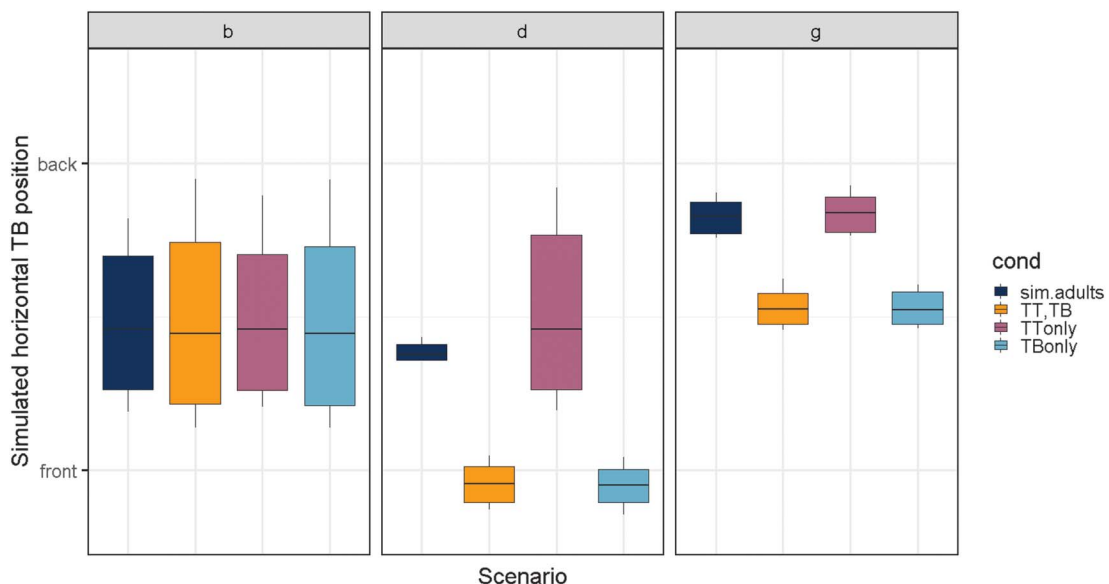


Figure 11. The normalized horizontal position of the highest points on the tongue body at six time points as observed in experimental data, and simulated data in three scenarios (“TT, TB”, “TTonly, and “TBonly”) for children.

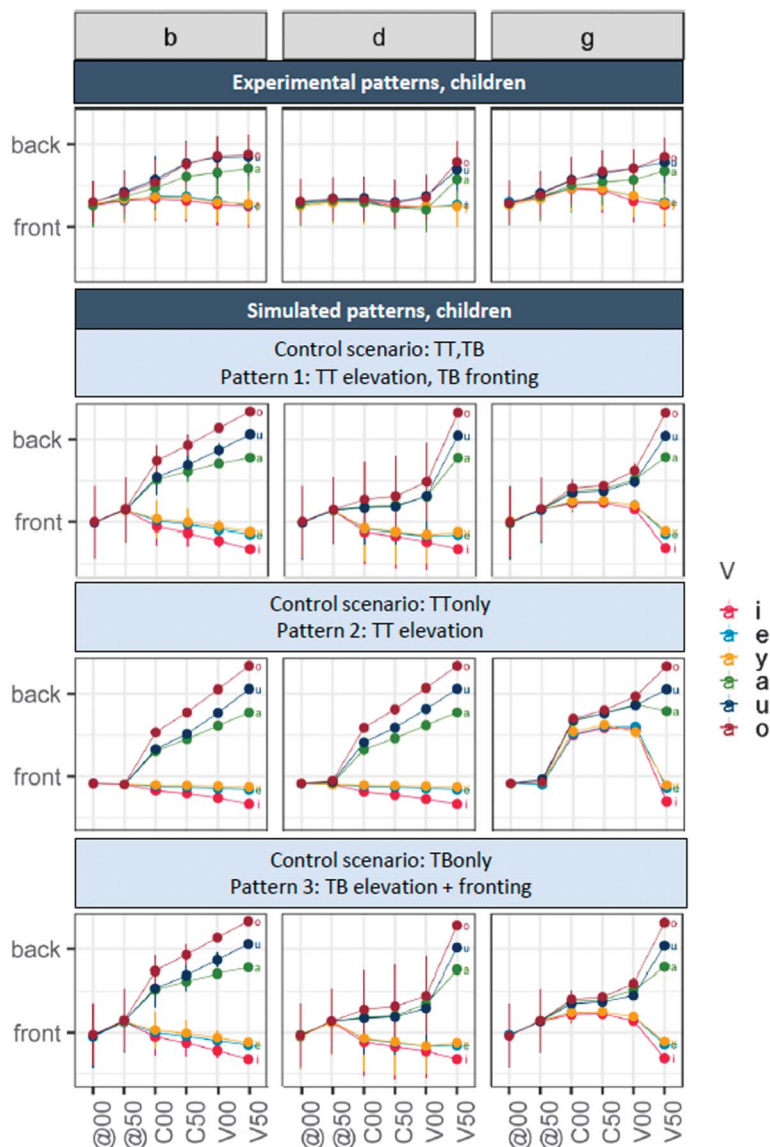


Table 6. The results of root-mean-square deviation (RMSE) analysis for three C1 and three simulated scenarios.

C1	Scenario	RMSE
/b/	TBonly	0.14
	TTonly	0.23
	TT, TB	0.14
/d/	TBonly	0.12
	TTonly	0.24
	TT, TB	0.13
/g/	TBonly	0.12
	TTonly	0.19
	TT, TB	0.12

comparisons, we needed more than two simulations of each word. For this reason, we took the final simulations from the previous sections that were identified as good approximations of child data and, based on them, we generated multiple simulations of each word. To introduce noise that would be produced by individual variability, we allowed the TB target for the consonant to covary with the target for the following vowel while keeping the allowed range of target variability the same for both adult and child simulations. CD is defined as the strength of the relationship between the tongue position at the consonant constriction and the tongue position at vowel constriction. In our previous studies, the tongue position was quantified

Table 7. The output of the linear regression predicting the horizontal positions of tongue body (TB) articulator at the consonant midpoint in simulated data based on the interaction between the horizontal positions of TB at vowel midpoint (TBx), condition (sim.adults – sim.child), and consonant context (C1).

Contrast	C1	Estimate	SE	df	T ratio	p value
TBxV–TBxC: sim.adults–sim.child	/b/	–0.09	0.07	92	–1.22	.2253
TBxV–TBxC: sim.adults–sim.child	/d/	–3.97	0.08	92	–52.88	< .0001
TBxV–TBxC: sim.adults–sim.child	/g/	–2.87	0.07	92	–38.37	< .0001

as the horizontal position of the highest point on the TB that was taken to represent the place of constriction. For the simulated utterances, the TB position at consonant and vowel constrictions was represented by the horizontal position of the TB point on the simulated tongue contour. The frames corresponding to the constrictions were determined from the simulated acoustic wave. Because the two simulated conditions, “TT, TB” and “TOnly,” were equally good in approximating the experimental productions, we pooled the output of both simulated scenarios under “sim.child.”

The difference in CD between the two conditions (sim.adults and sim.child) was compared by fitting linear regressions models in R. The response variable was the horizontal position of TB at the consonant midpoint (TBxC). The dependent variable was the interaction between the horizontal position of TB at vowel midpoint (TBxV), condition (factor with levels “sim.adults” and “sim.child”), and C1 (factor with levels b, d, and g). The marginal means were estimated using the “emmeans” function in the emmeans package, R (Lenth et al., 2018), and the “pairs” function in the same package was used to perform Tukey HSD post hoc tests of the linear relationship between TBxC and TBxV among conditions for each consonant.

Results

The resulting comparisons are presented in Table 7. The results show that the TBx position at vowel midpoint predicted the TBx position at consonant midpoint better in “sim.child” condition when compared to “sim.adults” condition for /d/ and /g/ ($p < .001$). There was no significant difference between conditions in /b/ context ($p = .23$).

Discussion

Developmental differences in CD have long been viewed as a window into the maturation of speech motor organization and its relation to the higher order linguistic organization. However, CD does not only characterize the temporal overlap of articulatory gestures, but also the spatial organization of coproduced speech segments (Fowler & Saltzman, 1993). In this study, we asked whether developmental differences in the articulatory organization of lingual

gestures could explain—at least in part—the developmental decrease in CD observed in various studies (e.g., acoustic: Gibson & Ohde, 2007; Nijland et al., 2002; Nittrouer et al., 1989; articulatory: Noiray et al., 2013, 2018).

To address this question, we analyzed a data set of lingual data collected in children from 3- to 7 years of age and adults and a set of comparable simulated data generated with TaDA application. First, the analyses of the empirical data set revealed differences in articulatory patterns between children and adults for alveolar consonant production. More specifically, we found a more fronted and elevated tongue body position associated with alveolar constriction in children compared to adults. This was an important first step. Indeed, while previous studies have argued for a protracted development of lingual motor control (Cheng et al., 2007; Murdoch et al., 2012), there was no direct evidence for differentiated control of the tongue in young children compared to adults. Naturally, there was also no evidence that such differences contribute to differences in CD. To address these issues, in the next step, we simulated child-like productions to test whether their articulatory patterns resulted from differentiated or undifferentiated control of the tongue (see Figure 1). To this end, we tested three hypothetical scenarios and found that child productions of alveolar consonants were best modeled by gestural specifications simulating greater anatomical coupling between TB and TT than in adults. Last, we found that these child-like simulated utterances resulted in patterns of CDs comparable to those in the experimental data set. Specifically, they elicited higher intersegmental CDs than adultlike simulated utterances in the context of the alveolar stop /d/ and velar stop /g/. However, CDs did not differ across groups in the /b/ context, unlike in the experimental data.

The independent contribution of differences in articulatory patterns into CD differences was assessed here with a model that allows for controlled manipulation of parameters. In essence, TaDA is shown here to be useful for testing hypotheses about speech development despite being originally designed to model adult speech. Another example of speech acquisition research using TaDA is provided by Nam et al. (2013) who successfully used simulations with TaDA to demonstrate that CV combination preferences in early babbling have a physiological basis and further compared two physiological accounts of the phenomenon.

The strength of the TaDA model resides in the possibility of changing specific parameters while controlling all the others. It can also help discard unlikely hypotheses and hence guide study design. While only the spatial interaction of temporally overlapping utterances was considered here, there are other potential reasons for the developmental differences in CD, for example, vocal tract maturation, the development of gestural timing (Tilsen, 2018), of children's expressive lexicon, and phonological representations (Nittrouer et al., 1989; Noiray, Popescu, et al., 2019; Popescu & Noiray, 2022). Modeling applications such as TaDA can help dissect the relative role of these factors with respect to the experimentally observed developmental changes in coarticulation patterns.

Implications for Theories of Speech Motor Development

The complementarity of empirical and simulated data allowed us to uncover children's articulatory patterns for alveolar production. The pattern experimentally observed in children as compared to adults—TB moving farther forward and higher relative to the whole tongue—was successfully simulated by two scenarios. Although one of the scenarios modeled the differentiated control of tongues functional subparts, and the other only modeled the control of TB with TT tagging along, both scenarios imply *greater coupling between TB and TT* in children as compared to those of adults: Even in the differentiated control scenario, TB moves forward to help TT to achieve the alveolar constriction goal.

The explanation for this finding of greater coupling between TB and TT for alveolar production in children remains to be established. On the one hand, children certainly need some time to learn to control such complicated systems as the tongue. On the other hand, the child tongue anatomy differs in many ways from adults. An important difference lies in the immaturity of children's superior longitudinal muscle that allows the tongue to bend and produce precise contractions and thus operates to separate movement from the TT and the other parts of the tongue (Denny & McGowan, 2012). The finding of flatter tongue shapes in children compared to adults exemplifies this lack of functional differentiation.

However, children may also simply not need to acquire differentiated control of their tongue because their vocal tract differs from adults. For instance, there is evidence that articulatory patterns correlate with hard palate shape in adult speech production (Brunner et al., 2009). Assuming children have a smaller and flatter hard palate than adults (Ciusa et al., 2007), producing intelligible alveolars without employing the TT may be optimal for children. In this case, later in development, when their hard

palate shape has become more domed or the relative size of the tongue and oral cavity have changed (Bosma, 1963; Bosma & Showacre, 1975), they may need to develop a differentiated control of the TT and TB, which they do. This explanation is in line with the motor equivalence phenomenon that was described by Kelso and Tuller (1983) as follows: "Within limits, people (and animals) can achieve the same 'goal' through a variety of kinematic trajectories, with different muscle groups and in the face of ever-changing postural and biomechanical requirements" (p. 217). In this perspective, the developmental changes in vocal tract would be an instant of ever-changing postural and biomechanical requirements.

It is important to note that even though tongue shape at /b/ midpoint differed across age, simulations in TaDA did not show differences in CD. This finding may imply that age differences in CD for /b/ are caused by factors other than articulatory patterns. However, this missing difference could also be caused by the implementation of the effects of long distance, across word coarticulation in TaDA.

Although we do not observe differences between 3-, 4-, and 5-year-olds in terms of their coarticulatory patterns, the causes behind their behavior may change. For example, physical growth may be more important for younger children, whereas cognitive factors may start to play a bigger role for older children. This would go along with the principle of dynamical development:

Likewise, developmental change is not planned but arises within a context as the product of multiple, developing elements. . . . Each component has its own trajectory of change. Some elements may be fully formed early in life but unseen because the supporting subsystems and processes are not ready. Other components may be comparatively delayed, and indeed one element may act as a "rate limiter," preventing the cooperative self-organization of the other component. Only when all the components reach critical functioning and the context is appropriate does the system assemble a behavior. (Thelen, 1995, p. 82)

Ultimately, understanding child-like articulatory patterns may help determine the nature of representations that underlie them. For example, it may help answer the question of whether children attempt to achieve auditory goals using the articulatory resources available, even if they do not match those of adults or whether they attempt to achieve adultlike articulatory gestures but with limited skill.

Implications for Clinical Practice

Our findings are also relevant for the research on SSDs of unknown origin. As the ambiguous term suggests,

children with SSDs of unknown origin are a heterogeneous group with various speech issues. To find a proper treatment for every child, the underlying mechanism needs to be established, to distinguish between cognitive/linguistic constraints and constraints at more peripheral levels of speech motor production and control (Gibbon, 1999). Previous research using EPG has shown that lingual consonants in the speech of school-age children with SSD of unknown origin have a distinctive articulatory characteristic: a high amount of tongue–palate contact, or UG (Gibbon, 1999). UGs are “interpreted as reflecting a speech motor constraint involving either delayed or deviant control of [the] functionally independent regions of the tongue” (Gibbon, 1999, p. 382). UGs have been studied only in a few school-aged children with particularly treatment-resistant SSDs, but Gibbon (1998) speculated that young typically developing children are likely to produce lingual consonants with increased tongue–palate contacts and suggested that investigation of articulatory differentiation in young typically developing children will assist in establishing whether undifferentiated gestures reflect delayed or deviant speech motor development (Gibbon, 1999). In our study, we provide evidence that typically developing German children indeed use less differentiated patterns for alveolar consonant production. However, while Gibbon (1999) found an effect of age for all consonant contexts, the children in our study showed an established alveolar/velar contrast already at the age of 3 years, and there was no age-related difference for /g/.

The relationship between coarticulation and motor control development has implications for the clinical practice, in particular, diagnostic of motor/neurological disorders such as childhood apraxia of speech (CAS). Children with CAS are not always easy to differentiate from children with phonological disorders. In recent years, considerable progress has been made in identifying diagnostic criteria. Lengthened coarticulation and disrupted coarticulatory transitions between sounds and syllables have gained acceptance as one of three key characteristics of the disorder (American Speech-Language-Hearing Association, 2007; Terband et al., 2019). However, children with CAS have been found to show both stronger (Nijland et al., 2002) and weaker (Nijland et al., 2002, 2003) coarticulation. A better understanding of the phenomena underlying coarticulation differences could improve the understanding of the nature of the disorder and enable more targeted diagnostic tools. TaDA can help to identify the underlying causes of different coarticulatory phenomena.

Conclusions

Developmental differences in coarticulation patterns are not well understood, partly because coarticulation is a

multifaceted phenomenon with various potential sources. This study provides evidence that age-related differences in inter-articulator CD are related to developmental differences in articulatory patterns for consonant production. This finding implies that a full account of age-related changes in coarticulatory patterns should include the role of the articulatory organization of lingual gestures. Further investigations are needed to uncover the relative contribution of articulatory synergies into age-related differences in coarticulatory patterns compared to factors, such as the development of coordinative control and vocal tract anatomy.

Author Contributions

Dzhuma Abakarova: Conceptualization (Lead), Formal analysis (Lead), Investigation (Lead), Methodology (Lead), Software (Lead), Visualization (Lead), Writing – original draft (Lead), Writing – review & editing (Equal). **Susanne Fuchs:** Funding acquisition (Lead), Project administration (Lead), Supervision (Lead), Writing – review & editing (Lead). **Aude Noiray:** Funding acquisition (Lead), Project administration (Lead), Resources (Lead), Supervision (Lead), Writing – review & editing (Lead).

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft GZ: NO 1098/2-1 and 255676067, the Leibniz Association, by the FIAS program and Collegium of Lyon, and by German Academic Exchange Service Short-Term Grants for Doctoral Students 2019/20 (57438025). We are grateful to the LOLA Lab team at the University of Potsdam for their assistance in collecting and processing the data. We would also like to thank Manfred Pastätter and Marianne Pouplier for providing us with their German version of Task Dynamics Application and Lucie Ménard for sharing the code for curvature position and degree calculations. We are grateful to Khalil Iskarous and Louis Goldstein for the helpful discussions. Finally, we would like to thank all participants for making this research possible.

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. <https://doi.org/10.1121/1.5047669>
- American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech*. <https://www.asha.org/policy/tr2007-00278/>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). *Parsimonious mixed models*. <http://arxiv.org/abs/1506.04967>

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Boë, L.-J., Sawallis, T. R., Fagot, J., Badin, P., Barbier, G., Captier, G., Ménard, L., Heim, J.-L., & Schwartz, J.-L. (2019). Which way to the dawn of speech?: Reanalyzing half a century of debates and data in light of speech science. *Science Advances*, 5(12), eaaw3916. <https://doi.org/10.1126/sciadv.aaw3916>
- Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (Version 6.1.40) [Computer software]. <http://www.praat.org/>
- Bosma, J. (1963). Oral and pharyngeal development and function. *Journal of Dental Research*, 42(1), 375–380. <https://doi.org/10.1177/00220345630420014301>
- Bosma, J. F., & Showacre, J. (1975). *Symposium on development of upper respiratory anatomy and function: Implications for sudden infant death syndrome*. U.S. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, John E. Fogarty International Center for Advanced Study in the Health Sciences.
- Bretz, F., Hothorn, T., & Westfall, P. (2008). Multiple comparison procedures in linear models. In P. Brito (Ed.), *COMPSTAT 2008* (pp. 423–431). Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-2084-3_35
- Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2), 201–251. <https://doi.org/10.1017/S0952675700001019>
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3–4), 155–180. <https://doi.org/10.1159/000261913>
- Brunner, J., Fuchs, S., & Perrier, P. (2009). On the relationship between palate shape and articulatory behavior. *The Journal of the Acoustical Society of America*, 125(6), 3936–3949. <https://doi.org/10.1121/1.3125313>
- Cheng, H. Y., Murdoch, B. E., Goozée, J. V., & Scott, D. (2007). Physiologic development of tongue–jaw coordination from childhood to adulthood. *Journal of Speech, Language, and Hearing Research*, 50(2), 352–360. [https://doi.org/10.1044/1092-4388\(2007\)025](https://doi.org/10.1044/1092-4388(2007)025)
- Ciusa, V., Dimaggio, F. R., Sforza, C., & Ferrario, V. F. (2007). Three-dimensional palatal development between 3 and 6 years. *The Angle Orthodontist*, 77(4), 602–606. <https://doi.org/10.2319/053106-221>
- Cleland, J., & Scobbie, J. M. (2021). The dorsal differentiation of velar from alveolar stops in typically developing children and children with persistent velar fronting. *Journal of Speech, Language, and Hearing Research*, 64(6S), 2347–2362. https://doi.org/10.1044/2020_JSLHR-20-00373
- Cychoz, M., Munson, B., & Edwards, J. R. (2021). Practice and experience predict coarticulation in child speech. *Language Learning and Development*, 17(4), 366–396. <https://doi.org/10.1080/15475441.2021.1890080>
- Denny, M., & McGowan, R. S. (2012). Implications of peripheral muscular and anatomical development for the acquisition of lingual control for speech production: A review. *Folia Phoniatrica et Logopaedica*, 64(3), 105–115. <https://doi.org/10.1159/000338611>
- D'Souza, H., Cowie, D., Karmiloff-Smith, A., & Bremner, A. J. (2017). Specialization of the motor system in infancy: From broad tuning to selectively specialized purposeful actions. *Developmental Science*, 20(4), e12409. <https://doi.org/10.1111/desc.12409>
- Dawson, K. M., Tiede, M. K., & Whalen, D. H. (2016). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical Linguistics & Phonetics*, 30(3–5), 328–344. <https://doi.org/10.3109/02699206.2015.1099164>
- Fowler, C., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36(2–3), 171–195. <https://doi.org/10.1177/002383099303600304>
- Fricke, M., & Johnson, K. (2012). Measuring coarticulation in spontaneous speech: A preliminary report. *UC Berkeley Phon-Lab Annual Report*, 8(8). <https://doi.org/10.5070/P79SM8J6DM>
- Gibbon, F. E. (1998). *Lingual articulation in children with developmental speech disorders* [Doctoral dissertation, University of Luton]. University of Bedfordshire Repository. <https://uobrep.openrepository.com/handle/10547/581284>
- Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, 42(2), 382–397. <https://doi.org/10.1044/jslhr.4202.382>
- 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. [https://doi.org/10.1044/1092-4388\(2007\)008](https://doi.org/10.1044/1092-4388(2007)008)
- Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In C. G. M. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 239–249). The Commercial Press. <https://hal.archives-ouvertes.fr/hal-03127293>
- Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43(1), 239–255. <https://doi.org/10.1044/jslhr.4301.239>
- 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. [https://doi.org/10.1044/1092-4388\(2002\)005](https://doi.org/10.1044/1092-4388(2002)005)
- Hanson, H. M., & Stevens, K. N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLSyn. *The Journal of the Acoustical Society of America*, 112(3), 1158–1182. <https://doi.org/10.1121/1.1498851>
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 26(3), 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Iskarous, K., Goldstein, L., Whalen, D., Tiede, M., & Rubin, P. (2003). *CASY: The Haskins Configurable Articulatory Synthesizer*. <https://www.semanticscholar.org/paper/CASY%3A-The-Haskins-Configurable-Articulatory-Iskarous-Goldstein/b698c5024137f41e3852c787f588691ab5cc5a12>
- Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C., Saltzman, E., & Whalen, D. (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. <https://doi.org/10.1121/1.4812855>
- Kabakoff, H., Harel, D., Tiede, M., Whalen, D. H., & McAllister, T. (2021). Extending ultrasound tongue shape complexity measures to speech development and disorders. *Journal of Speech, Language, and Hearing Research*, 64(7), 2557–2574. https://doi.org/10.1044/2021_JSLHR-20-00537
- Kelso, J. S., & Tuller, B. (1983). “Compensatory articulation” under conditions of reduced afferent information: A dynamic formulation. *Journal of Speech and Hearing Research*, 26(2), 217–224. <https://doi.org/10.1044/jshr.2602.217>
- Latash, M. L. (2012). The bliss of motor abundance. *Experimental Brain Research*, 217(1), 1–5. <https://doi.org/10.1007/s00221-012-3000-4>

- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version, 1*(1), 3.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America, 35*(11), 1773–1781. <https://doi.org/10.1121/1.1918816>
- Lindblom, B. (1983). Economy of speech gestures. *The Production of Speech, 217–245*. https://doi.org/10.1007/978-1-4613-8202-7_10
- 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. https://doi.org/10.1044/2017_AJSLP-16-0110
- Ménard, L., Aubin, J., Thibeault, M., & Richard, G. (2012). Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model. *Folia Phoniatrica et Logopaedica, 64*(2), 64–72. <https://doi.org/10.1159/000331997>
- Ménard, L., Davis, B. L., Boë, L.-J., & Roy, J.-P. (2009). Producing American English vowels during vocal tract growth: A perceptual categorization study of synthesized vowels. *Journal of Speech, Language, and Hearing Research, 52*(5), 1268–1285. [https://doi.org/10.1044/1092-4388\(2009/08-0008\)](https://doi.org/10.1044/1092-4388(2009/08-0008))
- Ménard, L., Schwartz, J., & Boë, L. (2004). Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research, 47*(5), 1059–1080. [https://doi.org/10.1044/1092-4388\(2004/079\)](https://doi.org/10.1044/1092-4388(2004/079))
- Ménard, L., Schwartz, J.-L., Boë, L.-J., & Aubin, J. (2007). Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics, 35*(1), 1–19. <https://doi.org/10.1016/j.wocn.2006.01.003>
- Mermelstein, P. (1973). Articulatory model for the study of speech production. *The Journal of the Acoustical Society of America, 53*(4), 1070–1082. <https://doi.org/10.1121/1.1913427>
- Murdoch, B. E., Cheng, H.-Y., & Goozée, J. V. (2012). Developmental changes in the variability of tongue and lip movements during speech from childhood to adulthood: An EMA study. *Clinical Linguistics & Phonetics, 26*(3), 216–231. <https://doi.org/10.3109/02699206.2011.604459>
- Nam, H., Goldstein, L. M., Giulivi, S., Levitt, A. G., & Whalen, D. H. (2013). Computational simulation of CV combination preferences in babbling. *Journal of Phonetics, 41*(2), 63–77. <https://doi.org/10.1016/j.wocn.2012.11.002>
- Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2004). TADA: An enhanced, portable Task Dynamics model in MATLAB. *The Journal of the Acoustical Society of America, 115*(5), 2430. <https://doi.org/10.1121/1.4781490>
- Nijland, L., Maassen, B., & van der Meulen, S. (2003). Evidence of motor programming deficits in children diagnosed with DAS. *Journal of Speech, Language, and Hearing Research, 46*(2), 437–450. [https://doi.org/10.1044/1092-4388\(2003/036\)](https://doi.org/10.1044/1092-4388(2003/036))
- Nijland, L., Maassen, B., van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics, 16*(6), 461–483. <https://doi.org/10.1080/02699200210159103>
- 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. <https://doi.org/10.1044/jshr.3605.959>
- Nittrouer, S., Studdert-Kennedy, M., & McGowan Richard, 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. <https://doi.org/10.1044/jshr.3201.120>
- Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research, 61*(6), 1355–1368. https://doi.org/10.1044/2018_JSLHR-S-17-0148
- 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. <https://doi.org/10.1121/1.4763983>
- Noiray, A., Popescu, A., Killmer, H., Rubertus, E., Krüger, S., & Hintermeier, L. (2019). Spoken language development and the challenge of skill integration. *Frontiers in Psychology, 10*. <https://doi.org/10.3389/fpsyg.2019.02777>
- Noiray, A., Ries, J., Tiede, M., Rubertus, E., Laporte, C., & Ménard, L. (2020). Recording and analyzing kinematic data in children and adults with SOLLAR: Sonographic & Optical Linguo-Labial Articulation Recording system. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11*(1), 14. <https://doi.org/10.5334/labphon.241>
- Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (2019). Back from the future: Nonlinear anticipation in adults' and children's speech. *Journal of Speech, Language, and Hearing Research, 62*(8S), 3033–3054. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0208
- Pastätter, M., & Pouplier, M. (2014). *The articulatory modelling of German coronal consonants using TADA*. Proceedings of the 10th International Seminar on Speech Production.
- Perrier, P., & Fuchs, S. (2015). Motor equivalence in speech production. In *The handbook of speech production* (pp. 223–247). Wiley. <https://doi.org/10.1002/9781118584156.ch11>
- Popescu, A., & Noiray, A. (2022). Learning to read interacts with children's spoken language fluency. *Language Learning and Development, 18*(2), 151–170. <https://doi.org/10.1080/15475441.2021.1941032>
- Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics, 15*(4), 299–312. [https://doi.org/10.1016/S0095-4470\(19\)30580-7](https://doi.org/10.1016/S0095-4470(19)30580-7)
- 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. <https://doi.org/10.1121/1.3089222>
- 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), Article e0203562. <https://doi.org/10.1371/journal.pone.0203562>
- Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333–382. https://doi.org/10.1207/s15326969eco0104_2
- Sanguineti, V., Laboissière, R., & Payan, Y. (1997). A control model of human tongue movements in speech. *Biological Cybernetics, 77*(1), 11–22. <https://doi.org/10.1007/s004220050362>
- Serrurier, A., Badin, P., & Neuschaefer-Rube, C. (2019). Influence of the vocal tract morphology on the F1–F2 acoustic plane. *Studentexte Zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung, 312–319*.
- Shim, J. K., Karol, S., Hsu, J., & de Oliveira, M. A. (2008). Hand digit control in children: Motor overflow in multi-finger pressing force vector space during maximum voluntary force production. *Experimental Brain Research, 186*(3), 443–456. <https://doi.org/10.1007/s00221-007-1246-z>

- Smith, A., & Zelaznik, H.** (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, *45*(1), 22–33. <https://doi.org/10.1002/dev.20009>
- Sweeney, K. L., Ryan, M., Schneider, H., Ferenc, L., Denckla, M. B., & Mahone, E. M.** (2018). Developmental trajectory of motor deficits in preschool children with ADHD. *Developmental Neuropsychology*, *43*(5), 419–429. <https://doi.org/10.1080/87565641.2018.1466888>
- Terband, H., Maassen, B., & Maas, E.** (2019). A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatrica et Logopaedica*, *71*(5–6), 216–227. <https://doi.org/10.1159/000499426>
- Thelen, E.** (1995). Motor development: A new synthesis. *American Psychologist*, *50*(2), 79–95. <https://doi.org/10.1037/0003-066X.50.2.79>
- The MathWorks, Inc.** (2020). *MATLAB*. (R2020b).
- Tilsen, S.** (2018). *Three mechanisms for modeling articulation: Selection, coordination, and intention*. Cornell University.
- 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. [https://doi.org/10.1044/1092-4388\(2007\)104](https://doi.org/10.1044/1092-4388(2007)104)
- Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-Lehouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S.** (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, *48*(3), 543–553. [https://doi.org/10.1044/1092-4388\(2005\)037](https://doi.org/10.1044/1092-4388(2005)037)
- Zharkova, N., Gibbon, F. E., & Hardcastle, W. J.** (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical Linguistics & Phonetics*, *29*(4), 249–265. <https://doi.org/10.3109/02699206.2015.1007528>
- Zharkova, N., Hewlett, N., Hardcastle, W. J., & Lickley, R. J.** (2014). Spatial and temporal lingual coarticulation and motor control in preadolescents. *Journal of Speech, Language, and Hearing Research*, *57*(2), 374–388. https://doi.org/10.1044/2014_JSLHR-S-11-0350