JSLHR

Research Article

Consonant Age-of-Acquisition Effects in Nonword Repetition Are Not Articulatory in Nature

Michelle W. Moore,[a,b] Julie A. Fiez,[b,c,d] and Connie A. Tompkins[b,d]

Purpose: Most research examining long-term-memory effects on nonword repetition (NWR) has focused on lexical-level variables. Phoneme-level variables have received little attention, although there are reasons to expect significant sublexical effects in NWR. To further understand the underlying processes of NWR, this study examined effects of sublexical long-term phonological knowledge by testing whether performance differs when the stimuli comprise consonants acquired later versus earlier in speech development.

Method: Thirty (Experiment 1) and 20 (Experiment 2) college students completed tasks that investigated whether an experimental phoneme-level variable (consonant age of acquisition) similarly affects NWR and lexical-access tasks designed to vary in articulatory, auditory-perceptual, and phonological short-term-memory demands. The lexical-access tasks were performed in silence or with concurrent articulation to explore whether consonant age-of-acquisition effects arise before or after articulatory planning.

Results: NWR accuracy decreased on items comprising later- versus earlier-acquired phonemes. Similar consonant age-of-acquisition effects were observed in accuracy measures of nonword reading and lexical decision performed in silence or with concurrent articulation.

Conclusion: Results indicate that NWR performance is sensitive to phoneme-level phonological knowledge in long-term memory. NWR, accordingly, should not be regarded as a diagnostic tool for pure impairment of phonological short-term memory.

Supplemental Materials: https://doi.org/10.23641/asha.5435137

Nonword repetition (NWR), which requires the immediate repetition of a spoken nonword, is a cornerstone task within the literature on communication science and disorders. It is widely used to diagnose language-learning difficulties (for a review, see Coady & Evans, 2008; Graf Estes, Evans, & Else-Quest, 2007) and is marketed as a "reliable indicator of short-term memory" ("Children's Test of Nonword Repetition," n.d.). However, this view of NWR neglects empirical findings and theoretical refinements that have emerged since the relationship between NWR and language learning was first studied by Gathercole and Baddeley (1990). Researchers now view NWR as a multidimensional measure affected by phonological short-term-memory ability as well as the storage and retrieval of linguistic information from long-term memory (Archibald & Gathercole, 2006; Gupta, 2006; Gupta & Tisdale, 2009; Rispens & Baker, 2012). Nearly all of the research examining long-term-memory effects on NWR has focused on lexical-level variables. Little attention has been paid to phoneme-level variables, although there are both empirical and theoretical reasons to expect significant effects. This article focuses on this question by testing whether performance on NWR, word-naming, and lexical-decision tasks differs when the items comprise consonants acquired later versus earlier in speech development.

The use of NWR in the literature on communication science and disorders has its roots in a landmark study by Gathercole and Baddeley (1990). This study found that six children with language impairment performed more poorly on NWR than an age-matched control group and a separate language-matched control group. To explain their findings, the authors proposed that children with language impairment have difficulty processing phonological information because of the limited capacity of their phonological short-term memory. With a limited capacity to maintain

---

[a] West Virginia University, Morgantown
[b] University of Pittsburgh, Pennsylvania
[c] Learning Research and Development Center, Pittsburgh, PA
[d] Center for the Neural Basis of Cognition, Pittsburgh, PA

Correspondence to Michelle W. Moore: mimoore@mail.wvu.edu

Editor: Julie Liss
Associate Editor: Amy Neel

Received September 9, 2016
Revision received March 29, 2017
Accepted May 8, 2017
https://doi.org/10.1044/2017_JSLHR-L-16-0359

Disclosure: The authors have declared that no competing interests existed at the time of publication.

phonological information in short-term memory, less online information could be available for the operation of the phonological processes involved in verbal functions such as word learning, sentence formulation, and so on. Thus, the authors reasoned that poor phonological memory could be an underlying mechanism that compromises language learning. In follow-up experiments, they used a serial-recall task and word-length manipulations to further probe the verbal-working-memory abilities of children with and without impaired language learning. On the basis of the different pattern of results obtained across the two groups, the authors concluded that the children with language impairment had a typically functioning articulatory rehearsal process but a deficient capacity for phonological storage. In addition, they replicated their original observation of impaired NWR in children with language impairment. These results were interpreted as evidence that NWR can be used to measure phonological-memory abilities that underlie language learning.

Since the original work of Gathercole and Baddeley (1990), the observed relationship between NWR performance and language impairment has been replicated numerous times. For example, in their meta-analysis of 23 studies, Graf Estes et al. (2007) showed that children with specific language impairment (SLI) performed an average of 1.27 *SD*s below children without SLI on NWR tasks. Data such as these are used to support claims that poor NWR is a clinical and phenotypic marker for children with language impairment (e.g., Bishop et al., 1999; Bishop, 2002; Bishop, North, & Donlan, 1996; Falcaro et al., 2008; SLI Consortium, 2002, 2004). However, it is important to recognize that replicating the association between NWR and SLI does not by itself further establish that poor phonological memory mediates the relationship.

Indeed, the emergence of NWR as a reliable measure for identifying children with language impairment has been paralleled by continued research into the underlying nature of NWR. This continued research has provided strong evidence that NWR performance is affected by factors associated with both phonological short-term memory and long-term phonological knowledge (e.g., Gupta, 2006; Gupta & Tisdale, 2009; Rispens & Baker, 2012). For example, Gupta and colleagues (Gupta, 2003; Gupta & MacWhinney, 1997; Gupta & Tisdale, 2009) have proposed a unitary framework for long- and short-term phonological information to account computationally for the relationship observed between NWR performance, immediate serial recall, and vocabulary acquisition. Their interactive model has both feed-forward and feedback processes between lexical, syllabic, and phonological or phonetic information, such that phonological short-term memory occurs via connections that are weighted on the basis of the network's experience. In other words, phonological short-term memory is functionally represented in the model, but it is influenced by and not structurally independent from long-term knowledge. Gupta and Tisdale (2009), importantly, were able to simulate many of the outcomes observed in behavioral studies reflecting causal effects of short- and long-term phonological memory in NWR. Thus, to understand the mechanisms that underlie the association between NWR and SLI, it is imperative that researchers gain a solid understanding of long-term-memory influences on NWR.

Research examining long-term-memory influences on NWR has focused primarily on lexical-level effects. This is a natural beginning point, because poor word learning (i.e., poor vocabulary acquisition) is a hallmark of SLI (Leonard, 2014). For example, Metsala and Chisholm (2010) investigated the effects of lexical status (word vs. nonword) and phonological-neighborhood density (dense vs. sparse—i.e., the number of similar-sounding words in the lexicon) on children's NWR performance. Their findings showed that within longer nonwords, children more accurately repeated real-word syllables than nonword syllables and more accurately repeated constituent syllables from dense neighborhoods than from sparse neighborhoods. The neighborhood-density effect was stronger in children who had lower vocabulary scores. On the basis of these findings, Metsala and Chisholm suggested that "changes in vocabulary knowledge and lexical organization are the pace setter for the observed association" between NWR and word learning (p. 502). Other examples of lexical-level long-term-memory effects on NWR include work demonstrating that vocabulary knowledge and NWR are linked (for a review, see Gathercole, 2006), as well as work showing that NWR performance is improved when the nonwords are more word-like (Gathercole, 1995; Graf Estes et al., 2007), when the stressed syllables of the nonwords are real words (Dollaghan, Biber, & Campbell, 1993, 1995), and when the nonwords have a higher phonotactic probability (Coady, Evans, & Kluender, 2010; Edwards, Beckman, & Munson, 2004; Munson, 2001; Munson, Kurtz, & Windsor, 2005).

Although refinements to theoretical accounts of NWR have included lexical-level influences from long-term memory, there has been a paucity of work directly manipulating phoneme-level variables. However, the small body of work that has explored phoneme-level influences in NWR has yielded positive results, thus warranting further consideration. One line of study has found effects of phonotactic probability on NWR performance (Coady et al., 2010; Edwards et al., 2004; Munson, 2001; Munson, Edwards, & Beckman, 2005; Munson, Kurtz, & Windsor, 2005). Of course, phonotactic probability—that is, the frequency of occurrence of a phoneme within a given word position (Edwards et al., 2004; Munson, Kurtz, & Windsor, 2005)—is by definition a measure that is intertwined with lexical-level information. Several studies have demonstrated that a phonotactic-probability effect in NWR is mediated by lexical knowledge (e.g., Edwards et al., 2004; Munson, Edwards, & Beckman, 2005; Munson, Kurtz, & Windsor, 2005; Roodenrys & Hinton, 2002). Thus, these findings alone do not clearly demonstrate phoneme-level influences from long-term memory on NWR.

More direct findings come from a second line of work. To be specific, phoneme-level effects in NWR performance have been found using a consonant age-of-acquisition (CAoA) manipulation. The manipulation takes advantage of the fact that children master the production of different phonemes at different ages. Consonant acquisition ranges from 3 to 9 years of age (using a 90% level-of-acquisition criterion), with phonemes such as /m/, /n/, and /p/ typically acquired by age 3 years and phonemes such as /s/, /z/, and /r/ acquired by age 7 years or later (Smit, Hand, Freilinger, Bernthal, & Bird, 1990). Moore, Tompkins, and Dollaghan (2010) explored the use of this CAoA manipulation with the goal of increasing task difficulty and thus avoiding ceiling effects when administering NWR to adolescents and adults. CAoA has been associated with articulatory complexity—that is, the motoric demands involved in speech sound production (e.g., Kent, 1992; Stokes & Surendran, 2005). The original motivation for use of the experimental manipulation by Moore et al. (2010), therefore, was to increase the articulatory complexity of the stimuli. They reasoned that this could in turn influence short-term-memory performance. For instance, increasing articulatory complexity could decrease the rate of speech production, thus making the short-term storage of phonological information more prone to decay (Baddeley, Thomson, & Buchanan, 1975). In their study with young adults, Moore et al. compared two NWR tasks, one with stimuli comprising only earlier-developing consonants and the other only later-developing consonants. As they predicted, participants performed more poorly on the task comprising later-developing phonemes. The two stimulus lists were made with an effort to control for the lexicality of constituent syllables, so that vocabulary knowledge could not easily explain the observed CAoA effect.

Although the results of Moore et al. (2010) achieved the goal of developing a more challenging NWR task, there are several reasons they could be regarded as unexpected. For instance, the effect was observed in a sample of native English-speaking college students who were highly skilled in the production of English phonemes, so it might be expected that developmental effects of consonant articulation would be negligible after so many years of speech-production practice. In addition, previous work assessing articulatory complexity with consonant class (Edwards & Lahey, 1998) and consonant clusters versus singletons (Archibald & Gathercole, 2006) showed no effect on NWR in children with typical development (although this was not always the case for the groups of children with language impairment). Last, Archibald, Joanisse, and Munson (2013) found that motorically constrained conditions of NWR (e.g., administration using a bite block) only modestly affected repetition performance in children. Hence, explaining the CAoA effect in NWR as a purely articulation-based influence on short-term memory seems incongruous with other findings.

One alternative explanation for the CAoA effect is that it reflects a long-term-memory influence on NWR. Theoretical models of speech production and semantic networks illustrate the possible mechanism for CAoA as an operationalized measure of long-term phonological knowledge. The current leading models of speech production, importantly, are convergent in depicting articulatory gestural information at a central level of processing (Dell, 1986; Guenther, Ghosh, & Tourville, 2006; Levelt, Roelofs, & Meyer, 1999). By extension, consonants that are more complex to articulate (i.e., later-developing phonemes) should involve the storage and encoding of a more complex articulatory gesture. Thus, the retrieval of phoneme-level information from long-term memory should be less efficient (i.e., slower or less accurate) for later-developing phonemes in various lexical-access tasks, even tasks that do not require articulation.

Although more tenuous, another consideration is that neural networks for phonological information develop similarly to semantic networks, and this affects the quality and retrieval of phonological representations. Steyvers and Tenenbaum (2005) have suggested that early-acquired semantic information has the advantage of becoming a "hub" (p. 43) from which other neural connections are established. Later-acquired information has decreased centrality and a smaller number of connections, and therefore is less prone for selection in networks. This neural organization provides a potential model in which typical speech development leads to long-term phonological representations of varied quality in the speech-language architecture. Lexical-access tasks should accordingly be sensitive to the CAoA effect. This should be true across variations in the perceptual input (e.g., auditory versus visual), articulatory demands (e.g., spoken versus recognition), and need to maintain phonological information in short-term memory.

In summary, although there are some reasons to find it surprising that developmental differences in consonant acquisition affect adult NWR performance, there are several plausible mechanisms that could explain the presence of CAoA effects. The purpose of this study was to determine whether prior CAoA effects can be replicated, and if so, to better understand whether they reflect either an articulatory or a central phonological influence on short-term memory. To be more specific, Experiment 1 probes for a CAoA effect in NWR and in lexical-access tasks that place minimal demands on phonological short-term memory and overt articulation. If the CAoA manipulation reflects a long-term-memory influence on phonological representation and activation within short-term memory, then CAoA effects across both NWR and lexical-access tasks should be observed. Experiment 2 takes this line of reasoning one step further by examining the lexical-access tasks under conditions in which articulatory rehearsal is suppressed by having subjects simultaneously perform a simple covert articulation task (i.e., engaging in concurrent articulation or articulatory suppression). If CAoA effects are linked to articulatory rehearsal, then the effects should be significantly reduced when a lexical-access task is performed under conditions of concurrent articulation. In contrast, the CAoA effects should persist if they arise from a central level of phonological representation.

Experiment 1

Stimuli comprising earlier- versus later-developing phonemes were contrasted to assess the influence of CAoA on NWR and other tasks that draw on the speech-language architecture. The stimuli were used in an NWR task and two other tasks (auditory lexical decision and nonword reading) that require minimal to no phonological short-term memory. The two tasks were selected to rule out potential confounds to the CAoA effect (Moore et al., 2010). Auditory lexical decision was selected because it eliminates overt articulatory demands; nonword reading was selected because it eliminates auditory-perceptual demands. If null effects are observed in auditory lexical decision and nonword reading, this would indicate that CAoA effects on NWR reflect purely short-term-memory mechanisms. In the alternative, if positive results are observed, this would indicate that phoneme-level knowledge in long-term memory may affect NWR.

Method

Participants

The participants were 30 undergraduate students (26 women, four men) from the University of Pittsburgh. They were 19–20 years of age, self-reported native English monolinguals without a history of a reading disorder or hearing problem. All passed a hearing and vision screening. All participants provided informed consent using procedures approved by the University of Pittsburgh Institutional Review Board, and they were all compensated $10 for their time and effort.

Technical Specifications

Participants were seated approximately 16 in. from the center of the computer monitor and viewed visual stimuli from the center of their visual fields. They listened to auditory stimuli through Sennheiser HD 280 Pro headphones. The presentation volume was set at a comfortable listening level and held constant across all participants. The visual stimuli were centrally presented in white Arial 30-point font against a black background. The auditory stimuli were digitally recorded samples produced by a trained female speaker of Standard American English. These auditory stimuli and participants' verbal responses were recorded using an Audio-Technica ATR 20 microphone and Adobe Audition 1.5 software (44100-Hz sampling rate and 16-bit resolution). Key-press responses were recorded using a serial response box (Psychology Software Tools, Sharpsburg, PA), with subjects using their left index finger to press the leftmost button and their right index finger to press the rightmost button.

Stimuli

Early- and late-developing phoneme groups. The stimuli for this study comprised either early- or late-developing consonant phonemes. The early and late consonant groups (seven phonemes per group: E7 and L7) were taken from the early, middle, and late consonant groups identified by Shriberg and Kwiatkowski (1994). In their work, 72 children with typical development ages 3–6 years correctly produced the Early-8, Middle-8, and Late-8 consonants with an average accuracy of 98%, 93%, and 42%, respectively (p. 1108). The "soft g" sound (/ʒ/, as in *beige* and *measure*) in Shriberg and Kwiatkowski's Late-8 group is difficult to represent in orthographic form and does not occur in the initial position of English words, so it was excluded from the set of late-developing phonemes used in this study. Consonants for the early group were selected from Shriberg and Kwiatkowski's Early-8 and Middle-8 groups in order for the E7 group to be more closely matched in articulatory-feature distribution to the L7 group, because featural differences were reported as a potential confound in previous work (Moore et al., 2010). The end result was an E7 group comprising the phonemes /m/, /n/, /p/, /d/, /t/, /f/, and /v/ and an L7 group comprising the phonemes /s/, /z/, /l/, /r/, /ʃ/, /θ/, and /ð/. The E7 and L7 phonemes were used to construct consonant–vowel (CV) and consonant–vowel–consonant (CVC) syllables in the three experimental tasks.

NWR stimuli. A total of 32 stimuli were constructed for an NWR task (see Supplemental Material S1). The stimuli were all nonwords ranging from one to four syllables in length. A CV structure was used for all nonfinal syllables and a CVC structure was used for final syllables, giving eight stimuli for each of the following structures: CVC, CVCVC, CVCVCVC, CVCVCVCVC. Primary stress was assigned to the second syllable of the four-syllable nonwords and the first syllable of all other nonwords. For half of the stimuli at each length, the consonant phonemes were all E7 phonemes, and for the other half they were all L7 phonemes. To construct the stimuli, the strict criteria used by Moore et al. (2010) were used along with additional criteria designed to further minimize potential confounds between lists. Factors that were considered included phoneme recurrence within a nonword and across the task, lexicality of constituent nonword syllables, phonotactic and biphone probability (Vitevitch & Luce, 2004), and average durations of the recorded E7 and L7 stimuli at each syllable length.

Phoneme recurrence within a nonword and across the NWR task was considered in order to control phoneme predictability. A phoneme was not used more than one time within a given nonword. Phoneme recurrence across the task was controlled as well. Each E7 and L7 consonant was used six to 10 times within the NWR task (E7: $M = 8$, $SD = 1.63$; L7: $M = 8$, $SD = 1.53$). Vowel distribution was identical across the E7 and L7 lists. For example, /o/ is used two times in both the E7 list and the L7 list of nonwords, /e/ is used four times in each list, and so on. In addition, there are 20 unique CV segments in both the E7 and L7 lists, with each CV segment used two times.

To minimize effects of wordlikeness (e.g., Dollaghan et al., 1993, 1995), no syllabic segment corresponded to a Standard American English word. Syllabic segments include all CVs and final CVCs. The constituent CV and VC of the final CVCs were considered as well. There were two exceptions in which a word was used in a syllabic segment—the E7 list contains the CV "nah" and the L7 list contains the VC "are." Phonotactic probability, biphone probability, and spoken duration of the recorded stimuli were controlled as well, so that there were no statistically significant differences between the E7 and L7 lists using independent-samples *t* tests (see Supplemental Material S2).

Nonword-reading stimuli. A total of 40 stimuli were constructed for a nonword-reading task (see Supplemental

Material S3). Each of the stimuli was one CVC syllable. For half of the stimuli, the consonant phonemes were all E7 phonemes, and for the other half they were all L7 phonemes.

Similar to the NWR task, phoneme recurrence within a nonword and across the nonword-reading task were considered in order to control phoneme predictability. No phoneme was repeated within a nonword. Phoneme recurrence across the task was generally balanced so that no phonemes were under- or overrepresented. The two sets of stimuli were balanced on a number of phonological and orthographic factors shown to affect reading performance: phonological-neighborhood density, phonotactic and biphone probability, number of letters, orthographic neighborhood, and mean bigram frequency (Balota et al., 2007; see Supplemental Material S4; see also Moore, 2012).

Because characteristics of the constituent components of stimuli have also been shown to affect reading performance, wordlikeness of constituent CVs and friends and enemies of constituent VCs were balanced in the nonword-reading stimuli as well (for a summary of the comparisons between E7 and L7 stimuli on these factors, see Supplemental Material S4). In order to assess the wordlikeness of the CVs, judgment ratings were obtained from nine native English speakers who were unaware of the purpose of the study. After the CV unit of each nonword was read aloud to the raters, they were asked to give a rating from 1 to 5, 1 being *not at all wordlike* and 5 being *a real word*. There were no significant differences in average wordlikeness ratings between the CVs in the E7 and L7 nonword lists.

Previous work has shown that the rhyme—that is, the VC units in the current word lists—has the greatest influence on pronunciation (Jared, McRae, & Seidenberg, 1990; Treiman, Goswami, & Bruck, 1990; Treiman & Zukowski, 1988). The consistency of the nonwords, or the degree to which a (non)word has the same pronunciation as similarly spelled words (Plaut, McClelland, Seidenberg, & Patterson, 1996), was considered using the rhymes of the nonwords in the current nonword-reading task. Treiman et al. (1990) found performance differences when the spelling pattern for the VC unit was shared with many words versus few or no words. For this reason, the number of *friends* (words with a shared spelling pattern and pronunciation) was compared between the E7 and L7 items in this task; no significant differences were found (see Supplemental Material S4).

Last, a consistency ratio was computed in order to account for the effects of both friends and *enemies* (words that have the same spelling pattern but different pronunciation; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Pattamadilok, Knierim, Kawabata Duncan, & Devlin, 2010; Plaut et al., 1996). The number of friends was divided by the sum of the total number of friends and enemies (Pattamadilok et al., 2010). There were no significant differences in the consistency ratio between the E7 and L7 lists. Note that the number of friends and the consistency ratio were considered separately because of the unique information that each measure provides. The consistency ratio factors in the effect of enemies, but for consistent words with no enemies it is 1. This is true whether a target nonword has two friends or 20. The use of a separate measure to compare the number of friends was motivated by the findings of Treiman et al. (1990).

Auditory lexical-decision stimuli. A total of 60 stimuli were constructed for an auditory lexical-decision task (see Supplemental Material S5). Each of the stimuli was one CVC syllable. Half the items were nonwords and half were words. For the nonwords, half (n = 15) were composed of E7 consonant phonemes and half were composed of L7 consonant phonemes. In contrast to the NWR task, in which constituent syllables were primarily nonwords, most of the nonwords for the auditory lexical-decision task were composed of constituent CV words (this was true for 12 and 11 of the E7 and L7 items, respectively). This was done to encourage participants to listen to all three phonemes before making a decision. The phoneme recurrence and phonological factors that were controlled in the nonword-reading stimuli were also controlled here, as well as word frequency for the constituent CVs (see Supplemental Material S6).

For the words, there were 15 items composed of E7 consonant phonemes, 12 composed of L7 consonant phonemes, and three composed of an L7 phoneme in the initial position of the word and an E7 phoneme in the final position of the word. These "mixed" words were used because relatively few of the English words that meet the selection constraints contain only L7 consonant phonemes. Using the nonwords as the basis for stimulus construction, a real word was created by changing either the vowel or the final consonant of a nonword (note one exception, the word *sill,* in which both the vowel and the final consonant were changed). The E7 and L7 word lists were balanced on the phonological variables mentioned previously as well as word frequency (see Supplemental Material S6).

Experimental Procedure

All procedures were administered to participants individually in a quiet testing room. After consenting to participate, subjects were asked to report any history or family history of speech, language, or reading disorder and then were given brief screenings for vision, hearing threshold, and diadochokinetic rates (Kent, Kent, & Rosenbek, 1987). Three experimental tasks were then administered on a computer using the E-Prime computer program (Schneider, Eschman, & Zuccolotto, 2002). NWR was administered first to all participants because it was the primary focus of the study, and we did not want any of our participants to have exposure to other nonwords before completing the task. Nonword reading and auditory lexical decision followed, with the two tasks counterbalanced across successive participants.

NWR task administration. For each trial of the NWR task, participants attempted to repeat aloud a presented auditory nonword. Three practice items were administered; then all eight single-syllable nonwords (four E7, four L7) were administered, followed by all two-syllable nonwords, and so on. Nonwords within each syllable length were presented in random order for each participant. Stimulus presentation was identical to the method described by

Moore et al. (2010). A red fixation cross was displayed on the computer screen 0.5 s prior to nonword presentation and remained on the screen throughout the duration of the auditory presentation of a nonword. After the nonword stimulus presentation, the red fixation cross turned green, prompting the participant to provide the spoken response. The green fixation remained on the screen for 2.5 s, followed by 0.5 s of the red fixation, the next nonword item, and so on. There was a total of 3 s between each nonword item and the next.

Scoring of the NWR task was completed according to the procedures described by Dollaghan and Campbell (1998). In brief, each phoneme was scored as correct or incorrect, and then the percentage of phonemes correct was computed for each item. In the few cases (0.2% of trials) in which a participant provided an incomplete response, the total number of phonemes repeated correctly was divided by the total number of scoreable phoneme targets.

Nonword-reading task administration. For the nonword-reading task, participants read aloud nonword stimuli as quickly and accurately as possible. To begin, three practice items were administered, and then all 40 nonwords were presented one at a time in random order. For each trial, a white fixation cross appeared and remained on the computer screen until the participant hit the space bar to elicit a nonword. When the space bar was pressed, the screen went blank for 250 ms, and then the nonword appeared and remained on the screen until the participant responded aloud. Via a button press, the examiner then recorded accuracy. With the examiner's button press, a white fixation cross appeared until the participant pressed the space bar to cue the next nonword.

Two measurements were recorded for this task: whole-word accuracy and reading latency. Responses were marked as correct if the participant pronounced the nonword identically to a target pronunciation (any legal pronunciation of the onset consonant and rhyme unit). Because of the ambiguity in determining the need for voiced or voiceless "th" in the initial position of nonwords, either phoneme was scored as correct when used in the initial position of any "th" nonword. Reading latency was measured as the duration from the appearance of the nonword to the start of phonation of the response using the spectral and waveform views of the recorded responses in Adobe Audition.

Auditory lexical-decision task administration. For the auditory lexical-decision task, participants pressed one of two keys to indicate whether each item was a word or nonword. To begin, a white fixation cross appeared and remained on the screen throughout the duration of the task. Then the first of four practice items was presented, followed by all 60 experimental items one at a time in random order. For each trial, the participant had unlimited time to respond, although participants were instructed to respond as quickly and as accurately as possible. Following each key-press response, there was a 1,500-ms intertrial interval before the onset of the next trial. Accuracy and reaction time (RT) were recorded. RT was recorded as the duration from the onset of the stimulus item to the onset of a participant response with the duration of the recorded stimuli subtracted out for each trial.

Scoring reliability. Inter- and intrarater reliability measures were obtained for judgments of accuracy in NWR and nonword reading. An undergraduate research assistant independently scored tasks for interrater reliability using participants' digital audio files. A subset of six participants (20% of the sample for each subset) was randomly selected for each task. Agreement for judgments of correctness was 90% or greater for E7 and L7 stimuli in both tasks. The primary scorer (the first author) randomly selected two different subsets of six participants' digital audio files to rescore (one subset per task) for intrarater reliability. The second round of scoring was completed approximately 5 months after the initial scoring, and the author was unaware of the participants' original scores. Agreement for judgments of correctness was 90% or greater for E7 and L7 stimuli in both tasks.

To measure reliability of the procedure used to obtain reading latencies for the nonword-reading task, an undergraduate research assistant independently marked the onset of phonation for the responses of four randomly selected participants. The average reading latencies of these four participants were nearly identical between the two scorers (Scorer 1: M = 681.0 ms, SD = 78.3; Scorer 2: M = 679.3 ms, SD = 77.9).

Results

NWR

A 2 × 4 (Phoneme Type × Syllable Length) repeated measures analysis of variance was used to analyze participants' performance on NWR. Results are reported in Table 1. Main effects were significant for each factor, indicating that performance was significantly decreased for L7 stimuli compared with E7 stimuli, $F(1, 29) = 76.85$, $p < .001$, $\eta_p^2 = .73$, and significantly different across syllable lengths, $F(2.24, 64.85) = 46.81$, $p < .001$, $\eta_p^2 = .62$ (with degrees of freedom corrected for violation of the sphericity assumption using Huynh–Feldt estimates). These findings for syllable length were analyzed further with post hoc pairwise comparisons using paired-samples t tests, corrected for multiple comparisons. Collapsed across phoneme type, there was no significant difference in performance between one- and two-syllable items ($p = .25$). For all other syllable-length contrasts, performance decreased as the number of syllables increased ($p \le .001$).

There was a significant Phoneme Type × Syllable Length interaction, $F(3, 87) = 9.87$, $p < .001$, $\eta_p^2 = .25$, indicating that performance on L7 items was significantly decreased compared with E7 items as a function of syllable length. As predicted, participants scored significantly lower on L7 nonwords at all syllable lengths and overall compared with E7 nonwords (one-tailed paired-samples t tests, $t \ge 3.29$, $p \le .002$, corrected for multiple comparisons; see Table 1 for all values). The phoneme-type difference increased from just under 4% for one-syllable items to just over 14% for four-syllable stimuli. The magnitude of the performance difference was large for all comparisons (Cohen's $d \ge 0.86$).

Table 1. Nonword-repetition percentage of phonemes correct at each nonword length for early-7 (E7) and late-7 (L7) items.

Nonword length (syllables)	*M* (*SD*)				
	E7	L7	*T*	*p*	Cohen's *d*
1	98.06 (3.58)	94.17 (5.43)	3.29	.002*	0.86
2	98.50 (2.68)	92.00 (5.81)	6.36	< .001*	1.46
3	94.52 (4.06)	86.07 (10.37)	4.26	< .001*	1.09
4	90.19 (7.81)	75.93 (11.56)	8.12	< .001*	1.47
Total	95.32 (3.07)	87.04 (5.87)	8.77	< .001*	1.80

*Significant at $p \leq .01$, correcting for multiple comparisons.

These results are consistent with the findings of Moore et al. (2010), in which there was a large effect size at all syllable lengths and overall (Cohen's $d \geq 1.18$) except for the one-syllable nonword scores (Cohen's $d = 0.01$).

Because we hypothesize that performance relates to consonant acquisition, a secondary analysis was completed to analyze participants' percentage of consonants correct (PCC) using a 2 × 4 (Phoneme Type × Syllable Length) repeated measures analysis of variance. The findings were similar to those of the primary analysis: Main effects were significant for each factor, indicating that PCC was significantly decreased for L7 stimuli compared with E7 stimuli, $F(1, 29) = 66.75$, $p < .001$, $\eta_p^2 = .70$, and significantly different across syllable lengths, $F(3, 87) = 37.53$, $p < .001$, $\eta_p^2 = .56$. These findings for syllable length were analyzed further with post hoc pairwise comparisons using paired-samples *t* tests, corrected for multiple comparisons. Collapsed across phoneme type, there was no significant difference in PCC between one- and two-syllable items ($p = .62$). For all other syllable-length contrasts, performance decreased as the number of syllables increased ($p \leq .01$). There was a significant Phoneme Type × Syllable Length interaction, $F(3, 87) = 20.96$, $p < .001$, $\eta_p^2 = .42$, indicating that PCC on L7 items was significantly decreased compared with E7 items as a function of syllable length. Participants scored significantly lower on L7 nonwords at all syllable lengths and overall compared with E7 nonwords (one-tailed paired-samples *t* tests, $t \geq 3.29$, $p \leq .002$, corrected for multiple comparisons).

Nonword Reading

The nonword-reading results show that participants were significantly less accurate reading nonword items containing late-developing phonemes (E7: $M = 79.00\%$, $SD = 13.98\%$; L7: $M = 72.67\%$, $SD = 14.55\%$; $t = 3.14$, $p = .004$, Cohen's $d = 0.45$). There were no significant differences in participants' reading latencies (for correct responses) between the early and late phoneme groups (E7: $M = 734.35$ ms, $SD = 195.26$; L7: $M = 736.51$ ms, $SD = 172.10$; $t = 0.16$, $p = .88$).

Error and item analyses were completed to further explain the accuracy results. There were 600 total E7 trials and 600 total L7 trials across participants. The number of vowel-pronunciation errors was relatively similar between the E7 and L7 trials: There were 121 E7 trials with vowel errors (20.2% of the total E7 trials) versus 101 L7 trials with vowel errors (16.8% of the total L7 trials). In contrast, the number of consonant-pronunciation errors was disproportionately high in the L7 trials: There were only six E7 trials that included consonant errors (1% of the total E7 trials), versus 90 L7 trials with consonant errors (15% of the total L7 trials). The few E7 consonant errors were idiosyncratic voicing (e.g., /vep/ for *fape*) or place (e.g., /zɔɪn/ for *voin*) substitution errors. Of the 90 L7 trials containing at least one consonant error, 3.3% contained consonant omissions (e.g., /zɑ/ for *zal*), 8.7% contained place substitution errors (e.g., /ʃiʃ/ for *seash*), and 89.1% contained voicing substitution errors (e.g., /ʃaɪθ/ for *shithe,* /zʌs/ for *suzz,* /ðaʊz/ for *thouse*).

An item analysis was used to evaluate which nonword-reading items were read with less than 50% accuracy across participants. There were two E7 nonwords which were read incorrectly by more than half of the participants: *voum* and *doum*. These nonwords do not share a rhyme unit with any English words. It is unsurprising, then, that nearly all errors occurred because of vowel mispronunciations (e.g., using /ɔ/, /o/, or /u/ for /aʊ/). There were four L7 nonwords which were misread by more than half of the participants: *luthe, sathe, shithe,* and *zal*. Of these four nonwords, *luthe* is the only one containing a rhyme unit that does not occur in English. Only 16.7% of the participants who read *luthe* inaccurately made vowel pronunciation errors. All participants who misread *luthe* ($n = 24$) made a consonant voicing error (/θ/ for /ð/). Misreading of *zal* was primarily due to a vowel substitution error (using /ɑ/ for /æ/). There were both consonant and vowel mispronunciations for *sathe* and *shithe*.

From these error and item analyses, it is apparent that the E7 and L7 trials have a relatively similar number of vowel-pronunciation errors, yet there is a disproportionate increase in consonant errors for the L7 trials. Vowel mispronunciations explained the error pattern in both of the E7 items that were read inaccurately by more than half of the participants, whereas the L7 items that were read inaccurately by more than half of the participants were the result of both vowel and consonant mispronunciations.

Auditory Lexical Decision

The auditory lexical-decision results show that participants were significantly less accurate on nonword items containing late-developing phonemes (E7: M = 91.56%, SD = 7.20%; L7: M = 82.00%, SD = 8.60%), $t(29) = 6.02$, $p < .001$. The effect size for accuracy was large (d = 1.23). Although the word items were not the primary focus of the analysis, the same general pattern was observed for these items ($p < .001$, d = 1.04). RT was analyzed for correct responses only, which excluded 8.44% of E7 nonword trials, 18% of L7 nonword trials, 10.22% of E7 word trials, and 21.67% of L7 word trials. Mean RT was not significantly different between E7 and L7 nonwords (E7: M = 458.76 ms, SD = 175.15; L7: M = 480.45 ms, SD = 193.75), $t(29) = 1.30$, $p = .20$. However, the average RT between E7 and L7 words was significantly different ($p = .003$, d = 0.41).

Experiment 2

In Experiment 1, the CAoA effect was observed both in NWR and in lexical-access tasks that placed little to no demands on phonological short-term memory and overt articulation. These findings suggest that CAoA effects are not purely due to the influence of articulatory complexity on articulatory rehearsal. However, Baddeley (2003) notes that articulatory rehearsal involved in short-term memory is not reliant on overt articulation, but could involve covert articulatory processes (see also Baddeley & Wilson, 1985). Thus, Experiment 2 takes this work one step further by examining the influence of CAoA effects in lexical-access tasks when articulatory rehearsal is suppressed. Concurrent articulation—that is, repeatedly saying aloud a rote word or phrase—is a classic method used to interrupt any covert articulatory processes (Baddeley, 1986). A 2 × 2 × 2 experimental design was used to examine the early–late phoneme contrast in auditory and visual lexical decision, with and without concurrent articulation. If CAoA effects are linked to articulatory rehearsal, then they should be eliminated when a lexical-access task is performed under conditions of concurrent articulation. In contrast, the CAoA effects should persist if they arise from a central level of phonological representation.

Method

Participants

The participants in this study were 20 undergraduate students, 18–22 years of age (12 women, eight men), recruited using the same method and criteria as for Experiment 1. Participants who completed Experiment 1 were ineligible to participate in Experiment 2.

Technical Specifications

The specifications for developing and presenting both auditory and visual stimuli were identical to those in Experiment 1. It is important to note that there were no differences in technical specifications between the E7 and L7 stimuli. Presentation volume of auditory stimuli was set at a comfortable listening level on the basis of feedback from pilot subjects. Visual stimuli were presented on the computer in 30-point font in uppercase letters. Participants' responses were recorded using the serial response box described for Experiment 1.

Stimuli

Four pairs of word–nonword lists were constructed for the lexical-decision tasks using one-syllable CVC stimuli (see Supplemental Materials S7–S10). No word or nonword was repeated across the lists. Phoneme recurrence was considered across lists, and the phonological and orthographic factors addressed in Experiment 1 were considered in Experiment 2 as well (see Supplemental Materials S11 and S12).

Each nonword list consisted of 30 nonwords, half (n = 15) comprising E7 consonant phonemes and half comprising L7 consonant phonemes. Average durations (in milliseconds) of the recorded E7 and L7 stimuli were similar across the four nonword lists ($F \leq 3.11$, $p \geq .08$). In each of the four word lists, there were 15 items comprising E7 consonant phonemes, 12 items comprising L7 phonemes, and three mixed items ($C_{L7}VC_{E7}$, as described for Experiment 1). Average durations (in milliseconds) of the recorded E7 and L7 stimuli were similar across the four lists ($F \leq 1.53$, $p \geq .21$).

Experimental Procedure

Administration. All procedures were administered individually in a quiet testing room. After consenting to participate, subjects were asked to report any history of speech, language, or reading disorder and then were given brief screenings for vision, hearing threshold, and diadochokinetic rates. Each participant completed four experimental conditions on the computer using E-Prime (Schneider et al., 2002): auditory lexical decision, auditory lexical decision with concurrent articulation, visual lexical decision, and visual lexical decision with concurrent articulation.

The task conditions were administered in pseudorandom order such that each subject was randomly assigned to a fixed condition order. The condition orders were counterbalanced so that each task condition was administered as the first condition five times, as the second condition five times, and so on. There was one exception due to examiner error: Auditory lexical decision was the first task condition six times and the second condition four times across participants; visual lexical decision was the first task condition four times and the second condition six times. The four word–nonword lists were counterbalanced across the four task conditions in pseudorandom order such that every subject received each list one time, and across subjects each list occurred in every task five times.

Participants repeatedly counted from one to four during conditions of concurrent articulation. We chose this phrase because these four numbers are relatively well balanced with both early- and late-developing phonemes and the phrase has been shown to produce similar effects as other commonly used phrases for concurrent articulation (Baddeley, 1986). Prior to the experiment, participants rehearsed the phrase approximately one time per second using an online metronome, and then practiced without the metronome to demonstrate that they could maintain the approximate pace.

The four task conditions were administered similarly to the procedures described for the auditory lexical-decision task in Experiment 1. Instructions were modified accordingly for the tasks using visual stimuli instead of auditory stimuli and for the tasks with concurrent articulation. A small break was added halfway through each task so that participants would have shorter intervals of concurrent articulation; the break was included in all tasks (even those without concurrent articulation) to maintain uniformity across the tasks.

Accuracy and RT were recorded. RT was recorded as the duration from the onset of the stimulus item to the onset of a participant response, subtracting the length of the digital file for the auditory trials. In the few cases (0.2% of all trials) in which a participant was not ready for a task to begin or for a task to resume after a break, the RT for that trial was not included in the analysis.

Results

In a 2 × 2 × 2 (Phoneme Type × Presentation Modality × Concurrent Articulation) repeated measures analysis of variance that was conducted to examine the lexical-decision accuracy for nonword items, main effects were significant for each factor, $F(1, 19) \geq 4.83$, $p \leq .04$, indicating that performance was significantly better for E7 stimuli compared with L7 stimuli, for visual items compared with auditory items, and without concurrent articulation compared with performance with concurrent articulation. There were no significant interactions ($p \geq .25$). For average nonword RT, a significant main effect was found for presentation modality, $F(1, 19) = 127.81$, $p < .001$, but not for phoneme type or concurrent articulation, $F(1, 19) \leq 0.04$, $p \geq .84$. There were no significant interactions for nonword RT ($p \geq .21$). Results for nonword accuracy and RT are shown in Figure 1. Although the word items were not the primary focus of the analysis, there was no significant effect of phoneme type for real-word accuracy ($p = .48$) nor real-word RT ($p = .26$). Otherwise, the general pattern of performance for presentation modality and concurrent articulation was similar to the pattern observed for nonword stimuli. There were no significant interactions for word accuracy or RT ($p \geq .16$).

Discussion

Over recent decades, nonword repetition has been a major focus of study in communication sciences and disorders. This is not surprising considering that “the ability to repeat multisyllabic nonwords . . . probably represents the most effective predictor of language learning ability that is currently known” (Gathercole, 2006, p. 513). The high sensitivity and specificity of NWR in distinguishing typical development from language impairment has motivated research into the underlying nature of NWR. In clinical practice, NWR is marketed and primarily used as a measure of phonological short-term memory. However, continued research has provided strong evidence of the task’s multidimensionality, particularly showing that NWR performance is affected by factors associated with both phonological short-term memory and long-term phonological knowledge (e.g., Gupta, 2006; Gupta & Tisdale, 2009; Rispens & Baker, 2012).

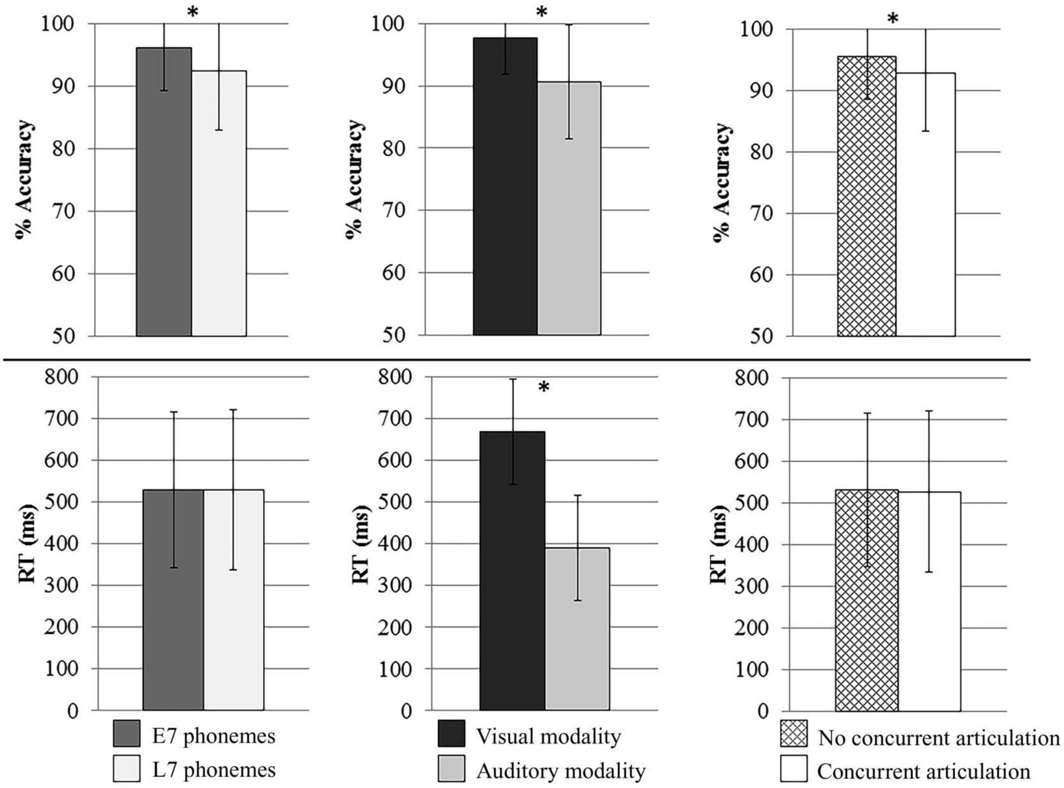
Nearly all of the previous research examining long-term-memory effects on NWR has focused on lexical-level variables. Little attention has been paid to phoneme-level variables, although there are both empirical and theoretical reasons to expect significant effects. Noteworthy are the findings by Moore et al. (2010) in which CAoA affected NWR performance, such that young adults repeated nonwords with decreased accuracy when the stimuli comprised only later-developing consonants compared with stimuli comprising only early-developing consonants. The purpose of this study was to determine whether the CAoA effect in the work of Moore et al. could be replicated, and to extend the previous work to better understand whether the CAoA effect reflects an articulatory influence in NWR or an influence from a central level of phonological representation.

Replication of CAoA Effects in NWR

Findings from this work replicated the CAoA effect in NWR observed by Moore et al. (2010). As in that study, in the current study significant early–late phoneme differences were observed for two-, three-, and four-syllable nonwords. One difference between the two studies is that Moore et al. did not observe significant CAoA effects between the one-syllable items. In contrast, the current work resulted in significant early–late phoneme differences between the one-syllable nonwords, and the effect size was large (albeit not as large as the effect sizes for the multisyllabic items). One possible explanation for the different results in one-syllable nonwords is the stimulus design. Moore et al. used tightly controlled stimuli; however, efforts were made in the current work to take the stimulus design one step further by controlling for additional extraneous variables such as phoneme recurrence, phonotactic and biphone probability, and lexicality of constituent syllables.

The replication of findings also weakens alternative explanations for the early–late phoneme contrast related to subject characteristics, for example dialectal experiences. Because regional American English dialects are based primarily on vowel contrasts, there are no a priori predictions that dialect would affect early- or later-developing stimulus items differently, but it is possible that individual items could be affected, potentially biasing one stimulus list over another. However, as mentioned already, the consistent findings across studies with different participant groups and different stimulus lists weaken this possibility. Further, in recent work, Moore and Smith (2014) observed the CAoA effect in both NWR and a word-learning task with college students from another university. Taken together, it seems unlikely that subject characteristics alone could account for the CAoA effect.

Figure 1. Main effects of phoneme type, presentation modality, and concurrent articulation for accuracy and response-time (RT) measures on nonword items in Experiment 2 lexical decision. RT is reported for correct responses only. Error bars show standard deviation. An asterisk indicates a significant main effect at $p \leq .04$.



CAoA Reflects a Central Level of Phonological Representation

One of the primary goals of this study was to test whether the CAoA effect reflects either an articulatory or a central phonological representational influence on short-term memory. Although CAoA typically has been associated with articulatory complexity (e.g., Kent, 1992; Stokes & Surendran, 2005), there are several reasons to favor a central phonological view over an articulatory view of the CAoA effect in NWR. For instance, it is unexpected to observe developmental effects of articulation in highly skilled adult speakers, yet the CAoA effect has been robust across three different groups of subjects (i.e., Experiments 1 and 2, and Moore et al., 2010). In addition, explaining the CAoA effect in NWR as a purely articulation-based influence on short-term memory seems incongruous with other findings that have examined articulatory and motoric demands in NWR performance and found modest or no effects (e.g., Archibald et al., 2013; Archibald & Gathercole, 2006; Edwards & Lahey, 1998). Further, there are several plausible mechanisms described earlier using models of speech production and the organization of neural networks to support the view that the CAoA effect in NWR arises from a central level of phonological representation.

The results of our experiments are consistent with the proposition that the CAoA effect arises from a central level of phonological representation, which in turn influences lexical access for a range of tasks that vary in articulatory, auditory-perceptual, and phonological short-term-memory demands. Across all tasks in Experiments 1 and 2, participants on average performed less accurately on nonword items comprising later-developing consonants than on those comprising early-developing consonants. The CAoA effect persisted in lexical-decision tasks even when concurrent articulation was used to minimize both overt and covert articulatory processes. The CAoA effect also persisted in nonword-reading and visual lexical-decision tasks that used visually presented stimuli to minimize auditory-perceptual and phonological short-term-memory demands.

Early–late phoneme differences, notably, were robust for accuracy measures but not for reading-latency and RT measures. These findings seem to work against the articulatory view of the CAoA effect. The basis of this articulatory view is that increased articulatory complexity decreases the rate of speech production during rehearsal, thus making short-term storage of phonological information more prone to decay. Under this view, early–late phoneme differences in response rate might be expected. Thus, the lack of early–late phoneme differences in reading-latency and RT measures seems to work against the articulatory view of the CAoA effect. Taken together, both the accuracy and response-rate results from this study are consistent with the view that CAoA effects reflect a central phonological representational influence on short-term memory.

Other observed patterns in the results can be explained from the central phonological view of the CAoA effect as well. For example, the CAoA manipulation interacted with word length such that the magnitude of the CAoA effect was larger for multisyllabic items than it was for the one-syllable items. As previously stated, from an articulatory view the CAoA effect might be explained as an influence of articulatory complexity that reduces the efficacy of articulatory rehearsal, thus making the short-term storage of phonological information more prone to decay. This articulatory effect would be exacerbated in longer words that already take longer to rehearse and are more prone to decay, subsequently creating an additive effect as word length increases. However, it is important to note that this articulatory view cannot account for the observed CAoA effect in the one-syllable NWR items, nor the presence of the effect in tasks with visually presented one-syllable CVC items (i.e., nonword reading and Experiment 2 visual lexical-decision conditions).

Alternatively, the central phonological view of the CAoA effect suggests that the long-term-memory effect of age of acquisition is pervasive across all of the stimuli in the lexical-access tasks. Under this view, error-rate and phoneme-distinctiveness effects illustrate two possible explanations for the interaction of word length and CAoA. For example, if each syllable has a 4% error rate, then the probability of an error at the item level increases as the nonword increases in syllable length. Another consideration is that the syllables in longer words are less discriminable than those in shorter words, and therefore are more prone to misselection errors (Acheson & MacDonald, 2009). In typical speech misarticulations, phoneme swaps occur in similar syllable positions of nearby syllables or are caused by feature overlap involving a phoneme competitor (Dell, 1986). If there is competition between phonemes in the same syllable position, a two-syllable word will have two onset phonemes that are susceptible to swaps, whereas with a three-syllable word there are three. According to this view, phonemes that are processed less efficiently in long-term memory (i.e., later-developing phonemes) could result in a greater number of phoneme swaps due to the decreased ability to resolve competing information. This central phonological viewpoint is compelling because it is able to account for both the interaction with word length and the observed CAoA effect in one-syllable stimuli in which there were little to no phonological short-term-memory demands.

The results also showed less-robust early–late phoneme differences for words compared with nonwords. There were statistically significant early–late phoneme differences with nonword stimuli in all tasks, whereas an early–late phoneme difference only reached significance with the word stimuli in one of the lexical-decision tasks. Computational work by Gupta and MacWhinney (1997) suggests that less-robust results for real words than nonwords are due in part to the stronger activation of word forms due to support from semantic representations. Lexical influences can help to prevent an error or may support an error when a phoneme misselection creates a valid lexical entry. In the case where lexical influences may support an error, it would be harder for the correct phoneme to be selected against the lexical bias created by the incorrect phoneme. Thus, with stronger activation of word forms, the lexical influence could have leveled the playing field and muted the CAoA effect. Future work could more directly test these ideas by directly manipulating both lexicality and CAoA in NWR. The lexical bias observed in this current work, importantly, is a common phenomenon observed in lexical-access tasks and does not preclude an articulatory nor a central phonological view of the CAoA effect in NWR.

In summary, the findings here are consistent with the view that a CAoA effect in NWR arises from a central level of phonological representation. Although CAoA typically has been associated with articulatory complexity (e.g., Kent, 1992; Stokes & Surendran, 2005), a purely articulation-based view of the CAoA effect seems to be inconsistent with both the findings reported here and previous work examining articulatory and motoric demands in NWR performance (e.g., Archibald et al., 2013; Archibald & Gathercole, 2006; Edwards & Lahey, 1998). In the current study, the CAoA effect persisted across lexical-access tasks that minimized articulatory, short-term-memory, and auditory-perceptual demands. Other patterns observed in the results, such as word length and lexical bias, are compatible with the central phonological view of the CAoA effect as well.

Alternative Explanations for the CAoA Effect

Other possible explanations for the early–late phoneme differences should be considered. One consideration within the central phonological view is that there is greater sublexical knowledge for the early-acquired consonants than the later-acquired consonants, so that the early-acquired phonemes may be accessed and retrieved in larger chunks (Jones, 2016, p. 81) to allow for more efficient performance in NWR. Jones describes a computational model in which new lexical information is segmented at the phoneme level, and because the model has more experience with the lexical information, it represents the information in larger "chunks" (e.g., biphones, words). Chunking allows more information to be held in phonological memory even with a fixed capacity. For example, with a phonological memory capacity of 4.5 items, if each item is a single phoneme then there would not be as much information held in memory as if each item were a biphone. It is possible that early-acquired phonemes are better learned through more articulatory experience and, thus, are represented in larger grain sizes within long-term knowledge, which could facilitate performance with the early-acquired items in lexical-access tasks.

Another consideration is whether the CAoA effect, like phonotactic probability, is intertwined with lexical-level representation on the basis of the frequency of phoneme occurrence in the language corpus. Evidence against this explanation is demonstrated in work showing that frequency of phoneme occurrence in English does not account for significant variance in the emergence or mastery of consonant sounds (Mader, 1954; Mines, Hanson, & Shoup, 1978; Stokes & Surendran, 2005). Further, a significant effect of CAoA has been observed in college students (Moore et al., 2010), whereas the phonotactic-probability manipulation has been shown to attenuate with age (Munson, 2001). Thus, the CAoA effect on NWR seems to reflect a phoneme-level influence from long-term memory that is relatively independent of lexical-level knowledge.

To further address this point, this study controlled potential confounding phonological and orthographic factors more rigorously than previous related work. In fact, at least six and as many as 11 potential confounding factors were controlled in each of the stimulus lists in Experiments 1 and 2, so that on average there were no statistically significant differences between the early- and late-developing stimuli on each lexical and sublexical factor. However, some of the *p* values were trending toward significance ($p < .05$), potentially drawing into question whether the early- and late-developing stimulus lists were comparable on those factors. In some instances, the mean values favored the early-developing stimuli and thus worked against the results that showed that participants have decreased performance on stimuli comprising later-developing consonants. In other instances, the mean values favored the later-developing stimuli and may be considered a potential confound to the results. The consistent findings across Experiments 1 and 2 with different stimulus lists weaken the threat of such confounds, but one way to avoid the nearly impossible task of perfectly balancing lists on a dichotomous variable is to treat CAoA as a continuous variable (e.g., Sosa & Stoel-Gammon, 2012) using a regression-based model for analysis. Work using this regression-based approach has been completed, and preliminary results are congruent to the findings presented here (McDonald & Moore, 2016).

Another consideration is the auditory-perceptual qualities of the stimuli, including the spoken duration of the auditory stimuli as well as the perceptual salience of individual phonemes. Spoken duration was controlled in each set of auditory stimulus lists in Experiments 1 and 2, so that on average there were no statistically significant differences in spoken duration between the early- and late-developing stimuli. Of the one-syllable stimulus sets in the NWR and lexical-decision tasks, some of the *p* values were trending toward significance ($p < .05$), potentially drawing into question whether the early- and late-developing stimulus lists were comparable in duration. Similar to the previous discussion, in some instances the mean durations were shorter for the stimuli with later-developing consonants and thus worked against the results—for example, if shorter durations place fewer demands on phonological memory capacity and therefore make task completion easier. In other instances, most notably the one-syllable CVC stimuli, the mean durations were longer for the L7 stimuli. However, one-syllable stimuli are not expected to tax the phonological memory capacity of typical young adults, so that it is unlikely that the small difference (≤ 51 ms) between mean durations of any set of E7 and L7 stimulus lists in question can fully account for the decreased performance on items with later-developing consonants. Further, in Experiment 2 there was no interaction between presentation modality (i.e., auditory versus visual stimulus presentation) and the CAoA variable. Taken together, it seems unlikely that the effect of spoken duration accounts for the decreased performance on items with later-developing consonants. Still, future work could digitally modify the spoken durations of the stimuli so that all of the stimuli within an E7–L7 stimulus set are the same length, so as to eliminate this potential confound.

The second consideration regarding the auditory-perceptual qualities of the stimuli is the perceptual salience of individual phonemes. Certain fricatives are easily confusable, such as /f/–/θ/ and /v/–/ð/, particularly when there is no verbal context (as in nonwords) and no visual support (as with audio-recorded stimuli; Miller & Nicely, 1955). However, it is important to note that each of these two highly confusable sets of fricatives contains an early- and a late-developing phoneme, so acoustic confusability posed no obvious disadvantage for one phoneme group over the other. Other consonant pairs may also be susceptible to perceptual-confusability errors, such as /m/–/n/, /s/–/f/, and /s/–/z/, but these pairs again were constituents of both the early- and late-developing phonemes. As stated, because there was no interaction between presentation modality (i.e., auditory versus visual stimulus presentation) and the CAoA variable in Experiment 2, the effect of perceptual salience does not appear to account fully for the decreased performance on items with later-developing consonants.

NWR Is Not a Pure Measure of Phonological Short-Term Memory

The results from this study indicate that nonword repetition is sensitive to phoneme-level influences from long-term memory that are relatively independent of lexical-level knowledge. These phoneme-level influences, importantly, are present in highly skilled young adult speakers. If task performance is sensitive to such fine-grained manipulations in skilled speakers, it is no surprise that NWR is sensitive to other variations in stimuli, language experience, and language ability, as observed across a broad range of groups including adults and children, first- and second-language learners, and individuals with and without a variety of communication disorders. The implications of this study, together with the multitude of work showing that NWR relies on multiple processes (e.g., Archibald & Gathercole, 2006; Briscoe, Bishop, & Norbury, 2001; Coady & Evans, 2008; Edwards & Lahey, 1998; Gathercole, 2006; Graf Estes et al., 2007; Gupta, 2006; Snowling, Chiat, & Hulme, 1991), underscore the need to move away from the use of this task as a specific measure of phonological short-term-memory storage. A more evidence-based adoption of the task would use its high clinical utility in identifying impaired performance, but its sensitivity to so many speech and language factors does not allow differential diagnoses or conclusions about specific underlying areas of deficit (e.g., phonological short-term-memory deficits).

Future work could utilize this study's battery of tasks in children with and without language impairment to determine the influence of phoneme-level long-term phonological memory on NWR in developing and impaired language systems. This work could help to adjudicate between the various theoretical accounts of deficits that have been associated with language impairment. For example, a positive effect of phoneme-level influences from long-term memory could suggest that language impairment in children may not be strictly from a limited capacity of short-term memory (compare Gathercole, 2006) or from limited vocabulary knowledge (e.g., Metsala & Chisholm, 2010). Many studies have used descriptive designs to identify the contribution of various speech and language factors involved in the NWR deficits observed in children with language and reading impairment, but there is a paucity of experimental research in this area. Thus, the systematic approach of this study has both potential theoretical and clinical implications.

Conclusion

The work presented here replicates the CAoA effect in NWR and further extends previous work to show the persistent CAoA effect across a diverse battery of lexical-access tasks varying in articulatory, phonological short-term-memory, and auditory-perceptual demands. The results are consistent with the perspective that CAoA effects on NWR reflect an influence from a central level of phonological representation. The results thus align with the body of literature on NWR demonstrating the sensitivity of repetition performance to multiple speech and language factors. This multidimensionality of the NWR task precludes it from being used as a stand-alone tool in identifying specific areas of deficits (such as phonological short-term-memory deficits). However, the NWR task has the potential to contribute to the study of the theoretical underpinnings and diagnosis of communication impairment when used in combination with other lexical-access tasks that also systematically manipulate the various speech and language factors that can influence NWR performance.

Acknowledgments

This project was funded by the University of Pittsburgh School of Health and Rehabilitation Sciences Research Development Fund and National Institutes of Health Grant 1R01HD060388, awarded to Julie A. Fiez.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135,* 50–68.

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research, 49,* 970–983.

Archibald, L. M. D., Joanisse, M. F., & Munson, B. (2013). Motor control and nonword repetition in specific working memory impairment and SLI. *Topics in Language Disorders, 33,* 255–267.

Baddeley, A. (1986). *Working memory*. Oxford, United Kingdom: Clarendon Press.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36,* 189–208.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589.

Baddeley, A., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language, 24,* 490–502.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459.

Bishop, D. V. M. (2002). The role of genes in the etiology of specific language impairment. *Journal of Communication Disorders, 35,* 311–328.

Bishop, D. V. M., Bishop, S. J., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *Journal of Speech, Language, and Hearing Research, 42,* 155–168.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *The Journal of Child Psychology and Psychiatry, 37,* 391–403.

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *The Journal of Child Psychology and Psychiatry, 42,* 329–340.

Children's Test of Nonword Repetition (CN REP) [Product description]. (n.d.). Retrieved from http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildMemory/ChildrensTestofNonwordRepetition(CNREP)/ChildrensTestofNonwordRepetition(CNREP).aspx

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders, 43,* 1–40. https://doi.org/10.1080/13682820601116485

Coady, J., Evans, J. L., & Kluender, K. R. (2010). Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International Journal of Language & Communication Disorders, 45,* 494–509.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech and Hearing Research, 36,* 1051–1054.

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics, 16,* 211–222.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136–1146.

Edwards, J., Beckman, M. E., & 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,* 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Edwards, J., & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics, 19,* 279–309.

Falcaro, M., Pickles, A., Newbury, D. F., Addis, L., Banfield, E., Fisher, S. E., . . . the SLI Consortium. (2008). Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in specific language impairment. *Genes, Brain and Behavior, 7,* 393–402. https://doi.org/10.1111/j.1601-183X.2007.00364.x

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23,* 83–94.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27,* 513–543.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29,* 336–360.

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50,* 177–195.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96,* 280–301. https://doi.org/10.1016/j.bandl.2005.06.001

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 1213–1236.

Gupta, P. (2006). Nonword repetition, phonological storage, and multiple determinations. *Applied Psycholinguistics, 27,* 564–568.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59,* 267–333. https://doi.org/10.1006/brln.1997.1819

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language, 61,* 481–502.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687–715.

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.

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65–90). Timonium, MD: York Press.

Kent, R. D., Kent, J. F., & Rosenbek, J. C. (1987). Maximum performance tests of speech production. *Journal of Speech and Hearing Disorders, 52,* 367–387.

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.

Mader, J. B. (1954). The relative frequency of occurrence of English consonant sounds in words in the speech of children in grades one, two, and three. *Speech Monographs, 21,* 294–300.

McDonald, T., & Moore, M.. (2016, November). Examining the effect of varied stimulus designs in phonological processing. Poster presented at the Annual Convention of the American Speech-Language-Hearing Association, Philadelphia, PA.

Metsala, J. L., & Chisholm, G. M. (2010). The influence of lexical status and neighborhood density on children's nonword repetition. *Applied Psycholinguistics, 31,* 489–506.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27,* 338–352.

Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language and Speech, 21,* 221–241.

Moore, M. W. (2012). *Differences between early-developing and late-developing phonemes in phonological processing* (Unpublished doctoral dissertation). University of Pittsburgh, PA. Retrieved from http://d-scholarship.pitt.edu/12046/

Moore, M., & Smith, E.. (2014, November). Examining the role of long-term phonological knowledge in learning new words. Poster presented at the Annual Convention of the American Speech-Language-Hearing Association, Orlando, FL.

Moore, M. W., Tompkins, C. A., & Dollaghan, C. A. (2010). Manipulating articulatory demands in non-word repetition: A "late-8" non-word repetition task. *Clinical Linguistics & Phonetics, 24,* 997–1008.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research, 44,* 778–792.

Munson, B., Edwards, J., & Beckman, M. E. (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,* 61–78.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48,* 1033–1047.

Pattamadilok, C., Knierim, I. N., Kawabata Duncan, K. J., & Devlin, J. T. (2010). How does learning to read affect speech perception? *The Journal of Neuroscience, 30,* 8435–8444.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research, 55,* 683–694.

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 29–33. https://doi.org/10.1037/0278-7393.28.1.29

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.

Shriberg, L. D., & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Research, 37,* 1100–1126.

SLI Consortium. (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *The American Journal of Human Genetics, 70,* 384–398.

SLI Consortium. (2004). Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *The American Journal of Human Genetics, 74,* 1225–1238. https://doi.org/10.1086/421529

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55,* 779–798.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics, 12,* 369–373.

Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research, 55,* 596–608.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29,* 41–78.

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,* 577–591.

Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition, 18,* 559–567.

Treiman, R., & Zukowski, A. (1988). Units in reading and spelling. *Journal of Memory and Language, 27,* 466–477.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36,* 481–487.