The acquisition of articulatory timing for liquids: evidence from child and adult speech

Phil J. Howson & Melissa A. Redford

Corresponding Author:

Phil Howson <philjhowson@uvic.ca>

Keywords: liquids, coarticulation, articulatory timing, acquisition, speech development

Funding Statement. This work was funded by NIH grant R01 HD087452 awarded to the second author, Melissa A. Redford.
Abstract

**Purpose:** Liquids are among the last sounds to be acquired by English-speaking children. The current study considers their acquisition from an articulatory timing perspective by investigating anticipatory posturing for /l/ versus /ɹ/ in child and adult speech.

**Method:** In Experiment 1, twelve 5-year-old, twelve 8-year-old, and 11 college-aged speakers produced carrier phrases with penultimate stress on monosyllabic words that had /l/, /ɹ/, or /d/ (control) as singleton onsets and /æ/ or /u/ as the vowel. Short-domain anticipatory effects were acoustically investigated based on schwa formant values extracted from the preceding determiner (= the) and dynamic formant values across the /ə#LV/ sequence. In Experiment 2, long-domain effects were perceptually indexed using a previously validated forward-gated AV speech prediction task.

**Results:** Experiment 1 results indicated that all speakers distinguished /l/ from /ɹ/ along F3. Adults distinguished /l/ from /ɹ/ with a lower F2. Older children produced subtler versions of the adult pattern; their anticipatory posturing was also more influenced by the following vowel. Younger children did not distinguish /l/ from /ɹ/ along F2, but both liquids were distinguished from /d/ in the domains investigated. Experiment 2 results indicated that /ɹ/ was identified earlier than /l/ in gated adult speech; both liquids were identified equally early in 5-year-olds’ speech.

**Conclusions:** The results are interpreted to suggest a pattern of early tongue body retraction for liquids in /ə#LV/ sequences in children’s speech. More generally, it is suggested that children must learn to inhibit the influence of vowels on liquid articulation to achieve an adult-like contrast between /l/ and /ɹ/ in running speech.
Introduction

The coarticulatory patterns of children’s speech are compatible with the hypothesis of whole word speech production (Nittrouer, Studdert-Kennedy, & McGowan, 1989; Redford, 2019). The hypothesis begs the question of how children acquire adult-like speech, which is characterized by the sequential production of distinct speech sounds. Whether the answer is gestural differentiation and tuning, as in Articulatory Phonology (Nittrouer et al., 1989; Studdert-Kennedy & Goldstein, 2003; Noiray et al., 2018), or the emergence of perceptual-motor units at trajectory crossings in the perceptual-motor map (Davis & Redford, 2019), the whole word production hypothesis implies that speech sound acquisition depends on mastering articulatory timing control. Although developmental studies of articulatory timing rarely address speech sound acquisition, articulatory dynamics and articulatory postures are inextricably linked through time and so information about one provides knowledge and context relevant to understanding the other. In the current study, we approached the acquisition of liquids from this perspective by investigating formant dynamics and visible kinematics in child and adult speech. Numerous studies with adults have shown that rhotics and laterals have distinctive articulatory timing patterns (e.g. Narayanan, Byrd, & Kaun, 1999; Proctor, 2011; Marin & Pouplier, 2014) and especially long-distance coarticulatory effects on adjacent sounds in comparison to other segments (West, 1999; Kochetov & Neufeld, 2013). Yet no studies we know of examine the acquisition of liquids from the perspective of articulatory timing. Our goal in doing so is to deepen the fundamental understanding of why these sounds are among the last to be acquired by English-speaking children (Prather, Hedrick, & Kern, 1975; Smit et al., 1990). Prior explanations have focused on the difficulty inherent to achieving the simultaneous anterior and posterior lingual constrictions needed to produce liquids (for relevant reviews, see McGowan, Nittrouer &
Manning, 2002; Denny & McGown, 2012). Other studies have linked delayed liquid acquisition to a general delay in the ability to form multiple simultaneous constrictions with anatomically connected articulators (Green, Moore, & Reilly, 2000; Studdert-Kennedy & Goldstein, 2003). Here, we consider this explanation in the context of anticipatory posturing for a liquid–vowel sequence in child and adult speech using a comparative approach. The specific study focus is on the acquisition of an /l/–/ɹ/ contrast in word-onset position by younger and older school-age children who are acquiring a west coast variety of American English in the typical manner.

**Child Development of Articulatory Timing**

In our view, the hypothesis of whole word production is a hypothesis about the speech plan: the coarticulatory patterns of children’s speech suggest holistic representations of articulatory (and perceptual) chunks, built up from communicative successes (Redford, 2015; 2019). More concretely, we observe that targeted speech sounds are more overlapped with adjacent targeted speech sounds in children’s speech production compared to adults’ production (Zharkova, Hewlett, & Hardcastle, 2011; 2012; Noiray et al., 2018; Rubertus & Noiray, 2018). Consider, for example, a recent ultrasound study by Rubertus and Noiray (2018) who elicited əCV sequences that spanned a lexical boundary (e.g., /aɪ ŋiːdə/ “a beeda”) within a syntactic constituent (i.e., a noun phrase) from 4 groups of German speaking children (3-, 4-, 5-, and 7-year-olds). They also elicited speech from college-aged adults. Their goal was to investigate the acquisition of anticipatory vowel effects on the production of schwa as a function of the coarticulatory “resistance” of the intervening consonant, which was the onset to a trochaically-stressed nonword. Following the Degree of Articulatory Constraints (DAC) model (Recasens et al., 1997), Rubertus and Noiray (2018) predicted that V-to-V effects would be weaker in əCV sequences where consonant articulation involved the tongue blade and body (/d/) and stronger
when the tongue body was free from the biomechanical constraint of tongue blade involvement (/b/ and /g/). They also expected that V-to-V effects would be stronger in younger children’s speech than in older children’s speech. The analyses focused on the tongue contour for [ə] as a function of the following consonant, stressed vowel, and speakers’ age. The results indicated anticipatory vowel effects on schwa production in speech produce by each of the 4 age cohorts of children and the adults. The effects were greatest when the intervocalic consonant was /g/, but only in adult speech. There was no effect of intervocalic consonant on the degree of anticipatory vowel articulation in children’s speech, though the pattern of results from 7-year-olds was at least in the same direction as the pattern from adults. Relatedly, V-to-V effects were strongest in the youngest children’s speech and declined with age.

Rubertus and Noiray (2018) interpreted their study results within the language of Articulatory Phonology to indicate a developmental trajectory from extensive “coproduction” (= gestural overlap) of sequential vowels in child speech to an adult-like pattern of consonant-mediated V-to-V coproduction. This explanation, which assumes a holistic speech plan (i.e., a gestural score), can be elaborated further with reference to the development of lingual motor control. As Noiray et al. (2018) observe in their ultrasound study of consonant-vowel coarticulation, the synergies between tongue body and blade movements required for the precise articulation of coronal consonants develops later than the synergies used to effect labial and dorsal articulations. The evidence from electropalatographic studies also suggests that functionally independent control over tongue blade/tip movements develops late relative to motor control over tongue body movement (Gibbon 1999; Cheng et al., 2007). For example, Cheng et al. (2007) reported that younger school-aged children produce coronal consonants with greater tongue-to-palate contact than older school-aged children and adults. They also reported a
maturational pattern of increasingly forward tongue placement during coronal articulation. These results fit with findings from acoustic studies, which show that consonantal articulations, especially those that involve anterior lingual constrictions, are less precise in children’s speech compared to adults’ speech (Nittrouer, 1995; Nissen & Fox, 2005; Zharkova et al., 2011; 2012; Holliday et al., 2015).

To summarize, children’s speech is more coarticulated than adults’ speech, and vowel-to-vowel effects are particularly strong. These patterns are consistent with holistic speech plans, where lingual trajectories are optimized for sequential vowel production with planning for consonantal articulations constrained by the protracted development of independent tongue tip and blade control relative to tongue body (and jaw) movement control. This developmental pattern of “tongue-body-postures-first” should have implications for the acquisition of timing control over the complex articulation of liquids.

The Complex Articulation of Liquids

Liquids are a large and diverse group of sounds that include both rhotics and laterals. They are articulatorily complex in that they require the coordination of a tongue tip and tongue body constriction (Sproat & Fujimura, 1993; Gick, 1999; Gick, Kang, & Whalen, 2002; Proctor, 2011). The tongue tip forms an anterior constriction and a posterior one is achieved with the tongue body. There are also timing differences between the anterior and posterior constriction formation. Rhotics often have an additional labial constriction (Alwan, Narayanan, & Haker, 1997) and may also include a tongue root constriction (Howson, 2018).

One well-known dimension along which liquids vary is in the degree of posterior tongue retraction: less retraction results in a clear (or light) articulation; greater retraction results in a dark (= pharyngealized) articulation. The clear-dark contrast is often the result of timing
differences between liquids and positional effects, although the clear-dark distinction may be
phonemic and constitute a primary point of contrast between the rhotic and lateral within a
language. For example, Carter and Local (2007) found that the Newcastle English in word-initial
positions the lateral had a higher F2 than the rhotic. The higher F2 is strongly correlated to more
tongue advancement, while the lower F2 correlates to a more retracted tongue. Thus, the results
indicated a clear lateral and dark rhotic in New Castle English. Leeds English showed the
opposite pattern, indicating a clear rhotic (= high F2) and a dark lateral (= low F2). The
difference between these dialects demonstrates the phonemic timing differences between
segments that can be points of contrast. The authors also found that there was positional variation
in F2 for the laterals. Irrespective of whether the lateral was clear (as in Newcastle) or dark (as in
Leeds), the F2 was lower in word-final position, indicating even more tongue body retraction
than was already present. The also found that for word-medial liquids, whether or not the foot
was iambic or trochaic had an effect for Newcastle speakers: F2 for the lateral remained higher
for the lateral in iambic feet, but the F2 contrast was eliminated between rhotics and laterals for
trochaic feet, resulting in the same F2 frequencies and trajectories. This effect was not observed
for Leeds speakers. In North American English, the clear–dark distinction may also vary with
dialect (Olive et al., 1993; Harris, 1994). The variation in timing across and within categories of
liquids contributes to their complexity and is something children must acquire.

In North American English, timing of tongue retraction for laterals famously varies as a
function of context resulting in a positional clear/dark distinction. Consider, for example, the
observation that laterals are dark intervocally and post-vocally and clear elsewhere in
many dialects of American English (Sproat & Fujimura, 1993; Olive et al., 1993; Hayes, 2000).
Sproat & Fujimura (1993) attribute the distinction to a timing difference: when a prevocalic /l/ is
clear, the anterior constriction is achieved before the posterior one; in postvocalic position, the posterior constriction is achieved first and involves more retraction than prevocally and /l/ is dark (see also Lawson et al., 2010). However, it is important to note that both clear and dark laterals do not always demonstrate a lag between the anterior constriction and posterior one. The specific timing patterns vary cross-linguistically, but there is an overwhelming tendency for a dark lateral to be accompanied by a later anterior constriction compared to the dorsal one (Gick, Campbell, Oh, & Tamburri-Watt, 2006). The position-dependent timing pattern of an early anterior constriction for the clear variant versus a later anterior constriction for the dark variant has been observed in other languages (Westerman & Ward, 1933; Recasens & Espinosa, 2005). For English rhotics, the positional variation in tongue dorsum retraction and timing has not been observed (Proctor, Walker, Smith, Szalay, Goldstein, & Narayanan, 2019).

Intriguingly, the position-dependent patterns of liquid articulation also appear to influence their acquisition in English. McGowan, Nittrouer, & Manning (2004) conducted a longitudinal acoustic study of /ɹ/ acquisition in speech produced by very young children between the ages of 14 months to at least 26 months. They observed that while all children developed the ability to articulate postvocalic and syllabic /ɹ/ in an adult-like manner by the end of the study period, none were able to articulate prevocalic /ɹ/ in an adult-like manner. The opposite pattern has been observed for /l/: its prevocalic articulation may be acquired before its postvocalic articulation (Prather et al., 1975; Dyson, 1988; Lin & Demuth, 2013; 2015), albeit in dialects where the prevocalic variant is perceived as dark (Lin & Demuth, 2013).

In sum, liquids are complex constrictions that involve a minimum of two lingual constrictions: an anterior one achieved with the tongue tip/blade, and a posterior one achieved with the tongue body. Within English, these constrictions are often differently timed. The timing
results in a clear versus dark distinction that contributes to the additional positional contrast in laterals and rhotics that appear in onset and offset position. The timing differences create allophonic variants of /l/ and /ɹ/ that children acquire at different rates. The position-dependent asymmetries of /l/ and /ɹ/ acquisition also suggest that the protracted development of adult-like timing contributes to the late acquisition of these sounds.

**Long-Domain Effects of Liquid Articulation**

The importance of articulatory timing to the acquisition of liquids is underscored by the observation that liquids have particularly strong coarticulatory effects on surrounding sounds (Tunley, 1999; Lawson et al., 2010; Proctor et al., 2019). For example, Kelly and Local (1986) found that British English liquids had anticipatory effects that spanned 2 syllables. Hawkins & Slater (1994) linked this effect to a F2 lowering, which can be interpreted as an early long-distance retraction for the tongue (see Mrayati, et al., 1988). Tunley (1999) showed that English /ɹ/ (realized as [ɹ]) conditions stronger anticipatory and carryover effects than English /l/. These effects influence the perceptual recovery of /l/ and /ɹ/ in running speech. West (2000) found that when the liquid itself was masked, participants could correctly identify an /l/ or /ɹ/ based on resonance characteristics of the preceding vowels. Rhotics could be identified up to 2 syllables before their articulation.

The longer distance coarticulatory effects of an /ɹ/ target compared to /l/ target might be attributable to a labial constriction, which—when coupled with the anterior and posterior lingual constrictions—greatly reduce F3. Support for this hypothesis comes from Kochetov & Neufeld (2013), who found that F3 lowering in the context of an upcoming /ɹ/ may extend over as many as 5 unstressed syllables, with no concomitant effect of F2 lowering. West (1999) also observed a very long-distance effect of F3 lowering in anticipation of /ɹ/. These very long-distance effects
were correlated with tongue and lip position differences in electromagnetic articulatory (EMA) sensor location. However, West (1999) indicates that because the magnitude of difference in tongue position is so small (often under 1mm) it is difficult to associate the observed acoustic differences with specific articulatory gestures because of the inherent limitations of EMA technology. But the data does confirm that both the tongue and lips are active in the observed long-distance effects. But this inference also assumes that tongue positioning is more influenced by the immediate segmental context than by larger chunks of speech, which may be less true of children’s speech than adults’ speech (Rubertus & Noiray, 2018). If tongue positioning is optimized within larger chunks of speech, then it is important to remember that a pharyngeal constriction will also contribute to F3 lowering (Mrayati et al., 1988). Thus, tongue root retraction in anticipation of an upcoming /ɹ/ could provide another source for the long-distance effects observed in perceptual studies of liquid coarticulation.

In sum, studies of liquid coarticulation suggest that tongue body retraction, associated with F2 lowering, extends through time. In the anticipatory direction, it may span up to two (unstressed) syllables. The lip rounding associated with English /ɹ/ may be initiated even earlier than tongue body retraction, but the extended domain of F3 lowering associated with /ɹ/ articulation may also reflect anticipatory tongue root retraction.

**Current Study**

The aim of the current study is to infer the development of articulatory timing control in the production of liquids by investigating the short- and long-domain effects of prevocalic /l/ and /ɹ/ targets in child and adult speech. In children with typical development, liquid articulation is immature when children enter school at age 5 years; in fact, it is not until age 8 years that the sound is considered to be fully acquired (Prather et al., 1975; Smit et al., 1990). We therefore
recruited a younger and older group of school-aged children (5- and 8-year-olds, respectively) to test for age effects on articulation. We also recruited a group of college-aged adults to compare liquids in children’s speech with the adult target. Overall, the goal was to infer a developmental trajectory from the immature production of liquids in 5-year-olds’ speech to the more advance productions in 8-year-olds’ speech to adult productions. All speakers produced sentences with disyllabic compound nouns in prefinal utterance position. The first lexeme of the compound was the monosyllable with /l/ or /ɹ/ as the target onset. To differentiate effects of anticipatory labial versus pharyngeal constriction on /ɹ/ articulation, the following vowel in the target rhyme was either a low-front unrounded vowel /æ/ or the high-back rounded vowel /u/. Several unstressed syllables preceded the target monosyllabic word. This word was the most prominent syllable of the utterance—receiving both lexical and sentential stress. In this way, supralexical prosodic units were created that included at least one preceding unstressed syllable and the target syllable (see Selkirk, 1996). The goal was to maximize the coarticulatory effect of the target syllable on adjacent unstressed syllables: prominent syllables transmit coarticulatory effects by “encroaching” on adjacent unstressed syllables (Fowler, 1981), presumably because they are hyperarticulated (de Jong, 1995). Short domain coarticulatory effects were investigated using acoustic measures. Long domain effects were investigated using a validated perceptual measure of coarticulation (Redford et al., 2018; Howson et al., 2020). The basic prediction was that children would employ a strategy of tongue-body-first articulation. Specifically, we predicted that children prioritize the tongue-body gesture over the tongue tip/blade when they lack the proper motor control to coordinate both. This prediction follows from previous research and children’s speech errors (overwhelmingly, liquids are substituted for [w], which prioritizes a tongue body gesture). In particular, we predicted that children’s poor independent control over
tongue tip/blade articulation would lead them to produce the lateral and rhotic liquids with a similarly retracted tongue body (relative to adults) in both the short and long domains, especially when the stressed vowel context further conditioned early tongue body retraction. The additional long domain prediction was that a targeted /l/ would be detectable earlier in the speech stream than a targeted /r/ in both adult and child speech in the /æ/ context based on lip rounding, but not necessarily in the /u/ context absent the presence of a tongue root constriction during [ɹ] articulation. Our long-domain predictions are tested with Experiment 2, which utilizes an AV Gated perception task (Redford et al., 2018; Howson et al., 2020).

**Experiment 1**

The first experiment compared anticipatory effects of an /l/ versus /r/ contrast in aCV sequences in child and adult speech. The sequences spanned a lexical boundary (i.e., a#CV). Acoustic measures were used to infer articulatory movement. Acoustic theories of speech production (e.g. Fant, 1960) explain that tongue height plays a major role in determining F1 frequency, while tongue advancement plays a major role in determining F2 frequency. Specifically, lower F1 frequencies are correlated with greater tongue height in the vocal tract, while lower F2 frequencies are correlated with greater tongue retraction. Sproat & Fujimura (1993) also connected the derived measure of F2-F1 with the degree to which there is tongue body retraction into the pharyngeal cavity: smaller differences between the two formants indicates greater retraction of the body into the pharynx. The specific predictions followed from the assumption of a tongue-body-first articulation strategy in young children’s speech and were as follows: F2 and F2-F1 values on the schwa preceding the target consonant would be relatively lower in anticipation of it in younger children’s speech compared to older children’s and adults’ speech; F2 and F2-F1 contrasts between the different coronal consonants would be reduced in
younger children’s speech relative to older children’s and adults’ speech; younger children would not produce any F2 and F2-F1 contrast for liquids due to position-dependent timing effects on anterior versus posterior constrictions. The stressed V was also predicted to influence schwa production, but much more so in children’s speech than in adults’ speech (Rubertus & Noiray, 2018). Also, the anticipatory effects of a targeted /-ud/ rhyme on schwa production was expected to further reduce any early contrast between /l/ and /ɹ/ in children’s speech, as this vowel context was expected to increase anticipatory tongue body retraction, making an anterior constriction that much more difficult to attain.

Two types of acoustic analyses were conducted: one based on static formant values and one based on formant dynamics. The static analyses were based on formant values extracted from schwa and used to investigate the effects of (alveolar) liquid versus stop articulation on anticipatory tongue body posturing across age groups. The dynamic analyses were based on values extracted along the length of the continuous formant trajectories in the schwa-liquid-vowel (əLV) portion of the determiner-monsoyllabic target sequence. These values were then analyzed using smoothing spline (SS)ANOVAs (Gu, 2002), allowing us to infer the effect of rhotic versus lateral articulations on the temporal dynamics of tongue movement in the sequence across age groups. Sequences with the stop consonant were omitted from these latter statistical analyses because the absence of formant values through the closure duration period disrupts the smoothing spline fits of the formant frequencies. The methods are introduced in detail below, followed by the results from the static formant analyses and then the dynamic formant analyses.

Methods

Participants
Participants were 11 college-aged adults, 12 8-year-olds, and 13 5-year-olds. College-aged adults ranged in age from 18.4 to 23.7 years ($M$: 21.2 years, $SD$: 1.5 years). They were recruited by word-of-mouth from the University of Oregon student body. The 5-year-old children ranged in age from 62 months to 74 months ($M$: 68.2 months, $SD$: 3.9 months); 8-year-old children ranged in age from 91 months to 106 months ($M$: 21.2 months, $SD$: 4.9 months). Children were recruited through Team Duckling, a developmental database maintained at the University of Oregon, and from summer camps run by the YMCA in Eugene, Oregon to take part in a large-scale project on speech rhythm acquisition. The speech elicited in the present study was therefore embedded in a longer protocol.

Typical speech and language development in college-aged adults was determined based on self-report. Typical development in children was determined based on parental report and based on an in-laboratory assessment of speech-language skills using the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd et al., 2002) and the Clinical Evaluation of Language Fundamentals – Fifth edition (CELF-5; Wiig, Semel, & Secord, 1992). Inclusion criteria were standardized scores within 1 standard deviation of the mean on both the DEAP and the CELF-5.

Participants first heard 1000 Hz at 20 dB in the right ear. If this elicited an appropriate response, participants heard a tone at 1000, 2000, and 4000 Hz at 20 dB in each ear (one ear at a time) as per the guidelines set by the American Academy of Audiology Task Force (2011).

Consent for participation was sought from all participants, including specific consent from parents and college-aged adults to show audio-visual (AV) clips to other college-aged students to collect perceptual judgment of coarticulation (see Experiment 2). Assent was also acquired from children via description in plain English of the activities they would complete during the study.
Families and college-aged adults were financially compensated for their time at the rate of $15 per hour. Children also earned a small prize at the end of the study.

**Stimuli**

The target words were two sets of real word minimal pairs with a lateral, rhotic, or alveolar stop onset: *rad, lad, dad* and *rude, lewd, dude*. The two sets were created to vary the vowel context in addition to the target consonant. The targets /æ/ and /u/ were chosen to produce maximally contrasting tongue postures: [æ] is a near-open front vowel and is produced with a fronted and lowered tongue posture, while [u] is a closed back vowel and is produced with a retracted and raised tongue posture (Stone & Lundberg, 1996). The /d/-onset was included as a control, alveolar consonant that is acquired early (<= 3 years; Smit et al., 1990).

Target words were embedded as the penultimate item in the frame sentences: *They said it could be the ___ house*. This sentence was created to encourage phrasal stress on the target item and to discourage final lengthening and devoicing of the target item, both of which occur in phrase-final position. The repeated sequence “said it could be the” was expected to be highly reduced as none of the elements were expected to receive primary stress. The unfooted determiner *the* was expected to be prosodified with the target word.

**Procedure**

**Elicitation.** Different pictures were created to depict the different types of houses that the target words defined. The pictures alluded to, rather than strictly adhered to, the definitions of the target words. They nonetheless served their purpose as memory prompts. Participants were first trained on each of the target words in association with the picture and encouraged to produce the whole noun phrase in the practice elicitations (e.g. *the rad house*). This training was the same for both adults and children. Participants were then told that they would say each item in a carrier
phrase. Full sentences were elicited with the picture prompt, which was shown at the same time as a recorded sentence prompt: “Look at this house, I think it could be the target house, what kind of house did I say it could be?” This prompt was played out-loud through speakers from a computer; no headphones were used. After being prompted, participants replied, “They said it could be the target house.” After a first repetition, participants were prompted by the experimenter to produce a second repetition with the phrase “What kind of house did they say it could be?” This was done 3 times in a randomized order for a total of 6 elicitations of each sentence, giving a total of 36 sentences per speaker for analysis. Adults and 8-year-olds did not experience any difficulties with the elicitation procedure; in contrast, some 5-year-olds produced some sentences with a prosodic break or by producing just the target word. When this happened, the experimenter reminded the child of the full target sentence, and the elicitation procedure was repeated. Participant’s speech was recorded at 44100 Hz with a Marantz PMD660 and a Shure ULXS4 microphone that was attached to the brim of a hat that each participant wore. Two experimenters were present throughout data collection. One performed instruction and elicitation and one managed the stimuli delivery, and the audio and visual recording devices. Audio-visual (AV) recordings were also made at the same time, as described under Experiment 2 below.

**Formant Extraction.** Praat (Boersma & Weenink, 2019) was used to extract formant frequency values were extracted at 40 equally spaced intervals across the determiner-target noun sequences (e.g., “the rad”) from the onset of schwa through to the offset of the vowel following the liquid. Schwa onset was determined as the point where closure and/or friction for /ð/ ended and where formant structure and a periodic waveform began. The offset of the target stressed vowel was taken to be the absence of higher formant structure at stop closure. Before the
formants for each token were extracted, the maximum frequency and number of tracked formants was adjusted so that the tracked formants matched a visual inspection of the spectrogram. Schwa offset in the əLV sequences was identified using a combination of spectral and amplitude cues. For the static formant analyses, only formant values at schwa midpoint were extracted to compare anticipatory effects of an upcoming stop with those of an upcoming liquid. All textgriding and formant extraction was performed by the first author.

**Vocal Tract Normalization.** In order to more directly compare the effect of the target consonant on its immediately preceding context as a function of the speaker’s age group, static formant values were normalized for vocalic tract length, using a modified bark difference metric (Syrdal & Gopal, 1986). Formants were first converted to the perceptual bark scale using the Kendall and Thomas (2018) vowels package in R (R Core Team, 2019). Normalized F1 and F2 values were then calculated by subtracting the bark-transformed F1 (= Z1) and F2 (= Z2) values from the bark-transformed F3 (=Z3) values. This vowel intrinsic normalization procedure results in low normalized F1 values for open vocal tract configurations and high normalized F1 values for more closed configurations. Similarly, low normalized F2 values indicate a more anterior constriction than higher normalized F2 values. Normalized F2-F1 values were also calculated by subtracting (Z3-Z2) from (Z3-Z1). This measure was incorporated because it correlates with retraction of the tongue dorsum/root (Sproat & Fujimura, 1993). Formant normalization was not performed for the SSANOVA analysis because the comparison is of formant trajectory patterns as a function of liquid type and vowel context across age groups; the absolute values of the formants was not of particular interest. In addition, we wanted to examine potential differences in the trajectory of F3 as a function of liquid and vowel within each age group and the formant normalization removes F3 (see above). Because the SSANOVA data is not normalized, formant
values can be interpreted more canonically; for example, lower unnormalized F2-F1 values correspond to a more retracted tongue. The normalized data require that this interpretation be inverted since these values reflect a difference from F3, so higher normalized F2-F1 values correspond with a more retracted tongue position than lower values.

**Analyses**

Linear mixed-effects modeling was used to test for the fixed effects of Age Group (3 levels: 5-year-olds, 8-year-olds, adults), Target (3 levels: /l, r, d/), and Vowel Context (2 levels: /æ, u/) on normalized schwa F1, F2, and F2-F1 (Sproat & Fujimura, 1993) using the lme4 package (Bates et al., 2015) in R (R Core Team, 2019). The model had a random intercept to account for the random effect of speaker. The R² metric was calculated to assess the model fit using the MuMin package (Barton, 2019).

The dynamic formant analyses used a smoothing spline (SS)ANOVA (Gu, 2002; Davidson, 2006) over the øLV sequence. A script to plot SSANOVA (Mielke, 2013) was used in R (R Core Team, 2019). The SSANOVA calculates a best fit contour of the input formant values with 95% confidence intervals for (non-normalized) F1, F2, F3, and F2-F1. A total of 8 plots were made to investigate significant differences in liquid articulation (= Target) as a function of vowel context within each age group. Analyses of coarticulatory effects on schwa production are presented first, followed by the analyses of formant dynamics across the schwa-liquid-vowel (øLV) sequence.

**Results**

**Static Formant Analyses**

The linear mixed effects modeling results for normalized schwa F1 revealed a main effect of Age Group \[F(2, 29) = 12.72, p < .001\], Target \[F(2, 1172) = 68.79, p < .001\], and Vowel
Context $[F(1, 1172) = 22.29, \ p < .001]$ on schwa articulation. There was a two-way interaction between Age Group and Target $[F(4, 1172) = 3.08, \ p = .016]$ and between Age Group and Vowel Context $[F(2, 1172) = 3.08, \ p = .046]$. There was no interaction between Target and Vowel Context $[F(2, 1172) = 1.54, \ NS]$ or Age Group, Target, and Vowel Context $[F(4, 1172) = 1.36, \ NS]$. The model R-squared was .38. The significant two-way effects are shown in Figure 1.

Recall that higher normalized F1 values indicate greater constriction along the height dimension than lower normalized F1 values. Given this, an examination of the direction of the main effects indicates the following: adults generally produced schwa with a more open vocal tract than children; schwa was articulated with a higher tongue body position before target /d/ than before the targeted liquids, consistent with anticipatory posturing for manner of articulation; schwa was articulated with a marginally more open vocal tract before the low vowel target, /æ/, than before the high vowel target, /u/, consistent with anticipatory posturing for the stressed vowel. Table 1 summarizes the differences on schwa with respect to age group, target phoneme, and vocalic environment.

Post hoc analysis of mean differences using pairwise comparison with Tukey p-value correction in the critical Age Group × Target interaction shown in Figure 1 (left) revealed no difference between 5-year-olds and 8-year-olds ($p > .05$), but both groups differed from adults in the degree of anticipatory posturing for the upcoming consonant ($p < .05$): children’s schwa production was less influenced by the target consonantal contrast than adults’ schwa production. A similar analysis of the significant Age Group × Vowel Context interaction shown in Figure 1 (right) again revealed no significant difference in the degree to which 5-year-olds and 8-year-
olds’ production of schwa was influenced by the following stressed vowel \((p > .05)\): neither group showed an effect of vowel context on schwa F1. In contrast, adult production of schwa varied systematically with vowel context \((p < .05)\).

The results for normalized F2 revealed a main effect of Target \([F(2, 1173) = 235.10, p < .001]\) on schwa production, but no main effect of Age Group \([F(2, 29) = 0.10, \text{NS}]\), or Vowel Context \([F(1, 1173) = 0.58, \text{NS}]\). As with F1, there was a two-way interaction between Age Group and Target \([F(4, 1173) = 8.744, p < .001]\), but not between Age Group and Vowel Context \([F(2, 1173) = 0.937, \text{NS}]\) or between Target and Vowel Context \([F(2, 1172) = 1.42, \text{NS}]\). The three-way interaction was also not significant \([\text{Target} \times \text{Age Group} \times \text{Vowel Context}, F(4, 1172) = 0.84, \text{NS}]\). The model \(R^2\) was .39. The significant effect of Age Group \(\times\) Target is shown in Figure 2.

[Figure 2 here]

Post-hoc tests revealed that there was a significant difference in normalized schwa F2 as a function of the target consonantal contrast \(/d/, M = 2.91, SD = 0.89; /l/, M = 4.30, SD = 1.22; /ɹ/, M = 3.20, SD = 1.15; p < .001). Intriguingly, 5-year-olds were found to produce schwa differently before /d/ than before /ɹ/ \((p = .004)\), while 8-year-olds and adults did not.

Results for normalized F2-F1 revealed a significant effects of Age Group \([F(2, 29) = 7.82, p = .002]\), Target \([F(2, 1153) = 164.43, p < .001]\), and Vowel Context \([F(1, 1153) = 16.77, p < .001]\) on schwa production. There was also a two-way interaction between Age Group and Target \([F(4, 1153) = 4.92, p < .001]\) and between Target and Vowel Context \([F(2, 1153) = 3.19, p = .042]\), but no interaction between Age Group and Vowel Context \([F(2, 1153) = 1.96, p = .142]\). The effect of Age Group \(\times\) Target \(\times\) Vowel Context was not significant \([F(4, 1153) = .}
The model R-squared was .41. Figure 3 shows the significant two-way interactions.

Post-hoc tests of mean differences revealed that many of contrasts in F2-F1 values that are visible in Figure 3 were significant ($p < .001$), including the difference between target /d/ and /l/ ($/d/, M = 8.68, SD = 1.55; /l/, M = -6.90, SD = 1.73), target /d/ and /ɹ/ ($M = -7.57, SD = 1.76$), and between /l/ and /ɹ/. In addition, the mean difference between target /æ/ ($M = -7.55, SD = 1.75$) and /u/ ($M = -7.91, SD = 1.90$) was also significant ($p < .001$). In general, the differences may indicate that schwa before target /l/ was produced with more tongue retraction than before either target /ɹ/ or /d/, and that schwa before target /ɹ/ was produced with more tongue retraction than before /d/. Alternatively, the differences may be due to degree of vocal tract constriction. This latter interpretation is consistent with the difference in schwa production before /æ/ and /u/.

Post-hoc tests on the effect of Age Group × Target differences revealed that 5-year-olds and 8-year-olds produced schwa before both target /l/ and target /ɹ/ with the same degree of bark-corrected F2-F1 (5yo /l/, $M = -7.24, SD = 1.48$; 8yo /l/, $M = -7.39, SD = 1.43$; 5yo /ɹ/, $M = -7.68, SD = 1.48$; 8yo /ɹ/, $M = -8.00, SD = 1.97$), but that adults produced [l] ($M = -5.60, SD = 1.47$) with higher bark-corrected F2-F1 than either 5-year-olds or 8-year-olds ($p < .001$). Although the F2-F1 difference between 5-year-old and adult schwa was not significantly different before /ɹ/ ($p = .117$; 5yo /ɹ/, $M = -7.67$; adult /ɹ/, $M = -6.69, SD = 1.47$), the F2-F1 difference between adults’ and 8-year-olds’ schwa before /ɹ/ was significant (8yo /ɹ/, $M = -8.00, SD = 1.97, p = .002$; adult /ɹ/, $M = -6.69, SD = 1.47$).

Post hoc tests on the effect of Target × Vowel Context differences revealed that schwa before target /l/ was produced differently as a function of vowel context (/æd/, $M = -6.57, SD =
1.63; /u/, $M = -7.23$, $SD = 1.76$; $p < .001$). The significant difference between schwa before /d/ as a function of vowel context (/æ/, $M = -8.51$, $SD = 1.51$; /u/, $M = -8.86$, $SD = 1.58$, $p = .006$) may support the interpretation that less negative values in normalized F2 minus normalized F1 result from the lower normalized F1 values (i.e., Z3-Z1) associated with a more open vocal tract, which has a higher non-normalized F1 value (but see Dynamic Formant Analyses below).

**Dynamic Formant Analyses**

The dynamic formant analyses were based on true (non-normalized) formant values. The relative change in formant frequencies through time was examined with reference to the 95% confidence intervals given. Only qualitative comparisons between age groups are made. Adult formant trajectories for the əLV sequence are presented first, followed by the 8-year-old trajectories, followed by the 5-year-old trajectories. This presentation is meant to facilitate developmentally relevant comparisons: where 8-year-old patterns deviate from adult patterns and 5-year-old patterns deviate from 8-year-old and adult patterns, we assume immaturity in the younger speakers’ articulatory timing of the sequence.

**Adult Speakers.** Figure 4 presents the SSANOVA results for F1, F2, and F3 from adults’ production of the əLV sequence in *the lad/rad house* (left) and *the lewd/rude house* (right). The figure shows similar overall trajectories across vowel contexts, with a similar early difference in F2 and F3 trajectories indicative of the liquid contrast: /l/ versus /ɾ/. In the low front vowel context, F1 increased somewhat towards the end of the schwa, then decreased during the constriction for the liquid, and increased again with the transition into the vowel. There were no visible differences in the F1 trajectories for target /l/ and /ɾ/, suggesting that both were produced with a similar degree of varying vocal tract constriction across the sequence. In contrast, the F2 trajectories for target /l/ and /ɾ/ differed substantially during the production of schwa and the
liquid until converging during the steady state portion of the stressed vowel. More specifically, the F2 trajectories were lower during the initial part of the əLV for target /l/ compared to target /ɻ/, consistent with a relatively “dark” articulation of the intervocalic lateral liquid. The F3 trajectories were, not surprisingly, even more distinct for target /l/ and /ɻ/ sequences. As expected, F3 dropped dramatically before and during the articulation of the rhotic liquid.

[Figure 4 here]

A similar pattern of results was evident in the high back vowel context. Again, the F1 trajectories did not differ significantly from one another as a function of the liquid contrast. They were however flatter, suggesting little in the way of overall height adjustments during schwa-liquid-vowel articulation. But, as in the case of the low front vowel, F2 decreased more with the coordinated constrictions for [l] and less so for [ɻ]. The differences between the F2 trajectories also disappeared earlier in the high back vowel sequence compared to the low front vowel context. The dramatic F3 drop associated with [ɻ] production was again evident across the sequence.

The SSANOVA results for F2-F1 in Figure 5 highlight the acoustic changes associated with tongue body retraction during production of the əLV sequence as a function of vowel context. The difference between the F1 and F2 trajectories for [l] decreased by nearly 500 Hz when the targeted constrictions are aligned in the high vowel context. The difference between F1 and F2 also fell with the realization of the coordinated constrictions during [ɻ] articulation before increasing again with the articulation of the stressed vowel. The trajectories for the target /l/ and /ɻ/ sequences in the high back vowel context were different from the low front vowel context, but the trajectory associated with target /l/ once again had a more dramatic fall then rise pattern than those associated with target /ɻ/. In the high back vowel context, the difference between the lateral
and rhotic persisted until the mid-point of the vowel. It then reappeared with the trajectory for target /l/ sequences maintaining slightly higher F2-F1 values than those associated with target /ɹ/ sequences.

[Figure 5 here]

**8-year-old Speakers.** Figure 6 presents the SSANOVA results for F1, F2, and F3 from 8-year-old children’s production of the target əLV sequences in *the lad/rad house* (left) and *the lewd/rude house* (right). The figure shows that, as in adults’ speech, older school-aged children produced the target /l/ and /ɹ/ sequences with overlapped F1 trajectories. Also, similar to adults’ speech, the F1 trajectories in the high back vowel context were flatter than in the low front vowel context. The shape of the F2 formant trajectories were also similar to those of adult speech. F2 decreased during the coordinated constrictions associated with liquid articulation, and more so for /l/ than /ɹ/ in the front vowel context. The F3 trajectories again distinguished /l/ from /ɹ/, with an expected dramatic decrease during [ɹ] articulation. Overall, 8-year-old children’s production of the əLV sequences differ in degree but not in kind from adults’ production of these sequences. In particular, the sustained differences between the different liquid F2 trajectories were smaller than in the adult case.

[Figure 6 here]

Figure 7 presents the SSANOVA results for F2-F1 in 8-year-old children’s speech. These results confirm that 8-year-olds distinguished /l/ from /ɹ/ along the F1 and F2 dimension. The shapes of the difference trajectories varied substantially as a function of vowel context, similar to the adult data (see Figure 5). A notable difference between the adult and 8-year-old contrasts of /l/ and /ɹ/ sequences is in the greater degree of initial overlap in F2-F1 trajectories compared to adults’ speech. The absence of difference is especially striking in the high vowel context. Also,
Figure 7 suggests a more gradual increase of F2-F1 values during the transition from [I] into the following low front stressed vowel in 8-year-olds’ speech compared to adults’ speech.

[Figure 7 here]

5-year-old Speakers. Figure 8 presents the SSANOVA results for F1, F2, and F3 from 5-year-old children’s production of the aLV sequences in the lad/rad house (left) and the lewd/rude house (right). As in 8-year-old children’s and adults’ speech, F1 trajectories are largely overlapped in the /l/ and /ʌ/ sequences. Unlike in speech produced by older children and adults, younger children’s F1 trajectories show more substantial movement across the sequence, including in the high back vowel context. This result suggests that younger children show larger changes in the open-close dimension of the vocal tract during aLV articulation compared to older children and adults.

[Figure 8 here]

An even more notable difference between the younger children’s speech and older children’s and adults’ speech is the almost complete lack of difference in the F2 trajectories for /l/ and /ʌ/ sequences. The F2 trajectory dips in the same way during the articulation of each liquids. The contrast between the two liquids is really only evident in F3, which is substantially lower for /ʌ/ compared to /l/.

Figure 9 presents the SSANOVA results for F2-F1, and confirms the observation that 5-year-olds barely distinguished between /l/ and /ʌ/ along the F1 and F2 dimensions. Nonetheless, there was a small, but significant, different between in /l/ and /ʌ/ F2-F1 values at the outset of the sequence. F2-F1 values also appeared lower for the lateral than for the rhotic during transition into the stressed front vowel, but this difference was not significant. There were no significant F2-F1 differences in [əlu] versus [aʊ], except for a brief period at the offset of the vowel.
Discussion

All speakers in the current study produced [ɹ] with lower F3 values than [l], especially during maximal constriction for the liquid (i.e., the maximum displacement of the articulators for the segment), but the combined results from static and dynamic formant analyses also indicated that young children did not otherwise contrast /l/ and /ɹ/ to the same degree as older children and adults. Whereas adults anticipated /l/ with a much lower normalized F2 and F2-F1 than /ɹ/ from the outset of the preceding unstressed vowel, older children were slower to contrast the liquids along these dimensions. Young children simply did not distinguish /l/ and /ɹ/ along F2, but some minor difference emerged in F2-F1 at the outset of the sequence. Comparison with ədV sequences indicated that adult speakers produced a darker /l/ and clearer /ɹ/ than children (see Figure 3): normalized F2 was especially low (= higher Z3-Z2) in adult speech when schwa preceded /l/ compared to when it preceded /ɹ/; schwa was not significantly different before /ɹ/ than before /d/. Comparison of schwa F2-F1 values before /ɹ/ and before /d/ also suggests that 8-year-olds produced a somewhat clearer version of /ɹ/ than 5-year-olds, whose F2-F1 values were statistically different for /ɹ/ and /d/, suggesting significantly greater tongue retraction during schwa for /ɹ/. This developmental difference between 5-year-olds’ and 8-year-olds’ speech was even more evident in the results from dynamic formant analyses. The dynamic F2 and F2-F1 trajectories suggested that 5-year-olds retract the tongue body to an equal degree in anticipation of both /l/ and /ɹ/. In contrast, 8-year-olds were more similar to adults in retracting the tongue body to a greater degree in anticipation for /l/ than for /ɹ/, but this timing difference in degree of retraction was less evident in the /u/ context compared to /æ/ context. Vowel context otherwise had less effect on children’s production of schwa compared to adults’ production of schwa,
contrary to the literature-based expectation of greater V-to-V effects in children’s speech (e.g. Rubertus & Noiray, 2018).

Taken together, the acoustic results suggest that 5-year-olds lack the motor skills necessary to distinguish liquids with respect to fine-grained differences in constriction targets (e.g. Green, Moore, & Reilly, 2000; Studdert-Kennedy & Goldstein, 2003). On the other hand, 8-year-olds have more complex motor skills, but still struggle with the coordination of liquids with different vocalic segments as indicated by the distinction between the /a/ and /u/ environments. The result reveals the children struggle with complex coordination of multiple segments that would indicate timing of liquids play a role in acquisition problems. However, long-distance timing relationships were not observable from this acoustic data.

**Experiment 2**

The results from Experiment 1 suggested that 5-year-olds produce both the lateral and rhotic liquid with the same degree of early tongue body retraction. In contrast, adults and 8-year-olds appear to distinguish /l/ and /ɹ/ in the degree to which the tongue body is retracted; specifically, /l/ is more retracted earlier in the speech stream than /ɹ/ in older children’s and adults’ speech, and /l/ is more retracted overall than /ɹ/. These results beg the question: How early in the speech stream does tongue body retraction occur in advance of liquid articulation? We used a forward-gated audiovisual (AV) speech prediction task to answer this question. Redford and colleagues (Redford, Kallay, Bogdanov & Vatikiotis-Bateson, 2018) have previously shown that this task, which leverages AV speech perception, can be used to accurately index the onset and strength of anticipatory coarticulation in the speech stream, and that this index is more sensitive to the details of articulatory posturing than acoustic indices. Moreover, Howson and colleagues (Howson, Kallay, & Redford, 2020) have shown that the forward-gated AV speech prediction
task can be generalized to index coarticulatory onset and strength in 5-year-old children’s speech.

To measure long-distance anticipatory effects, the gated AV speech task exploits the perception of difference in the articulation of a frame sentence that is held constant as a function of a minimal contrast in the targeted speech sound. Since we were interested here in the long-distance effect of liquids on tongue body retraction, the task used in the current study was the difference between /l/ and /d/ targets and between /ɹ/ and /d/ targets. Based on the short-domain results, the prediction was that both liquids would be detected/predicted earlier in children’s speech compared to adults’ speech. The effect of vowel context on tongue retraction was investigated separately by blocking the target consonant conditions by stressed vowel. The /u/-vowel context also allowed us to explore effects of lip rounding on the early detection of /l/ and /d/ in the speech stream. The prediction was that if the ultra-long-distance effects of /l/ on preceding speech sounds are due to lip rounding, then /l/ should be detected earlier in the speech stream in an /æ/ vowel context versus an /u/ vowel context, since anticipatory lip rounding for the /u/ is likely to interfere with detection of anticipatory lip rounding for /l/.

Methods

Participants

Sixty undergraduate students aged 20 to 24 years old served as perceivers in the forward-gated AV speech prediction task. These participants were recruited through the University of Oregon Psychology and Linguistics Human Subjects Pool and received course credit for their participation. They were given a link to the study, which was hosted by the online data collection platform Testable (www.testable.org). Howson et al. (2020) demonstrated that anticipatory coarticulation measures derived from online version of the forward-gated AV speech prediction
task were as reliable as those derived from an in-person version. Fifteen perceivers were randomly assigned to one of four experimental conditions (see below) to separately investigate the effect of vowel context on the articulatory timing of liquid articulation.

**Materials**

Audio-visual recordings were made at the same time as the audio-only recordings used in Experiment 1. Audio-visual (AV) speech was recorded using Panasonic AJ-PX270 with audio input from a Shure SM81 Condenser microphone placed within a foot of the participants. The video frame rate was 59.94 frames per second, which allowed for temporally precise cuts to the video (see below). Video dimensions were 1290 x 738 pixels, which provided a high-quality image for perceptual judgments. Gated AV speech stimuli were based on a random subsample of speakers whose speech was examined in the acoustic study. The choice to limit the indexing of long-distance coarticulation to just 3 speakers from each age group was made to reduce the number of overall stimuli per condition and therefore the length of the task for perceivers.

Regarding stimuli, the middle 3 of 6 repetitions for each target sentence were segmented from the AV recording for each of the 9 speakers (i.e., 3 repetitions × 6 target sentences × 9 speakers = 162 sentences). Next, 8 forward-gated AV speech stimuli were created for each repetition of every sentence by cutting the video at segmental boundaries (162 sentences × 8 cuts = 1,296 stimuli). The cuts were made at the offset of each segment, except for the final cut which was at the stressed vowel midpoint (e.g., *They said it cou*1|ld 2|b 3|e 4| th 5|e 6| r 7|a 8|ad house, where | indicates a segmental cut and the number refers to the corresponding gate). Figures 10–12 show the final frame from 8 gated AV speech stimuli derived from a single repetition of sentences containing 3 different targets: *lad, rad,* and *dad.*

[Figure 10 here]
Procedure

Participants were assigned to one of 4 conditions: an /ɹ/ vs. /d/ condition with /æ/ vowels in the target words (i.e., “rad” versus “dad”); an /ɹ/ vs. /d/ condition with /u/ vowels in the target words (i.e., “rude” versus “dude”); an /l/ vs. /d/ with /æ/ vowels in the target words (i.e., “lad” versus “dad”); and an /l/ vs. /d/ with /u/ vowels in the target words (i.e., “lewd” versus “dude”). Participants signed up for one and only one of these conditions through an online participant recruitment platform. Participants had no knowledge as the nature of the study or condition they were signing up for prior to participation.

During the study session, participants were first asked to calibrate their screen such that a bar on the screen was the size of an ID card so that the video appeared at a constant size. After informed consent and demographic information was collected, instructions were given. Participants were briefly introduced to the concept of anticipatory coarticulation with reference to the articulatory differences in onset between “st trop” and “street.” They were then instructed that their task was to predict, based on the fragment of a clip shown, if the speaker was intending an “l” or “d” sound (lateral condition) or an “r” or “d” sound (rhotic condition). After participants read through the written instructions, the experiment began. The gated sentence stimuli were blocked by speaker within condition and randomized during presentation. The speaker order was also randomized within condition. On each trial, perceivers were tasked with deciding whether the snippet of video they saw ended with a target an “l” or “d” word or with an “r” or “d” word. The stimuli would play through, then on the final frame, boxes containing either
an “l” or “d” or an “r” or “d” would appear. Once the perceiver had selected their response, the next stimulus would play.

**Analyses**

We included responses from all four conditions in a single linear model in order to compare the degree to which /ɹ/ had differences in predictability in contrast with /l/ and to test how the vowels /æ/ and /u/ interacted with that predictability. Prior to statistical analysis, accuracy (ACC; Metz, 1978) was calculated within participant for each speaker and every gate within each condition. ACC values served as the dependent variable in a linear mixed effects model, which included the following fixed effects: Age Group (3 levels: 5-year-olds, 8-year-olds, and Adults), Liquid (2 levels: /l/ or /ɹ/), Vowel Context (2 levels: /æ/ and /u/), and Gate (8 levels: 1-8). The model, which also included a random slope and intercept for the interaction between Perceiver and Speaker, was built in R (2019) using the lme4 package (Bates et al., 2015). Fixed effect and interaction p-values were calculated using Satterthwaite’s Method (Satterthwaite, 1946) with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). The $R^2$ metric (Devore, 2011) was also calculated using the MuMIn package (Barton, 2018). Tukey HSD correction was used for post-hoc comparisons (Tukey, 1949) in the emmeans package (Lenth, 2018). Figures were created using the ggplot2 package (Wickham, 2018).

**Results**

The results revealed simple effects of Liquid [$F(1, 528) = 48.17, p < .001$], Vowel Condition [$F(1, 528) = 40.09, p < .001$], and Gate [$F(7, 3696) = 443.73, p < .001$], but not of Age Group ($p = .415$); however, the two-way interaction between Age Group and Liquid was significant [$F(2, 528) = 6.54, p < .002$]. The interactions between Liquid and Gate [$F(7, 3696) = 8.65, p < .001$] and Vowel Context and Gate [$F(7, 3696) = 8.75, p < .001$] were also significant. The other two-
way interactions were not significant. Only the 3-way interaction between Liquid, Vowel Context, and Gate was significant \( F(7, 3696) = 2.30, p = .025 \). There was no 4-way interaction. Figures 13–15 show the different 2-way interaction effects.

[Figure 13 here]

[Figure 14 here]

[Figure 15 here]

Figure 13 shows that perceivers were better able to detect /ɹ/ versus /d/ overall than /l/ versus /d/ in sentences produced by adults \( p < .001 \) and 8-year-olds \( p < .001 \), but not in those produced by 5-year-olds \( p = .202 \). The Figure also indicates that the overall correct detection of /ɹ/ was higher in sentences produced by 5-year-olds than in those produced either by 8-year-olds or adults.

Figure 14 shows that perceivers correctly detected /ɹ/ earlier than /l/, except in sentences produced by the youngest speakers where both /l/ and /ɹ/ were detected equally early. These effects were confirmed in post hoc tests, which showed that, across age groups, correct detection of /ɹ/ was reliably above chance at the offset of the “be” vowel in the frame sentence (They said it could be the ___-house). Correct detection of /l/ also occurred at least by the offset of “be” in 5-year-olds’ speech, but not until the onset of schwa in 8-year-olds’ and adults’ speech. These systematic age-related differences likely have to do with the timing of tongue body retraction. This conjecture is supported by the interaction between Liquid and Vowel Context, shown in Figure 15: /l/ was harder to distinguish from /d/ before /u/ than before /æ/, suggesting that anticipatory tongue body retraction for the vowel partially obscured cues associated with the identification of /l/. The effect of vowel context on /ɹ/ detection extended even further.
Presumably, a combination of anticipatory lip rounding and tongue body retraction for /u/ obscured the cues perceivers used to distinguish /u/ from /d/ early in the speech stream.

**Discussion**

The results from the gated AV speech prediction task align with previous findings for English (e.g. West, 1999, 2000; Kochetov & Neufeld, 2013) in that anticipatory coarticulation for liquids was found to extend across multiple syllable and word boundaries. The results also align with previous reports of longer anticipatory coarticulation domains for /ɹ/ than for /l/. At the same time, the results provide new information about the development of anticipatory coarticulation for liquids. In particular, our results strongly suggest that young children initiate tongue body retraction for /l/ earlier than do older children and adults. The results also suggest that tongue retraction for /l/ is timed more or less similarly to early constriction gestures associated with /ɹ/ production. The evidence for these claims is in the earlier correct detection of /l/ versus /d/ in 5-year-old speech than in 8-year-old and adult speech, and the negative impact that anticipatory posturing for the high back vowel had on detection. We additionally found similar degrees of prediction accuracy for /ɹ/ in 5-year-olds’ speech as in adults’ speech (as far back as d in the preceding word could was above chance for both groups). However, prediction accuracy in 8-year-olds’ speech was only above chance in the /ɹ/ versus /d/ conditions as far back as the /i/ in “be.” In addition, the overall effect of vowel context on liquid detection suggests that long-distance effects of an upcoming /ɹ/ on speech sound articulation is due at least in part to anticipatory posturing of the tongue body, especially perhaps of the tongue root, and not only to an associated labial constriction. Taken together, the results suggest that the fine tuning of timing necessary to produce adult-like liquids takes longer to acquire than the constriction locations themselves.
General Discussion

The hypothesis of whole word speech production entails that speech sound acquisition should be studied from an articulatory timing perspective. The current study adopted this perspective to investigate the acquisition of liquids in younger and older school-aged children. Although the children who were included in this study had already acquired perceptually acceptable /l/ and /ɹ/ as determined by a standardized perceptual assessment (i.e., the DEAP), the results indicate that even the older school-age children had not yet achieved adult-like mastery over the articulatory timing of these sounds. This finding is consistent with the observation that liquids are mastered late in development (Prather, Hedrick, & Kern, 1975; Boyce, Hamilton, & Rivera-Campos, 2016), which is explained with reference to their complex articulation: liquids require an anterior and posterior lingual constriction, and children are slow to develop independent control over the tongue blade and body (Gick, Kang, & Whalen, 2002; McGowan et al., 2004). The coarticulatory patterns described in the current study expand on this explanation to provide preliminary insight into how children approach the problem of a double (or triple) liquid articulation in time. Two such insights are discussed below. Both are presented as hypotheses to be tested in future work using direct articulatory measures.

First, the acoustic results from the present study are consistent with the hypothesis that children adopt a tongue-body-first articulation strategy. The dynamic formant measures indicated a stronger anticipatory effect of vowel context on children’s coarticulatory patterns than adults’ patterns. This effect suggests that anticipatory posturing for tongue body targets is prioritized over posturing for tongue blade targets. The suggestion is further supported by that finding that, unlike adults, children did not distinguish /l/ from /ɹ/ along F2 prior to maximal constriction of
the vocal tract. Also, the normalized F2-F1 results suggest that, like adults, children retract the tongue in advance of maximal constriction for liquids.

The hypothesis of a tongue-body-first strategy for liquid articulation situates the articulatory complexity explanation for the slow acquisition of liquids in a running speech context. It is not that school-aged children have difficulty with anterior constrictions per se, they just privilege posterior ones in an intervocalic context. The specific suggestion is that motor planning for a vowel-liquid-vowel sequence results in early tongue body retraction for liquid articulation because the surrounding vowels contribute strength to tongue body activation. This activation strength may overwhelm the control signal for an anterior constriction, which is in any case more difficult to attain with a concurrently retracted tongue body. This suggestion links the problem of liquid acquisition to the observation of greater vowel-to-vowel (i.e., αCV) coarticulation in children’s speech than in adult speech (Nijland, et al., 2002; Rubertus & Noiray, 2018) and to a theory of speech production where consonantal articulation is overlaid on a substrate of sequential vowel articulation (Öhman, 1966; Fowler, 1981; Fowler & Saltzman, 1993; Carré & Chennoukh, 1995; Birkholz, 2013). As a fundamentally context-dependent explanation, it also predicts systematic differences in the realization of liquids; for example, the activation strength for the posterior constriction may be significantly diminished in the presence of other anterior constrictions, such as in clusters, which predicts that the immature realization of liquids in this context may involve the elision of the posterior constriction. This prediction is consistent with cluster simplification strategies observed in very young children’s speech where liquids are often realized as glides (McLeod, Doorn, & Reed, 2001). The acquisition of timing differences may also explain why children acquire onset laterals earlier than postvocalic laterals in dialects of English (e.g. Lin & Demuth, 2015). The different timing of constrictions for dark (coda) and
clear (onset) laterals likely puts additional learning load on the child and results in a positional delay in acquisition.

The idea that context differentially influences the activation strength of constrictions associated with liquid articulation suggests that speech sound acquisition also depends on the selective modulation or inhibition of articulatory control signals during the sequential attainment of the sound targets. This suggestion is compatible with the following hypothesis derived from the long-domain perceptual results: the acquisition of adult-like /l/ articulation entails more clearly separating its articulation from the vocalic surround in a running speech context. More specifically, we suggest that children learn to inhibit early tongue body retraction for /l/-onset articulation to achieve, over developmental time, a more nearly simultaneous anterior and posterior constriction closer to the maximal constriction target. This suggestion is based on the finding that perceivers were able to distinguish /l/ from /d/ earlier in the gated speech stream for sentences produced by 5-year-old children compared to those produced by 8-year-old children and adults (see Figure 14).

More specifically, the long-domain perceptual results indicated earlier discrimination of /ɹ/ versus /d/ compared to /l/ versus /d/ in 8-year-old child and adult speech, but similarly early discrimination of /l/ versus /d/ and /ɹ/ versus /d/ in 5-year-old child speech. Recall that 5-year-old children only distinguished /l/ from /ɹ/ along F3, suggesting a contrast based on lip rounding for /ɹ/ (Fant, 1960; Lindblom & Sundberg, 1971), but also perhaps a contrast due to lower pharyngeal constriction for /ɹ/ than for /l/ (Al-Tamimi & Heselwood, 2011) - a suggestion compatible with other descriptions of rhotics as critically specified by a low pharyngeal constriction (Delattre & Freeman, 1968; Howson, 2018). Recall also that the different vowel context conditions allowed for a preliminary test of the effect of rounding on early /ɹ/ detection.
Although the results indicated an overall reduction in /ɹ/ versus /d/ accuracy in the /u/ environment, the rhotic was still detected above chance at the offset of /i/ in “be” of the frame sentence “They say it could be the ___-house.” Furthermore, /l/ versus /d/ accuracy was also lower in the /u/ environment, suggesting that the cue conflict was due to greater tongue retraction in the presence of an upcoming high back vowel than to lip rounding for that vowel. Taken together, the perceptual results therefore suggest that the long-domain effect of both rhotic and lateral articulation in the present study was due, at least in part, to anticipatory tongue retraction for a pharyngeal constriction. This interpretation echoes a similar interpretation of the long-domain effect of liquid articulation on resonances in American English (Hall et al., 2017).

The long-domain results also suggest that articulatory timing provides additional complexity that must be accounted for in the acquisition of liquids. We suggest that the pattern of results for 5-year-olds reveals that children begin with a more neutral timing pattern between liquids. Our results suggest 8-year-olds have matched their timing closer to adults due to observed delay in the onset of /l/. However, the onset of /ɹ/ for 8-year-olds was also delayed for that age group, which caused a mismatch between their timing and adults’ timing. It is possible that 8-year-olds inappropriately delay /ɹ/ as a result of having to delay /l/. In light of these results, we suggest delays in acquiring liquids extends past the age of 8-years-of-age and that multiple constriction locations are not the sole reason for the delay. Rather, children must learn the appropriate timing differences between segments in order to produce adult-like speech.

Overall, the results from the current study suggest stronger context effects on the articulation of liquids in children’s speech compared to adults’ speech. In so far as liquids are thought to have a high degree of coarticulatory resistance (Recasens & Pallarès, 1999), the present results suggest that this resistance takes time to master. The articulatory timing perspective adopted here
to study liquid acquisition suggests that resistance may be an artifact of learning to separate the postures associated with liquid articulation from the vocalic surround. More generally, the results from the current study demonstrate that an articulatory timing perspective on liquid acquisition adds information about what type of constriction children are likely to realize first, both developmentally and sequentially, when attempting doubly (and triply) articulated liquids in running speech.

Acknowledgements

This research was wholly supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) under grant R01HD087452 (PI: Redford). The content is solely the authors’ responsibility and does not necessarily reflect the views of NICHD.

References.


Figure 1. The mean accuracy with 95% confidence intervals for Age Group by Target (left) and Age Group by Vowel Context (right) on normalized F1 of the preceding schwa vowel.

Table 1. Summary of the F1 finding (standard deviation in brackets)

<table>
<thead>
<tr>
<th></th>
<th>5-year-olds</th>
<th>8-year-olds</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ə/</td>
<td>11.33 (1.34)</td>
<td>11.61 (1.21)</td>
<td>10.34 (1.11)</td>
</tr>
<tr>
<td>/d/</td>
<td>/l/</td>
<td>/t/</td>
<td>/s/</td>
</tr>
<tr>
<td>/ə/</td>
<td>11.60 (1.14)</td>
<td>11.22 (1.30)</td>
<td>10.79 (1.42)</td>
</tr>
<tr>
<td>/æ/</td>
<td>/æ/</td>
<td>/æ/</td>
<td>/æ/</td>
</tr>
<tr>
<td>/ə/</td>
<td>11.05 (1.34)</td>
<td>11.35 (1.31)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2. The mean accuracy with 95% confidence intervals for each Age Group by Target for normalized F2 of the preceding schwa vowel.

Figure 3. The mean accuracy with 95% confidence intervals for the effect of Age Group by Target (left) Target by Vowel Context (right) on normalized F2–F1 from the preceding schwa vowel.
Figure 4. Formant trajectories in adult speech for the different \(\partial V\) sequences in the front vowel (left) and back vowel (right) contexts. Dotted lines indicate 95% confidence intervals.

Figure 5. \(F_2-F_1\) trajectories in adult speech for the different \(\partial V\) sequences in the front vowel (left) and back vowel (right) contexts. Dotted lines indicate 95% confidence intervals.
Figure 6. Formant trajectories in 8-year-olds’ speech for the different əLV sequences in the front vowel (left) and back vowel (right) contexts. Dotted lines indicate 95% confidence intervals.

Figure 7. F2-F1 trajectories in 8-year-olds’ speech for the different əLV sequences in the front vowel (left) and back vowel (right) contexts. Dotted lines indicate 95% confidence intervals.
Figure 8. Formant trajectories in 5-year-olds’ speech for the different əLV sequences in the front vowel (left) and back vowel (right) contexts. Dotted lines indicate 95% confidence intervals.

Figure 9. F2-F1 trajectories in 5-year-olds’ speech for the different əLV sequences in the front vowel (left) and back vowel (right) contexts. Dotted lines indicate 95% confidence intervals.
Figure 10. The final frame from each of 8 gated AV speech stimuli derived from a child’s production of the sentence “They said it could be the lad-house.” Gates were placed at the segmental landmarks shown using acoustic and kinematic cues in the AV recording.
Figure 11. The final frame from each of 8 gated AV speech stimuli derived from a child’s production of the sentence “They said it could be the rad-house.” Gates were placed at the segmental landmarks shown using acoustic and kinematic cues in the AV recording.

Figure 12. The final frame from each of eight gated stimuli derived from a child’s production of the sentence “They said it could be the dad-house.” Gates were placed at the segmental landmarks shown using acoustic and kinematic cues in the AV recording.
Figure 13. The mean accuracy with 95% confidence intervals for the interaction between Age Group and Target.

Figure 14. The mean accuracy with 95% confidence intervals for both liquids in each of the Age Groups: 5-year-olds (left), 8-year-olds (center), and adults (right). The interaction with Age Group did not reach significance, but post-hoc tests confirm important Age Group differences on prediction accuracy.
Figure 15. The mean accuracy with 95% confidence intervals for /l/ (left) and /ɹ/ (right) for each Gate in each of the vowel contexts.