



## Research Article

## The representation and execution of articulatory timing in first and second language acquisition

Melissa A. Redford<sup>a,\*</sup>, Grace E. Oh<sup>b</sup><sup>a</sup> University of Oregon, United States<sup>b</sup> Konkuk University, South Korea

## ARTICLE INFO

*Article history:*

Received 15 December 2015

Received in revised form 3 January 2017

Accepted 12 January 2017

Available online 9 February 2017

*Keywords:*

Vowel duration

Temporal variability

Temporal patterns

Speech motor control

Speech acquisition

## ABSTRACT

The early acquisition of language-specific temporal patterns relative to the late development of speech motor control suggests a dissociation between the representation and execution of articulatory timing. The current study tested for such a dissociation in first and second language acquisition. American English-speaking children (5- and 8-year-olds) and Korean-speaking adult learners of English repeatedly produced real English words in a simple carrier sentence. The words were designed to elicit different language-specific vowel length contrasts. Measures of absolute duration and variability in single vowel productions were extracted to evaluate the realization of contrasts (representation) and to index speech motor abilities (execution). Results were mostly consistent with a dissociation. Native English-speaking children produced the same language-specific temporal patterns as native English-speaking adults, but their productions were more variable than the adults'. In contrast, Korean-speaking adult learners of English typically produced different temporal patterns than native English-speaking adults, but their productions were as stable as the native speakers'. Implications of the results are discussed with reference to different models of speech production.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

Articulatory timing refers to the coordination of speech articulators in time to achieve motor goals in sequence. Given this definition, timing can be thought of either as a motor speech skill or as a language behavior: stable coordination patterns emerge with neuromotor maturation and speech motor practice; goal sequencing emerges with the acquisition of language. Whereas children are slow to acquire stable coordination patterns (Smith & Zelaznik, 2004), language-specific sequencing is acquired fairly early (Stoel-Gammon & Dunn, 1985). These observations suggest a dissociation between the representation and execution of timing information, consistent with a theoretical distinction between competence (representation) and performance (execution). The current study tested this dissociation hypothesis against an alternative interaction hypothesis by investigating the effect of language-specific vowel length contrasts on production in first and second language acquisition.

## 1.1. Development of timing control

Skilled action includes patterns of movement coordination that are acquired for functional ends (i.e., goals). Change in the duration and variability with which goal-directed movements are executed is thought to reflect neuromotor maturation and/or motor learning (see Smith, 1992). Whatever the underlying explanation, both duration and variability are observed to decrease as coordinated articulatory movements become faster and more stable (Smith & Zelaznik, 2004).

The earliest studies to link acoustic duration and temporal variability in children's speech to motor skill development focused on linguistic units of various sizes, including segments, syllables, and words (Kent & Forner, 1980; Smith, 1978; Tingley & Allen, 1975). For example, Smith (1978) compared the mean acoustic duration and standard deviation of repeated word productions in 2- and 4-year-old children's speech to adults' speech. He found that children's word durations were greater than adults', and that 2-year-olds' repetitions of the same word were more variable than adults'. Kent and Forner (1980) found that even 6-year-olds produced more variable phrase, word, and segment durations than adults. Noting the correlation between mean duration

\* Corresponding author at: Linguistics Department, 1290 University of Oregon, Eugene, OR 97403, United States.

E-mail address: [redford@uoregon.edu](mailto:redford@uoregon.edu) (M.A. Redford).

and standard deviation in their data, Kent and Forner examined whether age-related differences would persist if standard deviations were mean normalized. They did, leading the authors to conclude that duration and temporal variability were independent markers of motor skill in children's speech. This conclusion has since been echoed many times in developmental studies of speech production (e.g., Lee, Potamianos, & Narayanan, 1999; Redford, 2014; Smith, 1992; Smith, Sugarman, & Long, 1983), and is consistent with the broader literature on motor learning (see, e.g., Rosenbaum, 2009: Ch. 4).

Developmental studies have also found that children's speech is more acoustically variable than adult speech until age 12 years (e.g., Lee et al., 1999); kinematic differences persist until age 14 years (e.g., Green, Moore, Higashikawa, & Steeve, 2000; Sharkey & Folkins, 1985; Smith & Goffman, 1998; Smith & Zelaznik, 2004). These findings indicate that speech motor development is protracted. Despite this, children produce linguistically-relevant temporal patterns accurately from a very early age. For example, in the aforementioned study on speech timing in 2-year-old children, Smith (1978) also investigated the effects of place of articulation and voicing on the children's production of mean stop closure duration, VOT duration, and vowel duration. He found that whereas absolute duration values differed in child and adult speech, the proportional duration of these intervals varied with linguistic factors in the same way across all age groups. He concluded from these and other results from the same study that "even prior to age three, children recognize important temporal parameters of the language they are learning and incorporate them into their phonological system—a system which, despite certain limitations, seems quite sophisticated (p. 65)."

Subsequent acoustic–phonetic studies on early child language have confirmed the idea that children acquire temporal information early as part of their language grammar or abstract word form representations. For example, a number of studies on lexical stress production in young children have shown that children use duration to distinguish stressed from unstressed syllables in English as early as 2 years of age (e.g., Kehoe, Stoel-Gammon, & Buder, 1995; Pollock, Brammer, & Hageman, 1993; Schwartz, Petinou, Goffman, Lazowski, & Cartusciello, 1996). Studies on stop production indicate that English-speaking children use voice onset time to convey a voicing contrast from an early age (Bond & Wilson, 1980; Imbrie, 2005), even if during the earliest period (prior to age 2 years) the contrast is not perceptible to adults (Macken & Barton, 1980). Two year old children also use vowel duration to reliably signal voicing in stop codas (Buder & Stoel-Gammon, 2002; Song, Demuth, & Shattuck-Hufnagel, 2012).

In sum, studies on speech motor development have shown that children's speech is slower and more variable than adults' speech, and that this difference persists until at least age 12. In contrast, studies on early child phonology indicate that language-specific temporal patterns are mostly acquired by 3 years of age. Some difficulties in specific sound or cluster production persist until children have begun school at age 5 (see, e.g., Stoel-Gammon & Dunn, 1985), but resolve soon thereafter and certainly well before speech motor abilities are adult-like.

## 1.2. Timing control in second language acquisition

Whereas children acquire language-specific timing patterns early during first language (L1) acquisition, adult second language (L2) learners often fail to achieve native-like timing in their L2; instead, they produce L1 influenced patterns. For example, adult Spanish-speaking learners of English produce voiceless stops in English with shorter voice onset times (VOT) than do native English speakers (Flege, 1991), presumably because Spanish voiceless stops are characterized by shorter VOTs than English voiceless stops in syllable positions where these are released. Adult Korean-speaking learners of English produce less contrastive vowel durations to signal differences in coda stop voicing than native English speakers (Cho & Shin, 2013), presumably because the Korean voicing contrast for stops is neutralized in final position. Zsiga (2003) cites many similar examples and goes on to show that specific patterns of Russian word-to-word timing influence Russian learner's production of English patterns. She also reports that English learners' of Russian produce unmarked articulatory timing patterns that do not occur in either English or Russian. Zsiga interprets the former results to support the notion of cross-language transfer and the latter to support the idea of distinct second language representations that may also reference universal phonological processes (cf. Selinker, 1972).

It is the interpretation of the L2 findings that is especially relevant to our present interest in a dissociation between the representation and execution of articulatory timing. Non-native timing patterns in second language speech are nearly always explained with reference to representational factors; not motoric ones. This is true even when the observed patterns cannot be explained either in terms of the L1 or L2 patterns, as in the Russian learner results reported in Zsiga (2003; see also Cebrian, 2000). Moreover, Flege (1991:406) explicitly rejects the idea that adult learners are less able than early learners "to motorically implement their perceptual representations for sounds," noting with reference to his VOT data that there "is no *a priori* reason to think that it is somehow easier for late learners to produce partial modification of previously established articulatory patterns than to produce a *complete* modification that would enable them to match native speakers of English (emphasis in the original)." The implication is that motor factors have no impact on L2 representations, in keeping with the dissociation hypothesis. Relatedly, the two main theories of second language speech acquisition, the Speech Learning Model (SLM; Flege, 1995) and Perceptual Assimilation Model (PAM; Best, 1995), are models of perceptual learning; specifically, they are models of how pre-existing phonological categories influence and are influenced by second language speech perception. Motor learning and control are not considered in the models.

In sum, the mainstream assumption in adult second language acquisition research is that the motor system faithfully executes whatever "interlanguage" representation has been established. This is likely both because foreign accents tend to be stable over time and because we often think of speech motor skills in maturational terms.

### 1.3. Questioning the divide: First language acquisition

The assumption that motor factors are irrelevant to second language acquisition has implications for how to interpret the robust finding that adult L2 speech is slower than L1 speech (Lennon, 1990; Munro & Derwing, 1995; Munro & Derwing, 1998; Trofimovich & Baker, 2006). Not surprisingly, this finding is explained with reference to cognitive factors in the second language acquisition literature. For example, Trofimovich and Baker (2006:25) hypothesize that slower L2 speech compared to L1 speech results either from competition between L1 and L2 representations or from an overreliance on declarative memory during learning. Note the contrast between this type of explanation for slow articulation rates and the explanation offered in studies of first language acquisition: in first language acquisition, slower articulation rates in children compared to adults is attributed to immature speech motor control (e.g., Lee et al., 1999; Tilsen, 2016).

The reason that slow articulation rates (i.e., overall longer acoustic durations of segments and syllables) can be explained differently in first and second language acquisition is because so many factors are known to influence duration. In addition to the previously mentioned motor and cognitive factors, utterance length, phonological context, dialectal variation, and even personality can influence the rate at which speech sounds are produced (see Redford, 2014:2952–53). It is perhaps because of the wide variety of influences on duration that temporal variability in the acoustic domain and spatial–temporal variability in the kinematic domain have become the measures of choice in developmental studies of speech motor control. Like duration, variability may be attributed to a wide range of factors; but, unlike duration, almost all of these reference motor factors rather than linguistic or other cognitive factors.

For example, Smith (1992:2172) cites explanations for temporal variability that range from changing biomechanical properties of the articulators and on-going development of the peripheral and central nervous system to incomplete routinization of an action pattern (see also Smith & Zelaznik, 2004:31). The biomechanical and neurophysiological explanations are consistent with a maturational view of speech motor development, as well as with the observations that movement variability increases with senescence (Morris & Brown, 1994; Smith, Wasowicz, & Preston, 1987) and is greater in adult populations with motor speech disorders than in healthy populations (Seddoh et al., 1996; Ackermann & Hertrich, 2000). By contrast, the explanation of incomplete routinization is consistent with a view that attributes decreases in variability over developmental time to motor learning (e.g., Lee et al., 1999); specifically, to the emergence of functional synergies, or units of coordinated muscle activity, “that reduce the degrees of freedom and provide stable collectives from which motor behaviors emerge (Smith & Zelaznik, 2004:22–23).”

In speech, synergistic muscle activity (a coordinative structure) emerges over developmental time in service of speech sound articulation. Coordinative structures thus represent inter-articulatory coordination at a single point in time. Of course, a complex motor skill like speech requires not only the accurate production of individual sounds, but also appropriate sequencing of these sounds. Sequencing represents

inter-articulatory coordination through time. Insofar as sequencing depends on being able to implement individual sounds in context, emergent mastery over single sound production is inseparable from mastery over language-specific sequencing. Moreover, as Smith and Zelaznik (2004:32) note, “(t)he units of movement production for speech are unknown. Syllable, phoneme, and gestural units have been proposed, and it seems likely that there are multiple units of production operating in parallel.” If we assume that these linguistic units emerge over developmental time (see, e.g., Tilsen, 2016), then we might hypothesize that the execution of articulatory timing patterns would influence their abstraction (i.e., their representation) and vice versa. This hypothesis of interaction is an alternative to the disassociation hypothesis. It is also in keeping with the general view that language acquisition interacts with speech motor development (Goffman, 2010; Nip, Green, & Marx, 2009; Redford, 2015; Smith & Goffman, 2004).

Work on the interaction between speech motor development and language acquisition has focused on how linguistic representations influence execution. To take just one example, a number of studies report that production variability increases with phrase length and morphosyntactic complexity, and more so in children’s speech than in adults’ speech (MacPherson & Smith, 2012; Maner, Smith, & Grayson, 2000; Sadagopan & Smith, 2008). In the current study, we test for influences of representation on execution in acquisition by investigating whether production variability varies as a function of language-specific vowel length contrasts and speakers’ age.

The hypothesis of an interaction between speech motor development and language acquisition also predicts that the execution of language-specific sound patterns should influence the structure of linguistic representation. This prediction is more difficult to test experimentally than the effect of representation on execution, but it is nonetheless consistent with the interpretation of child language phonology as motorically constrained. For example, Vihman (2014) details a theory of how well-practiced motor patterns evolve into stable phonological representations (templates) that a child then leverages to facilitate lexical acquisition. In the current study, we indirectly test the influence of timing execution on its representation in first and second language acquisition by investigating whether age- or experience-related differences in production variability are associated with age- or experience-related differences in the realization of language-specific temporal patterns.

### 1.4. Questioning the divide: Second language acquisition

If speech motor development interacts with language acquisition, then we might also expect that production variability would be high whenever language learning occurs, including in adults. So far, the evidence for this prediction is mixed. The number of relevant studies is also very limited. Findings from three that we know of conflict with one another.

Chakraborty, Goffman, and Smith (2008) investigated movement duration and variability in bilingual Bengali-English speakers who had different levels of proficiency in (L2) English. As in previous studies, they found that movement durations were greater in L2 compared to L1, especially in speakers with lower English proficiency. They also found that movement variability was the same in L1 and L2, regardless

of proficiency level. Nip and Blumenfeld (2015) investigated movement variability in native English-speaking learners of Spanish and found significant effects of language experience on production in the expected direction: more variability in L2 than in L1. Finally, Smith and Hayes-Harb (2016) investigated acoustic target attainment and temporal variability in vowels produced by non-native learners of English, including native speakers of Mandarin, Korean, and Spanish. They found that more than half of learners' productions matched native speaker productions in both target attainment and temporal variability. The next most common scenario in their data was for learners' productions to be non-native-like in target attainment ("foreign accent"), but native-like in temporal variability.

The mixed results from studies that have investigated second language effects on intra-speaker and intra-item variability in production might be due to differences in the quantity and quality of participants' practice with the L2. Second language speakers in the Chakraborty et al. (2008) and Smith and Hayes-Harb's (2016) studies were immersed in their L2, while learners in Nip and Blumenfeld's (2015) study experienced their L2 in a classroom setting. Perhaps the effects of language on speech motor learning (and vice versa) are only evident in adults who have very little daily practice in their L2 (viz. Flege, 1980).

### 1.5. Current study

The current study investigated vowel production in first and second language acquisition to test for a dissociation between the representation and execution of articulatory timing. Participants were two groups of native English-speaking children (5- and 8-year-olds), two groups of adult English learners (native Korean speakers), and a control group of native English-speaking adults. The participants were asked to repeatedly produce real English words in a simple carrier sentence. The words were phonologically controlled to elicit inherent and context-dependent vowel length contrasts specific to English. Mean vowel durations and variability in vowel durations were calculated across repetitions of an item and analyzed as a function of age and the target length contrasts in first language acquisition and as a function of language experience and the target length contrasts in second language acquisition. The effect of age and experience on vowel production was used to index execution, which was expected to be slower and more variable in children (i.e., immature) and slower but stable in adult second language learners (i.e., mature). The effect of the target length contrasts on vowel duration was used to index the linguistic representation of timing, which was expected to be native-like in children and non-native like in adult second language learners. The dissociation hypothesis also predicts no interaction between age/experience and the length contrasts, whereas the interaction hypothesis predicts that the effects of length contrasts on production will vary with age and language experience.

## 2. Methods

### 2.1. Participants

A total of 60 speakers participated in the study: two groups of 12 American English-speaking children, aged 5 and 8 years

old; two groups of 12 adult Korean learners of English, with different proficiency levels in English; and one group of 12 American English-speaking adults. The mean age of children in the 5-year-old group was five years, seven months (=5;7). The range was from 5;2 to 6;3. The mean age in the 8-year-old group was 8;1. The range was from 7;7 to 8;8. All children spoke a west coast dialect of American-English. All had typical hearing and typical speech-language development for their age, as determined by a hearing screen, parental report, and age-normalized scores on the PPVT-4 (Dunn & Dunn, 2007). Half of the 5-year-olds, and 7 of the 8-year-olds were female.

The Korean learners of English and native English-speaking adults were all college-aged. Most of the Korean learners of English were female (10 and 9 in each group), 7 of the adult native English speakers were female. Regarding the English language skills of the two groups of second language learners, one group had been exposed almost exclusively to classroom English, albeit from middle school onwards. The other group had experienced between 3 and 5 years of language immersion in an English-speaking country in addition to the same amount of classroom English as the first group. We refer to the first group as the no-immersion group and the second as the immersion group. The no-immersion group had an average Test of English as a Foreign Language (TOEFL) score of 91.33 ( $SD = 10.25$ ). The immersion group's average score was 111.42 ( $SD = 5.43$ ). An independent samples t-test confirmed that the mean difference of 20.1 points between the learner groups' scores was significant,  $t(22) = 5.99$ ,  $SE = 3.35$ ,  $p < 0.001$ .

### 2.2. Stimuli

The stimuli were designed to elicit the production of 3 different length contrasts in American English: one due to vowel quantity; one due to final consonant voicing; and one due to word length. The real word stimuli used to encode each contrast are shown in the Tables 1–3. Note that in all cases the target (stressed) vowel occurred between two stop consonants.<sup>1</sup>

Unlike in many other languages, English vowel quantity is not phonemic (i.e., there is no singleton/geminate distinction). Instead, vowel quality is considered the defining aspect of English vowel categories.<sup>2</sup> Nonetheless, there are well-known categorical differences in the inherent duration of English vowels: diphthongs are especially long, tense vowels are also long, but may be shorter than diphthongs, and lax vowels are short. Children and second language speakers must learn to reproduce the quantity contrast to sound like native speakers of American English.

Vowel duration also varies systematically with phonological context in English; for example, with the voicing status of coda stop consonants. Vowels that precede a final voiced stop consonant are longer on average than those that precede a final voiceless stop consonant. Although some have explained this particular length contrast in mechanistic terms (Malécot, 1970), cross-linguistic evidence strongly suggests that the contrast is

<sup>1</sup> The intervocalic stops in disyllabic and multisyllabic words (Table 3) were produced with the allophonic tap.

<sup>2</sup> Note that nothing of interest hinges on vowel quality categorization in the present study, and that item is treated as a random factor in the analyses.

**Table 1**

Stimuli used to investigate a length contrast due to vowel quantity in a /bV/ frame.

Diphthong	Tense	Lax
/aɪ/	/i/	/ɪ/
"bite"	"beat"	"bit"
/eɪ/		/ɛ/
"bait"		"bet"
/aʊ/	/ɑ/	/ʌ/
"bout"	"bought"	"but"

**Table 2**

Stimuli used to investigate a length contrast in vowel production due to final consonant voicing.

	Voiced	voiceless
bæT	"bad"	"bat"
bæK	"bag"	"back"
k <sup>h</sup> æP	"cab"	"cap"
k <sup>h</sup> æT	"cad"	"cat"

**Table 3**

Stimuli used to investigate a length contrast in stressed vowel production due to word length (i.e., polysyllabic shortening).

	Mono-syllabic	Di-syllabic	Tri-syllabic	Quadra-syllabic
'bæ__	"bat"	"batty"	"battery"	
'k <sup>h</sup> æ__	"cat"	"catty"	"catalogue"	"caterpillar"

language-specific and so must be learned (Flege & Port, 1981).

Language-specific systematic variation in vowel duration is also prosodic in nature. For example, in American English, stressed vowels are longer than unstressed vowels, vowels in accented words are longer than those in unaccented words, and vowels in phrase-final syllables are longer than those in phrase-medial or phrase-initial syllables. Polysyllabic shortening is another prosodic effect on vowel duration. Vowels are longer in shorter words than in longer words. This has been explained with reference to abstract, universal temporal structures, including constraints on word duration (Lehiste, 1972), but White and Turk (2010) have argued that the effect is due to language-specific prosodic factors. They also cite Suomi's (2007) finding that polysyllabic shortening is absent in Finnish to support their argument.

### 2.3. Elicitation procedure

The stimuli were recorded with a number of similar additional items (*boot*, *boat*, *batter*, *catter*, *batterless*) by a female speaker of west coast American English in the frame sentence "I said \_\_\_ again." The sentences were then aggregated and presented auditorily to participants one at a time by an experimenter. Participants responded to the stimulus sentence with the phrase, "She said \_\_\_ again." The change in the frame sentence was meant to make the speech task more meaningful. Auditory presentations were used to control for age-dependent differences in reading level.

The items were presented in random order once per block, but repetitions of a single item were elicited when a speaker mis-heard or mis-spoke the target. The experiment included six blocks with a break between each block to prevent undue fatigue. Each block took up to 7 min to complete. In this way, 6 non-consecutive and correct repetitions of each item were elicited.

Participants' speech 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 or headband that the speaker wore.

### 2.4. Outcome measures

The elicitation procedure yielded 8280 items for segmentation (138 per speaker). Recordings were displayed both as oscillograms and spectrograms in Praat (<http://www.fon.hum.uva.nl/praat/>). Word intervals and stressed vowels were marked out on "tiers" based on the acoustic landmarks noted below. Durations were then extracted automatically.

Word intervals were defined as the time between the offset of the vowel in "said" and the onset of voicing in "again" for all 8280 items. Interval durations were extracted and then examined by speaker and word set (quantity, final voicing, word length). Target word interval durations that were extreme outliers for a particular word set produced by a particular speaker were noted and excluded from the analyses. The goal was to reduce variability due to disfluency or frame sentence prosodification, and to define both objectively. Extreme outliers were always greater than the reference median duration and so indicated productions that were much longer than normal. The assumption was that these especially long word interval durations indicated either word prolongation or the presence of a preceding or following pause. A total of 27 items were identified as extreme outliers and excluded from the 8280 produced: 13 of the excluded items were produced by 5-year-olds, 5 by 8-year-olds, 6 by the no immersion learner group, 2 by the immersion learner group, and 1 by the adult native speakers.

Once extreme outliers were excluded, stressed vowels were segmented in each of the target words. Vowel onsets were identified at the onset of voicing after the release of the initial /b/ or /k/ consonant. Vowel offsets were defined by the loss of energy that accompanies oral closure. Durations were again extracted based on the segmentation. The mean and standard deviation were calculated across the 6 repetitions of a particular target word item. The standard deviation in duration was divided by the mean to produce a normalized measure of temporal variability (i.e., the coefficient of variation). These two measures, the mean absolute duration and variability of absolute duration, were the dependent variables in the analyses on speech timing across all groups. Vowel durations are reported in milliseconds; variability values were the coefficient of variation multiplied by 100, which we will refer to as *scaled covar*.

Reliability was assessed on 23% of the data (1,902 tokens), which represented all six repetitions of 317 randomly selected words (an average of 5 words per speaker). The same rater blindly re-segmented the stressed vowels in these data following the criteria described above. Intra-rater reliability was extremely high,  $r^2 = 0.988$ . This result is not at all surprising given that the stimuli were designed with stops flanking the vowel so that segmentation would be easy.

### 2.5. Statistical analyses

The analyses were split by acquisition type (first versus second) and by stimulus set (quantity, final voicing, word length). The native English-speaking adults served as a reference

group in both the first and second language acquisition analyses. Thus, the fixed effect of group was a 3-level factor defined by age (5-year-olds, 8-year-olds, adults) in the analyses of first language acquisition and by experience (no immersion, immersion, native speakers) in the analyses of second language acquisition. Length contrast was the other fixed effect in the analyses. The levels of this factor varied with the word set. In the word set that encoded a vowel quantity contrast it was a 3-level factor defined by vowel type (diphthong, tense, and lax); in the set that encoded a subphonemic length contrast due to voicing of the coda stop consonant it was a 2-level factor defined by voiced versus voicelessness; and in the set that encoded the length contrast associated with polysyllabic shortening it was a 4-level factor defined by the number of syllables in a word (an ordinal variable). Generalized linear mixed-effects models (GLMM in SPSS) were used to investigate the contribution of the fixed factors and their interaction on the outcome measures, mean vowel duration and temporal variability. All models included a random intercept for speaker. Item, nested within length contrast, was treated as a within-subjects random effect. A diagonal covariance structure was used. The output of the models included ANOVA tables. The F values from those tables are reported here. Parameter estimates were reported when significant differences between levels of a factor were of interest.

### 3. Results

The first set of analyses investigated the fixed effects of age group and length contrast on vowel duration and temporal variability in the word sets that encoded contrasts due to quantity, final voicing, and word length. The second set investigated the fixed effects of language experience and length contrast on the same dependent variables in the same three word sets. Results for each word set are presented separately and discussed with reference to the competing disassociation and interaction hypotheses.

#### 3.1. First language acquisition

##### 3.1.1. Vowel quantity

The random effects of speaker and item accounted for 22.3% and 25.3% of the variance in the model of vowel

duration and 5.2% and 1.2% of the variance in the model of temporal variability. The analysis on vowel duration indicated a significant effect of quantity,  $F(2, 279) = 8.64$ ,  $p < 0.001$ , but no main effect of age group. The interaction between the factors was also not significant. Both children and adults produced the same expected temporal pattern: diphthongs and tense vowels were longer than lax vowels (diphthong vs. lax: mean difference = 54.8 ms.,  $t = 3.81$ ,  $p < 0.001$ ; tense vs. lax: mean difference = 37 ms.,  $t = 2.45$ ,  $p = 0.015$ ). The data are shown in the left hand panel of Fig. 1.

In contrast to the analysis on vowel duration, the analysis on temporal variability indicated a significant effect of group,  $F(2, 279) = 5.31$ ,  $p = 0.005$ , but no effect of quantity on the dependent variable. Children's repetition of the same vowel was typically more variable than adults', as shown in the right hand panel of Fig. 1. This result held both for 5-year-olds versus adults (mean difference = 5.4 scaled covar,  $t = 2.63$ ,  $p = 0.009$ ) and for 8-year-olds versus adults (mean difference = 5.7 scaled covar,  $t = 2.47$ ,  $p = 0.014$ ). There was no interaction between group and quantity.

Together, the results on vowel duration and temporal variability indicate that children acquire an adult-like representation of English vowel quantity by age 5 years, even though they also have immature speech motor skills. The absence of an effect of quantity on temporal variability or of an interaction between age and quantity suggests a dissociation between the execution and representation of the quantity contrast in first language acquisition.

##### 3.1.2. Final voicing

As in the word set that encoded the vowel quantity contrast, there was no effect of group on vowel duration in words that encoded a length contrast due to the voicing identity of the coda consonant. Still, parameter estimates in the model indicated a nearly significant difference between 5-year-old and adult vowel durations in the expected direction; that is, longer in children than adults (mean difference = 16.1 ms.,  $t = 1.93$ ,  $p = 0.055$ ). The effect of final voicing on vowel duration was significant,  $F(1, 282) = 20.51$ ,  $p < 0.001$ . Vowels that preceded a voiced consonant were longer than those that preceded a voiceless consonant ( $M = 241.1$  ms.,  $SD = 40.9$  versus  $M = 174.2$  ms.,  $SD = 28.0$ ). There was no significant interaction between group and length contrast, indicating once again

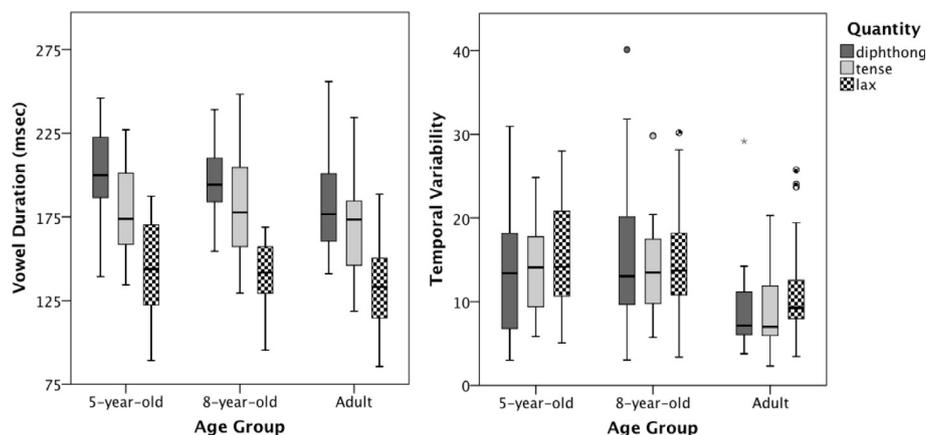


Fig. 1. Box plots show vowel duration (left) and temporal variability (right) as a function of speakers' age group and vowel quantity targets.

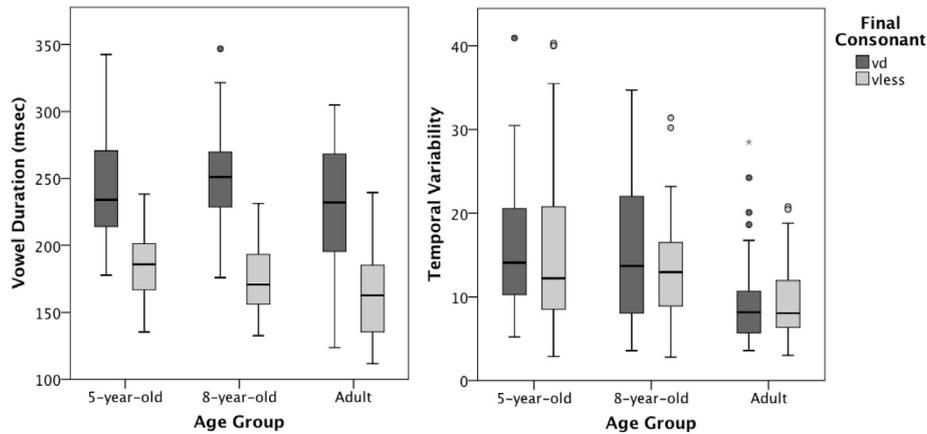


Fig. 2. Box plots show vowel duration (left) and temporal variability (right) as a function of speakers' age group and a target length contrast due to the voicing status of the coda consonant.

that children produced the same temporal pattern as adults. The vowel duration data are shown in the left hand panel of Fig. 2.

The right hand panel of Fig. 2 shows the temporal variability data. As before, vowel production was less stable in children than in adults,  $F(2, 282) = 6.04$ ,  $p = 0.003$ , though parameter estimates indicated that the differences between 5-year-olds versus adults and between 8-year-olds versus adults were relatively small (5-year-olds vs. adults mean difference = 5.9 scaled covar,  $t = 1.96$ ,  $p = 0.051$ ; 8-year-olds vs. adults mean difference = 6.3 scaled covar,  $t = 1.90$ ,  $p = 0.059$ ). The analysis indicated no effect of final voicing on temporal variability, nor any interaction between age and final voicing. Thus, the results are consistent both with the assumption of immature speech motor skills in children, and with the hypothesis of a dissociation between the execution and representation of the vowel length contrast due to final voicing.

Regarding the random effects of speaker and item, these again accounted for more of the variance in the model of vowel duration (10.7% and 10.5%, respectively) compared to the model of temporal variability (3.0% and 3.2%, respectively).

### 3.1.3. Word length

In contrast to the other 2 word sets, stressed vowel duration was found to vary systematically with speakers' age in the

word set that encoded a contrast due to word length,  $F(2, 240) = 4.93$ ,  $p = 0.008$ . The parameter estimates indicated, however, that only 5-year-olds produced longer vowels than adults (mean difference = 20.8 ms.,  $t = 2.33$ ,  $p < 0.001$ ). Also, unlike in the other word sets, the main effect of contrast on vowel duration was not significant. There was no significant interaction between group and length contrast.

The absence of a main effect of word length on vowel duration is at odds with the data shown in the left hand panel of Fig. 3. For example, the parameter estimates from the model indicated that vowels in monosyllabic words were systematically longer than in the reference category, 4-syllable word (mean difference = 75 ms.,  $t = 3.39$ ,  $p = 0.021$ ). The inconsistency between the overall model results and distributions shown in the figure may be due to the high proportion of variance explained by item in the model (32.5% versus 13% for speaker). This possibility is supported by the finding that when item is removed as a random effect, the overall effect of word length on vowel duration was significant,  $F(3, 240) = 129.7$ ,  $p < 0.001$ .

As before, the effect of age on temporal variability was significant,  $F(2, 240) = 12.40$ ,  $p < 0.001$ , though parameter estimates indicated that only 5-year-old productions were more variable than adult productions (mean difference = 6.9 scaled covar,  $t = 2.87$ ,  $p = 0.005$ ). Other results from the analysis on

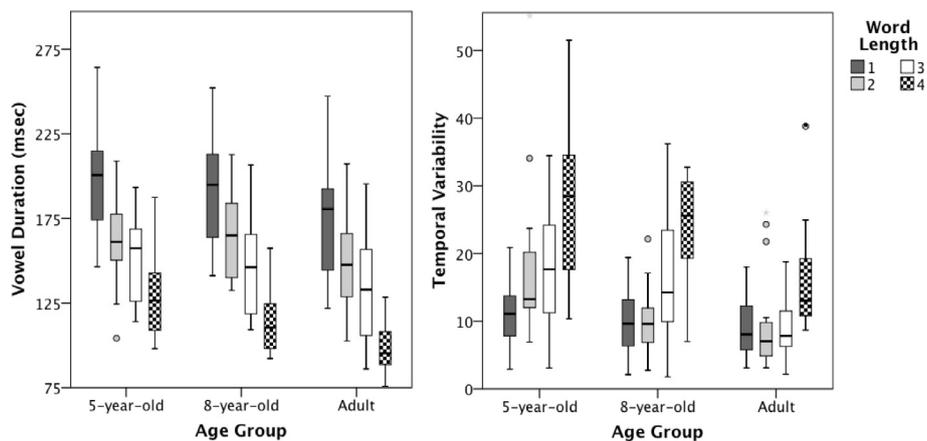


Fig. 3. Box plots show vowel duration (left) and temporal variability (right) as a function of speakers' language experience and word length in syllables.

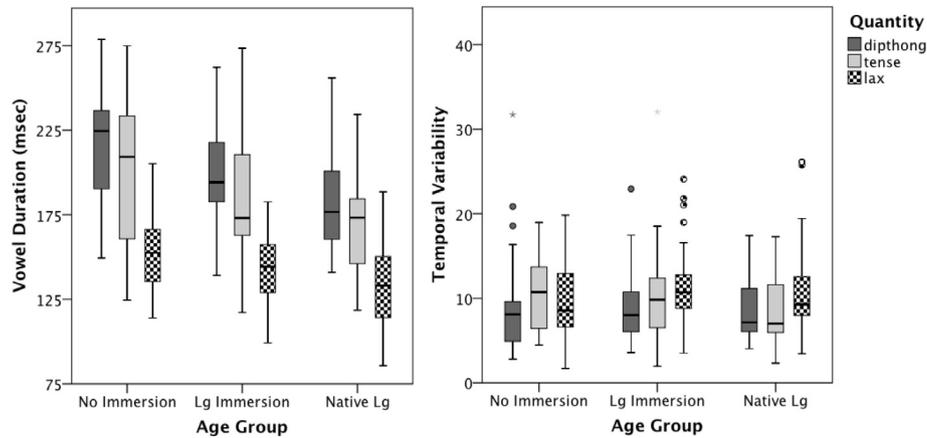


Fig. 4. Box plots show vowel duration (left) and temporal variability (right) as a function of adult speakers' English language experience and vowel quantity targets.

temporal variability in this word set were somewhat different from those in other word sets. In particular, there was a significant interaction between age and word length,  $F(6, 240) = 3.23$ ,  $p = 0.005$ , which can be seen in the data shown in the right hand panel of Fig. 3. The interaction was due to a stronger effect of word length on temporal variability in children's speech compared to adults' speech. Individual differences were such, though, that the overall effect of the contrast on temporal variability was not significant.

Overall, the results on vowel duration and temporal variability in words that ranged from one to four syllables in length are consistent with the assumption of immature speech motor skills in children. Unlike the results on the production of vowel quantity and final voicing, the present results were inconsistent with the dissociation hypothesis. Instead, the results on the production of vowels as a function of word length suggest an interaction between representation and execution, especially in children's speech. Children produce the expected language-specific temporal pattern, but their execution of shorter targets (i.e., stressed vowels in 3- and 4-syllable words) is more variable than their execution of longer targets (i.e., stressed vowels in 1- and 2-syllable words).

### 3.2. Second language acquisition

#### 3.2.1. Vowel quantity

Stressed vowel duration varied systematically both with language experience,  $F(2, 279) = 6.172$ ,  $p = 0.002$ , and with the target length contrast. The interaction between experience and quantity was also significant,  $F(4, 279) = 4.36$ ,  $p = 0.002$ . This interaction was likely due to a larger mean difference between diphthong duration and lax vowel duration in the no immersion language group compared to the difference between these vowel types produced by native speakers (i.e., mean difference of 63.5 ms. in no immersion group versus 52 ms. in native speaker group). Second language speakers with less experience (i.e., no immersion group) also produced longer vowel durations on average than native English speakers (mean difference = 27 ms.,  $t = 2.47$ ,  $p = 0.014$ ). The relevant data are shown in the left hand panel of Fig. 4.

The temporal variability data are shown in the right hand panel of Fig. 4. Analyses indicated no effect of language experience on variability, but the effect of quantity was significant,

$F(2, 279) = 3.71$ ,  $p = 0.026$ . There was no interaction between the factors.

The main effect of quantity on variability was likely due to the difference between lax vowels ( $M = 10.9$  scaled covar,  $SD = 6.2$ ) compared to diphthongs ( $M = 8.7$  scaled covar,  $SD = 4.6$ ) and tense vowels ( $M = 9.9$  scaled covar,  $SD = 4.9$ ). Although none of the parameter estimates associated with levels of the factor reached statistical significance, Fig. 4 and the mean values suggest that lax vowels were produced with more temporal variability than other vowels by all speakers.

Regarding the random effects, these accounted for very little of the overall variance in vowel duration (item = 0.9% and speaker = 0.7%) and temporal variability (item = 0% and speaker = 2.7%).

Overall, the results are consistent with the non-native representation of English vowel quantities in second language speakers of English. The results also suggest an unexpected experience-independent effect of vowel length on execution.<sup>3</sup> The data shown in Fig. 4 suggest that this effect is inversely correlated with vowel duration, and so more likely the manifestation of execution intrinsic factors (e.g., coarticulation) than the result of representational ones.<sup>4</sup>

#### 3.2.2. Final voicing

Vowel duration also varied systematically with the voicing status of coda consonants,  $F(2, 282) = 12.26$ ,  $p = 0.001$ . The effect of language experience was not significant, even though the interaction between experience and final voicing was,  $F(2, 282) = 4.77$ ,  $p = 0.009$ . The data shown in the left hand panel of Fig. 5 suggest that the significant interaction was due a greater difference in vowel duration between the two length categories in the least experienced language group (= 75.1 ms) compared to the difference produced by native speakers (= 64 ms).

There were no significant effects of the fixed factors on temporal variability. These data are shown in the right hand panel of Fig. 5. The random effects accounted for a fair proportion of the overall variance in vowel duration (speaker = 14% and

<sup>3</sup> In particular, the effect of quantity on variability is significant in speech produced by native English-speaking adults,  $F(2, 93) = 7.91$ ,  $p = 0.001$ .

<sup>4</sup> This suggestion is in keeping with the hypothesis of duration-dependent undershoot (Lindblom, 1990): absent adjustments in stiffness, target accuracy decreases under increasing time pressure.

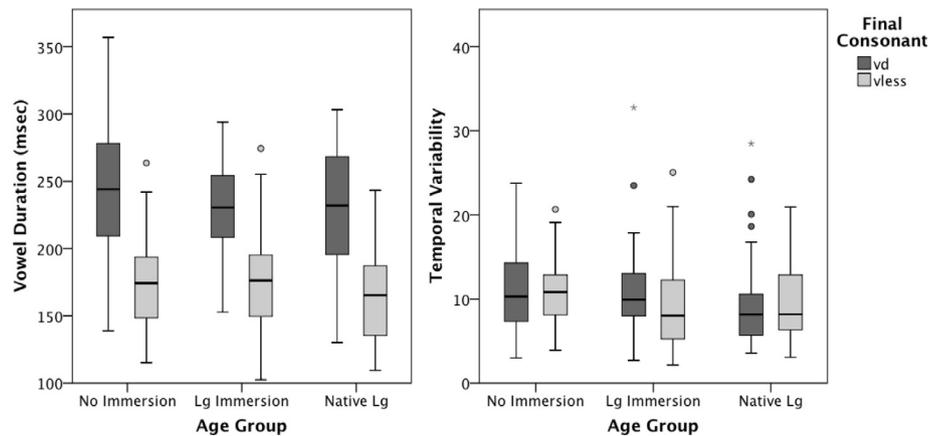


Fig. 5. Box plots show vowel duration (left) and temporal variability (right) as a function of adult speakers' English language experience and a target length contrast due to the voicing status of the coda consonant.

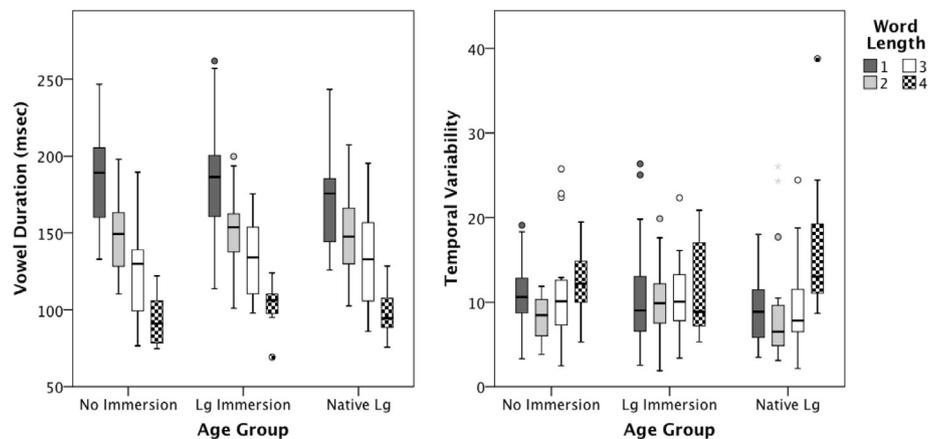


Fig. 6. Box plots show vowel duration (left) and temporal variability (right) as a function of adult speakers' English language experience and word length in syllables.

item = 16.9%) and for much less of the variance in temporal variability (speaker = 5.7% and item = 0.5%).

The absence of significant effects on temporal variability, coupled with the significant effect of language experience on vowel duration and an interaction between language experience and final voicing suggest mature execution of the non-native timing pattern; that is, a dissociation between execution and representation.

### 3.2.3. Word length

Finally, neither experience nor the target length contrast had a significant effect on vowel duration in the word set designed to elicit polysyllabic shortening, though the effect of contrast trended towards significance,  $F(3, 240) = 2.47$ ,  $p = 0.062$ . The relevant data are shown in the left hand panel of Fig. 6.

As in the first language data, vowel duration was significantly longer in monosyllabic words than in 4-syllable words (mean difference = 87.7 ms.,  $t = 2.27$ ,  $p = 0.024$ ). Once again, the random effect of item accounted for high proportion of the variance in the model (29.1%) compared to the effect of speaker (0.7%). As before, when the effect is removed word length was significant,  $F(3, 240) = 206.12$ ,  $p < 0.001$ .

The temporal variability data are shown in the right hand panel of Fig. 6. Note that there were no effects of either

experience or contrast on variability, consistent with the dissociation hypothesis. The random effects of speaker accounted for 2.4% of the overall variance in temporal variability; the effect of item accounted for 0.8%.

## 4. General discussion

The goal of the current study was to test for a dissociation between the representation of articulatory timing and its execution. The alternative hypothesis was that the representation of language-specific temporal patterns is influenced by their execution and vice versa given that articulatory timing control emerges in the context of language acquisition. The results were mostly consistent with a dissociation. Children realized all language-specific length contrasts in the same way as adults, but their productions were often slower and always more variable. In contrast, adult second language learners of English produced different temporal patterns from native English-speaking adults, but their repeated productions of the same vowel was as stable as native speakers' productions across length contrasts. These findings are consistent with immature control (execution) over adult-like specification of sequential motor goals (representation) in children, and mature control (execution) over the realization of non-native specification

of sequential goals (representation) in adult second language learners; that is, a double dissociation between the execution and representation of articulatory timing.

The foregoing summary of results elides the finding that age and word length interacted in their effect on production variability (see Fig. 3). The data shown in Figs. 3 and 6 also suggest a main effect of length on production variability: even adult vowel production was more variable in longer words compared to shorter words. Together these results parallel the finding that spatial–temporal variability in children’s and adults’ speech increases with phrase length and morphosyntactic complexity (MacPherson & Smith, 2012; Maner et al., 2000; Sadagopan & Smith, 2008). How should we understand this effect of linguistic complexity and/or length on production given the other evidence for a dissociation between representation and execution? Two alternative explanations are discussed in the remainder of this section: (1) the representation of temporal patterns is disrupted during speech planning under cognitive load; (2) variability is inversely related to automaticity in production.

The explanation that disrupted output representations result in increased production variability follows from models that envision speech planning as a process during which abstract, atemporal phonological representations become phonetically specified. This process, known as phonetic encoding, would seem to imply a role for verbal working memory in speech planning. In particular, phonetic encoding entails that abstract forms remain active while more detailed output representations are built according to language-specific phonetic rules, including those that would specify timing relations (see, e.g., Keating, 1990; Levelt, 1999; Turk & Shattuck-Hufnagel, 2014). Increases in the length or complexity of an abstract representation requires additional working memory resources because it increases the amount of information that must be encoded and thus the amount of time that the representation must remain active. Given that working memory resources are capacity limited, the required additional resources amount to a tax that could disrupt the encoding process by introducing noise. Noise introduced during the encoding process would result in degraded speech plan representations, which would in turn result in more variable productions. Developmental differences in working memory capacity could then account for why variability in children’s productions increases in a linear fashion with word length, but variability in adult productions is only greater in especially long words.

A problem with this explanation of disrupted phonetic encoding is that there is little evidence to support the idea that verbal working memory affects speech production (Gathercole & Baddeley, 1993: Ch. 4; Lee & Redford, 2015). The lack of evidence may be one reason that Baddeley (2003) divides the phonological loop component of his verbal working memory model into “separate storage and articulatory components” (p. 197), with the articulatory component dependent on “language habits” (p. 192). The invocation of language habits in turn suggests a model in which speech plan representations for individual lexical items are not actively built, contra the encoding hypothesis.

The alternative to a model in which timing information is specified during an encoding stage is one where planning is conceived of as a retrieval process; specifically, the retrieval of representations that include information about the relative

timing of speech motor goals. Articulatory Phonology posits exactly this type of representation in the form of gestural constellations (Browman & Goldstein, 1995). These representations automatically generate specific interval timing information for the speech output system based on the relative phasing of gestures (i.e., motor goals) within the constellation. In this way, the constellation is both a linguistic representation and a unit of production that is automatically executed.

A model of production that incorporates linguistic representations of the kind imagined in Articulatory Phonology might explain the effect of length/linguistic complexity on production variability in terms of motor learning. This explanation depends on the assumption that production stability indexes automaticity (see, e.g., Smith & Zelaznik, 2004) and on the assumption that production units increase in size over developmental time with speech motor practice (Redford, 2015; Tilsen, 2016). Given these assumptions, the explanation for the interaction between age and word length on temporal variability is that children have robust representations of monosyllabic word forms as units of production, and execute these automatically (i.e., without feedback control), but their representation of production units decreases in strength with increasing length, and thus so too does the automaticity with which longer words are executed.

Note that this explanation from motor learning predicts second language effects on variability, and so is consistent with Nip and Blumenfeld’s (2015) finding that English-speaking second language learners of Spanish, with (it seems) minimal daily practice in Spanish, produced more variable speech in their L2 compared to their L1. The explanation is also compatible with the findings that increasing phrase length and morphosyntactic complexity affects production variability in adult speech (Maner et al., 2000; Sadagopan & Smith, 2008), since it is unlikely that units of production extend beyond the word.

A motor learning explanation for language effects on variability also predicts that such learning should influence linguistic representation. This prediction is derived from the interaction hypothesis, which is also a hypothesis about the acquisition of representations. By hypothesis, representations like gestural constellations that include relative timing information are low-dimensional specifications of highly practiced articulatory routines. Viewed in this way, it becomes possible to imagine that immature timing control would have specific effects on the representations that emerge with practice. For example, greater temporal variability in the production of individual targets might lead children to aim for more extreme targets than adults. Repeated realization of more extreme targets could result in representations that code for greater contrastive specification of temporal patterns.

Of course, the present results provide no evidence for the influence of immature execution on school-age children’s representation of temporal patterns. There were no interactions between age and the length contrasts on vowel duration, even though children’s productions were more variable than adults’. This result, and the inverse pattern in second language acquisition, suggest that if linguistic representations are also feed-forward plans for articulation, these plans incorporate more information than that which is abstracted from motor practice. One possibility is that, under normal speaking conditions, speech plans are based on the unification of two lexically-linked holistic representations: an abstract practice-based

relative timing representation, which serves as a sort of skeleton for optimal speech action given a particular sequence; and a detailed experience-based phonetic representation, which drives the parameterization of the plan. Though more complicated than a model where plans are more or less directly accessed (e.g., Articulatory Phonology), incorporation of perceptual information into the production process via an experience-based phonetic representation allows both for the dissociation of representation and execution as well as for perceptually-based motor learning and control.

## 5. Conclusion

The aim of the present study was to test for a dissociation between the representation and execution of articulatory timing. The findings presented here from first and second language acquisition largely support such a dissociation. The exception was the finding of an interaction between age and length on variability in vowel production, which suggests effects of representation on execution that are especially pronounced in first language acquisition. We sought to understand this finding in the context of different models of speech production. In models where speech planning is viewed as an active (generative) process, the interaction might be explained as due to working memory-related disruptions in phonological or phonetic encoding. In models where speech plans are retrieved rather than built, influences of language on execution might be explained with reference to practice. Insofar as retrieval models also imply representations of relative timing abstracted over experience, they predict effects of execution on representation, which were not found in the present study. To account for this dissociation, a retrieval model must incorporate some version of the hypothesis that perception guides production.

## Acknowledgments

This work was supported in part by award number R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (PI: Redford) and in part by a fellowship to the first author from the European Institutes for Advanced Study (EUR-IAS), co-funded by the European Commission (Marie-Sklodowska-Curie Actions COFUND Programme FP7). The content is solely our responsibility and does not necessarily reflect the views of our sponsors. We are grateful to Ulrich Mayr and Volya Kapatsinski for discussion and feedback on the earliest version of this manuscript, and to several anonymous colleagues for their careful reviews and extensive comments on subsequent versions of the paper.

## References

- Ackermann, H., & Hertrich, I. (2000). The contribution of the cerebellum to speech processing. *Journal of Neurolinguistics*, 13, 95–116.
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189–208.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.
- Bond, Z. S., & Wilson, H. F. (1980). Acquisition of the voicing contrast by language-delayed and normal-speaking children. *Journal of Speech, Language, and Hearing Research*, 23, 152–161.
- Browman, C. P., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 175–193). Cambridge, MA: MIT Press.
- Buder, E. H., & Stoel-Gammon, C. (2002). American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *Journal of the Acoustical Society of America*, 111, 1854–1864.
- Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition*, 22, 1–26.
- Chakraborty, R., Goffman, L., & Smith, A. (2008). Physiological indices of bilingualism: Oral-motor coordination and speech rate in Bengali-English speakers. *Journal of Speech, Language, and Hearing Research*, 51, 321–332.
- Cho, H., & Shin, J.-A. (2013). L2 Learners' production of the voicing contrast in English word-final stops. *Korean Journal of English Language and Linguistics*, 13, 695–716.
- Dunn, L. M., & Dunn, D. M. (2007). PPVT-4: Peabody picture vocabulary test. Pearson.
- Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, 30, 117–134.
- Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, 89, 395–411.
- Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, 24, 125–146.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). Timonium, MD: York Press.
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Taylor & Francis.
- Goffman, L. (2010). Dynamic interaction of motor and language factors in normal and disordered development. In B. Maassen & P. H. Van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 137–152). Oxford University Press.
- Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43, 239–255.
- Imbrie, A. K. K. (2005). Acoustical study of the development of stop consonants in children (Doctoral dissertation, Massachusetts Institute of Technology).
- Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics*, 18, 321–334.
- Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, 38, 338–350.
- Kent, R. D., & Forner, L. L. (1980). Speech segment duration in sentence recitations by children and adults. *Journal of Phonetics*, 8, 157–168.
- Lee, O., & Redford, M. A. (2015). Verbal and spatial working memory load have similarly minimal effects on speech production. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow. Paper number 0798.
- Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105, 1455–1468.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018–2024.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417.
- Levelt, W. J. (1999). Producing spoken language: A blueprint of the speaker. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford University Press.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Kluwer Academic.
- Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7, 41–74.
- Malécot, A. (1970). The lenis-fortis opposition: Its physiological parameters. *Journal of the Acoustical Society of America*, 47, 1588–1592.
- Maner, K., Smith, A., & Grayson, L. (2000). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research*, 43, 560–573.
- McPherson, M. K., & Smith, A. (2012). Influences of sentence length and syntactic complexity on the speech motor control of children who stutter. *Journal of Speech, Language, and Hearing Research*, 56, 89–102.
- Morris, R. J., & Brown, W. S. (1994). Age-related differences in speech variability among women. *Journal of Communication Disorders*, 27, 49–64.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73–97.
- Munro, M. J., & Derwing, T. M. (1998). The Effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, 48, 159–182.
- Nip, I. S., & Blumenfeld, H. K. (2015). Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners. *Journal of Speech, Language, and Hearing Research*, 58, 653–668.
- Nip, I. S., Green, J. R., & Marx, D. B. (2009). Early speech motor development: Cognitive and linguistic considerations. *Journal of Communication Disorders*, 42, 286–298.
- Pollock, K. E., Brammer, D. M., & Hageman, C. F. (1993). An acoustic analysis of young children's productions of word stress. *Journal of Phonetics*, 21, 183–203.
- Redford, M. A. (2014). The perceived clarity of children's speech varies as a function of their default articulation rate. *Journal of the Acoustical Society of America*, 135, 2952–2963.
- Redford, M. A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics*, 53, 141–152.
- Rosenbaum, D. A. (2009). *Human motor control*. Academic press.
- Sadagopan, N., & Smith, A. (2008). Developmental changes in the effects of utterance length and complexity on speech movement variability. *Journal of Speech, Language, and Hearing Research*, 51, 1138–1151.

- Schwartz, R. G., Petinou, K., Goffman, L., Lazowski, G., & Cartusciello, C. (1996). Young children's production of syllable stress: An acoustic analysis. *Journal of the Acoustical Society of America*, 99, 3192–3200.
- Seddoh, S. A., Robin, D. A., Sim, H. S., Hageman, C., Moon, J. B., & Folkins, J. W. (1996). Speech timing in apraxia of speech versus conduction aphasia. *Journal of Speech, Language, and Hearing Research*, 39, 590–603.
- Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, 10, 209–232.
- Sharkey, S. G., & Folkins, J. W. (1985). Variability of lip and jaw movements in children and adults: implications for the development of speech motor control. *Journal of Speech, Language, and Hearing Research*, 28, 8–15.
- Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, 6, 37–67.
- Smith, B. L. (1992). Relationships between duration and temporal variability in children's speech. *Journal of the Acoustical Society of America*, 91, 2165–2174.
- Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, 41, 18–30.
- Smith, A., & Goffman, L. (2004). Interaction of motor and language factors in the development of speech production. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 227–252). Oxford University Press.
- Smith, B., & Hayes-Harb, R. (2016). Non-native speakers' acoustic variability in producing American English tense and lax vowels. *Journal of the Acoustical Society of America*, 139, 2163.
- Smith, B. L., Sugarman, M. D., & Long, S. H. (1983). Experimental manipulation of speaking rate for studying temporal variability in children's speech. *Journal of the Acoustical Society of America*, 74, 744–749.
- Smith, B. L., Wasowicz, J., & Preston, J. (1987). Temporal characteristics of the speech of normal elderly adults. *Journal of Speech, Language, and Hearing Research*, 30, 522–529.
- Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45, 22–33.
- Song, J. Y., Demuth, K., & Shattuck-Hufnagel, S. (2012). The development of acoustic cues to coda contrasts in young children learning American English. *Journal of the Acoustical Society of America*, 131, 3036–3050.
- Stoel-Gammon, C., & Dunn, C. (1985). Normal and disordered phonology in children. Pro Ed.
- Suomi, K. (2007). On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics*, 35, 40–55.
- Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, 53–77.
- Tingley, B. M., & Allen, G. D. (1975). Development of speech timing control in children. *Child Development*, 186–194.
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28, 1–30.
- Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1658). <http://dx.doi.org/10.1098/rstb.2013.0395>.
- Vihman, M. M. (2014). *Phonological development: The first two years*. Wiley-Blackwell.
- White, L., & Turk, A. E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, 38, 459–471.
- Zsiga, E. C. (2003). Articulatory timing in a second language. *Studies in Second Language Acquisition*, 25, 399–432.