

The perceived clarity of children's speech varies as a function of their default articulation rate

Melissa A. Redford^{a)}

Department of Linguistics, University of Oregon, 1451 Onyx Street, Eugene, Oregon 97403-1290

(Received 7 December 2012; revised 7 March 2014; accepted 18 March 2014)

The current study investigated whether variation in children's default articulation rate might reflect individual differences in the development of articulatory timing control, which predicts a positive correlation between rate and perceived clarity (motor skills hypothesis), or whether such variation is better attributed to speech external factors, which predicts that faster rates result in poorer target attainment (undershoot hypothesis). Two different speech samples were obtained from 54 typically developing children (5;2 – 7;11). Six utterances were extracted from each sample and measured for articulation rate and segmental duration. Fourteen adult listeners rated the utterances for clarity (enunciation). Acoustic correlates of perceived clarity, pitch, and vowel quality were also measured. The findings were that age-dependent and individual differences in children's default articulation rates were due to segmental articulation and not to suprasegmental changes. The rating data indicated that utterances produced at faster rates were perceived as more clearly articulated than those produced at slower rates, regardless of a child's age. Vowel quality measures predicted perceived clarity independently of articulation rate. Overall, the results support the motor skills hypothesis: Faster default articulation rates emerge from better articulatory timing control.

© 2014 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4869820>]

PACS number(s): 43.70.Ep, 43.70.Bk, 43.70.Mn [LK]

Pages: 2952–2963

I. INTRODUCTION

Speaking rate refers to the rate at which units such as phones or syllables are produced in a given period of time. When calculated without pauses, the rate of phone or syllable production is referred to as “articulation rate.” Although articulation rate controls to some extent for the factors that contribute to pausing (e.g., planning and breathing), it nonetheless also varies with factors that are clearly external to the specifics of articulation and, more generally, of speech. These factors include, but are not limited to, the length of an utterance, phonological structure, the type of speech task used, a speaker's sex, and dialectal variation (Crystal and House, 1990; Jacewicz *et al.*, 2010).

In addition to sources of systematic variation, articulation rate also varies extensively across individuals within a single age group and speech community when speech materials are controlled. Whereas systematic variation is explained with reference to the manipulated variables, individual variation is often treated as random noise in the data; that is, we frequently average across individuals in order to investigate group effects. The goal of the current study was to explain variation in articulation rates at the level of the individual in the context of group effects. The specific focus was on individual differences in children's articulation rates since these differences are extensive enough to obscure the effect of age in smaller samples (see, e.g., Pindzola *et al.*, 1989; Walker and Archibald, 2006). The working hypothesis was that variation in children's articulation rates is not idiosyncratic, but instead it emerges either from continuous

differences in the development of the motor skills that underlie sequential articulation (i.e., articulatory timing) or from a speech external source, similar to the linguistic and social factors mentioned above, with consequences for children's attainment of articulatory targets.

A. Articulation rate and motor skill development

The sequential movements that constitute speech require precise control over the timing of articulatory movement into and out of patterns of articulatory coordination associated with different acoustic goals. Adopting the vocabulary of Articulatory Phonology (Browman and Goldstein, 1992), we might refer to the patterns of interarticulatory coordination as gestures and to the precisely timed movements into and out of gestures as a property of the plan. Assuming that both the gestures and their relative timing are acquired with words, lexical representations will provide the instructions for serial action in speech. So, where is the motor skill in all of this?

It turns out that it is one thing to have a plan and another to execute it with maximal speed and efficiency. Although children presumably acquire gestures and the appropriate (word-specific) phasing of gestures as they learn words, their segmental articulation remains considerably slower and more variable than adults' until early adolescence (Kent and Forner, 1980; Smith *et al.*, 1983; Lee *et al.*, 1999). In addition to being slow and variable, children's speech movements are also of larger relative amplitude than adults' movements (e.g., Green *et al.*, 2002; Cheng *et al.*, 2006). The larger amplitude speech movements also affect the rate of sequential articulation. Gestural overlap is reduced in children's speech compared to adult speech in both consonant-vowel (Serenio *et al.*, 1987; Katz *et al.*, 1991) and consonant-consonant

^{a)}Author to whom correspondence should be addressed. Electronic mail: redford@uoregon.edu

sequences (Gilbert and Purves, 1977; Cheng *et al.*, 2006).¹ Insofar as segmental and sequential articulation are attributed to articulatory timing control (i.e., intra- and inter-gestural coordination), the effect of age on articulation rate is probably best understood as an effect of motor skill development on rate.

B. Individual differences in articulation rate

In the adult literature, speaker-specific articulatory strategies are often documented in the context of inter-speaker rate differences (e.g., Ostry and Munhall, 1985; Adams *et al.*, 1993; Matthies *et al.*, 2001), but the different strategies are typically treated as idiosyncratic rather than as emergent from differences in an underlying common skill set or from an interaction with this skill set. Work by Tsao and Weismer (1997; Tsao *et al.*, 2006a) provides an exception to this generalization. These authors, adopting a theory of extrinsic timing control, propose that a neural clock (i.e., timekeeping function), common to all motor action not just speech action (Kozhevnikov and Chistovich, 1965; Wing and Kristofferson, 1973), controls articulation rate. Tsao *et al.* (2006a) hypothesize that inter-speaker variation in articulation rate emerges from individual differences in the biological setting of this clock. Specifically, they argue that the clock runs fast in speakers with habitually fast articulation rates and runs slow in those with habitually slow articulation rates. The term “habitual” refers here to the default rate that people use absent any instruction to speak more or less slowly or quickly.

Although Tsao and Weismer’s (1997; Tsao *et al.*, 2006a) results are consistent with individual variation in the default setting of an extrinsic timing mechanism, alternative explanations are not ruled out. For example, it is also possible that default articulation rates emerge from individual differences in speech motor skills; that is, from factors intrinsic to speech production. This proposal is similar to the idea that a speech motor skills continuum underlies individual differences in speech fluency (Peters *et al.*, 2000). With regard to rate variation, the hypothesis is that people with superior skills speak faster because they can do so without adversely affecting speech quality. Insofar as a default rate emerges from the act of articulation, the hypothesis is consistent with a theory of intrinsic timing control.

The hypothesis that individual differences in speech motor skills explain individual differences in articulation rate is especially attractive in the context of a developing system. As noted, the idea that age-dependent increases in the articulation rate reflect the development of articulatory timing control is widely accepted (e.g., Kent and Forner, 1980; Smith *et al.*, 1983; Lee *et al.*, 1999). Thus, rate variation in children’s speech may merely reflect normal variation in the development of relevant speech motor skills: Children with faster default articulation rates may have superior articulatory timing control than their peers with slower default articulation rates. Since timing control is key to interarticulatory coordination, better developed control will have consequences for target attainment in that individual gestures will be realized more precisely (i.e., with less variability). By

hypothesis, then, children who speak more quickly than their peers are also likely to be perceived as speaking more clearly than their peers.

The alternative hypothesis follows from a theory of extrinsic timing control: Individual variation in children’s articulation rates reflects pressure from a biological clock, the setting of which may vary across individuals (Tsao and Weismer, 1997) or be driven by some other well-established individual difference factor such as personality (e.g., extraversion-introversion; Dewaele and Furnham, 2000). Whether the source for articulation rate differences is the biological setting of an extrinsic timing mechanism, personality, or some other individual difference factor, if rate control is external to speech, then faster rates will negatively impact articulatory precision resulting in poorer speech clarity—assuming, of course, that children of the same age who speak at different default rates have equally developed speech motor skills. The prediction of an inverse relationship between rate and clarity follows straightforwardly from a model of duration-dependent undershoot (Lindblom, 1963). In this model, an articulatory configuration is not fully achieved when the command to execute the next configuration is issued prior to the realization of the current one, as in fast speech or when the acoustic target that is being attempted is particularly short (e.g., unstressed vowels). The model has also been used to explain the differences in articulation associated with so-called clear and casual speech (Lindblom, 1990).

C. Articulation rate and perceived clarity

The two hypotheses outlined above predict a correlation between articulation rate and speech clarity in children’s speech, albeit in opposite directions. In both cases, the mediating link between rate and clarity is articulatory precision, defined here in terms of target attainment. A relationship between articulatory precision and perceived speech clarity is supported in the literature on unintentional clear speech.

Although many studies on clear speech rely on some kind of instruction or within-subject manipulation in order to affect speaking style (see Lam *et al.*, 2012), a number of studies have also investigated individual differences in the perceived clarity of speech that is produced absent specific instruction (Bond and Moore, 1994; Bradlow *et al.*, 1996; Hazan and Markham, 2004). For example, Hazan and Markham (2004) asked 45 British English speakers (adults and children) to produce a set of target words in a frame sentence (“The next three words are __, __, __ it”) by having them read the sentences off a computer monitor. No other special instruction was given. Different groups of adult and child listeners were then asked to identify the words. The number of identification errors was recorded. The findings were that all listeners made more errors when listening to some speakers than when listening to others. The number of identification errors were then used to select the six most intelligible (“intrinsically clear”) and six least intelligible speakers, whose speech was rated along a number of dimensions by university students with phonetics training. The ratings revealed that (perceived) articulatory precision and

voice dynamics were important predictors of talker intelligibility. Hazan and Markham followed up on these results with acoustic measurements. The measurements showed that the intrinsically clear speech of the most intelligible talkers shared many features with deliberately clear speech; for example, more hyper-articulated vowels and higher, more variable pitch patterns.

Hazan and Markham (2004) were specific about the link between articulatory precision and perceived speech clarity in typical populations, but this link is also assumed in the clinical literature. In particular, intelligibility is a measure that is frequently used to assess the severity of involvement to speech production in persons with speech motor disorders (see, e.g., Kent *et al.*, 1989). Other researchers have directly investigated whether the acoustic correlates of perceived clarity, especially measures of vowel quality, predict intelligibility in clinical populations. For example, Turner *et al.* (1995) found that the size of the $F1 \times F2$ vowel space predicted 45% of the variance in intelligibility rating of speech produced (at different rates) by individuals with amyotrophic lateral sclerosis.

D. Current Study

The current study investigated the relationship between individual differences in children's default (i.e., unstructured) articulation rates and perceived speech clarity to test between our alternative hypotheses regarding individual differences in children's articulation rate. A positive relationship between rate and clarity will be taken to support the hypothesis that individual differences in children's articulation rates emerge from differences in underlying speech motor skills with faster rates emergent from superior, more mature articulatory timing control. A negative relationship will be taken to support the hypothesis that default articulation rates are motivated by one or more factors external to speech production and children therefore achieve faster rates at the expense of target attainment, as predicted by a model of duration-dependent undershoot. Pitch and vowel formant measures were used to explore the relationship between articulation rate and these acoustic correlates of perceived clarity as well as to ensure that perceived clarity was in fact associated with aspects of articulation.

II. METHODS

A. Participants

Fifty-four native American-English speaking children (30 female, 24 male) provided the speech sample used in the current study. All spoke the west coast variety of American English common to Eugene, OR. All were typically developing, as determined by parental report, and all were in school at the time of the study. The children ranged in age from 5;2 yr (i.e., 62 months) to 7;9 yr (i.e., 93 months) in a continuous fashion: 18 (7 female) were 5 yr old ($M = 66$ months, $SD = 2.6$ months); 18 (14 females) were 6 yr old ($M = 76$ months; $SD = 3.5$ months); and 18 (9 females) were 7 yr old ($M = 88$ months, $SD = 2.63$ months).

Fourteen undergraduate students from introductory psychology and linguistics classes at the University of Oregon

rated the children's speech in return for course credit. All 14 students were also native American-English speakers, and all had resided on the west coast of the United States for most of their lives.

B. Speech samples

Since rate is known to vary with task type (e.g., Jacewicz *et al.*, 2010), it was decided that evidence for one of the two primary hypotheses would be stronger if the same basic pattern was found to hold across tasks. Accordingly, speech samples were drawn from a narrative task and a sentence repetition task. Speech in both tasks was digitally recorded onto a Marantz PMD660 (with a sampling rate of 44 100 Hz) using a Shure ULXS4 standard wireless receiver and a lavalier microphone, which was attached to a baseball hat that the child wore.

The narrative task was designed to elicit structured spontaneous speech. Children and an accompanying caregiver were asked to select from among four wordless picture books (the frog stories illustrated by Mercer Mayer). Once each had chosen a book, they familiarized themselves with the story, and then each told the story to the other. The tellings alternated from caregiver to child or child to caregiver, then back again for a second telling. The narrative speech had the advantage of naturalness but had the disadvantage of uncontrolled phonological material.

The sentence repetition task eschewed naturalness, but it provided control over phonological material. The repeated sentences were from the Recalling Sentences subtest from the Clinical Evaluation of Language Fundamental (CELF-4) test (Semel *et al.*, 2003). This test was administered in the usual way: Children heard a sentence, then repeated it back to the tester, who kept track of any mistakes. Since the test was being used to elicit children's speech, two modifications to the normal procedure were made. First, regardless of age, all children started with test item 1 (normal assessment procedure is to start with later items for older children). Second, the tester played pre-recorded versions of the sentences rather than modeling the sentences for the child. The recorded sentences were produced by a female speech-language pathologist who spoke very clearly and fluently, with no pauses or other prosodic breaks within the sentences of interest.

The spontaneous speech sample consisted of six prosodically complete, fluent, pause-delimited utterances selected from the second storytelling in the narrative task. Prosodic completeness and fluency was defined by coherent intonational contours, including the presence of clear phrase-initial and phrase-final boundary tones that marked a completed thought. Phrase-final boundaries were also marked by perceptible lengthening of the final syllable. Such judgments are not difficult to make, even for the linguistically naive, who would render the same with appropriate punctuation including an initial capital letter and a final period or question mark, and with commas in cases where a prosodic break within an utterance is tonally marked to indicate thought continuation (see Redford *et al.*, 2012, for further details). In addition to prosodic completeness, utterances were selected from the middle

of the story. The goal was to avoid the stereotyped language associated with the beginnings and endings of stories. Also, utterances were only selected if they were roughly 2 s in duration. The goal was to provide some control over utterance length. On occasion it was necessary to include sentences with a small internal prosodic break so as to find 6 utterances that conformed to the duration criterion. That is, in some instances, instead of a single intonational phrase, some of the utterances were produced with two intonational phrases, usually delimited by a short utterance-internal pause (more on this below).

To verify that the selected samples were representative of the child's default articulation rate during the narrative task, we also extracted a continuous 30 s sample from the middle of the children's stories, and correlated the articulation rate calculated from this interval with the mean rate calculated from the selections. The correlation was robust and statistically significant [$r(52) = 0.708$, $p < 0.001$], but not perfect. The imperfect correlation between the articulation rate in the selected utterances and the rate in the continuous samples is because the length of an utterance affects its rate (see, e.g., Crystal and House, 1990), and there was more control over utterance length in the selected utterances than in the continuous sample.

To rule out the possibility that a child's articulation rate may have been influenced by the caregiver's rate, we selected a continuous 30 s speech sample from caregiver stories and correlated the articulation rate calculated from this interval with that obtained from the child's 30 s sample. This correlation was not significant [$r(51) = 0.089$; $p = 0.535$], which suggests that children's rate was not imitative.

The first six sentences of the Recalling Sentences subtest were selected to represent articulation rate in the repetition task. These sentences were selected because they were the easiest of the set and so the most likely to be repeated back fluently and correctly. Those researchers familiar with the test know that sentence number 6 is the typical starting point for children aged 9 to 13 yr. It is worth noting, though, that a 5-yr-old child who does not get beyond this sentence will receive a raw score of no more than 18 points on the test, which would put even the youngest 5-yr-old at the 16th percentile, 1 standard deviation below the population mean. The raw Recalling Sentences scores for the youngest children in our study (62 to 66 months, $N = 11$) ranged from 30 to 52, with a mean of 42. The very youngest child had a raw score of 43. Older children had higher scores. In brief, all children were average to above average in their language abilities as measured by the Recalling Sentences subtest of the CELF-4, and almost all produced the first six sentences in Recalling Sentences without any error, though not all children produced all sentences with maximal fluency (see below).

The selection procedure resulted in 648 utterances for measurement. Twenty-seven of the 324 utterances from the repetition task were excluded from analysis due to incorrect repetition and/or a disfluency that involved either the truncation of a word and a repetition restart or simply a pause and repetition restart. Of the remaining 297 utterances, 15 were correctly rendered, but produced with fairly long (>250 ms) pauses. These were nonetheless included for measurement

because we did not want to further diminish the sample size and because pausing is a natural part of speech and so could not be considered in the same class as the outright errors that disqualified the other sentences from the repetition sample.

In total, an average of 28 s ($SD = 324$ ms) of speech was examined for each child. Table I presents average utterance durations and utterance lengths for each age group and task, along with the number of utterances included for analysis.

C. Rating task

The 621 utterances were amplitude normalized to 72 dB and blocked by task. Within each block, utterances were presented in random order over headphones to adult listeners. Blocks were counterbalanced across listeners. The listeners, who sat in front of a computer screen, were asked to rate each utterance for clarity on a 7-point Likert scale. The scale was presented as a series of boxes labeled with the numbers 1 through 7 and ordered sequentially from left to right. Above the boxes, the scale was anchored on the left with the instruction "1 = Totally Unclear. Hard to understand." and on the right with the instruction "7 = Extremely Clear. Very well enunciated." Listeners heard an utterance, rated it for clarity by clicking on one of the 7 boxes, and then clicked "OK" to advance to the next screen and listen to the next utterance. Each listener heard each utterance only one time. The rating task took approximately 45 min for listeners to complete. Listeners were offered a break between each block of sentences.

D. Acoustic measurement

1. Rate and duration measurements

Each utterance was displayed as an oscillogram and spectrogram in Praat (Boersma and Weenink, 2011) and an associated Textgrid was generated. The utterance was orthographically transcribed on one tier of the Textgrid, word boundaries established on another tier, and the number of syllables counted. A third tier was used to indicate intervals associated with the production of consonants, vowels, and pauses when these occurred. A final tier of only vowels was inserted for utterances drawn from the repetition task to facilitate the spectral analysis of vowels (more on this below).

Segmentation of the acoustic waveform into consonants and vowels used auditory judgment and visual cues. For vowels adjacent to obstruents, the most relevant cue was the presence of higher formants that continued with an unbroken

TABLE I. Mean utterance durations in seconds, length in syllables, and number of utterances for each age group and task. Standard deviations are given in parentheses.

Task	Utterance	5 years old	6 years old	7 years old
Narrative	Duration	2.29 (0.45)	2.37 (0.52)	2.29 (0.43)
	Length	8.2 (2.1)	8.6 (2.5)	8.8 (2.3)
	Number	108	108	108
Repetition	Duration	2.69 (0.83)	2.43 (0.48)	2.20 (0.42)
	Length	8.8 (1.7)	8.9 (1.7)	8.9 (1.7)
	Number	96	95	106

trajectory. For vowels adjacent to nasals, the primary cue was an abrupt spectral change, including an anti-resonance in the mid-frequencies (1 to 1.5 kHz). For vowels adjacent to liquids or glides, the primary cue used was a decrease in energy, especially at higher frequencies, associated with greater vocal tract obstruction. Other spectral changes (e.g., F2 and F3 transitions) were secondary to this primary cue, and no boundary was marked absent an energy decrease at higher frequencies. Other vowel/consonant boundary decisions followed standard practices in the literature; for example, diphthongs were considered to represent a single vocalic interval, and stop releases and aspiration were considered to be part of the consonantal interval. As for silent intervals, these were delimited as pauses if they were perceived as a prosodic break by the coder. If silence occurred adjacent to a fricative, sonorant, or vocalic segment, then the pause interval was taken to be the entire interval of silence. If a silent interval occurred with a vowel-initial stop consonant and/or an unreleased final stop, then it was only delimited as a pause if (1) it was heard as a pause, and (2) the interval was longer than necessary to accommodate stop closure, which was arbitrarily but consistently defined as 150+ ms. Our procedure for identifying pauses has been used successfully in other published work (Redford, 2013) and the 150 ms interval is within the range of consonant durations (especially, stop closure + release durations) observed in school-aged children's running speech (see Fig. 1).

Two trained research assistants and the author completed the segmentations. The author checked all utterances that she and the others had segmented to ensure that a consistent procedure was followed. Figure 1 illustrates the segmentation of obstruents, sonorants, vowels, and pauses based on amplitude and spectral cues. Recall, that auditory judgments were also used. The spectrogram shows a portion of a 5-yr-old boy's spontaneous production "then the other frog pushed him away," which was selected for measurement because (a) it was produced under a coherent contour that was marked for completion at the final utterance boundary, and (b) it had a total duration that was close to 2 s (precisely, 2.316 s). Note that the tier with the full utterance transcription has been removed and that the others tiers were reordered in order to provide the reader with the best visual representation of the segmentation procedure.

The segmentation procedure yielded a total of 13 288 intervals for consideration, 48% of which were from the spontaneous speech samples. Durations were automatically extracted over the intervals, and several temporal measures were calculated. First, total utterance duration was calculated as the sum of total vowel and consonant duration (i.e., total duration excluding pause duration). This total was divided by the number of syllables in the utterance to yield a measure of articulation rate for each speaker and each task, which was then expressed in syllables per second. The other temporal measures calculated for each speaker and task were mean vowel duration, mean consonant duration, proportion of vowel duration to total duration, and mean normalized measures of variance in vowel and consonant duration (i.e., the coefficient of variation). Proportion of vowel duration and the measures of temporal variance across vowels and consonants are influenced by suprasegmental patterns, including syllable structure and language rhythm (e.g., Ramus *et al.*, 1999). These measures were thus included to investigate possible suprasegmental effects on rate. Such effects are known to vary within speaker in adults; specifically, studies of intra-speaker rate control in adults show that fast speech is achieved primarily through vowel, and especially stressed vowel, compression (Gay, 1978, 1981). This is likely because, by default, adults maximally reduce unstressed vowels and overlap consonants in consonant sequences, which means that only stressed vowels can be reduced further at faster rates. Since younger school-age children do not reduce unstressed vowels or overlap consonants in sequences to the same degree as adults (Kehoe *et al.*, 1995; Pollock *et al.*, 1993), the prediction was that differences in rate would be explained by both consonant and vowel durations rather than by the smaller variance in vowel durations that would follow from a strategy of stressed vowel compression.

2. Frequency measurements

Acoustic correlates of speech clarity were also measured to validate the use of perceived clarity as a measure of speech motor skill, and to determine whether articulation rate also varies systematically with these correlates. Pitch and vowel formants were measured in utterances selected

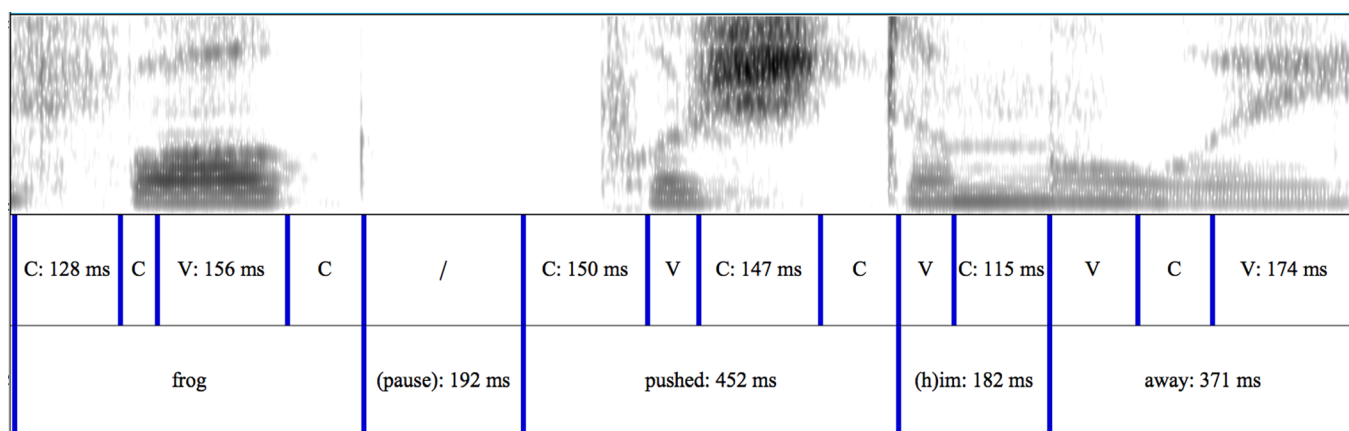


FIG. 1. (Color online) An example segmentation of the first portion of an utterance produced by a 5-yr-old boy in the narrative task.

from the repetition task where phonological material was controlled across speakers. Measurements were restricted to the prominent syllables in the sentences. Prominent syllables included all lexically stressed syllables in all content words in the utterance (3 or 4 syllables) as well as the vowels in pronouns since these were also prominent in the children’s speech. F0, F1, and F2 frequencies were measured at the midpoint of all monophthongal vowels in the prominent syllables, and at the midpoints of the steady-state regions for each of the two vowel qualities in diphthongal vowels. The total number of intervals identified for measurement was 1350, which includes the two measurements taken on diphthongs. The average number of intervals measured per child was 25 ($SD = 4$), with variation due either to a sentence being excluded from the study or to diphthongs being realized as monophthongs, which happened especially for the mid-front vowel /e/ in words like “cage” and “base.” Figure 2 illustrates vowel segmentation and the pitch and formant tracks associated with the first 2 prominent vowels in the phrase “The boy fell and hurt himself.”

The frequency measurements were extracted based on the automatic pitch and LPC formant tracking algorithms in Praat. The autocorrelation method was adopted for pitch tracking, and the range set for between 75 and 600 Hz; the window size to 0.01 s. Very low, very high, and “undefined” F0 values were hand-checked, and corrected by calculating cycles per second from the oscillogram if possible and necessary. Of the 1350 vowel intervals, 13 were largely voiceless, and so F0 could not be calculated.

As for the vowel quality measurements, formant tracks were overlaid on a wideband spectrogram that displayed frequency up to 5000 Hz. A Gaussian analysis window was used as was the default bandwidth setting of 250 Hz. By default, five formants were tracked between the range of 0 to 6000 Hz. However, this setting provided only a starting point for measuring formant frequencies. Every formant track was visually inspected and the settings adjusted if the tracks were off, and if this failed, midpoint frequency was measured by hand. Values were only extracted for the first and second formants because these carry most of the salient information for vowel quality, and because the third formant was less reliably tracked or visible on the spectrogram (see, e.g., the

initial portion of the vowel in “boy” in Fig. 2). F1 and F2 frequency measurements were taken on 11 of the 13 voiceless vowels.

Once extracted, the values were converted from Hertz to Bark using the formula proposed in Traunmüller (1990). Formant values were then normalized using the vowel intrinsic method proposed in Syrdal and Gopal (1986), which relies on bark transformed F0, F1, and F2; specifically, $F1 - F0$ for information regarding degree of vocal tract stricture and $F2 - F1$ for information regarding tongue advancement. Many other vowel intrinsic normalization methods rely on F3 (Vorperian and Kent, 2007) and so were eschewed here. Analyses used the speaker mean and standard deviation of the normalized values as predictors, as well as the speaker mean and standard deviation of F0 in bark. The final predictor was a measure of $F1 \times F2$ dispersion, also calculated for each speaker. This measure was simply the mean Euclidean distance of each vowel from the vowel centroid, defined as the mean of $F1 - F0$ and $F2 - F1$ of all vowels from all prominent syllables.

III. RESULTS

A. Articulation rate

Children’s articulation rates ranged from a low of 2.69 syllables per second in the narrative task to a high of 4.99 syllables per second, with a mean rate of 3.72 syllables per second ($SD = 0.52$). The range of articulation rates was narrower in the repetition task, with a low of 2.80 syllables per second to a high of 4.62 syllables per second. The mean rate in this task was 3.70 syllables per second ($SD = 0.42$). It is perhaps worth noting that adult articulation rates, measured from a continuous 30 s sample of spontaneous speech elicited in the narrative task, ranged from a low of 2.35 syllables per second to a high of 5.93 syllables per second, with a mean of 4.10 syllables per second ($SD = 0.69$). The articulation rate of the pre-recorded sentences from the repetition task was 4.03 syllables per second ($SD = 0.43$), very close to the overall adult mean in the narrative task. These descriptive data indicate that children’s articulation rates were slower on average than adults’ articulation rates, but that

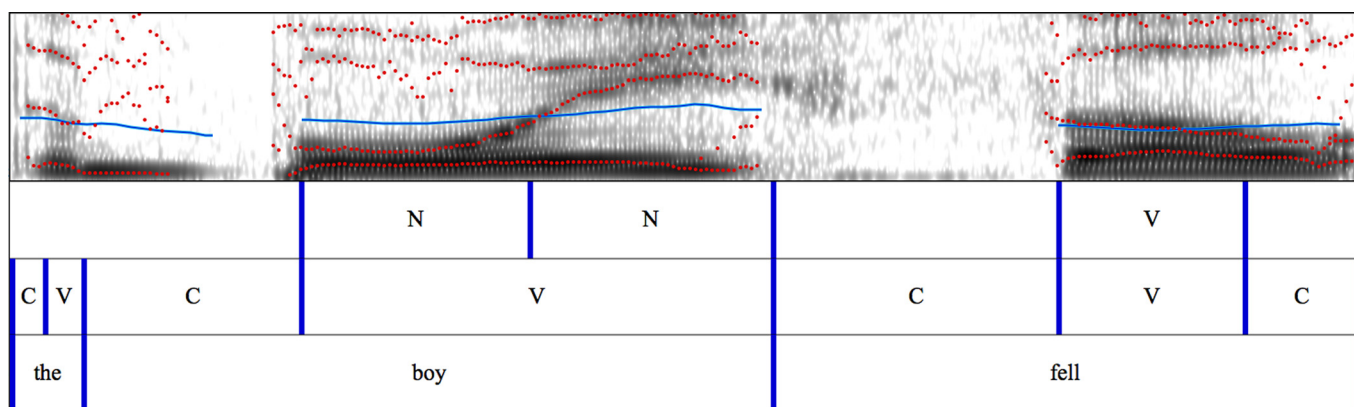


FIG. 2. (Color online) An example segmentation used for vowel measurement, and the pitch and formant tracks on which they were based. The range for pitch tracking was set from 75 to 600 Hz. The range for the default tracking of five formants was set from 0 to 6000 Hz.

many of the children also achieved adult-like articulation rates.

The first set of analyses investigated the extent to which articulation rate varied systematically with any of the fixed factors in the study. Generalized linear mixed effect modeling was used to test for the within-subject effect of task (narrative vs repetition) and the between-subject effects of sex (M vs F) and age (in months) on children's mean rate of articulation. Speaker and utterance length were treated as random effects. The results indicated significant effects of task [$F(1, 100) = 9.93, p = 0.002$] and age [$F(1, 100) = 9.47, p = 0.003$] as well as a significant interaction between these two factors [$F(1, 100) = 8.58, p = 0.004$]. A follow-up paired t -test indicated that articulation rates in the two tasks were positively correlated [$r(54) = 0.34, p = 0.011$], and that the mean rates were not in fact different [narrative rate, $M = 3.72$ syll/s ($SD = 0.52$) versus repetition rate, $M = 3.70$ syll/s ($SD = 0.42$); $t(53) = 0.34, p = 0.737$]. The contradiction between the overall model and the t -test suggests an important role for random effects in the model. When the data were split by task, and the overall analysis rerun, the effect of age held only for speech elicited in the repetition task [$F(1, 52) = 18.96, p < 0.001$]. Overall, these results underscore the point that articulation rates are highly variable across individual children as shown in Fig. 3.

The next analysis was aimed at understanding the different temporal patterns that give rise to individual differences in articulation rates. The prediction was that segmental durations would provide the best predictors of rate, following the hypothesis that the rate is tied to the development of speech motor skills in children; but suprasegmental influences were also examined since adult strategies for increasing rate involve disproportionate compression of vowels and especially of stressed vowels (Gay, 1978, 1981). Specifically, the following variables were used to predict articulation rate in a linear regression model: Mean vowel and consonant durations, the proportion of vowel to total utterance duration, and the coefficient of variation for vowel and consonant duration.

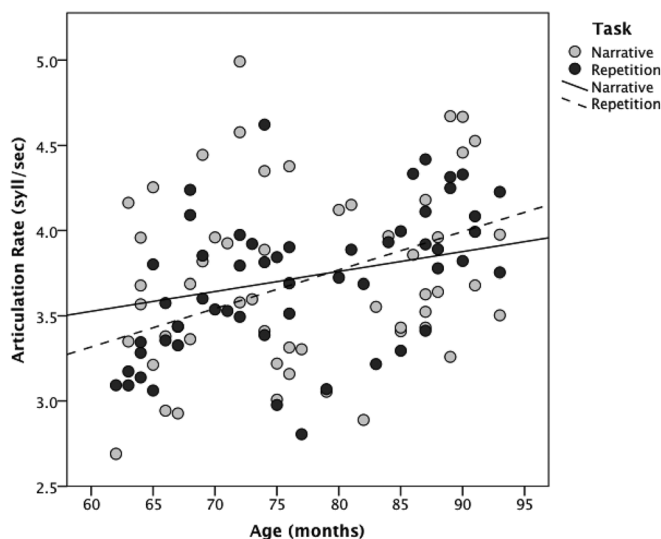


FIG. 3. Mean articulation rates for each speaker in each task.

Age in months and utterance length were entered as control predictor variables.

First, inter-correlations between the predictors were examined. The largest correlation was between mean vowel duration and mean consonant duration in the repetition task [$r(54) = 0.75, p < 0.001$]. Collinearity was a problem in this instance as the variance inflation factor (VIF) was above 10 (Neter *et al.*, 1989, p. 409). As mean vowel duration was also strongly correlated with mean consonant duration in the narrative task [$r(54) = 0.51, p < 0.001$] and with proportion of vowel duration in both tasks [narrative: $r(54) = 0.46, p = 0.001$; repetition: $r(54) = 0.66, p < 0.001$], the problem of collinearity was solved by dropping mean vowel duration as a predictor of articulation rate. A stepwise procedure with a removal criterion based on the probability of F greater than 0.06 was then used to reduce the remaining predictor variables (including the controls) in the regression model to the smallest set of significant variables.

The results, shown in Table II, were that mean consonant duration and proportion of vowel duration best predicted articulation rate across both tasks. The control variable, utterance length, was also a significant predictor of rate. Together the variables accounted for 78% of the overall variance (i.e., raw R^2) in the narrative task and 91% in the repetition task.

As indicated by the coefficients presented in Table II, mean consonant duration and proportion of vowel duration were both inversely related to articulation rate in the regression model, but consonant duration was more strongly so than proportion of vowel duration. By itself, consonant duration accounted for 59% of the rate variance in the narrative task and 82% in the repetition task (Fig. 4). By contrast, proportion of vowel duration did not account for any of the variance in the narrative task by itself, and only 11% of the variance in the repetition task.

The relative strength of the different temporal predictors might indicate that individual differences in articulation rate depend more on differences in the articulation of consonants than in the articulation of vowels. However, this possibility is contradicted by the robust correlation between consonant and vowel duration and by the strength of vowel duration as a sole predictor of articulation rate: Vowel duration

TABLE II. Results are shown for the best-fit models of children's articulation rates.^a

Model: Narrative, Adjusted $R^2 = 0.76$			
	β	t	p
(Constant)		13.89	0.000
Utterance Length	0.48	6.16	0.000
Mean Consonant Duration	-1.06	-7.23	0.000
Proportion of Vowel Duration	0.64	-2.20	0.033
Model: Repetition, Adjusted $R^2 = 0.90$			
(Constant)		13.21	0.000
Utterance Length	0.05	2.24	0.030
Mean Consonant Duration	-0.88	-20.05	0.000
Proportion of Vowel Duration	0.77	-6.20	0.000

^aDependent variable: Articulation rate for each speaker averaged across utterances.

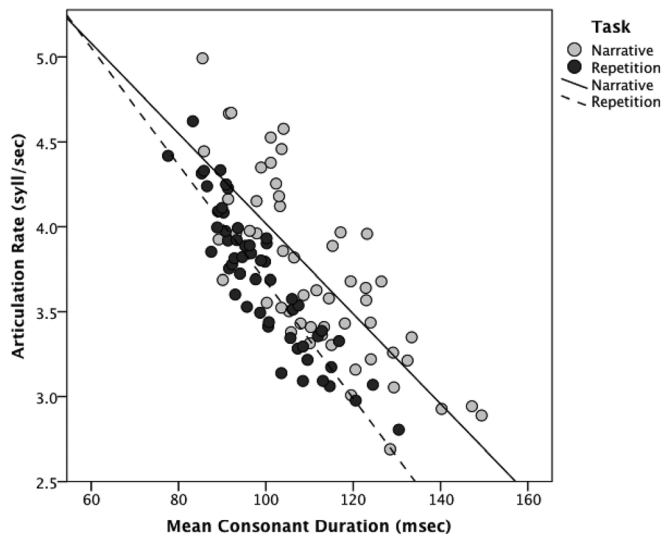


FIG. 4. Mean articulation rates for each speaker in each task plotted against mean consonant duration.

accounted for 61% of the rate variance in the narrative task and 82% in the repetition task. When vowel duration is substituted for proportion of vowel duration in the regression model collinearity is no longer a problem, and the fit of the best model is improved over the model with consonant duration and proportion of vowel duration. In particular, consonant duration, vowel duration, and utterance length combine to account for 86% of the variance in the narrative task and 94% in the repetition task.

In summary, articulation rate varied predictably with the independent factors of age and task, but substantial individual variation in rate was nonetheless observed. The variation in the articulation rate was equally well accounted for by consonant and vowel durations: As sole predictors, consonant duration accounted for 59% and 82% of the variance and vowel duration for 61% and 82% in the narrative and repetition tasks, respectively. Thus, unlike in adult studies of rate control where faster articulation rates are associated with a disproportionate reduction in vowel duration, faster rates in children's speech were due to shorter consonant and vowel durations.

B. Perceived clarity and articulation rate

We turn now to the question of whether faster articulation rates were associated with better or worse articulatory precision. These alternatives were tested using the perceived clarity of children's speech as a measure of articulatory precision. First, clarity ratings were standardized (z-scores) within listener, and a two-way random effects intraclass correlation was conducted to estimate agreement on the perceived clarity of utterances from the narrative and repetition tasks. The analysis indicated moderate absolute agreement for individual listeners [$ICC(2, 1) = 0.51$], and very high mean agreement [$ICC(2, 10) = 0.94$]. Given this high mean agreement between listeners and our interest in speaker differences, analyses on the perceived clarity of children's speech were conducted using the mean rating for each speaker.

TABLE III. Results are shown for the model of the perceived clarity of children's speech.^a

Model: Clarity, Adjusted $R^2 = 0.47$	β	t	p
(Constant)		-2.80	0.007
Age in months	0.50	4.45	0.000
Utterance length	-0.33	-2.60	0.012
Articulation rate	0.44	3.25	0.002

^aDependent variable: Clarity ratings for each speaker averaged across utterances, listeners, and tasks.

Next, we investigated whether perceived clarity varied systematically with any of the independent factors in the study. Once again, a generalized linear mixed effects model was used to test for the within-subject effect of task (narrative vs repetition) and the between-subject effects of sex (M vs F) and age (in months) on the mean perceived clarity of children's speech. Speaker and utterance length were again treated as random effects. The analysis indicated a significant effect of age [$F(1, 99) = 34.52, p < 0.001$], but no other significant main effects or interactions.

The next analysis investigated the relationship between perceived clarity and articulation rate using linear regression. Given the absence of an effect of task on perceived clarity, and the weakness of this effect on articulation rate, the analysis used values that were averaged across sentences and tasks. In addition to articulation rate, the regression model included age in months and mean utterance length as control predictor variables. The results, shown in Table III, indicate a positive relationship between clarity and rate, such that perceived clarity was higher at faster rates of articulation (Fig. 5). This result is consistent with the hypothesis that individual differences in children's default articulation rates emerge from differences in the relevant underlying speech motor skills.

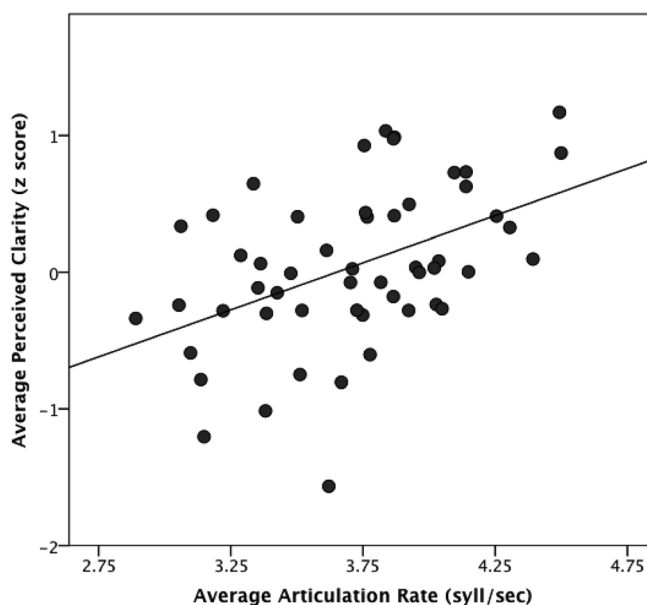


FIG. 5. Perceived clarity of children's utterances as a function of mean articulation rate.

To test the assumption that perceived clarity in the present study did in fact index articulation, other known correlates of clarity were incorporated into the regression model. The added predictor variables were the mean and standard deviation of F0 (in bark), normalized F1 ($F1 - F0$), and normalized $F2 - F1$ measures for each speaker as well as the measure of vowel dispersion (mean Euclidean distance from vowel centroid).

First, inter-correlations between the spectral measures were examined. The normalized F1 measure was highly correlated with mean $F2 - F1$ [$r(54) = -0.72, p < 0.001$], but collinearity was not a problem in this instance ($VIF < 10$). Collinearity was a problem for the standard deviation of $F2 - F1$ and the measure of dispersion ($VIF > 10$). These two measures were nearly perfectly correlated [$r(54) = -0.91, p < 0.001$]. Since the measure of dispersion was also highly correlated with the standard deviation of F1 [$r(54) = 0.78, p < 0.001$], albeit in the opposite direction, the decision was made to exclude this variable from the analyses. As before, a stepwise procedure with a removal criterion based on the probability of F greater than 0.06 was used to reduce the remaining predictors variables (including the controls) to the smallest set of significant variables.

The resulting best-fit model accounted for 66% of the variance in perceived clarity. Table IV presents the complete results from this model. These indicate that clarity was associated with a higher mean F1, more variability in F1, and a greater difference between F1 and F2. Thus, as in previous studies of intrinsically clear speech, vowel articulation significantly influenced ratings. Here, higher perceived clarity (good enunciation) was related to more open articulations, more differentiation between high and low vowels, and greater differentiation between F1 and F2 (relevant for front vowels). The results also suggest that articulation rate was independent of these other measures in that it remained a significant predictor in the model.

A follow up analysis directly investigated the correlation between rate and the different vowel measures. This analysis confirmed the independence between rate and the spectral measures; rate correlated only with mean F0 [$r(54) = -0.30, p = 0.026$], and this was presumably due to the relationship between age and F0 [$r(54) = -0.37, p = 0.007$].

The data shown in Fig. 6 also make the point that rate and vowel articulation were independent. The figure shows

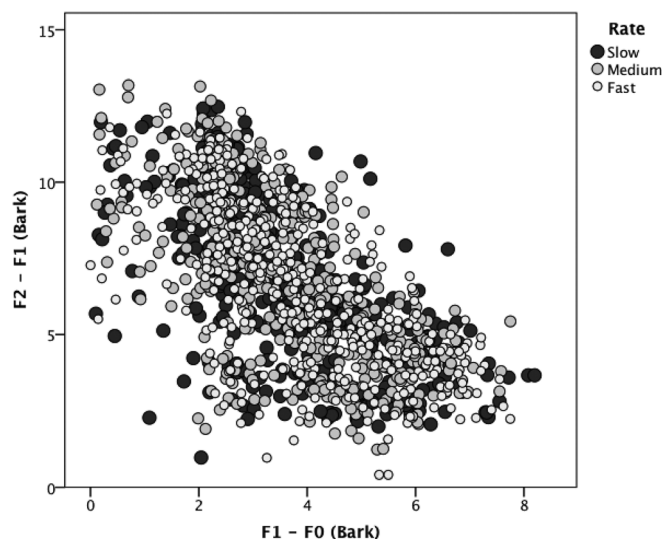


FIG. 6. The dispersion of vowels measured as a function of categorically defined articulation rates.

vowel production as a function of rate expressed in terciles. The substantial overlap in the distribution of vowels across these groups indicates that fast, medium, and slow child speakers all attained a range of vowel targets. The overlap also suggests that the space defined by the $F1 - F0$ and $F2 - F1$ extrema did not vary with rate. This suggestion is supported by the absence of a significant effect of categorically defined rates (slow, medium, fast) on our measure of vowel dispersion, tested in an analysis of variance (ANOVA) with age in months entered as a covariate.

To determine whether perhaps there were group differences in overall target attainment, the standard deviation of the dispersion measure was computed for the speakers in each tercile of articulation rate. There were no obvious differences between the groups: Mean dispersion in the group with the slowest articulation rates was 2.63 bark ($SD = 0.37$); mean dispersion in the group with medium rates was 2.73 bark ($SD = 0.37$); mean dispersion in the group with the fastest rates was 2.72 bark ($SD = 0.35$).

In sum, the perceived clarity of children's speech varied systematically with articulation rate. Acoustic correlates of vowel production also predicted perceived clarity, consistent with the assumption that clarity indexes articulation. The spectral correlates of clarity did not vary systematically with articulation rate: On average, vowels were similarly dispersed in the $F1 \times F2$ space regardless of whether the child spoke fast or slow by default. Together, the results are consistent with the idea that individual differences in articulation rate can be explained with reference to articulatory timing control.

IV. DISCUSSION

The current study examined individual differences in articulation rate in the context of a developing system. The goal was to understand whether variation in children's default rates of articulation might reflect individual differences in the development of articulatory timing control or whether rate variation might be better attributed to a speech

TABLE IV. Results are shown for the best-fit model of perceived speaker clarity in the repetition task.^a

Model: Clarity (Repetition Task), Adjusted $R^2 = 0.61$	β	t	p
(Constant)		-5.31	0.000
Age in months	0.42	3.78	0.000
Articulation rate	0.27	2.47	0.017
Mean $F1 - F0$	0.47	3.54	0.001
SD $F1 - F0$	0.24	2.53	0.015
Mean $F2 - F1$	0.27	2.07	0.045
SD F0	-0.20	-1.95	0.057

^aDependent variable: Clarity ratings for each speaker averaged across utterances and listeners.

external factor with implications for articulatory precision, construed here as target attainment. We used perceived clarity as the principal measure of articulatory precision, and found that utterances produced at faster rates were generally perceived to be clearer than those produced at slower rates. This finding was taken to support the hypothesis of individual differences in the acquisition of adult-like articulatory timing control. Such a hypothesis may have implications not only for understanding articulation rate variation in adults, as suggested in Sec. I, but also for understanding articulation rate variation in clinical versus typical populations, at least in children. Specifically, we might imagine a continuous distribution of speech motor skill attainment in children with and without speech and language disorder rather than a dichotomy between atypical and typical motor skill development. In addition, the hypothesis has implications for theories of rate control since, by hypothesis, rate is an emergent property of articulatory timing patterns, which are acquired with practice and the development of motor skills. In other words, the interpretation of the current results is more in keeping with theories of intrinsic timing control than with theories of extrinsic timing control. This is only true, however, if we assume that articulatory effort is constant across children.

A. Speech motor skill versus differences in articulatory effort

There are a number of studies that have investigated vowel space as a function of articulation rate in adults (e.g., Gay, 1978; Fourakis, 1991; Tsao *et al.*, 2006b). As in the present study, the motivating question for many of these studies has been whether or not increases in articulation rate entail decreases in target attainment. The hypothesis that it does is due to the fact that if one holds energy (force/effort resulting in speed) constant, smaller movements will be completed in less time than larger movements. This means that one strategy for increasing rate is to reduce articulatory displacement. If articulatory displacement is reduced, then target attainment is less likely, especially in coarticulatory contexts that involve substantial movement of the same articulator from target to target.

The hypothesis that a speed-accuracy trade-off applies to rate control is a generalization of Lindblom's (1963) duration-dependent undershoot model of vowel reduction. The model, first proposed to account for differences in the articulation of lexically stressed versus unstressed vowels, has also been invoked to explain articulatory differences resulting from polysyllabic vowel shortening (Moon and Lindblom, 1994) as well as from style shifting (Lindblom, 1990). The fundamental assumption behind the model is that speakers aim to conserve energy under most speaking conditions. Later modifications to the model (e.g., Lindblom, 1990) emphasized that speakers have a choice: Under some speaking conditions, speakers will aim to preserve acoustic targets regardless of cost. The assumption of speaker choice opens the door to variability. Variability is consistent with the literature on rate control, which frequently alludes to individual strategies of control (e.g., Ostry and Munhall,

1985; Adams *et al.*, 1993; Matthies *et al.*, 2001). Some speakers decrease displacement at faster rates, and others increase movement velocity (i.e., effort) to maintain displacement at faster rates. Individual variability in rate control strategies may also explain why some studies find an effect of rate on vowel space (e.g., Fourakis, 1991), while others do not (e.g., Gay, 1978; Tsao *et al.*, 2006b).

Given documented individual differences in rate control strategies, the question that arises in the context of the current study is whether and to what extent the differences that characterize control over one's own articulation rate apply to variability in children's default rates. That is, does the "choice" of sacrificing articulatory precision to preserve energy or expending energy to preserve target attainment apply to default articulation rates? If so, then rate and clarity may covary in children's speech not because of individual differences in motor skill attainment, but because of individual differences in the effort expended during speaking (possibly a speech external factor). Children who expend more energy may speak faster and more clearly than those who expend less energy. Unfortunately, the results from the current study are not sufficient to reject this alternative explanation in favor of the one advanced so far. Instead, the explanation one prefers for the relationship between default articulation rate and perceived clarity in children's speech is a matter of fundamental assumption.

In summary, we have followed Lindblom (1963, 1990) in assuming that a speaker's default behavior is to conserve energy, and so we have advanced the explanation that differences in children's articulation rates emerge from individual differences in motor skills, and specifically with articulatory timing control. Whereas this explanation is consistent with widespread observations of individual differences in developmental schedules of all kinds (e.g., first steps, first words, or even the onset of puberty), we also acknowledge that the assumption of energy conservation may be more relevant to the adult than to the child: In order for children to develop at all in the motor or cognitive domains, they must expend energy attaining new postures and routines, assimilating new sensations in the context of developing meaning and synthesizing information across domains. Future work that first manages to define articulatory effort could perhaps distinguish between an explanation for rate variation based on speech motor factors (intrinsic) versus one based on effort (possibly extrinsic). The contribution of the present study is to rule out the possibility that such variation in children's speech is random and to show that faster speaking children are in fact better at speaking in the sense that their speech is perceived as more well enunciated than the speech of slower speaking children.

V. CONCLUSION

The current study indicates that age-dependent and individual differences in children's default articulation rates are due to changes in both consonant and vowel production, with faster rates of speech associated with shorter consonant and vowel durations. The finding that perceived clarity varies systematically with default articulation rates in

school-aged children suggests that both developmental and individual differences in the temporal aspects of consonant and vowel production can be attributed to maturation in articulatory timing control. This interpretation of the results is consistent with the hypothesis that individual differences in default articulation rates emerge from differences in the speech motor skills underlying the sequential production of speech sounds. Note that such a hypothesis assumes that articulation rate is an emergent property of speech production and so supports a theory of intrinsic timing, at least at the level of word production. Intrinsic timing implies that the abstract representation of sequential speech action encodes relative timing information. We expect that this abstract representation evolves with motor practice and with the acquisition of increasingly efficient and precise control over articulatory movements into and out of speech sound targets.

ACKNOWLEDGMENT

This research was supported by Award Number R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD). The content is solely the responsibility of the author and does not necessarily reflect the views of NICHD. The author is grateful to Faïre Holliday and Jessica Stine for their significant help with data collection and segmentation.

¹Note that overlap is being used here in the sense of coproduction/coarticulation. We recognize that some studies find more coarticulation in young children's speech than in older children's speech (e.g., Nittrouer *et al.*, 1989), but the young children investigated in those studies may not yet have acquired the notion of a gesture as a speech goal (i.e., they have not attained "segmental independence"). When coarticulation is studied in children who have acquired this notion, the spectral and temporal evidence suggests that the younger ones coarticulate less than the older ones (see literature cited in main text).

- Adams, S. G., Weismer, G., and Kent, R. D. (1993). "Speaking rate and speech movement velocity profiles," *J. Speech Hear. Res.* **36**, 41–54.
- Boersma, P., and Weenink, D. (2011). "Praat: Doing phonetics by computer (version 5.2.11) [computer program]," <http://www.praat.org> (Last viewed January 18, 2011).
- Bond, Z. S., and Moore, T. J. (1994). "A note on the acoustic-phonetic characteristics of inadvertent clear speech," *Speech Commun.* **14**, 325–337.
- Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic characteristics," *Speech Commun.* **29**, 255–272.
- Browman, C. P., and Goldstein, L. (1992). "Articulatory phonology: An overview," *Phonetica* **49**, 155–180.
- Cheng, H. Y., Murdoch, B. E., Goozée, J. V., and Scott, D. (2006). "Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults," *J. Speech Lang. Hear. Res.* **50**, 375–392.
- Crystal, T. H., and House, A. S. (1990). "Articulation rate and the duration of syllables and stress groups in connected speech," *J. Acoust. Soc. Am.* **88**, 101–112.
- Dewaele, J.-M., and Furnham, A. (2000). "Personality and speech production: A pilot study of second language learners," *Personality Individ. Diff.* **28**, 355–365.
- Fourakis, M. (1991). "Tempo, stress, and vowel reduction in American English," *J. Acoust. Soc. Am.* **90**, 1816–1827.
- Gay, T. (1978). "Effect of speaking rate on vowel formant movements," *J. Acoust. Soc. Am.* **63**, 223–230.
- Gay, T. (1981). "Mechanisms in the control of speaking rate," *Phonetica* **27**, 44–56.
- Gilbert, J. H. V., and Purves, B. A. (1977). "Temporal constraints on consonant clusters in child speech production," *J. Child Lang.* **4**, 417–432.
- Green, J. R., Moore, C. A., and Reilly, K. J. (2002). "The sequential development of lip and jaw control for speech," *J. Speech Lang. Hear. Res.* **45**, 66–79.
- Hazan, V., and Markham, D. (2004). "Acoustic-phonetic correlates of talker intelligibility for adults and children," *J. Acoust. Soc. Am.* **116**, 3108–3118.
- Jacewicz, E., Fox, R. A., and Wei, L. (2010). "Between-speaker and within-speaker variation in speech tempo of American English," *J. Acoust. Soc. Am.* **128**, 839–850.
- Katz, W. F., Kripke, C., and Tallal, P. (1991). "Anticipatory coarticulation in the speech of adults and young children: Acoustic, perception, and video data," *J. Speech Hear. Res.* **34**, 1222–1232.
- Kehoe, M., Stoel-Gammon, C., and Buder, E. H. (1995). "Acoustic correlates of stress in young children's speech," *J. Speech Hear. Res.* **38**, 338–350.
- Kent, R. D., and Forner, L. L. (1980). "Speech segment durations in sentence recitations by children and adults," *J. Phon.* **8**, 157–168.
- Kent, R. D., Weismer, G., Kent, J. F., and Rosenbek, J. C. (1989). "Toward phonetic intelligibility testing in dysarthria," *J. Speech Hear. Res.* **54**, 482–499.
- Kozhevnikov, V., and Chistovich, L. (1965). *Speech: Articulation and perception* (Joint Publications Research Service, U.S. Department of Commerce, Washington, DC).
- Lam, J., Tjaden, K., and Wilding, G. (2012). "Acoustics of clear speech: Effect of instruction," *J. Speech Lang. Hear. Res.* **55**, 1807–1821.
- Lee, S., Potamianos, A., and Narayanan, S. (1999). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *J. Acoust. Soc. Am.* **105**, 1455–1468.
- Lindblom, B. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* **35**, 1773–1781.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, The Netherlands), pp. 403–439.
- Matthies, M., Perrier, P., Perkell, J. S., and Zandipour, M. (2001). "Variation in anticipatory coarticulation with changes in clarity and rate," *J. Speech Lang. Hear. Res.* **44**, 340–353.
- Moon, S. J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," *J. Acoust. Soc. Am.* **96**, 40–55.
- Neter, J., Wasserman, W., and Kutner, M. H. (1989). *Applied Linear Regression Models* (Irwin, Homewood, IL), Chap. 11, pp. 386–432.
- Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). "The emergence of phonetic segments: Evidence from the spectral structure of fricative vowel syllables spoken by children and adults," *J. Speech Hear. Res.* **32**, 120–132.
- Ostry, D. J., and Munhall, K. G. (1985). "Control of rate and duration of speech movements," *J. Acoust. Soc. Am.* **77**, 640–648.
- Peters, H. F. M., Hulstijn, W., and van Lieshout P. H. H. M. (2000). "Recent developments in speech motor research into stuttering," *Folia Phoniatr. Logop.* **52**, 103–119.
- Pindzola, R. H., Jenkins, M. M., and Lokken, K. J. (1989). "Speaking rates in very young children," *Lang. Speech Hear. Serv. Sch.* **20**, 133–138.
- Pollock, K. E., Brammer, D., and Hageman, C. (1993). "An acoustic analysis of young children's productions of word stress," *J. Phon.* **21**, 183–203.
- Ramus, F., Newport, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," *Cognition* **73**, 265–292.
- Redford, M. A. (2013). "A comparative analysis of pausing in child and adult storytelling," *Appl. Psycholing.* **34**, 569–589.
- Redford, M. A., Dilley, L. C., Gamache, J. L., and Weiland, E. A. (2012). "Prosodic marking of continuation versus completion in children's narratives," in *13th Annual Conference of the International Speech Communication Association 2012 (INTERSPEECH 2012, Portland)*, pp. 2493–2496.
- Semel, E., Wiig, E. H., and Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals*, 4th ed. (The Psychological Corp. A Harcourt Assessment Co., Toronto, Canada).
- Sereno, J. A., Baum, S. R., Marean, G. C., and Liberman, P. (1987). "Acoustic analyses and perceptual data on anticipatory labial coarticulation," *J. Acoust. Soc. Am.* **81**, 512–519.

- Smith, B. L., Sugarman, M. D., and Long, S. H. (1983). "Experimental manipulation of speaking rate for studying temporal variability in children's speech," *J. Acoust. Soc. Am.* **74**, 744–749.
- Syrdal, A. K., and Gopal, H. S. (1986). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," *J. Acoust. Soc. Am.* **79**, 1086–1100.
- Traunmüller, H. (1990). "Analytical expressions for the tonotopic sensory scale," *J. Acoust. Soc. Am.* **88**, 97–100.
- Tsao, Y.-C., and Weismer, G. (1997). "Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component," *J. Speech Lang. Hear. Res.* **40**, 858–866.
- Tsao, Y.-C., Weismer, G., and Iqbal, K. (2006a). "Interspeaker variation in habitual speaking rate: Additional evidence," *J. Speech Lang. Hear. Res.* **49**, 1156–1164.
- Tsao, Y.-C., Weismer, G., and Iqbal, K. (2006b). "The effect of intertalker speech rate variation on acoustic vowel space," *J. Acoust. Soc. Am.* **119**, 1074–1082.
- Turner, G. S., Tjaden, K., and Weismer, G. (1995). "The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis," *J. Speech Lang. Hear. Res.* **38**, 1001–1013.
- Vorperian, H. K., and Kent, R. D. (2007). "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," *J. Speech Lang. Hear. Res.* **50**, 1510–1545.
- Walker, J., and Archibald, L. (2006). "Articulation rate in preschool children: A 3-year longitudinal study," *Int. J. Lang. Commun. Disord.* **41**, 541–565.
- Wing, A. M., and Kristofferson, A. B. (1973). "Response delays and the timing of discrete motor responses," *Percept. Psychophys.* **14**, 5–12.