

ARTICLE

A lexical advantage in four-year-old children's word repetition

Margaret CYCHOSZ^{1*}, Michelle ERSKINE², Benjamin MUNSON³, and Jan EDWARDS²

¹University of California, Berkeley, ²University of Maryland, College Park, and ³University of Minnesota, Twin Cities

*Corresponding author: Margaret Cychosz, 1203 Dwinelle Hall, Department of Linguistics, University of California, Berkeley. E-mail: mcychosz@berkeley.edu

(Received 1 May 2019; revised 30 September 2019; accepted 29 January 2020)

Abstract

This study examined a potential lexicality advantage in young children's early speech production: do children produce sound sequences less accurately in nonwords than real words? Children aged 3;3–4;4 completed two tasks: a real word repetition task and a corresponding nonword repetition task. Each of the 23 real words had a paired consonant–vowel sequence in the nonword in word-initial position (e.g., 'su' in ['sutkes] 'suitcase' and ['sudras]). The word-initial consonant–vowel sequences were kept constant between the paired words. Previous work on this topic compared different sequences of paired sounds, making it hard to determine if those results were due to a lexical or phonetic effect. Our results show that children reliably produced consonant–vowel sequences in real words more accurately than nonwords. The effect was most pronounced in children with smaller receptive vocabularies. Together, these results reinforce theories arguing for interactions between vocabulary size and phonology in language development.

Keywords: phonology; nonword repetition; vocabulary size

Introduction

A critical question in child development is how language and speech representations develop throughout childhood. In a traditional view of speech development, children acquire phonology from the bottom up, beginning with the production and perception of individual sounds. Children are predisposed, whether through a Universal Grammar or another innate system of linguistic knowledge to perceive individual phones which they can combine together to construct words and build a vocabulary (Berent, 2013; Dinnsen & Gierut, 2008; Jakobson, 1941/1968; Jusczyk, Smolensky & Alocco, 2002; Tesar & Smolensky, 1998). Alternative accounts suggest that children construct speech representations gradually, by generalizing over language chunks such as words or syllables (Edwards, Beckman & Munson, 2004; Ferguson & Farwell, 1975; Fowler, 1991; Gathercole, Willis, Emslie & Baddeley, 1991; Metsala & Walley, 1998; Ngon et al., 2013; Walley, 1993; Vihman, 2017; Vihman &

© Cambridge University Press 2020

Velleman, 1989). This developmental pattern relies on a mix of bottom-up and top-down learning. If children rely on lexical items and sublexical sequences such as syllables to construct speech representations, we should anticipate that children will repeat sound sequences contained in real words more accurately than sequences contained in nonwords. Furthermore, we also anticipate an interaction between production accuracy with children's vocabulary size (Edwards *et al.*, 2004; Metsala, 1999). Specifically, if speech segments emerge from generalizations made over children's lexicons, then children with larger vocabularies should have more abstract segmental representations. This is because children with larger vocabularies have more diverse words (*i.e.*, types) from which they can infer how important a given sound segment may be for their native language(s). We test these hypotheses in four-year-old children who completed nonword and real word repetition tasks and found that children produced consonant-vowel (CV) sequences in real words more accurately than in nonwords. Performance interacted with children's vocabulary size: the disparity between CV sequences contained in real words and nonwords was less apparent in children with larger receptive vocabularies, a correlation that we take as evidence for the primacy of the lexicon in early phonological development.

Traditional views on phonological development, often espoused in formal linguistic theories, suggest that phonology begins with the production and perception of individual sounds. Infants have access to a universal inventory of segments that are pruned according to what they do or do not hear in their ambient language. Eventually, children string these segments together to construct words and build a vocabulary (Berent, 2013; Dinnsen & Gierut, 2008; Jakobson, 1941/1968; Juszyk *et al.*, 2002; Tesar & Smolensky, 1998). Young learners anticipate strings of segments because they are predisposed to parse the phone as the most basic unit of language. This view argues that domain-general perceptual-motor constraints are insufficient to explain cross-linguistic similarities in language production over the first years of life. Instead, an internalized algebraic mechanism, present from birth, may combine base units of language, phones, into words (Berent, 2013). This could explain universal predispositions to partition speech into phone-sized segments and combine them via ordered rules. For example, neonates (mean 1.14 days) can discern sequence-final and sequence-initial sequences (*e.g.*, AAB and ABB) from random co-occurrences in an artificial grammar-learning task, suggesting a universal ability to construct phonological rules (Gervain, Berent & Werker, 2012).

Complementary evidence that a segmentalized phonology may be present early in development comes from the psychology tradition. Infants' perceptual narrowing to their native phoneme categories over the first year of life has shown that, by six months, infants are sensitive to many individual vowels and consonants in their native language (Kuhl, Conboy, Padden, Nelson & Pruitt, 2005; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008; Maye, Werker & Gerken, 2002, *inter alia*). However, infants' performance may be more variable than adults' (Polka & Bohn, 1996). Furthermore, the mispronunciation paradigm has been used to show that children as young as 14 months are sensitive to even one-feature changes in familiar words (Swingle & Aslin, 2002).

The hypothesis that segments are a basic unit in speech perception does not imply that the segment is innate. This bottom-up approach nevertheless does invite further explanation. Unlike traditional theories in linguistics, research from the psychology tradition proposes that statistical, domain-general learning, and not predisposed language learning and parsing mechanisms, is what explains infants' patterns. For

example, infants may learn the phones of their native language on the basis of phone and syllable co-occurrence patterns in the ambient speech stream (e.g., Saffran, 2003). In English, some syllables, such as /li/, are much more likely to occur word-finally (i.e., [lʌvli] 'lovely') than word-initially (i.e., [lɪmə] 'lemur') or -medially (i.e., [kə'lɪdʒəl] 'collegial'). On the basis of this information, a young learner could suppose that the syllable or phone that followed /li/ in the speech stream marked the beginning of a word. If that word formed a member of a minimal pair, the child could begin to delimit a segmental phonology. However, much of this research nevertheless suggests that infants rely on segments, more than words or even syllables, to construct a phonology (Newport, Weiss, Wonnacott & Aslin, 2004).

An alternative account of speech development proposes that children gradually build a segmental phonology from holistic chunks of language such as syllables and words (Edwards et al., 2004; Ferguson & Farwell, 1975; Fowler, 1991; Gathercole et al., 1991; Lindblom, 1992; Metsala & Walley, 1998; Walley, 1993; Vihman, 2017; Vihman & Velleman, 1989), chunks resultant from incorrect word segmentation (Ngon et al., 2013), or multi-word phrases (Arnon, 2010; Arnon & Christiansen, 2017). Over time, these chunks individuate, and gradually resemble the segments that characterize adult phonology. Specifically, children calculate the statistical properties from their speech stream and form phonological categories around segments that are the most meaningful (Pierrehumbert, 2003). A meaningful segment could come in several different forms. It could occur frequently in the speech directed to the child (Edwards & Beckman, 2008; Zamuner, 2009). A meaningful segment could help the child contrast many different words in their ambient language/lexicon (i.e., functional load [Stokes & Surendran, 2005; Van Severen, Van Den Berg, Molemans & Gillis, 2012]). Or a meaningful segment could occur in a dense phonological neighborhood, necessitating that the child analyze the internal structure of the word forms in the phonological neighborhood to differentiate them from one another (Charles-Luce & Luce, 1990; Storkel, 2002). For these reasons, children with larger vocabularies, who have more word types to analyze and derive phonological statistics over, have been shown to have more segmental speech (Noiray, Popescu, Killmer, Rubertus, Krüger & Hintermeier, 2019). Consequently, while much language acquisition research has traditionally delimited the child's lexicon and phonology to separate areas of study, this proposal assumes interdependence between the areas (Sosa, 2008; Stoel-Gammon, 2011; Storkel & Morissette, 2002).

At least two findings support lexical-phonological interactions during speech development. First, at age 2;0, precocious talkers (children with significantly larger vocabularies than the norm) show different patterns of speech development relative to other children aged 2;0 (age-matched peers) and similar patterns of speech development relative to children aged 2;6 (vocabulary-matched peers). At age 2;0, precocious talkers have larger phonetic inventories and more complex syllable structures than their age-matched peers, but similar inventories and syllable structures to their vocabulary-matched peers (Smith, McGregor & Demille, 2006). Elsewhere, two-year-old children who were late talkers, characterized as having "limited lexical ability" (Rescorla & Ratner, 1996:154), showed a similar pattern: those children had smaller phonetic inventories and less complex syllable structures relative to their age-matched, typically-developing peers (e.g., Pharr, Ratner & Rescorla, 2000; Rescorla & Ratner, 1996). Finally, this predictive effect of vocabulary

size has also been observed in perception tasks: at 14 months, infants' ability to recognize minimal pair differences is related to the size of the receptive vocabularies (Werker, Fennell, Corcoran & Stager, 2002).

Second, children's speech production is susceptible to a variety of word and phonological frequency effects. Three- to 8-year-old children produce high frequency biphone sequences more accurately than low frequency sequences (Edwards *et al.*, 2004; Munson, 2001). These high frequency biphone sequences were also shorter in duration than the low frequency counterparts. Duration of biphone sequences, as the authors suggest, is an acoustic measure that reflects how quickly speech is accessed and articulated. Consequently, shorter sequence duration should indicate a more fluent production. So the finding that children were faster in their productions of high frequency sequences suggested that these frequent sequences were mastered more, a process that was gradual and emerged as the children had more opportunity to use the high-frequency sequences (see also Zamuner, 2009). High frequency sequences are more common in words that children have heard and produced, thus affording them additional practice relative to less-familiar, low-frequency sequences (Edwards, Beckman & Munson, 2015; Zamuner, Gerken & Hammond, 2004).

Edwards *et al.* (2004) also found that the effect of phonotactic transition probability on repetition accuracy was correlated with vocabulary size: the effect of probability on production was smaller for English-speaking children with larger vocabularies than those with smaller vocabularies. Furthermore, the cross-linguistic finding that children learning different languages master the same consonants at different rates is partly attributed to the type, but not token, frequencies of words containing these sounds in the respective lexicons of each language (Beckman & Edwards, 2010). Finally, the predictive role of lexical frequency has been attested in speech perception tests. Second and fourth grade students had higher phone awareness (*i.e.*, the ability to remove a speech sound from a word 'cat' → 'at'), for sounds from dense lexical neighborhoods compared to sounds from sparse lexical neighborhoods (Hogan, Bowles, Catts & Storkel, 2011).

Present research

The current study is designed to extend previous research on interactions between word and speech development by directly examining how existing lexical representations support speech sound production. The secondary interest is to explore how this lexicality effect might interact with vocabulary size. We tested these questions by comparing children's production accuracy of CV sequences in real words and nonwords. The nonword repetition (NWR) task, which asks participants to repeat nonwords after a model speaker, has been employed extensively in child language research (Chiat & Roy, 2007; Dispaldro, Deevy, Altoé, Benelli & Leonard, 2011; Gathercole *et al.*, 1991; Metsala & Chisholm, 2010; Roy & Chiat, 2004). The difficulty of nonword production stems from the demands of encoding and executing a sequence of sounds, the nonword, without the support of a lexical representation (Chiat & Roy, 2007; Gathercole *et al.*, 1991).

Previous studies on this topic have not directly compared children's productions in real words and closely-controlled nonwords. Chiat and Roy (2007) tested real word and nonword production accuracy in children aged 2;0-4;0, but the phone sequences in test items were not consistent between word types. For example, the corresponding nonword for *banana* might be created by scrambling consonants (*e.g.*, *banana* > /nə'nabə/).

Furthermore, Chiat and Roy did not examine whether there was an interaction between the effect of lexicality and vocabulary size. Keren-Portnoy, Vihman, DePaolis, Whitaker and Williams (2010) compared two-year-olds' ($n = 15$) production accuracy of real words, nonwords containing consonants familiar to the child, and nonwords containing consonants unfamiliar to the child. The authors found that children produced the real words more accurately than either set of nonwords. However, for the central analysis, the consonants were not controlled between word types. Furthermore, when target consonants were controlled between real words and nonwords in a subanalysis (2010:1287-1288), the consonants were selected from each child's idiosyncratic phonological repertoire, making the results difficult to generalize and replicate.

We remedy these issues in an experimental design consisting of two tasks, real word repetition and nonword repetition, to evaluate the representation of phonology in English-speaking children aged 3;3-4;4. We tested an age group that both fell within the range of previous research (Chiat & Roy, 2007) and reliably produced a variety of sounds in word-initial position. We examined the children's production of the same CV sequences in nonwords and corresponding real words. In doing so, the articulatory difficulty of the sounds was controlled and we could isolate a lexicality effect. This comparison can help put the words versus sounds debate to the test: if children are more accurate at producing the CV sequence in the real word (e.g., [taigəɪ] 'tiger') than the corresponding nonword (e.g., [tablor]), after we control for stress, prosodic structure, and phonotactic transitional probability, then this LEXICAL ADVANTAGE could be evidence that these four-year-old children are not reliably constructing their phonology from individual sounds. Instead, they tend to rely on language chunks such as words and syllables to construct a phonology (Beckman & Edwards, 2010; Gathercole et al., 1991; Munson, Edwards & Beckman, 2011; Stoel-Gammon, 2011). Consequently, the current study is able to contrast strict, bottom-up theories (e.g., Berent, 2013, Tesar & Smolensky, 1998) with top-down, word-first theories but cannot demarcate between interactive bottom-up theories that permit representations beyond biphoneme sequences. According to a theory where children rely on chunks to acquire speech categories, children should find nonwords more difficult to process and repeat because children do not have any memory traces of the nonwords. As a result, they cannot rely on entire episodic traces during lexical retrieval and production. Furthermore, this theory predicts that even though subcomponents of nonwords may have some mental representation – because they form parts of real words that the children know – those representations are likely fainter and the memory traces less established because the sequences are less frequently heard and produced.

Finally, in line with previous work (Edwards et al., 2004; Metsala, 1999; Storkel & Morissette, 2002), we hypothesized that the lexical advantage would be more pronounced for children with smaller vocabularies. This is because these children have fewer word tokens over which they can generalize to calculate statistics concerning which segments are most meaningful for their language and construct a segmental phonology, as is consistent with many emergentist models of language development (McAllister Byun & Tessier, 2016; Menn, Schmidt & Nicholas, 2013; Pierrehumbert, 2003; Vihman, 2017). In addition, because the lexicons of children with smaller vocabularies tend to revolve around denser phonological neighborhoods (Storkel, 2004; Stokes, 2010), these children also have access to fewer word types to construct a phonology and complete the current experimental tasks.

Methods

Participants

Eighty-six children aged 3;3 to 4;4 (years;months, mean = 3;7, 51 girls, 35 boys) participated in this study. Through the universities' IRB-approved process, all families consented to participate in the research upon their initial visit to the lab. All participants were monolingual speakers of English and passed a hearing screening in at least one ear at 25dB for 1000, 2000, and 4000Hz prior to the study. All children had normal speech and hearing development, as reported by the child's primary caregiver.

Caregivers also reported socioeconomic status, quantified as mother's education level. Maternal education level was binned into seven levels: 1) less than high school, 2) high school equivalent certificate (e.g., General Education Development [GED]), 3) high school diploma, 4) technical-associate degree, 5) some college (2+ years)/trade school, 6) college degree, and 7) graduate degree. We attempted to include children with caregivers from a diverse educational background. To accomplish this, participants were recruited from a database of families at the University of Wisconsin or University of Minnesota. Participants were additionally recruited at community centers with demographics that showed a high percentage of children from diverse backgrounds. We additionally made connections with local community partners including beauticians, religious organizations, local libraries, and community programs who specialized in supporting children from diverse language and cultural backgrounds, and through social media accounts. Still, 38% (n = 33) of mothers had a graduate degree, 35% (n = 30) a college degree, 16% (n = 14) some college (2+ years)/trade school, 0% a technical/associate degree, 7% (n = 6) a high school diploma, 1% (n = 1) a GED or high school diploma equivalent certificate, and 1% (n = 1) had less than a high school diploma (one caregiver did not report).

Tasks

Prior to the experimental tasks, children completed a battery of standardized, language-related assessments. The language assessments were conducted at the University of Wisconsin or the University of Minnesota. Receptive vocabulary inventory was measured via the *Peabody Picture Vocabulary Test, 4th edition* (PPVT-4) (Dunn & Dunn, 2007) (standard score mean = 120.5, SD = 17, range = 76-151). Expressive vocabulary inventory was measured via the *Expressive Vocabulary Test, 2nd edition* (EVT-2) (Williams, 2007) (standard score mean = 119, SD = 18, range = 75-156). For the current study, the PPVT-4 and EVT-2 standard scores were employed, instead of growth scale values which transform standard scores to correspond to age-related development, as child age is included as a parameter in the statistical modeling. Additional tasks that the children completed included other measures related to vocabulary and speech development. Children were participating in a longitudinal study. All of these data, the experimental tasks and tests of vocabulary size, were collected during the second of three time-points, all of which were spaced approximately one year apart.

In the experimental phase, children completed two picture-prompted, verbal production tasks: a nonword repetition (NWR) and real word repetition (RWR) task, where participants repeat nonwords or real words after a model speaker. The experiments and pre-task assessments were completed in two 1-hour test sessions on

different days, usually about one week apart. In all cases, children completed the real word repetition task during the first testing session and the nonword repetition task during the second. The word repetition tasks were carried out in sound treated rooms at the University of Wisconsin and the University of Minnesota.

The nonword and real word repetition tasks were separated for several reasons. First, nonword and real word tasks are conventionally given separately (Chiat & Roy, 2007). Second, given the participants' age, child fatigue and restlessness were real concerns so we needed to limit experiment duration. Third, combining real words and nonwords into a single experiment risks real words priming their corresponding nonwords or vice versa. Recall that for four-year-old children, many words encountered are novel and the representations even of the real words that they know may still revolve around just a few exemplars. In the experiment, children were instructed to repeat the "silly" words (nonwords) because we did not want children to attempt to adapt the nonwords to more familiar real words. In our experience working with speech production in children of this age, we have found that children may misperceive high-phonotactic transition probability nonwords as real words and that children do not easily switch back and forth between real words and nonwords. Nevertheless, it is clear that the nonword task was different from the real word task and we address this point in more detail below.

Materials

Word stimuli came from recordings of real words and nonwords elicited using the picture- and audio-prompted repetition task described in Edwards & Beckman (2008). The images used for the RWR task were color photographs of the object names. The images used for the nonword task were color photographs of unfamiliar tools, plants, animals, etc.

The presentation of visual stimuli with the real words presents children with another way to access the real word form, in addition to the auditory stimulus. For example, if a child misheard a real word, the visual stimulus could reinforce the real word. However, the visual stimuli for the nonwords do not assist lexical retrieval. This makes the real word and nonword tasks less equivalent. Still, this confound is likely a limitation of any study pitting real words and nonwords; the alternative, not presenting images, would make the task difficult for toddlers to complete. All stimuli images and sound files are archived on our Open Science Framework project: <https://osf.io/dcb9p/> and are available to use for replication.

The NWR task consisted of 67 nonwords, including the 44 nonwords from Edwards et al. (2004), and an additional 23 nonwords used in the current analysis. Each of the 23 nonwords for the current study had a corresponding real word elicited in the RWR task (e.g., *taiblor* [taibloɹ], *tiger* [taɪgəɹ]) (Table 1) for a total of 46 words used in the current study. IPA transcriptions of real words are included in appendix A. In each word, the target CV sequence was in word-initial position. All words were bisyllabic with penultimate stress. The CV sequence was always an open syllable followed by a consonant (CV.C(C)).

Each of the 23 real words was chosen from lists such as the "Toddler Says" portion of the MacArthur Bates Communicative Development Inventory (Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007), to ensure familiarity to the majority of children aged 3;3-4;4. Combined with training trials, the entire NWR task consisted of 73 trials (6 training) and the RWR task 94 trials (4 training). There are more real words

Table 1. Nonword and corresponding real word stimuli employed in repetition tasks

Real word	Nonword	Real word	Nonword
candle	kæmig 'ka-mig'	sidewalk	saiɹɒt 'sigh-prote'
chicken	tʃimig 'chih-mig'	sister	sɪplɒk 'sih-plok'
coffee	kʌsep 'kah-sep'	suitcase	sudras 'soo-dras'
cutting	kʌfɪm 'kuh-feem'	sunny	sʌbiθ 'suh-bith'
kitchen	kɪpɒn 'kih-pon'	tiger	taiɹlɒr 'tie-blɒr'
rabbit	ræpɔɪn 'ra-poin'	toaster	tozel 'toe-zell'
raisins	reɪθ 'rae-bith'	toothbrush	tugraɪf 'too-grafe'
reading	rɪfrʌs 'ree-fras'	waiting	weməg 'way-mahg'
rocking	rʌlɪd 'rah-lide'	washer	wʌkræd 'wah-krad'
running	rʌglok 'ruh-glok'	water	wʌpɒt 'wah-prote'
sandwich	sæmel 'sa-mell'	window	wɪmel 'wih-mell'
sharing	ʃevʌs 'shae-vahs'		

than nonwords because we also used the real words for a fine-grained acoustic analysis of particular phonological contrasts that are irrelevant for this investigation. The largest concern about having more items to repeat in the real word task is that the real word task is then longer. This longer session could potentially have induced more fatigue in the children than the shorter NWR task. However, the only consequence of this possible fatigue would be to reduce the size of the lexicality effect, and not the direction of the difference between NWR and RWR accuracy, or the influence of other measures on the size of this effect.

Though the CV sequence was constant between paired real words and nonwords (e.g., [t-ai] 'tiger', [t-ai] [taiɹlɒr]), the phonotactic transition probability between the target CV sequence and the first segment of the remaining word varied between real words and nonwords. For example, the transition between /a/ and /ʃ/ in the real word /wʌʃəɪ/ is less phonotactically probable (represented by the log-transformed probability, -10.47) than the transition between /a/ and /k/ in the corresponding nonword /wʌkɪæd/ (-9.88), where a value closer to 0 indicates that the transition is more phonotactically probable.

We computed the phonotactic transition probability between the V of the initial CV sequence and the onset C of the second syllable (e.g., transitional probability between /a/ and /ʃ/ in the real word /wʌʃəɪ/) using the Hoosier Mental Lexicon Database (Pisoni, Nusbaum, Luce & Slowiaczek, 1985). We did not take stress into account when making the transitional probability calculations from the database. The phonotactic transition probability was only computed for the 23 matched real words and nonwords. Again, a value closer to 0 (i.e., the log of 1) indicates more phonotactically probable.

The mean phonotactic transition probability for real words was greater than the phonotactic transition probability for nonwords (real words: -7.88 , nonwords: -8.22) but the median probability was greater for nonwords (real words: -7.84 , nonwords: -7.72). Consequently, we included Phonotactic Transition Probability as a

parameter in our baseline statistical model fitting procedure to ensure that any effect of word type was independent of transitional probability. Phonotactic Transition Probability did not improve on baseline model fit suggesting that repetition was independent of transition probability; see Results for details. The script used to calculate phonotactic transition probability is included in the accompanying OSF project; see Edwards et al. (2004) for further details about calculation.

In addition to the phonotactic transition probability, the frequency of the second syllable in the nonwords and real words also varied, potentially explaining repetition accuracy between lexical items. To ensure that second syllable frequency did not predict repetition accuracy, we computed the log frequency of the second syllable of each real word and nonword using the Hoosier Mental Lexicon Database (Pisoni et al., 1985). For example, we computed the frequency of [kɪŋ] from the real word 'rocking' and [laɪd] from the corresponding nonword [ralaɪd] in the database. These frequencies were then log-transformed. The second syllables in the real words were, on average, more frequent than the second syllables in the nonwords (real word mean: 2.96, median: 2.94, range: 0-6.4; nonword mean: 1.11, median: 0.69, range: 0-3.83).¹ This mismatch between second syllable frequency in the real words and the nonwords was due, in large part, to the frequent nominalizer [ə] which was present in syllables such as [stə] in 'toaster'. As a result of the difference between the test items, the parameter second syllable frequency was incorporated into the statistical modeling. Second syllable frequency mildly improved baseline model fit (models fit without parameters of interest such as word type [real versus nonword]). However, the effect of word type outweighed the effect of second syllable frequency, suggesting that repetition accuracy was not best explained by the frequency of the second syllable of the words; see Results for details.

A young female speaker of Mainstream American English or African American English provided the recordings for the word stimuli used in the tasks. All recordings were digitized at a sampling frequency of 44,100 Hz using a Marantz PMD671 solid-state recorder. Amplitude was normalized within and between the real- and nonwords. The speech stimuli were naturalistic so duration of the CV sequence was not normalized between the real- and nonwords. We decided to use naturalistic stimuli instead of synthesized speech as we hypothesized that part of the advantage children would exhibit in real word activation and production might be due to the coarticulatory and durational cues present in real words produced in naturalistic speech (Mahr, McMillan, Saffran, Ellis Weismer & Edwards, 2015). Furthermore, children perform worse on myriad speech perception tasks constructed with synthesized and modified speech stimuli than naturalistic (Coady et al., 2007; Evans et al., 2002).

We still wanted to be able to attribute children's repetition accuracy to the lexical status of the stimuli, however. So we measured the duration of the CV sequence and the coarticulation between the consonant and vowel in the CV sequence in real words and nonwords in the Mainstream American English and African American English stimuli (Table 2). Given the small units of measurement, we quantified

¹Ten of the second syllables in the nonwords were not present in the Hoosier Mental Lexicon Database, making it impossible to log-transform their frequencies. Thus, for these items, the frequency was untransformed (kept at 0). Additionally, the real word lexical items 'raisins' and 'sidewalk' were not present in the Hoosier Mental Lexicon database. For the purposes of the second syllable frequency calculation, the frequency of those real word lexical items was set at 1 (before log transformation).

Table 2. Durational and coarticulatory properties of CV sequences in nonwords and real words

	Mainstream American English		African American English	
	Mean (SD)	Median	Mean (SD)	Median
CV DURATION				
REAL WORDS	305.68(73) ms	296.41	245.67(55)	247.56
NONWORDS	332.50(87)	318.52	264.99(64)	261.47
CV COARTICULATION				
REAL WORDS	29.83(12) ms	30.00	25.77(10)	25.50
NONWORDS	29.48(12)	30.00	26.23(9)	28.50

coarticulation as the transition duration between the consonant and vowel with the onset of the transition beginning at the point of maximum acoustic similarity between the consonant and vowel (see Gerosa, Lee, Giuliani & Narayanan, 2006 and Cychosz, Edwards, Munson & Johnson, 2019 for further details on the coarticulation calculation including validation of the measure on these lexical stimuli).

In both dialects, the mean and median durations of the CV sequence were slightly shorter in the real words than the nonwords. However, a paired two-tailed t-test demonstrated that this difference did not approach significance: African American English ($t_{(22)} = 0.93$, $p = .36$) and Mainstream American English ($t_{(22)} = 1.33$, $p = .20$). The CV sequences were not reliably shorter in the real words than the nonwords in either dialect. Another paired two-tailed t-test also showed that the difference between CV coarticulation in nonwords and real words was insignificant: African American English ($t_{(22)} = .13$, $p = .90$) and Mainstream American English ($t_{(22)} = -.10$, $p = .92$). There may still be some differences between the CV sequences in the nonword and real words that the children can use as a perceptual cue. However, as neither duration nor CV coarticulation varied, we confidently proceeded with the naturalistic experimental stimuli.

Children received the task in their native dialect (either mainstream American English [$n = 77$] or African American English [$n = 8$]), as determined by observing the mother-child interaction at the beginning of the first session in the longitudinal study. No specific features of AAE were included in the stimuli. Stimuli differences were in intonation and voice quality rather than in segmental or morphosyntactic features. However, we did ensure that diphthongs were not monophthongized as monophthongs in the AAE stimuli would have amplified differences between the dialects. There were also not any relevant dialect differences for the initial CV sequences in children's productions. We did not include dialect as a parameter in our statistical modeling as it was confounded with SES and vocabulary level.

Task procedure

Each participant was seated in a quiet room with at least two experimenters who guided the task. For the RWR task, a photo of each word appeared on a computer screen in front of the child while the word played simultaneously over external speakers. At word offset, the child repeated the word that they heard. An identical procedure was

used for the NWR task: a photo of an unfamiliar object appeared on the screen while the nonword played over external speakers. The experimenter instructed the child to repeat the “silly” name of the object on the screen as best as possible. Children were encouraged to respond on the first trial. A second trial was allowed only if the child did not attempt the word on the first trial. 3.54% of the scored real words were second trial repetitions and 4.10% of the scored nonwords were second trial repetitions. After each trial, the experimenter manually advanced to the subsequent trial. Stimuli (see Table 1) were presented randomly using E-prime software (Schneider, Eschman & Zuccolotto, 2012).

Data coding and analysis

All words were first segmented in Praat (Boersma & Weenink, 2018). Each CV sequence was transcribed by a trained phonetician who is a native speaker of American English. Transcription was conducted auditorily and by reviewing the associated acoustic waveform. Given the children’s young age, some children needed to repeat a word to render a clear recording. Still, every effort was made to use the first word produced for the analyses to avoid a practice effect.

Scoring was conducted in a feature-based system for the individual segments of the CV sequence. The use of features was meant to provide a more fine-grained description of children’s production accuracy and does not reflect a theoretical stance. The alternative, coding binarily, would mark both the productions of [putkes] and [zutkes] for ‘suitcase’ as incorrect, though the latter approximates the target ‘suitcase’ more closely.

The features of each segment (3 features each for consonants and monophthongs [e.g., [a], 4 features for diphthongs [e.g., [aɪ]) were binarily evaluated as [+/- correct] in the following way for consonants: 1) place, 2) manner, and 3) voicing and for monophthongal vowels: 1) length 2) height, and 3) backness. For diphthongal vowels an additional point was evaluated for backness of the off-glide. Then, structural accuracy was binarily evaluated: 1) were the correct number of syllables produced? [yes/no], 2) was the consonant of the initial CV sequence produced in the correct position? (e.g., C in word-initial position) [yes/no], and 3) was the vowel of the initial CV sequence produced in the correct position [yes/no]. The result was a maximum score of 9 for CV sequences with monophthongal vowels, and a maximum score of 10 for CV sequences with diphthongs.

To ensure transcription accuracy, a second transcriber, also a trained phonetician and native American English speaker, transcribed a 10% subset of the original words. An intraclass correlation (ICC) statistic assessed inter-rater agreement. ICC is regarded as an improvement upon traditional Cohen’s Kappa score, as the latter does not account for chance agreement given that the measurement assumes discrete dependent variables. This artificially inflates the agreement score (Mandrekari, 2011). The ICC between raters was 0.881, which was significantly greater than chance ($F(374,375) = 15.9, p < .001, 95\% \text{ CI} = 0.86, 0.90$) and in the range that Cicchetti (1994) considers excellent. The transcribers were not blind to the purpose of the study. However, to annotate each CV sequence, a script separately queried the two sounds of the target sequence rather than the whole word. For consonants, transcribers separately judged place, voice, and manner in an iterative process. Similarly, for vowels, transcribers separately judged height, frontness-backness, and

tense-lax-diphthong quality. At no time did transcribers simply judge sounds as correct or incorrect. This process should reduce any possible transcriber bias.

Results

A mixed effects linear regression model was fitted to evaluate the predictors of children's production accuracy using lme4 (Bates, Maechler, Bolker & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff & Christensen, 2017) packages using the R programming language (R Core Team) in the Rstudio computing environment (Version 1.2.1335). The significance of potential model parameters was determined using a combination of log-likelihood comparisons between models, AIC estimations, and p-values procured from lmerTest. The alpha level for log-likelihood comparisons was corrected to .0125 to account for the multiple comparisons (.05/4 for four planned tests, including interactions).

First we fit the base model which included only the random effects of WORD and individual PARTICIPANT to account for the repeated measures design. Then, the following fixed parameters were examined in a forward testing procedure: PHONOTACTIC TRANSITION PROBABILITY, SECOND SYLLABLE FREQUENCY, AGE (in months)², MATERNAL EDUCATION LEVEL, GENDER (reported by caregiver), WORD TYPE (real word versus nonword), EXPRESSIVE VOCABULARY SIZE (EVT-2 standard score), and RECEPTIVE VOCABULARY SIZE (PPVT-4 standard score). Predictors that did not improve upon model fit – a likelihood test or AIC value did not indicate that the model fit was significantly improved – were removed from further analysis.

Maternal Education Level and Gender are demographic variables that have been shown to affect vocabulary size. Including these variables in the model before adding vocabulary size allows us to assess whether any effects of vocabulary size on repetition performance are secondary to one or both of these demographic variables. Though we attempted to capture a range of maternal education levels in our sample, mothers with lower education levels were underrepresented. Our original 6-level **Maternal Education Level** variable, as described in Methods, did not improve model fit. Still, because lower educational levels were underrepresented, we wanted to ensure that the insignificance of **Maternal Education Level** was not due to sampling error. Consequently, we planned an additional model comparison where we condensed **Maternal Education Level** from 6 levels to 4 levels in the following manner: less than high school/high school equivalent certificate/high school diploma (low), some college/trade school (middle), college degree (high), and graduate degree (highest). **Maternal Education Level** still did not improve upon model fit so we confidently excluded it from all further analyses.

We also wanted to ensure that any effect of **Word Type** was independent from the effects of transitional probabilities between the CV sequence and the first consonant of the carrier word, as well as the frequency of the second syllable in the stimulus word, as described in the Methods. The parameters **Phonotactic Transition Probability** and **Second Syllable Frequency** were added to the baseline model before our variables of interest – **Word Type** and the measures of vocabulary. **Phonotactic Transition Probability** did not improve model fit indicating that the difference between repetition accuracy on nonwords and real words could not be predicted by the transitional probability between CV sequences and the first consonant of the carrier

²Refers to the age (in months) that the child took the PPVT-4 test.

word. Furthermore, it also demonstrated that the effect of **Word Type** was independent of the transitional probabilities between CV sequences and the first consonant of the second syllable of the carrier word.

Second Syllable Frequency improved upon a baseline model with only random effects of **Word** and **Participant** suggesting that children repeated words better when the second syllable was more frequent ($\beta = 0.76$, $t = 3.99$, $p < .001$). However, once the parameter **Word Type** was added to the model, **Second Syllable Frequency** was no longer significant ($>.05$). Additional log-likelihood testing showed that a model with **Second Syllable Frequency** only marginally improved upon a model with **Word Type**, **Receptive Vocabulary Size** and **Age** ($\chi^3 = 4.09$, $p = .04$), an insignificant comparison under the 0.0125 alpha criterion for multiple comparisons. We thus conclude that although **Second Syllable Frequency** may be positively correlated with repetition accuracy, **Word Type** – real word versus nonword – remains the best predictor of repetition accuracy for this task.

The final, best-fitting model included the main effects of **Word Type**, **Receptive Vocabulary Size**, and the interaction of **Word Type** with **Receptive Vocabulary Size**. We additionally included **Age** (in months) to control for age-correlated articulatory maturity. Absent the **Receptive Vocabulary Size** parameter, **Expressive Vocabulary Size** likewise improved model fit. However, following the addition of the interaction between **Receptive Vocabulary Size** and **Word Type**, **Expressive Vocabulary Size** did not further improve fit. Furthermore, **Receptive Vocabulary Size** improved upon the model containing **Age** and **Gender** more than **Expressive Vocabulary Size** (AIC = -5392.2 for Receptive versus AIC = -5386.2 for Expressive). On the basis of this analysis we confidently concluded that **Receptive Vocabulary Size** was a better predictor of repetition accuracy than **Expressive Vocabulary Size**. The finding that receptive vocabulary size predicted repetition accuracy better than expressive vocabulary size contrasts with several previous studies that determined that expressive vocabulary was the better predictor (Edwards & Lahey, 1998; Edwards et al., 2004). We discuss this difference further in the Discussion.

No additional predictors improved model fit and the **Word Type** by **Receptive Vocabulary Size** was the only pre-planned interaction tested. Although it did not improve upon model fit, **Age** (in months) was included in the final model to control for age-correlated maturity. Analysis and modeling scripts are archived and publicly available on our Open Science Framework project: <https://osf.io/dcb9p/>.

Table 3 lays out a summary of the final model predicting children's word repetition accuracy. Continuous predictors were mean-centered to facilitate model interpretation. There is a significant main effect of **Word Type**, with nonword as a reference level, indicating that children produce CV sequences in nonwords less accurately than in real words. There is a significant main effect of **Receptive Vocabulary Size**, indicating that accuracy for CV production in both word types increases as the size of children's receptive vocabulary increases. Finally, there is a significant interaction between **Word Type** and **Receptive Vocabulary Size**, indicating that the effect of **Word Type** decreases as receptive vocabulary size increases. **Figure 1** illustrates how the lexicality effect varies as a function of vocabulary size: children with smaller receptive vocabularies show a larger disparity in CV repetition accuracy between nonwords and real words than children with larger receptive vocabularies.

³Beta coefficients, standard errors, and confidence intervals multiplied by 100 to ease model interpretability.

Table 3. Model fit predicting CV repetition accuracy in nonwords and real words

	β	S.E.	t	p	CI
Intercept	96.67	0.56	171.88	<.001***	95.58, 97.76
Word type[nonword]	-3.19	0.62	-5.13	<.001***	-4.39, -2.00
Receptive vocabulary size (PPVT-4)	0.06	0.03	2.37	0.019*	0.01, 0.11
Word type[nonword] *PPVT-4	0.14	0.02	6.35	<.001***	0.10, 0.18
Age	0.02	0.11	0.20	0.840	-0.19, 0.23

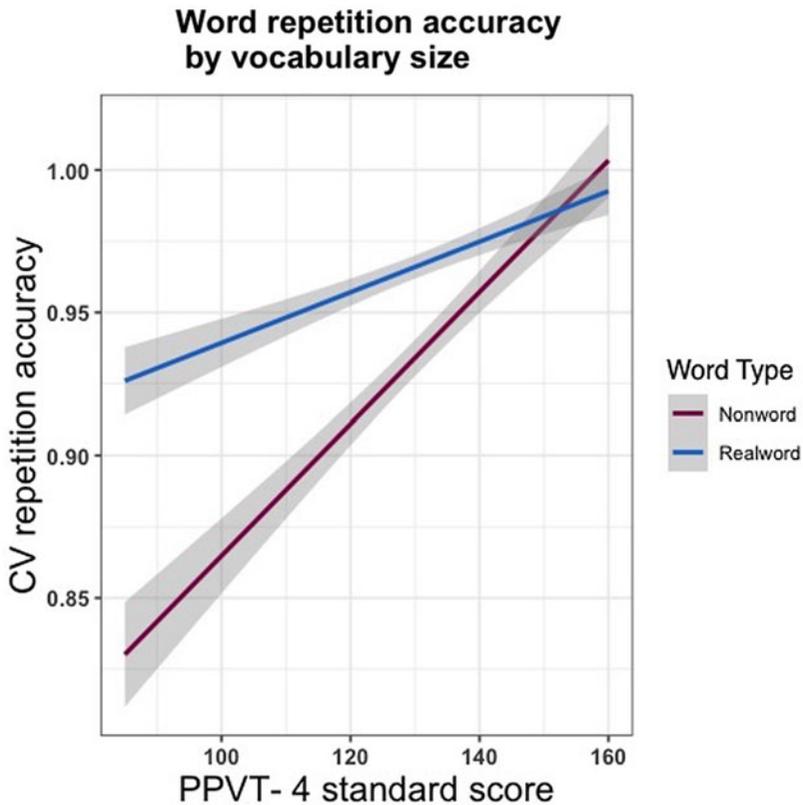


Figure 1. Consonant-vowel repetition accuracy by receptive vocabulary size

The **Gender** covariate, with female as the reference level, approached significance in the model (female: $\beta = .02$, $t = 2.19$, $p = .03$), but was not ultimately included as a factor. However, [Figure 2](#) displays a trend showing that female children tended to be more

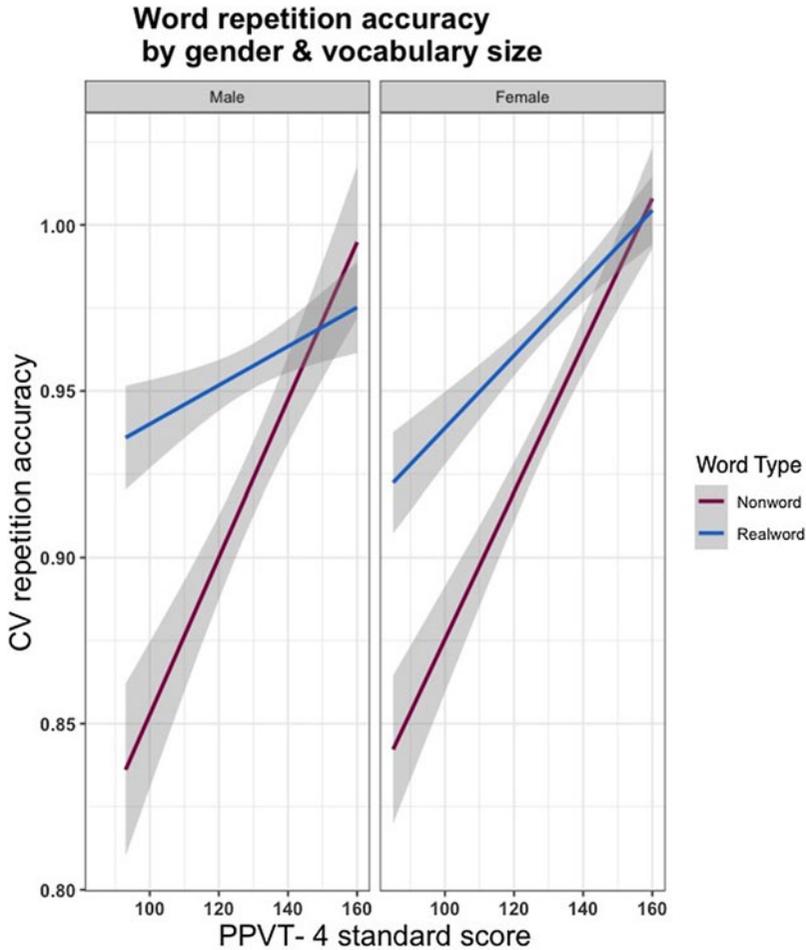


Figure 2. Consonant-vowel repetition accuracy by gender and receptive vocabulary size

accurate at the repetition tasks than male children. However, this effect by gender was not strong enough to warrant inclusion in the final model.

Finally, though individual words were controlled for in the random effect structure of the model, Figure 3 displays the variation in repetition performance by each CV sequence. Graphs represent a median split of receptive vocabulary size with the left graph displaying repetition performance from children with larger vocabularies (mean = 134.70, range = 123–151) and the right graph displaying performance from children with smaller vocabularies (mean = 106.26, range = 76–122). The dotted line in the figures represents parity between nonword and real word conditions – a location above the dotted line indicates that children repeated the CV sequence better in a nonword than a real word. For the smaller vocabulary group, on average, no CV syllables were more accurately produced in the nonword condition. However, for the children with larger vocabularies, four syllables, [ka], [su], [sɪ], and [wɪ], were

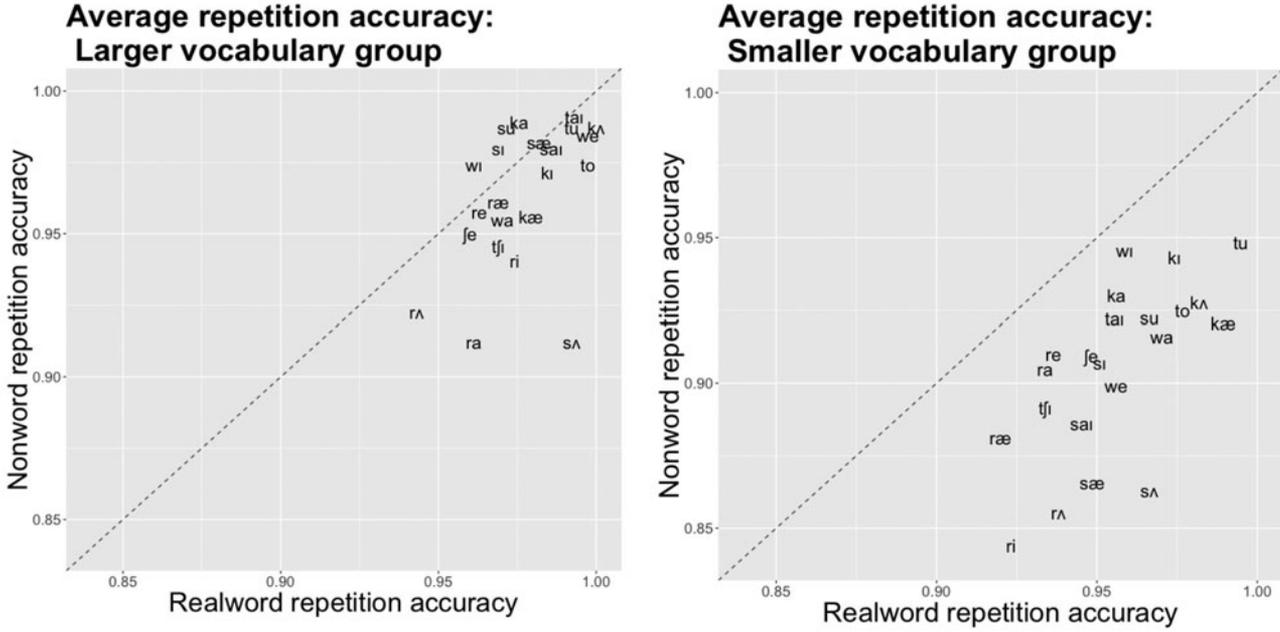


Figure 3. Consonant-vowel repetition accuracy in real words and nonwords: Median vocabulary split

produced more accurately in the nonword condition than in the real word condition. Additionally, [sæ] was produced equally well between real words and nonwords. To avoid confusion due to the idiosyncrasies of English orthography, sequences are plotted in the International Phonetic Alphabet; see endnotes for mapping of the phonetic symbols to common English words.⁴

Discussion

A critical component of children's vocabulary development is the ability to map previously-learned sounds and sound combinations to new words. We employed a word repetition task in four-year-old children to test for a potential lexical advantage in early speech production, hypothesizing that children would produce the same CV sequence more accurately in a real word than a nonword (e.g., [sutkes] 'suitcase' and [sudras]). In doing so, we attempted to control for the articulatory demands of the sounds, as well as stress, prosodic structure, and phonotactic transitional probability, as all of these factors were consistent between nonwords and real words. In doing so, we attempted to isolate a lexical effect upon children's speech production.

We demonstrated that four-year-old children repeat CV sequences more accurately when the sequences are contained in real words rather than nonwords. This may suggest that children's early speech sounds are not produced via arithmetic concatenations of abstracted speech segments (cf. Berent, 2013; Dinnsen & Gierut, 2008; Tesar & Smolensky, 1998). The four-year-olds studied may instead have been more accurate at producing these CV sequences in real words because the children have not fully abstracted the speech segments away from the contextual environments in which they originally learned to produce the sounds. This lack of complete, adult-like representations makes it difficult for young children to generalize to novel words, such as the nonwords tested in the word repetition task in this study.

Nevertheless, the children did repeat the CV sequences in nonwords correctly SOME of the time. Accuracy on the nonword task varied by vocabulary size, but average performance ranged from 82% to even 100% accuracy for those children with the largest receptive vocabularies. The children's success at the nonword task is the result of both a typical language development trajectory and the experimental design. First, as the size of children's receptive vocabularies increases, the lexical advantage fades. So we anticipate that some children, particularly those with larger vocabularies, will be fairly successful at the nonword repetition task. A second reason why the children were fairly successful at the nonword repetition task concerns our experimental design. Recall that we limited evaluation of children's repetition accuracy to just the first sequence. If we had evaluated the entire word, accuracy rates might have been lower. Anecdotally, children did seem to perform worse on the second syllable of the words. This is not surprising, as the second syllables were word-final and unstressed, rendering them less acoustically salient. However, we did not balance the second syllable between all nonwords and real words, precluding any analysis of overall repetition accuracy.

Furthermore, the fact that the children successfully repeated the nonwords at all is actually evidence that the participants do have a segmental, or at least sub-lexical,

⁴/æ/ as in the vowel in "cat"; /i/ as in "eat"; /ʌ/ as in "cut"; /e/ as in "ate"; /ɪ/ as in "hit"; /u/ as in "shoo"; /a/ as in "blah"; /aɪ/ as in "sigh"; /ʃ/ as in "sh"; /tʃ/ as in "ch"

phonology. If there were no abstraction at all from the original lexical context, then children would be entirely incapable of repeating nonwords. Decades of research into nonword repetition demonstrate that this is not the case (e.g., Edwards & Lahey 1998; Edwards et al., 2004; Gathercole et al., 1991; Metsala, 1999). In addition, recent work by Szewczyk et al. (Szewczyk, Marecka, Chiat & Wodniecka, 2018) carefully analyzes myriad sublexical and lexical predictors of nonword repetition – ngram phonemic frequency, ngram subsyllabic frequency, receptive vocabulary size, adult-evaluated wordlikeness – in children aged 4;5-6;10. The authors find that both sublexical factors, such as phonemic n-gram frequency (the frequency of segments in sequences 2–8 units in length), and receptive vocabulary size explain children’s nonword repetition performance in Polish. Conclusions from Jones (2016)’s computational model of nonword repetition in children (2;1–6;1) likewise support reliance on both lexical and segmental representations. In that study, children’s language exposure, quantified as the child-directed speech received between ages 2;0–6;0, positively predicted nonword repetition accuracy. Evidence for segmental representations came from the fact that the model’s learning mechanism began learning at the individual segment, not holistically. However, evidence for lexical representations came from the positive correlations between vocabulary size and nonword repetition results leading to the conclusion that, “lexical phonological knowledge is therefore pervasive in many of the processes involved in nonword repetition” (Jones, 2016:86).

If the conclusion in the current study is correct, we should anticipate an interaction of production accuracy with children’s vocabulary size (Edwards et al., 2004; Metsala, 1999). Specifically, if speech segments emerge from generalizations made over children’s lexicons, then children with larger vocabularies should have more abstract segmental representations. The production data presented here bore out this prediction.

Interestingly, though both receptive and expressive vocabulary size predicted repetition accuracy, receptive vocabulary size was a better predictor of accuracy than expressive vocabulary size. The primacy of receptive vocabulary over expressive vocabulary contrasts with some previous studies that found that expressive vocabulary better predicted repetition accuracy (Edwards & Lahey, 1998; Edwards et al., 2004). Edwards & Lahey (1998) found that expressive, and not receptive, vocabulary predicted nonword repetition accuracy in children with SLI and typically-developing children aged 4;6-9;8. The authors concluded that receptive vocabulary size was not relevant for the six nonwords used in the study. Edwards et al. (2004) found that both receptive and expressive vocabulary sizes were correlated with nonword repetition accuracy in children aged 3;2-8;10. However, only expressive vocabulary was explicitly tested as a predictor of repetition accuracy so it is unclear which measure of vocabulary would have been the stronger predictor. Finally, Munson et al. (Munson, Edwards & Beckman, 2005) found that in children aged (3;0-6;0) with phonological disorder, receptive and expressive vocabulary were correlated with mean accuracy in a nonword repetition task, though expressive vocabulary was the more reliable predictor.

In all, expressive and receptive vocabulary have both been shown to predict nonword repetition accuracy. The effect can vary by population studied (e.g., typically-developing children, children with phonological disorder) or the nonword stimuli employed. For the current study, receptive vocabulary may have been the stronger predictor because, unlike those previous works, the current study measured both nonword and real word repetition. It could be that the words that a child recognizes, but does not

necessarily produce, assist them with both real word and nonword repetition whereas words that a child articulates may assist more with nonword repetition accuracy. Still, it is important to reiterate that the difference between the model fit with receptive vocabulary was only marginally better than the model with expressive vocabulary. More research is needed to conclude if receptive vocabulary is consistently a better predictor of real and nonword repetition.

Regardless of the vocabulary measure that ultimately proves to be the better predictor of word repetition accuracy, the abstraction of phonological segments from the lexicon could unfold similarly from expressive or receptive vocabularies. Specifically, statistical properties of speech that children hear (and produce) such as how many minimal pairs a sound contrasts (i.e., functional load [Stokes & Surendran, 2005; Van Severen et al., 2012]), how frequently a given sound occurs (i.e., type frequency [Edwards & Beckman, 2008; Zamuner, 2009]), and how much competition exists between perceptually-confusable words (i.e., phonological neighborhood density [Charles-Luce & Luce, 1990; Storkel, 2002]), allow children to begin to make generalizations about the importance of certain sounds and sequences in their language. The more words that a given sound contrasts, the more likely a child is to begin to form an autonomous segmental-sized representation of that sound. The more frequently that a sound or sound sequence occurs (across distinct lexical items), the more likely that a child will generalize the sound or sequence away from its original lexical context to be used in novel environments. With each accumulated exemplar trace, the category – syllabic, segmental, or otherwise – that the child is constructing becomes stronger (Pierrehumbert, 2003), a phenomenon illustrated by the protracted development of speech categories well into adolescence (Hazan & Barrett, 2000; McMurray, Danelz, Rigler & Seedorff, 2018) and cross-linguistic differences in consonant production in children as young as 2;0 (Edwards & Beckman, 2008).

While all children produced the CV sequences in real words more accurately than nonwords, children with larger receptive vocabularies showed less of a disparity by word type than children with smaller receptive vocabularies. Again, this is not to say that those children with smaller vocabularies entirely lack a segmental phonology. They repeated – with relatively good success – the nonwords, suggesting that they can produce speech sounds in novel lexical environments. However, we argue that this finding indicates that children with larger receptive vocabularies, relative to children with smaller receptive vocabularies, were better at abstracting segmental representations away from the word level and generalizing them to novel words and contexts.

In an emergentist model of child phonology, a conclusion so strongly contingent upon vocabulary size is unsurprising. Production differences between speakers are anticipated as children construct unique phonological grammars based on their experienced language. In this way, rather than abandoning abstraction in child language altogether, we display how an array of language experiences, manifesting in children's speech production abilities, demonstrates the emergence of abstract speech.

One limitation of this study might be that nonwords and real words were presented separately, on different days. The order of presentation was also not counterbalanced between the children. However it is difficult to see how a different experimental design would resolve this issue because combining the two word types introduces alternative confounds such as a possible priming effect and children's propensity to produce nonwords as real words. Furthermore, even in a combined real word-nonword task, the nonword *TRIALS* would still be more difficult. This may just be a

confound with lexicality. Nevertheless, it would be of interest to see if the same effect would be observed if the real words and nonwords were interwoven in the same task.

In addition, the nonword task is arguably more demanding than the real word task. Nonword repetition requires that speakers encode and execute a sequence of sounds without the scaffold of a lexical representation or previous experience concatenating these syllables into a single word. The real word task, on the other hand, asks that participants activate a semantic representation and execute a sequence of sounds that they have likely already heard or produced. The nonword task could also be more difficult because the visual stimuli presented with the real words can assist children in the lexical retrieval of real words. The visual stimuli do not help with nonword retrieval, however. While the real word and nonword tasks differ along some dimensions, there may not be a tractable way to address the disparity. The experimenter could intentionally make the real word task more difficult for the children by embedding real words in noise, for example. However, all of the myriad ways to make real word repetition more difficult introduce additional variables into the comparison between nonwords and real words.

This research raises questions about the timeline of phoneme abstraction. If the mental representation of speech sounds differs between adults and children, and even between children of the same age, when do children approximate adults? That English-speaking adults contrast speech at the segmental level is well-documented (Steriade, 2007; Vitevich & Luce, 1999). But to account for well-known findings in exemplar-theoretic models of speech perception, adults must likewise represent words somewhat holistically, encoded with great acoustic detail (Goldinger, 1996; 1998; Johnson, 2006; Strand *et al.*, 1999). So adults sometimes incorporate knowledge of segmental units, and sometimes whole words. Fisher and colleagues demonstrate something similar in two- and three-year-old children who repeated a series of nonwords more accurately after only two brief, auditory exposures (Fisher, Hunt, Chambers & Church, 2001). However, the children also showed this priming effect for just PARTS of those novel words (syllables) which they accurately produced in novel environments. This suggested abstraction, as well as specificity, in the children's representations. These findings demonstrate that preschool children have abstracted phonological units away from their original lexical contexts somewhat.

Yet the lexical advantage found here, the performance disparity between CV sequences contained in real words and nonwords, means that, by age 4, children, particularly those with smaller receptive vocabularies, are still in the process of abstracting these adult-like segments. Consequently, future investigations should test a similar real word/nonword repetition paradigm in slightly older children to evaluate how abstraction progresses throughout childhood. Any future work with older children may want to complicate the current task – use a more demanding syllable structure or trisyllabic words – as some children here approached a ceiling effect. Since literacy will inevitably influence older children's phoneme awareness as they begin formal schooling, future work must additionally measure phonological awareness or evaluate speech emergence in children from non-literate populations or bilingual populations without a literary tradition in one of their spoken languages. Only through this cross-linguistic and cross-experiential lens can we truly understand if our segmental biases in speech are given or emerge through usage.

Acknowledgements. The authors thank the participating families and Learning to Talk lab members, especially Rebecca Higgins and Michele Liquori. Additional thanks to Sharon Inkelas and two anonymous reviewers for extremely constructive comments. This work benefitted from feedback from

audiences at Macquarie University, UC Berkeley, and BUCLD 43. Research was supported by NIDCD Grant No. R01 02932 to J.R.E., B.M., and Mary E. Beckman, a UC Berkeley Dissertation Completion Fellowship to M.C., and the Raymond H. Stetson Scholarship in Phonetics and Speech Science to M.C.

References

- Arnon, I., & Christiansen, M. H.** (2017). The Role of Multiword Building Blocks in Explaining L1–L2 Differences. *Topics in Cognitive Science*, 9(3), 621–636.
- Arnon, I.** (2010). Starting big: *The role of multi-word phrases in language learning and use*. Unpublished PhD dissertation, Stanford University.
- Bates, D., Maechler, M., Bolker, B., & Walker, S.** (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Beckman, M., & Edwards, J.** (2010). Generalizing over lexicons to predict consonant mastery. *Laboratory Phonology*, 1(2), 319–343.
- Berent, I.** (2013). *The phonological mind*. Cambridge University Press: Cambridge, MA.
- Boersma, P., & Weenink, D.** (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.42, retrieved 15 August 2018 from <http://www.praat.org/>
- Charles-Luce, J., & Luce, P. A.** (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17(1), 205–215.
- Chiat, S., & Roy, P.** (2007). The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*, 50(2), 429–443.
- Cicchetti, D. V.** (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- Dinnsen, D. A., & Gierut, J. A.** (2008). Optimality Theory: A Clinical Perspective. In M. J. Ball, M. R. Perkins, N. Müller, and S. Howard (Eds.), *The Handbook of Clinical Linguistics* (pp. 439–451). Blackwell.
- Dispaldro, M., Deevy, P., Altoé, G., Benelli, B., & Leonard, L. B.** (2011). A cross-linguistic study of real-word and non-word repetition as predictors of grammatical competence in children with typical language development. *International journal of language & communication disorders*, 46(5), 564–578.
- Dunn, L. M., & Dunn, D. M.** (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.
- Edwards, J., & Lahey, M.** (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics*, 19(2), 279–309.
- Edwards, J., Beckman, M., & Munson, B.** (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2), 421–436.
- Edwards, J., & Beckman, M.** (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning & Development*, 4(2), 122–156.
- Edwards, J., Beckman, M., & Munson, B.** (2015). Frequency effects in phonological acquisition. *Journal of Child Language*, 42(2), 306.
- Cychoz, M., Edwards, J., Munson, B., & Johnson, K.** (2019). Spectral and temporal measures of coarticulation in child speech. *Journal of the Acoustical Society of America-EL*, 146(6), EL516–EL522.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J.S., & Bates, E.** (2007). *MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual, Second Edition*. San Diego, CA: Singular.
- Ferguson, C. A., & Farwell, C. B.** (1975). Words and sounds in early language acquisition. *Language*, 51(2), 419–439.
- Fisher, C., Hunt, C., Chambers, K., & Church, B.** (2001). Abstraction and Specificity in Preschoolers' Representations of Novel Spoken Words. *Journal of Memory and Language*, 45(4), 665–687.
- Fowler, A. E.** (1991). How early phonological development might set the stage for phoneme awareness. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: a tribute to Isabelle Y. Liberman*. (pp. 97–117). Hillsdale, NJ: Erlbaum.
- Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D.** (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, 12(3), 349–367.

- Gerosa, M., Lee, S., Giuliani, D., & Narayanan, S. (2006). Analyzing Children's Speech: An Acoustic Study of Consonants and Consonant-Vowel Transition. 2006 *IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, 1, 393–396.
- Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience*, 24(3), 564–574.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166–1183.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396.
- Hogan, T. P., Bowles, R. P., Catts, H. W., & Storkel, H. L. (2011). The influence of neighborhood density and word frequency on phoneme awareness in 2nd and 4th grades. *Journal of Communication Disorders*, 44(1), 49–58.
- Jakobson, R. (1941/1968). Kindersprache, aphasie, und allgemeine lautgesetzte. In A. R. Keiler (Ed., trans.), *Child language, aphasia, and phonological universals*, The Hague: Mouton.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4), 485–499.
- Jones, G. (2016). The influence of children's exposure to language from two to six years: The case of nonword repetition. *Cognition*, 153, 79–88.
- Juszyk, P. W., Smolensky, P., & Alallocco, T. (2002). How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition*, 10(1), 31–73.
- Keren-Portnoy, T., Vihman, M. M., DePaolis, R. A., Whitaker, C. J., & Williams, N. M. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research*, 53(5), 1280–1293.
- Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the “critical period”. *Language Learning and Development*, 1 (3–4), 237–264.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1493), 979–1000.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26.
- Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C.A. Ferguson, L. Menn & C. S. Gammon (Eds.), *Phonological development: Models, research, implications*. (pp. 131–163). Timonium, MD: York.
- Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, 142, 345–350.
- Mandrekar, J. N. (2011). Measures of interrater agreement. *Journal of Thoracic Oncology*, 6(1), 6–7.
- Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111.
- McAllister Byun, T., & Tessier, A. M. (2016). Motor influences on grammar in an emergentist model of phonology. *Language and Linguistics Compass*, 10(9), 431–452.
- McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, 54(8), 1472–1491.
- Menn, L., Schmidt, E., & Nicholas, B. (2013). Challenges to theories, charges to a model: The Linked-Attractor model of phonological development. *The emergence of phonology: Whole-word approaches and cross-linguistic evidence*, (pp. 460–502). Cambridge: Cambridge University Press.
- Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91(1), 3.
- Metsala, J. L., & Chisholm, G. M. (2010). The influence of lexical status and neighborhood density on children's nonword repetition. *Applied Psycholinguistics*, 31(3), 489–506.
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, New Jersey: Erlbaum.
- Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*, 44(4), 778–792.

- Munson, B., Edwards, J., & Beckman, M. (2005). Relationships Between Nonword Repetition Accuracy and Other Measures of Linguistic Development in Children With Phonological Disorders. *Journal of Speech, Language, and Hearing Research*, 48(1), 61–78.
- Munson, B., Edwards, J., & Beckman, M. (2011). Phonological representations in language acquisition: Climbing the ladder of abstraction. In A. C. Cohn, C. Fougerson & M. K. Huffman, *The Oxford Handbook of Laboratory Phonology*, (pp. 288–309). Oxford: Oxford University Press.
- Newport, E., Weiss, D. J., Wonnacott, E., & Aslin, R. (2004). *Statistical learning in speech: Syllables or segments?* Paper presented at the 29th Annual Boston University Conference on Language Development, Boston, MA.
- Noiray, A., Popescu, A., Killmer, H., Rubertus, E., Krüger, S., & Hintermeier, L. (2019). Spoken language development and the challenge of skill integration. *Frontiers in Psychology*, 10(2777), 1–17.
- Pharr, A. B., Ratner, N. B., & Rescorla, L. (2000). Syllable structure development of toddlers with expressive specific language impairment. *Applied Psycholinguistics*, 21(4), 429–449.
- Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2-3), 115–154.
- Pisoni, D., Nusbaum, H., Luce, P., & Slowiaczek, L. (1985). Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, 4, 75–95.
- Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, 100(1), 577–592.
- Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech, Language, and Hearing Research*, 39(1), 153–165.
- Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, 47, 223–34.
- Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12(4), 110–114.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2012). E-Prime 2.0 reference guide manual. Pittsburgh, PA: Psychology Software Tools.
- Smith, B. L., McGregor, K. K., & Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics*, 27(3), 355–375.
- Sosa, A. V. (2008). *Lexical effects in typical phonological acquisition*. Unpublished doctoral dissertation, University of Washington, Seattle.
- Steriade, D. (2007). *Contrast*. In P. de Lacy (Ed.) *The Cambridge Handbook of Phonology*, (pp. 139–156), Cambridge: Cambridge University Press
- Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, 38(1), 1–34.
- Stokes, S. F., & Surendran, D. (2005). Articulatory Complexity, Ambient Frequency, and Functional Load as Predictors of Consonant Development in Children. *Journal of Speech Language and Hearing Research*, 48(3), 577.
- Stokes, S. F. (2010). Neighborhood Density and Word Frequency Predict Vocabulary Size in Toddlers. *Journal of Speech, Language, and Hearing Research*, 53(3), 670–683.
- Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, 29(2), 251–274.
- Storkel, H. L., & Morrisette, M. L. (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech, and Hearing Services in Schools*, 33(1), 24–37.
- Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, 25(02).
- Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, 18(1), 86–100.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
- Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, 179, 23–36.
- Tesar, B., & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry*, 29(2), 229–268.

- Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review, 13*(3), 286–350.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*(1), 1–30.
- Williams, K. T. (2007). Expressive Vocabulary Test. 2nd edition Pearson Assessments; Minneapolis, MN.
- Van Severen, L., Van Den Berg, R., Molemans, I., & Gillis, S. (2012). Consonant inventories in the spontaneous speech of young children: A bootstrapping procedure. *Clinical Linguistics & Phonetics, 26*(2), 164–187.
- Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology, 108*(1), 1–27.
- Vihman, M. M., & Velleman, S. L. (1989). Phonological reorganization: A case study. *Language and Speech, 32*, 149–170.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*(3), 374–408.
- Zamuner, T. S. (2009). Phonotactic probabilities at the onset of language development: Speech production and word position. *Journal of Speech, Language, and Hearing Research, 52*(1), 49–60.
- Zamuner, T. S., Gerken, L., & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language, 31*(3), 515–536.