Journal of Experimental Child Psychology 183 (2019) 276–294

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Children's suffix effects for verbal working memory reflect phonological coding and perceptual grouping

Joanna H. Lowenstein [*], Courtney Cribb, Popy Shell, Yi Yuan, Susan Nittrouer

Department of Speech, Language, and Hearing Sciences, University of Florida, Gainesville, FL 32610, USA

A R T I C L E I N F O

Article history:
Received 11 September 2018
Revised 7 March 2019

Keywords:
Verbal working memory
Suffix effect
Phonological loop
Perceptual grouping
Children
Developmental

A B S T R A C T

When listeners recall order of presentation for sequences of unrelated words, recall is most accurate for first and final items. When a speech suffix is appended to the list, however, the advantage for final items is diminished. The usual interpretation is that listeners recover phonological structure from speech signals and use that structure to store items in a working memory buffer; the process of recovering phonological structure for a suffix interrupts that processing for the final list item. Although not mutually exclusive, another hypothesis suggests that perceptual grouping of list items and suffix based on common acoustic structure is necessary for the effect to occur. To evaluate these accounts as well as potential age-related differences, adults and 8-year-old children were asked to recall order of presentation for a closed set of nouns in five suffix conditions: none, auditory *go,* lipread *go,* a tone, and a colored circle. Overt articulation was prohibited, but attention to the suffix was mandated. Children's serial recall was generally poorer than that of adults, but patterns across list positions were similar for both age groups. Participants showed stronger effects for speech suffixes than for nonspeech suffixes regardless of whether suffixes were seen or heard, but effects were not restricted to final list items. And although effects of heard and lipread suffixes were similar for early list items, heard speech exerted greater effects on late

---

[*] Corresponding author.
E-mail address: jlowenstein@phhp.ufl.edu (J.H. Lowenstein).

https://doi.org/10.1016/j.jecp.2019.03.003

list items. Outcomes suggest that some effect of heard and lipread speech suffixes arises from their shared phonological structure, but this effect is strongest when perceptual grouping occurs.

Introduction

The ability to hold information in a memory buffer just long enough to perform an operation on that information is essential to cognitive functioning and to many kinds of learning. For example, this skill is needed to perform the mental arithmetic a person might use while shopping to estimate what a set of items would cost at checkout. Students listening to lectures that present new material must be able to hold briefly in memory early-presented information in order to integrate it with later-presented information. Thus, this ability—typically termed *working memory*—is important to daily functioning. It has been extensively studied, and several models have been proposed to help explain various aspects of its operations, including the nature of the representation of speech signals in the short-term buffer. For example, a dual-component model originally described by Baddeley and colleagues (e.g., Baddeley, 1992, 2007; Baddeley & Hitch, 1974) helps to explain the coding of signals in that buffer as well as subsequent processing. The two components of this model are a front end that recovers some form of structure from the sensory input and deposits that structure into a temporary store and a central processor that performs operations on that recovered and temporarily stored structure. The study reported here focused on the first of these components: the front end responsible for recovering structure from the sensory input and depositing that structure in a buffer.

Traditional versions of the dual-component model of working memory (e.g., Baddeley & Hitch, 1974) describe two different front ends: a phonological loop that is recruited into operation when the input signal is speech and a visuospatial sketchpad that operates when the input signal is visual. The current study was designed to extend our understanding of verbal working memory and so focused on the phonological loop only. This particular front end—also known as the articulatory loop—is presumed to operate by recovering phonological structure from the speech signal and using that structure to code words into a temporary memory buffer for processing (e.g., Baddeley, Gathercole, & Papagno, 1998; Morra, 2015). Support for this presumption derives from experiments showing that when the phonological structure comprising the signal is more distinctive, items are stored in the memory buffer more durably. For example, words presented in a list are recalled with greater accuracy when those words do not rhyme rather than when they rhyme (Baddeley, 1966; Botvinick & Plaut, 2006; Conrad & Hull, 1964; Macnamara, Moore, & Conway, 2011; Salame & Baddeley, 1986). However, other experiments have demonstrated that when the acoustic structure of the words presented is degraded in some manner, memory storage is diminished even if the words are non-rhyming. For example, Nittrouer and Lowenstein (2014) presented non-rhyming consonant–vowel–consonant (CVC) words that were either unprocessed, and so clear in acoustic structure, or noise vocoded, and so spectrally smeared. The vocoded signals were recognized with perfect accuracy by both adults and children when presented one at a time for matching to pictures. Nonetheless, when those same words were presented in multi-item sequences, these same listeners were significantly less accurate at recalling presentation order for the vocoded words compared with the unprocessed words. Because the signal processing resulted in degraded signals while maintaining recognition, the possibility arises that the acoustic structure of the signals being stored may affect durability.

Serial recall and the suffix effect

One experimental method that has proven to be useful in the examination of verbal working memory involves the presentation of strings of words auditorily, untethered by syntactic structure. In serial recall tasks, participants are asked to recall the order in which words were presented. Outcomes for these tasks with verbal materials typically show a U-shaped function, with recall being more accurate

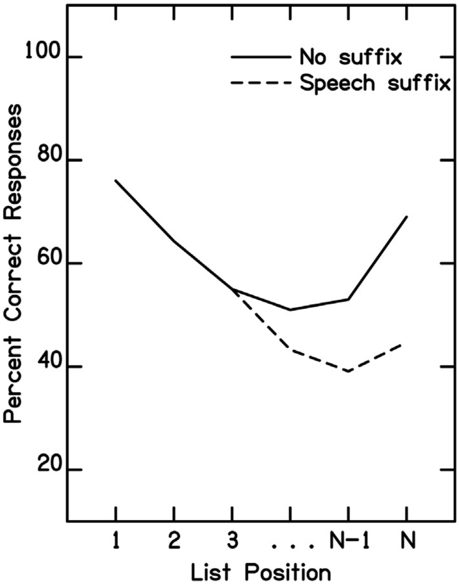

Fig. 1. Mean percent correct recall for a sample working memory task.

for early list and late list positions than for positions in the middle (i.e., the primacy and recency effects). This pattern is illustrated by the solid line in Fig. 1, which presents characteristic outcomes for experiments involving serial recall. The traditional interpretation for this outcome is that early list items have been completely analyzed into phonological structure and durably stored in the memory buffer; although not as thoroughly analyzed in terms of phonological structure, the auditory trace of late list items is fresh, so these items can be recalled relatively well based on acoustic structure (e.g., Murdock, 1962).

A variation of this serial recall paradigm can inform us regarding the nature of the code used to store words in the temporary memory buffer: Adding a suffix that does not need to be recalled to the end of a sequence can result in diminishment of the recency effect under certain circumstances. An example of an effective suffix could be the word *go* appended to the end of the sequence. The dashed line in Fig. 1 illustrates that such suffixes take their greatest toll on recall of the very last item in the list. Investigators typically interpret their outcomes as evidence that the suffix is analyzed into its phonological form even though it does not need to be recalled (Campbell & Dodd, 1980; Crowder, 1983; Salter & Colley, 1977; Spoehr & Corin, 1978). The process of that analysis interferes with the ongoing recovery of phonological structure for the last list item. Nonspeech suffixes fail to evoke the effect, presumably because there is no phonological structure to be recovered.

This interpretation has been supported by experiments in which various sorts of speech suffixes were appended to word lists. For example, Spoehr and Corin (1978) presented lists of digits and used *zero* as the suffix. This zero suffix was presented in many forms, including a spoken version with no availability of lipread information, a silently articulated version that made available only lipread information, a version presented as the numeral written on a card, and a version presented with the word written on a card. Results showed that both the spoken and lipread versions of the word evoked the suffix effect, but the written versions did not. Numerous studies subsequently replicated this fundamental outcome. In particular, other investigators demonstrated that the suffix effect was not elicited when an acoustic tone similar in length to a spoken word was appended to a list of spoken digits or words even when a spoken suffix appended to the same lists successfully elicited the effect (de Gelder & Vroomen, 1992; Nairne & Crowder, 1982; Rowe & Rowe, 1976). In fact, the suffix did not need to be heard or seen; instead, it could be silently articulated by the research participants themselves and the effect was observed (Nairne & Crowder, 1982; Nairne & Walters, 1983). Thus, a general conclusion reached as a consequence of these studies was that the suffix must be articulatory in nature, as are the items in the sequence, in order to impose its effect. The finding that heard and lipread suffixes operate in the same manner bolstered both the position that the nature of the coding of linguistic items in working memory is phonological and the position that lipreading and hearing a spoken suffix provide identical access to phonological structure. These positions found support more generally from evidence that sensory inputs from seen and heard speech integrate seamlessly during the perceptual process (McGurk & MacDonald, 1976) and that both sensory modalities provide transparent access to articulation, the basis of phonological coding (Fowler, 1996).

This direct perception of articulatory events across modalities is presumed to be an essential property of perception and so should be available from birth. Indeed, empirical studies have demonstrated that infants as young as 2 months can match visual and auditory speech signals (Kuhl & Meltzoff, 1982; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983) even without experiencing pairings of these signals for specific phonemic sequences in their native language environments (Danielson, Bruderer, Kandhadai, Vatikiotis-Bateson, & Werker, 2017). Accordingly, if the effects of heard and lipread speech suffixes are due to their articulatory similarity, children should show the predicted effect for spoken suffixes as well as similarity to adults in the nature and magnitude of that effect across heard and lipread suffixes.

Alternative account

For decades, the dual-component model with its phonological front end for speech inputs has remained a predominant view of verbal working memory, thereby providing tacit support for the usual account of the suffix effect as phonologically based. But there have been challenges to that account. Some investigators have contended that a suffix must be perceptually grouped with the list items in order to exact its toll on recall; suffixes that are perceptually grouped with the list items diminish the recency effect, and those suffixes that do not get grouped with list items fail to do so (Bloom, 2006; Maidment, Macken, & Jones, 2013). In one demonstration of this effect, Nicholls and Jones (2002) presented strings of digits with the auditory suffix *go* to participants under two conditions: with no concurrent signal and with the word *go* presented repeatedly as a concurrent string. In both cases, *go* was positioned at the same temporal location at the end of the sequence. Results showed that the suffix effect was evoked only when there was no concurrent train of the word *go* accompanying the presentation of to-be-recalled list elements. The interpretation of this finding was that the concurrent train of *go* caused a separate perceptual stream consisting of that word to be created, thereby preventing the occurrence of the word at the end of the sequence from being grouped with the list of digits.

The phenomena of perceptual grouping of auditory signals and of its counterpart of stream segregation both are part of auditory scene analysis. This set of processes describes how listeners separate the incoming acoustic signal according to its independent sources and then organize the components associated with each source (Bregman, 1990). According to this model, there are ‘‘primitive” functions, defined as those present at birth, so requiring no experience, as well as ‘‘schema-based” functions, defined as those that are dependent on experience. The principle that would presumably be responsible for perceptual grouping of a suffix with a list of words should be primitive in nature, according to the model. Specifically, separate parts of a complex signal occurring at disparate frequencies are grouped together if they share a common fundamental frequency, meaning that harmonic structure is consistent across the spectrum. Because this phenomenon matches the description of primitive auditory functions, infants should demonstrate the phenomenon in their perception. And indeed, several studies have found evidence that infants do exhibit perceptual grouping (or its counterpart, segregation) similar to that of adults (Demany, 1982; Fassbender, 1993). However, methodological constraints limit interpretation, and studies with older children fail to support that conclusion.

In addition to promoting integration across the spectrum, an important function of the shared fundamental frequency is that it provides a basis for masking release in speech perception: If the target talker has a substantially different fundamental frequency than the speech doing the masking, recognition thresholds are better than they are otherwise—at least for adults (Darwin, Brungart, & Simpson, 2003; Lee & Humes, 2012; Summers & Leek, 1998). But in one experiment, children as old as 11 years failed to show the same extent of masking release for separation in fundamental frequency between the target talker and masking speech (Leibold, Buss, & Calandruccio, 2018), suggesting that they might not actually group disparate spectral components of speech signals based on a common fundamental frequency. In a direct demonstration of that suggestion, Nittrouer and Tarr (2011) found that adults failed to integrate the low- and high-frequency components of isolated vowels if those components did not share a common fundamental frequency; for 5- and 8-year-old children, similarity in fundamental frequency had no effect on the extent of spectral integration of speech signals demonstrated in their perception. Where nonspeech signals are concerned, evidence is more limited, but one study found that children as old as 8 years failed to exhibit mature perceptual streaming, a phenomenon closely related to grouping; compared with older children and adults, these younger children needed greater frequency separations to report hearing two distinct auditory streams (Sussman, Wong, Horvath, Winkler, & Wang, 2007). Thus, the preponderance of evidence indicates that the perceptual grouping of speech signals based on a common fundamental frequency is something that develops across childhood. Consequently this potential explanation for the speech suffix effect in serial recall tasks should lead to different predictions for children than the phonologically based explanation. According to the perceptual grouping explanation, children might not show suffix effects similar in kind or magnitude to the effects demonstrated by adults.

Children and the suffix effect

Only a handful of studies have previously investigated the suffix effect in children. Although all found evidence of this effect (Engle, 1977; Engle, Fidler, & Reynolds, 1981; Frank & Rabinovitch, 1974; Gillam, Cowan, & Day, 1995), they all were limited in the breadth of conditions examined, making interpretation of this effect difficult. In particular, all of these studies used digits for the list items and presented only a spoken suffix—not a lipread one. Furthermore, none of the studies compared effects of the spoken suffix with any other kind of suffix such as a nonspeech tone or a nonverbal image. Thus, the full range of variables needed to determine whether children demonstrate a similar effect for auditory and lipread speech suffixes, and whether that effect is different from effects for nonspeech suffixes, was not implemented. Finally, response mode always involved speech production, which is problematic because it means that storage of late list items could be negatively affected by the explicit generation of phonological forms in the recall of early list items. The current experiment built on these earlier studies with children while broadening the range of conditions employed. A nonverbal response mode was implemented as a way to minimize the effect of recalling early list items on the recall of late list items.

The current study

The overarching goal of the current study was to investigate the nature of coding in verbal working memory: Is it explicitly phonological, or does perceptual grouping based on acoustic structure play a role? To address that goal, we designed an experiment to see whether greater reductions in recall accuracy would be observed when the suffix was speech, rather than a nonspeech tone or visual image, and to see whether adults and children would demonstrate similar effects for spoken suffixes across modalities of presentation, both auditory and visual. If phonological structure is the basis of coding in verbal working memory, then children should show similar effects as adults because cross-modal speech perception has been observed in infants. If, however, perceptual grouping based on acoustic (harmonic) structure is critical, then children might not show as strong effects in the auditory speech condition because that appears to be developmental in acquisition.

This experiment was similar to experiments conducted in the past but included some modifications. In particular, several suffixes were used in several different conditions. The task itself involved the presentation of six words in a closed set presented in different serial orders. The task facing participants was to recall the order of presentation, and the response method was to touch pictures presented on a computer monitor in the order recalled, so articulation was not involved. This prohibition on articulation meant both that overt rehearsal was not available to participants and that spoken recall of early list items could not override storage of late list items. All words were simple nouns that could be concretely represented with pictures because it has been shown previously that words that cannot be concretely represented require more cognitive effort to store and recall (Allen & Hulme, 2006; Nittrouer, Caldwell-Tarr, Low, & Lowenstein, 2017; Walker & Hulme, 1999).

Four different suffixes were used. The word *go* was presented in auditory-only format. Because this suffix was produced by the talker who produced the words comprising the lists, it shared the same fundamental frequency as those list items. A silent video of the same talker saying the word *go* was also used as a suffix. This condition was the critical one for evaluating whether visual speech signals provide similar phonological codes as auditory signals for short-term storage. However, two other suffixes were presented as control conditions. First, a nonspeech auditory suffix was presented: a 1000-Hz tone of similar length to the word *go.* Second, a green circle, indicating *go* as in a traffic light, was presented. This image was present on the monitor for the same amount of time as the image of the talker saying the word *go.* Finally, a condition with no suffix was presented to participants.

A modification in experimental protocol made in this study was that participants were required to attend to the suffix. This requirement was imposed by presenting alternative suffixes in each condition that signaled that participants should not respond. For the auditory and visual *go* suffixes, the "no-go" alternative was the word *stop.* For the auditory tone, the alternative was an interrupted tone. For the green circle, the alternative was a red circle. The role of attention in sustaining the suffix effect is equivocal, with some studies observing that it is not necessary (e.g., Greenberg & Engle, 1983; Nairne & Crowder, 1982; Penney, 1989), but with one study reporting that the effect of visual spoken suffixes is attenuated if participants are not attentive (Hitch, 1975). Given this uncertainty, it seemed to be important to ensure sustained attention to the suffix across participants, especially because visual suffixes were employed in some conditions. Furthermore, this procedure of requiring attention seemed to be a way to encourage grouping of the suffix with the list items. At the very least, this procedure ensured that for any conditions failing to elicit a suffix effect, that failure could not be attributed to a lack of attention on the part of participants. In particular, if adults showed suffix effects in some conditions but children did not, we worried that a reasonable concern could be that children were not attending as closely as adults. The use of no-go conditions helped to alleviate that concern.

In summary, three questions were asked by this work, with stronger predictions made for the first two questions than for the third one. First, we examined developmental changes in serial recall and expected 8-year-old children to demonstrate generally poorer recall than adults because that finding has been ubiquitous in previous studies. Second, we explored differences in the effects of speech and nonspeech suffixes. We predicted that speech suffixes would show stronger effects than nonspeech suffixes if the latter showed any effects at all. This speech-based effect was predicted to be similar for adults and 8-year-olds because the matching of visually and auditorily presented speech signals is a skill that appears to be present at birth. Finally, we examined whether a speech suffix that shared the same fundamental frequency, and so the same harmonic structure, would exert an especially strong effect due to perceptual grouping of list items and suffix, an effect that could not be accomplished with a visual signal. This account has been less thoroughly explored than the speech-based account, so we were less certain in our predictions. Even if it were observed for adults, it was not clear that it would similarly be observed for children.

Method

Participants

In total, 40 adults aged 18 to 31 years and 40 8-year-old children participated. The 8-year-olds ranged in age from 7 years 11 months to 8 years 11 months (M = 8 years 4 months). All were native speakers of American English, and none had any history of speech, language, or hearing disorder. Although these negative histories served as the primary indicators that all participants had typical speech, language, and hearing, screening measures were administered as additional checks. Adults were given the reading subtest of the Wide Range Achievement Test 4 (WRAT; Wilkinson & Robertson, 2006) and needed to demonstrate better than a 12th-grade reading level. Children were given the Goldman–Fristoe Test of Articulation 3 (Goldman & Fristoe, 2015) and were required to score at or better than the 30th percentile for their age in order to participate. The mean score for children was at the 70th percentile (*SD* = 10); the lowest-scoring child was at the 42nd percentile. Finally, all participants needed to pass hearing screenings consisting of the pure tones 0.5, 1, 2, 4, and 6 kHz presented at a 20-dB hearing level to each ear separately.

Equipment

All testing took place in a sound-treated booth. Hearing was screened with a Welch Allyn TM 262 audiometer using TDH-39 headphones. Stimuli were stored on a computer and presented through a Creative Labs Sound Blaster sound card using a 44.1-kHz sampling rate and 16-bit digitization. That signal was then passed through a Samson C-Que 8 headphone amplifier and AKG-K141 headphones. This system has a flat frequency response and low noise. Custom-written software presented all stimuli. Computer graphics of equal size were used to represent each word. Responses were collected via a 22-inch widescreen touchscreen monitor (HP E220t). For stimulus generation, speech samples were collected from a male talker with a fundamental frequency of 109 Hz using a Shure MX195 microphone and a Marantz PMD661 solid state recorder. Video samples were recorded using a Sony HDR-XR550V video recorder.

Participants' responses to the WRAT or Goldman–Fristoe test instruments were recorded using a Sony HDR-XR550V video recorder, and participants wore Sony FM transmitters to ensure good sound quality on the recordings. Receivers for these FM systems were connected directly to the audio input of the video recorder. These recordings were used for offline scoring of the WRAT and Goldman–Fristoe tests.

Stimuli

The words used in the serial recall task have been used previously (Nittrouer, Caldwell-Tarr, & Lowenstein, 2013; Nittrouer & Lowenstein, 2014). These words consisted of six non-rhyming nouns that were transparently represented with pictures: *dog, ham, rake, pack, soap,* and *teen.* Five tokens of each word were recorded, and the tokens that best matched selected tokens of the other words on duration, fundamental frequency, and intonation were used.

Additional stimuli were created for each of four suffix conditions: an auditory speech suffix, a visual speech suffix (lipread), an auditory nonspeech suffix (tone), and a visual nonspeech suffix (image). All suffixes were roughly 1 s in length. For the auditory speech suffix, the words *go* and *stop* were recorded using the male talker who provided the test words. For the lipread suffix, the same talker was video-recorded saying *go* and *stop.* In these productions, *go* was produced with liprounding throughout, but *stop* had no liprounding, so the two stimuli differed through their entirety. The video was cropped so that the man's face filled the frame vertically, and the audio was removed. During testing, the video was presented at 720 × 480 pixels in the center of the monitor. For the tone suffix, two 1000-Hz tones were used: one continuous tone that was 1000 ms long and two 400-ms tones separated by 200 ms of silence. Finally, for the image suffix, a green circle and a red circle were used. Each circle was 420 pixels in diameter and was presented on a white background in the center of the monitor.

Procedures

All testing took place in a single session lasting 45 min, and all procedures were approved by the local institutional review board.

General procedures

After obtaining consent (for adult participants) or parental consent and child assent (for child participants), the hearing screening was administered. Three of the five serial recall tasks were then administered, followed by the WRAT or Goldman–Fristoe test. Finally, the last two serial recall tasks were administered. The serial recall tasks were presented via headphones at a peak intensity of a 68-

dB sound pressure level. The monitor was positioned directly in front of the participants, 9 inches from the edge of the table.

There were five serial recall tasks administered: a no-suffix condition (none), a spoken audio suffix condition (auditory speech), a soundless video suffix condition (lipread), a visual image suffix condition (image), and a tone suffix condition (tone). Conditions were randomized such that if either of the two auditory suffixes (auditory speech or tone) was the first condition, then one of the visual suffixes (lipread or image) was the second condition. Similarly, if a visual suffix condition was presented first, then an auditory suffix condition followed. The order was further restricted such that the speech-based conditions (auditory speech and lipread) or nonspeech conditions (tone and image) were not presented back to back. The no-suffix condition (none) was always the third condition, followed by the remaining two suffix conditions. These restrictions resulted in a total of eight presentation orders that were randomized across participants.

Task-specific procedures

The first task during testing was designed to measure baseline response times. Participants saw a series of six blue squares and needed to tap on them from left to right as quickly as possible. The average time across five trials served as the baseline. Next, participants were trained to match the pictures on the monitor to the test words by seeing the pictures at the top of the monitor and hearing each word in isolation. Participants needed to tap the picture representing each word after it was presented to demonstrate recognition. The procedure was done before and after testing to ensure accurate recognition.

During each test condition, the six words were played at a rate of one per second. If a suffix was included, it was presented immediately following the last list item at the same rate. Then, all pictures appeared across the top of the monitor in random order. Participants touched the pictures in the order recalled. As each image was touched, it dropped to the bottom of the monitor into the next position going from left to right. The order of pictures could not subsequently be changed.

Two suffixes were generated for each of the four suffix conditions: one to indicate that participants should respond and one to indicate that participants should not respond. This procedure was undertaken to ensure that participants were paying attention. If the suffix indicated that participants should respond (auditory speech *go,* lipread *go,* the green circle image, or the continuous tone), then participants needed to touch the pictures in the order recalled. If the suffix indicated that participants should not respond (auditory speech *stop,* lipread *stop,* the red circle image, or the interrupted tone), then participants needed to refrain from responding. For the four suffix conditions, there were 10 lists presented that required responding and 3 lists that required refraining from responding. For the none condition, 10 lists were presented. Order of presentation of words, as well as order of respond–refrain lists, was randomized by the software.

Participants were instructed to keep their hands flat on the table in front of the monitor during audio presentation. They were told that there could be no articulatory movement of any kind (voiced or silent) during presentation of the words or between hearing the words and touching the pictures. The software recorded both the order of presentation and the order of participants' responses. The software also computed the time elapsed between the appearance of the pictures and participants' touching of the last picture; this latency was the *response time.*

After testing, the software automatically compared the order in which words were recalled with the order actually presented for each suffix condition. A word was considered wrong if it was recalled in the wrong list position. Percent correct scores were calculated by multiplying the proportion of correct responses (across the 10 list presentations) by 100. That value was used to index accuracy. Mean response time across conditions was used to index cognitive effort, as others have done previously (e.g., Cooper-Martin, 1994; DeLeeuw & Mayer, 2008). The software also reported whether participants responded to the no-go trials, but no one ever did.

Analyses

All statistical analyses were conducted using SPSS software (Version 24). Multi-way analyses of variance (ANOVAs) were performed as the first step for all dependent measures. In addition, analyses were done separately for adults' and children's data to examine the patterns of effects independently for each group so that these patterns could be compared across groups. In all analyses, a traditional alpha level of .05 was used, but Bonferroni corrections were applied when analyses had multiple comparisons.

Results

Screening of the data showed that all measures had normal distributions and homogeneity of variances. Although alpha was set to .05, precise p values are reported when $p < .10$. When $p > .10$, outcomes are reported as not significant (*ns*).

General observations

Fig. 2 shows mean percent correct recall for words presented in each list position for each of the five conditions. Outcomes for adults are shown on the left, and those for 8-year-old children are shown on the right. Mean standard errors across the five conditions were 2.4% for adults and 2.0% for children. As predicted, children performed more poorly than adults, and that difference is similar to previously observed outcomes (e.g., Nittrouer & Lowenstein, 2014).

Both groups show primacy and recency effects of similar magnitudes for the none condition; percent correct recall values for Position 1 and Position 6 were nearly identical for each group. Thus, the quintessential U-shaped function was obtained. When it comes to the traditional suffix effect—that is, the influence specifically on recall of the last list position—both groups show a strong effect for the auditory speech condition. The other conditions (image, tone, and lipread) seem to have had a small effect on recency for adults but a slightly larger effect for children. It appears that the presence of any suffix affected children to at least some extent. Children and adults also showed reduced primacy effects in the lipread and auditory speech conditions, with children showing a larger reduction than adults.

To examine these impressions, a three-way repeated-measures ANOVA was performed with suffix condition and list position as the within-participants factors and age group as the between-participants factor. Results are presented in Table 1 and show that all three main effects were significant. These main effects indicate that one or more suffixes affected recall of serial order, recall differed across list positions, and children were generally poorer than adults at recall.

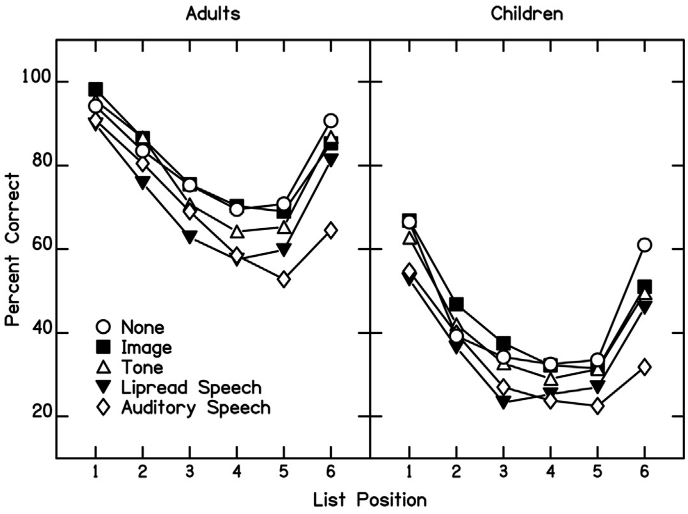

Fig. 2. Mean percent correct recall for the verbal working memory task by position and suffix condition for adults and children.

Table 1
Outcomes of a three-way repeated-measures ANOVA performed on percent correct recall scores for serial recall of words across suffix condition and list position.

	F	*df*	*p*	η^2
Main effects				
Condition	26.26	4, 312	<.001	.252
Position	150.60	5, 390	<.001	.659
Age group	202.21	1, 78	<.001	.722
Two-way interactions				
Condition × Position	6.22	20, 1560	<.001	.074
Condition × Age	*ns*	*ns*	*ns*	*ns*
Position × Age	3.13	5, 390	.009	.039
Three-way interaction				
Condition × Position × Age	*ns*	*ns*	*ns*	*ns*

There were also two significant two-way interactions: Condition × Position and Position × Age. The Condition × Age interaction was not significant, indicating that adults and children showed similar response patterns across suffix conditions. The Condition × Position × Age three-way interaction was similarly not significant. The two significant two-way interactions required further analyses so that they could be better understood. Analyses reported below serve to elucidate the bases of these interactions.

Immediate recency

The typical way to examine suffix effects is to measure immediate recency, which is defined as the difference in recall accuracy for the final list position and the penultimate position. In this experiment, that means that the difference in accuracy between Position 6 and Position 5 was calculated. In earlier work (e.g., de Gelder & Vroomen, 1992; Nairne & Crowder, 1982; Rowe & Rowe, 1976; Spoehr & Corin, 1978), immediate recency was found to be significantly reduced for speech suffixes compared with no-suffix conditions. Nonspeech suffixes have failed to show this reduction in immediate recency. In this experiment, a two-way repeated-measures ANOVA was performed to examine immediate recency with suffix condition as the within-participants factor and age group as the between-participants factor. Results showed a significant main effect of suffix condition, $F(4, 312) = 6.338$, $p < .001$, $\eta^2 = .075$, but neither the main effect of age group nor the Age × Condition interaction was significant.

Next, paired *t* tests were conducted comparing immediate recency between the none condition and the four suffix conditions for each age group separately. Outcomes are shown in Table 2. The only significant comparison for adults was for none versus auditory speech, although the comparison did not meet significance after Bonferroni corrections were applied. For children, three comparisons were significant: image, lipread, and auditory speech. However, only the auditory speech comparison remained significant after Bonferroni corrections were applied.

Table 2
Outcomes of paired *t* tests for immediate recency across suffix conditions for adults and children separately.

Comparison	Adults			Children		
	t	*p*	*d*	*t*	*p*	*d*
None vs. image	1.196	*ns*		2.221	.032	.41
None vs. tone	−0.426	*ns*		1.816	.077	
None vs. lipread	−0.378	*ns*		2.062	.046	.45
None vs. speech	2.404	.021	.53	4.735	<.001*	1.00

Note. The degrees of freedom value is 39 for all tests. Cohen's *d* values are given for significant within-group results.

* Significant with Bonferroni correction for four comparisons.

In spite of appearances in Fig. 2 of suffix effects on immediate recency, only one condition was found to be significant and only for children. But in Fig. 2 it appears that there were effects on earlier list elements of both the auditory and lipread speech suffixes for both adults and children. Those potential effects were examined in further analyses.

All list positions

Fig. 3 shows mean recall accuracy for each condition collapsed over all positions for adults and children separately. Table 3 shows outcomes of paired *t* tests for percent correct recall across list positions for adults and children separately.

For both adults and children, no significant differences were observed for recall accuracy across the nonspeech conditions (none, tone, and image), but recall in each of those conditions was significantly different from that in the auditory speech and lipread conditions. The auditory speech and lipread conditions were not statistically different from each other. Thus, recall accuracy was similar based on the speech or nonspeech status of the suffix. This matches one of the original predictions for this study.

Closer inspection of Fig. 2, however, suggests that the effects of the speech suffixes may be different for early and late list positions, depending on whether that speech suffix was lipread or heard. Specifically, recall accuracy for early list items appears to be similarly affected by the lipread and auditory speech suffixes, whereas recall accuracy for the late list items appears to be more deleteriously affected by the auditory speech suffix than by the lipread suffix. These apparent position effects were examined next.

Early list items

To examine the observed primacy effects more closely, recall accuracy for early list items was calculated as mean percent correct recall across Positions 1 to 3. These results are displayed in the top panel of Fig. 4. A repeated-measures ANOVA was conducted with suffix condition as the within-participants factor and age group as the between-participants factor. Results were significant for suffix condition, $F(4, 312) = 13.65$, $p < .001$, $\eta^2 = .149$, and for age group, $F(1, 78) = 201.59$, $p < .001$, $\eta^2 = .721$. The Age × Condition interaction was not significant.

Next, a series of paired *t* tests was performed on data from adults and children separately to explore the source of the condition effect. Outcomes are presented in Table 4. For neither age group was a difference in performance observed for these early list items across the none and nonspeech

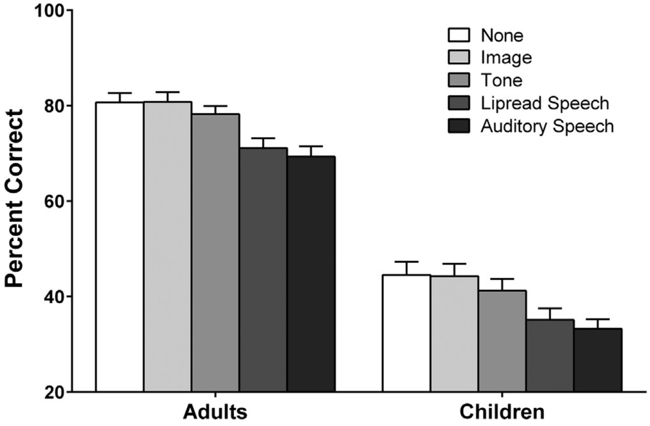

Fig. 3. Mean percent correct recall for the verbal working memory task collapsed across position for each suffix condition for adults and children. Error bars are standard errors of the mean.

Table 3
Outcomes of paired *t* tests for percent correct recall across list positions for all pairs of suffix conditions given for adults and children separately.

Comparison	Adults			Children		
	t	*p*	*d*	*t*	*p*	*d*
Comparisons among nonspeech suffixes						
None vs. image	−0.074	*ns*		0.084	*ns*	
None vs. tone	1.558	*ns*		1.313	*ns*	
Image vs. tone	1.359	*ns*		1.592	*ns*	
Comparisons between nonspeech and speech suffixes						
None vs. lipread	5.235	<.001*	.75	4.085	<.001*	.57
Image vs. lipread	4.463	<.001*	.74	4.218	<.001*	.59
Tone vs. lipread	5.104	<.001*	.59	3.504	.001*	.40
None vs. speech	6.398	<.001*	.87	4.467	<.001*	.73
Image vs. speech	5.297	<.001*	.86	5.211	<.001*	.76
Tone vs. speech	5.231	<.001*	.72	3.304	.002*	.56
Comparison between speech suffixes						
Lipread vs. speech	0.836	*ns*		0.756	*ns*	

Note. The degrees of freedom value is 39 for all tests. Cohen's *d* values are given for significant within-group results.
* Significant with Bonferroni correction for 10 comparisons.

suffix conditions. Comparisons of the nonspeech and speech suffix conditions, however, reveal some significant differences. In particular, adults and children alike showed differences in performance between the lipread suffix condition and both the none and nonspeech suffix conditions. When it comes to comparisons involving the auditory speech suffix, the groups differed in magnitude of the effects. For adults, all of these comparisons were statistically significant, although only the comparison with the image suffix remained so when Bonferroni corrections were applied. For children, all three comparisons resulted in $p < .10$, but only the comparison with the image suffix was significant; it remained so when Bonferroni corrections were applied. There was no difference in performance between the lipread and auditory speech conditions for adults or children. Thus, these results confirm that both speech suffixes affected accuracy of recall for early list items.

Late list items

Next, recall accuracy for the late list items was examined by calculating mean percent correct recall across Positions 4 to 6. These results are shown at the bottom of Fig. 4. A repeated-measures ANOVA was conducted with suffix condition as the within-participants factor and age group as the between-participants factor. Results were significant for suffix condition, $F(4, 312) = 31.68$, $p < .001$, $\eta^2 = .289$, and age group, $F(1, 78) = 139.15$, $p < .001$, $\eta^2 = .641$. The Age × Condition interaction was not significant. A series of paired *t* tests was also performed to examine where the differences resulting in the condition effect occurred. Outcomes are presented in Table 5.

For these comparisons, children and adults showed a significant difference in responding for one nonspeech suffix condition: none versus tone. However, these differences did not remain significant when Bonferroni corrections were applied. The comparisons between nonspeech and speech suffix conditions again showed the same differences across age groups. For adults, all of these comparisons were significant, with large effect sizes, even when Bonferroni corrections were applied. Thus, both lipread and auditory speech suffixes served to diminish the recency effect for adults.

For children, all comparisons between the auditory speech suffix condition and nonspeech suffix conditions were significant and remained so when Bonferroni corrections were applied. For the lipread suffix, differences existed when scores were compared with those of the nonspeech suffix conditions, but they were not as strong; only the comparison between the none and lipread conditions reached significance when Bonferroni corrections were applied. Finally, both groups showed significant differences in effects of lipread and auditory speech suffixes. This difference met the criterion

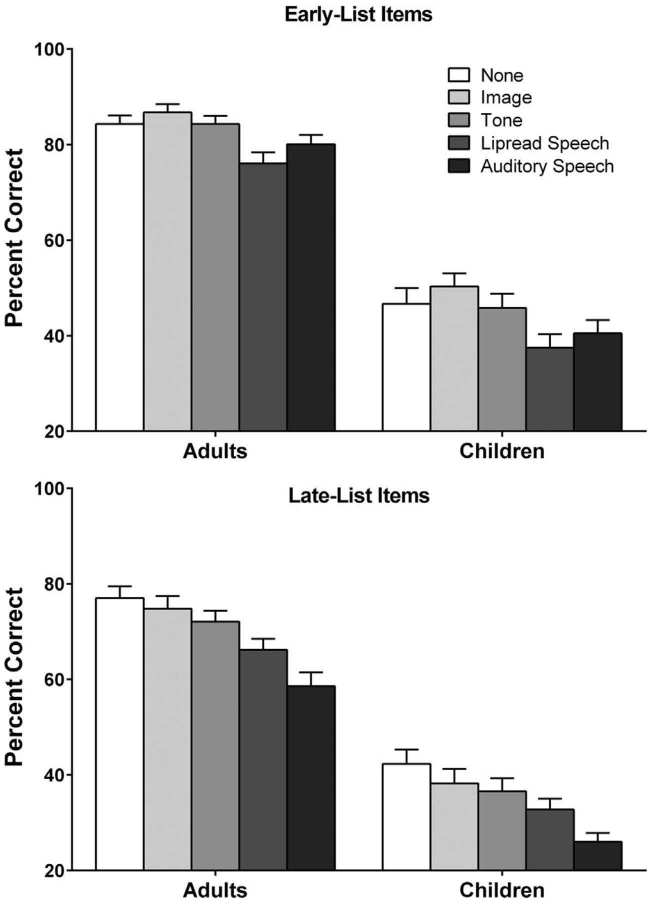

Fig. 4. Mean percent correct recall. The top panel presents results for early list items. The bottom panel presents results for late list items. Error bars are standard errors of the mean.

for significance with a Bonferroni correction for adults, but not quite for children. Thus, unlike the early list items, the two speech suffixes had different effects on the late list items, with auditory speech suffixes exerting stronger effects. This finding supported our last prediction, and the magnitudes of difference in these effects were similar for adults and children.

Response time

The mean baseline for adults was 1.5 s ($SD = 0.2$), and the mean baseline for children was 1.7 s ($SD = 0.3$). A t test performed on these baselines was significant, $t(78) = 4.426$, $p < .001$, Cohen's $d = .99$, indicating that children were slower at responding than adults. Consequently, response time was evaluated separately for each age group.

Mean response time for each condition is shown in Fig. 5. Separate repeated-measures ANOVAs with suffix condition as the within-participants factor were performed on response time for each group. Results were significant for both groups: adults, $F(4, 156) = 4.821$, $p = .001$, $\eta^2 = .110$; children, $F(4, 156) = 2.475$, $p = .047$, $\eta^2 = .060$. Consequently, a series of paired t tests was performed, and outcomes are presented in Table 6.

Table 4
Outcomes of paired *t* tests for percent correct recall across early list items given for adults and children separately.

Comparison	Adults			Children		
	t	*p*	*d*	*t*	*p*	*d*
Comparisons among nonspeech suffixes						
None vs. image	−1.365	*ns*		−1.121	*ns*	
None vs. tone	0.000	*ns*		0.256	*ns*	
Image vs. tone	1.249	*ns*		1.692	*ns*	
Comparisons between nonspeech and speech suffixes						
None vs. lipread	3.852	<.001*	.64	2.992	.005*	.47
Image vs. lipread	4.104	<.001*	.83	4.708	<.001*	.73
Tone vs. lipread	4.598	<.001*	.65	3.527	.001*	.45
None vs. speech	2.270	.029	.36	1.952	.058	
Image vs. speech	2.969	.005*	.57	3.788	.001*	.56
Tone vs. speech	2.385	.022	.37	1.848	.072	
Comparison between speech suffixes						
Lipread vs. speech	−1.776	.084		−0.927	*ns*	

Note. The degrees of freedom value is 39 for all tests. Cohen's *d* values are given for significant within-group results.
* Significant with Bonferroni correction for 10 comparisons.

Table 5
Outcomes of paired *t* tests for percent correct recall across late list items given for adults and children separately.

Comparison	Adults			Children		
	T	*p*	*d*	*t*	*p*	*d*
Comparisons among nonspeech suffixes						
None vs. image	1.008	*ns*		1.500	*ns*	
None vs. tone	2.616	.013	.33	2.376	.022	.32
Image vs. tone	1.110	*ns*		0.746	*ns*	
Comparisons between nonspeech and speech suffixes						
None vs. lipread	4.714	<.001*	.71	3.653	.001*	.57
Image vs. lipread	3.916	<.001*	.55	2.285	.028	.32
Tone vs. lipread	3.393	.002*	.41	1.859	.071	
None vs. speech	8.077	<.001*	1.07	5.929	<.001*	1.04
Image vs. speech	5.853	<.001*	.92	4.336	<.001*	.78
Tone vs. speech	5.995	<.001*	.81	3.841	<.001*	.72
Comparison between speech suffixes						
Lipread vs. speech	2.959	.005*	.46	2.735	.009	.51

Note. The degrees of freedom value is 39 for all tests. Cohen's *d* values are given for significant within-group results.
* Significant with Bonferroni correction for 10 comparisons.

For adults, the comparisons of lipread with none, image, and auditory speech were significant, as were the comparisons of none with auditory speech and tone. However, only the comparison of the none and lipread suffixes remained significant after Bonferroni corrections were applied. For children, the comparisons of image with both lipread and auditory speech were significant, but the comparisons did not remain so after Bonferroni corrections were applied. Thus, although some differences across conditions were observed, these trends were neither consistent across age groups nor especially large.

Relationship between time and accuracy

Finally, an analysis was performed to see whether individual differences in cognitive effort, as indexed by response time, could account for individual differences in recall accuracy. To this end, average response times and average accuracy across all five conditions were calculated for each participant and compared. A scatterplot of the results is shown in Fig. 6.

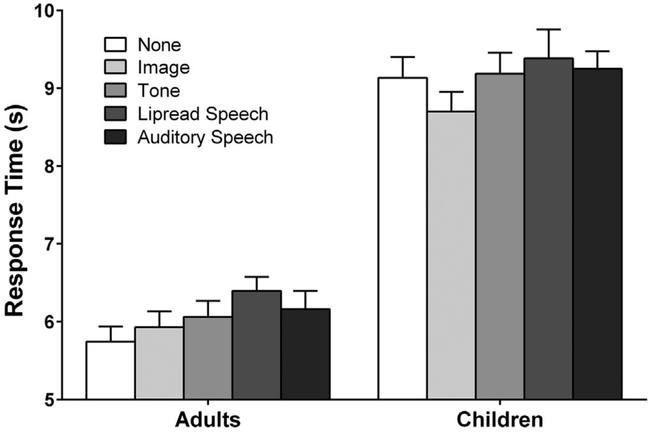

Fig. 5. Mean response time (in seconds) presented for each suffix condition for adults and children. Error bars are standard errors of the mean.

Table 6
Outcomes of paired *t* tests for response time given for adults and children separately.

Comparison	Adults			Children		
	t	*p*	*d*	*t*	*p*	*d*
Comparisons among nonspeech suffixes						
None vs. image	−1.348	*ns*		1.491	*ns*	
None vs. tone	−2.113	.041	−.25	−0.184	*ns*	
Image vs. tone	−0.895	*ns*		−1.497	*ns*	
Comparisons between nonspeech and speech suffixes						
None vs. lipread	−4.007	<.001*	−.49	−1.563	*ns*	
Image vs. lipread	−2.618	.013	−.34	−2.412	.021	−.48
Tone vs. lipread	−2.300	.027	−.25	−1.710	.095	
None vs. speech	−2.501	.017	−.31	−0.549	*ns*	
Image vs. speech	−1.378	*ns*		−2.347	.024	−.35
Tone vs. speech	−0.621	*ns*		−0.215	*ns*	
Comparison between speech suffixes						
Lipread vs. speech	1.529	*ns*		1.146	*ns*	

Note. The degrees of freedom value is 39 for all tests. Cohen's *d* values are given for significant within-group results.
* Significant with Bonferroni correction for 10 comparisons.

Pearson product–moment correlation coefficients were computed for adults and children separately. Adults demonstrated a correlation coefficient of −.349, which was significant ($p = .027$). For children, the correlation coefficient was −.221, which was not significant. Thus, a weak relationship was observed for adults only. Overall cognitive effort was relatively constant across conditions and was not strongly related to accuracy of serial recall.

Discussion

This study was undertaken to examine the coding of words in verbal working memory by using a suffix paradigm. Numerous earlier experiments demonstrated that some suffixes can diminish the enhanced performance for final list items typically observed in serial recall for adults, but agreement regarding the explanation for that suffix effect has remained elusive. An especially perplexing aspect of the effect has been that it is stronger for speech suffixes than for nonspeech suffixes. Furthermore,

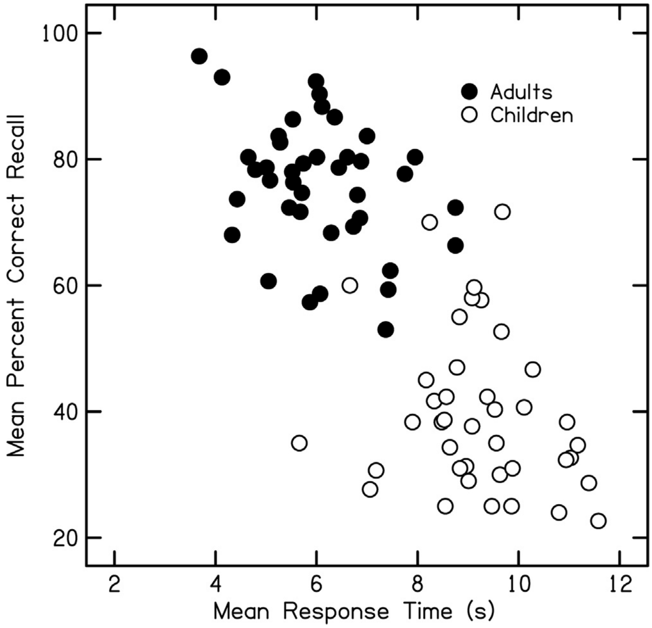

Fig. 6. Mean response time (in seconds) versus mean percent correct recall.

the effect has been observed for speech suffixes regardless of whether they are presented auditorily, through heard speech, or visually, through lipreading.

Although still not unanimously agreed on, the explanation offered most often for the speech suffix effect relates to the most common model regarding how verbal working memory functions. According to this model, there exists a dedicated front end that retrieves phonological structure from the sensory input and deposits words, using that phonological structure, into a temporary memory store for processing by a central executive. It has been hypothesized that both heard and lipread speech can provide input that is articulatory in form to that front end. It is the articulatory nature of the input that is shared by these two suffixes and the quality that both of these suffixes shares with list items in verbal working memory tasks. And it is that articulatory structure that supports the operations of the phonological loop. But as each item is being entered into the short-term memory store, it effectively interrupts, for the preceding item, the process of recovering phonological structure and using it to deposit that item into short-term store. It is that reiterative interference across list items that is responsible for the lower middle portion of the quintessential U-shaped function observed in serial recall tasks. And just as that effect unfolds for list items, so does the suffix function as an additional item, interrupting processing of the final list item if the suffix shares the property of providing the kind of input from which phonological structure can be recovered.

That explanation for the suffix effect predicts that young children would demonstrate similar suffix effects as adults because perception of articulatory events by ear or eye should be direct and compulsory for all observers from birth. However, there is another explanation that has received some support. This alternative account suggests that the suffix effect is realized only when the suffix is grouped auditorily with the list items rather than separated into a distinct perceptual stream. Although originally described as primitive in nature—and so not dependent on experience—empirical evidence has shown that perceptual grouping based on a common fundamental frequency is developmental and so must be acquired. Thus, young children would be predicted to demonstrate weaker suffix effects for heard speech than adults demonstrate if this alternative account is correct.

Outcomes of the study reported here support the articulatory account of the suffix effect. Although children were less accurate than adults overall, relative suffix effects across conditions were consistent for both age groups. These patterns showed effects for auditory speech and lipread suffixes but showed no effects for nonspeech suffixes.

However, the effects evoked by the two speech suffixes did not perfectly match previously described outcomes. Earlier studies of the suffix effect described only a diminishment in recall accuracy for the last list item even when authors examined earlier list positions for potential effects (Spoehr & Corin, 1978). In the current study, both heard speech and lipread speech evoked an effect on the last list item but also evoked effects on earlier items. In particular, Fig. 2 reveals that recall across the three early list items was poorest in the presence of the two speech suffixes for both adults and children. It is tempting to attribute this effect on early list items to the fact that a decision needed to be made in this experiment regarding whether or not to respond. It might be that the cognitive load of needing to reach a decision would disrupt recall at the start of that recall activity. But the same decision needed to be made across all suffixes, yet only the speech suffixes demonstrated a diminishment in accuracy of recall for early list items.

The finding that the effect of spoken suffixes on recall accuracy was similar for early list items, regardless of whether the suffixes were heard or seen, may shed light on the source of the effect. Traditional accounts (e.g., Murdock, 1962) suggest that the primacy effect relies on participants completing the phonological coding of these early list items into the temporary memory store. By the time recall is initiated, these items have already acquired a permanent, or at least semi-permanent, form of storage. Only the auditory speech and lipread suffixes consisted of phonologically structured information, so only these suffixes could exact a toll on the recall of items already coded phonologically. This explanation further supports the suggestion that heard and lipread speech function similarly as suffixes, because they both are articulatory in nature and so support the operations of the phonological loop.

The finding, however, that auditory speech exerted slightly greater effects on recall of late list items than lipread speech suggests that these two kinds of suffixes exert different influences on recall for late list items. It may be that the late list items have not been fully recoded into phonological structure by the time recall is initiated. Thus, listeners rely to some extent on sensory traces for recall of these late list items, and those sensory traces are modality specific. This suggestion also matches traditional hypotheses about the source of the U-shaped function in serial recall (Murdock, 1962): Because the input modality was the same for the list items and auditory speech in this experiment, this suffix exerted the largest effect on late list items. Of course, it cannot simply be that the auditory speech suffix and the list items were auditorily presented because if that were all there was to the effect, the tone suffix would have exerted a similar influence on responding as did the heard speech. However, the tone did not consist of harmonic structure determined by a fundamental frequency. A future study will need to examine whether a nonspeech complex tone that shares a common fundamental frequency with the spoken list items exerts a suffix effect. The finding that the difference in magnitude of effect for the auditory speech and lipread suffixes was roughly the same for adults and children suggests that children were grouping list items and the auditory speech suffix together based on the shared fundamental frequency, just as adults were. Thus, if this effect is developmental, it appears to have reached mature status by 8 years of age.

Limitations of the current study

One potential limitation of the current study could have been the implementation of the go/no-go task. This procedure was used to ensure that all participants were attending to the suffix. Auditory signals can be presented to the sensory system whether participants are attending or not. However, a participant who is not looking at the computer monitor cannot be presented with a visual signal. This go/no-go procedure provided confirmation that attention was consistent across suffix conditions. Nonetheless, concern might exist that the need to make a decision interfered with accessing information coded in the memory buffer. Fortunately, evidence to assuage that concern is provided by response times, which were similar regardless of suffix condition and regardless of whether there was a suffix presented or not.

Another potential limitation of this study was that the conclusion was reached that the auditory speech suffix exerted a stronger effect than the lipread suffix because of perceptual grouping based on a common fundamental frequency. However, a broader variety of experimental conditions is needed to answer this question with certainty. For example, speech suffixes with a range of fundamental frequencies could be presented. In addition, nonspeech signals with the same harmonic structure could be used.

The limited age range of the children in the study may also be seen as a limitation. These children demonstrated suffix effects similar to those of adults even though one potential source of those effects was presumed to be present from birth (matching of auditory and visual speech signals) and another was presumed to be developmental (perceptual grouping based on a common fundamental frequency). Children covering a wider range of ages are needed to examine what the developmental course of each effect might be.

Summary

The goal of this study was to explore the operations of verbal working memory, especially in regard to the nature of the code used for storage. To accomplish that goal, the effects of a suffix on serial recall for words was examined. Two alternative accounts have been offered to explain the commonly observed suffix effect: one suggesting that it is the shared phonological nature of the coding for list items and suffix that is responsible for the effect and the other suggesting that perceptual grouping across list items and suffix are required for the effect to be realized. The first account offers an effect that should be intrinsic and present from birth; the second account describes a strategy that would likely be developmental. To evaluate these accounts, adults and 8-year-old children were asked to recall order of presentation for a closed set of nouns in five suffix conditions. Results showed that children's serial recall was generally poorer than that of adults, but patterns were largely similar across age groups and somewhat different than anticipated. Participants showed poorer recall in the presence of speech suffixes than in the presence of nonspeech suffixes regardless of whether suffixes were seen or heard; that aspect of the outcomes was anticipated. However, these effects were not restricted to final list items. Instead heard and lipread suffixes also exerted effects on early list items, which were similar in magnitude. But for late list items, the auditory speech suffix exerted a stronger effect on recall than the lipread suffix. Outcomes suggest that the effects of auditory speech and lipread suffixes arise primarily from their shared phonological coding, but perceptual grouping can facilitate the effect.

Acknowledgments

This work was supported by research funds from the College of Public Health and Health Professions at the University of Florida. We thank Donal G. Sinex for help with software design and Lindsay Arena for help with data collection.

References

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64–88.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362–365.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556–559.

Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford, UK: Oxford University Press.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). New York: Academic Press.

Bloom, L. C. (2006). Two-component theory of the suffix effect: Contrary evidence. *Memory & Cognition, 34*, 648–667.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*, 201–233.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology, 32*, 85–99.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429–432.

Cooper-Martin, E. (1994). Measures of cognitive effort. *Marketing Letters, 5*, 43–56.

Crowder, R. G. (1983). The purity of auditory memory. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 302*, 251–265.

Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development, 42*, 37–48.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America, 114*, 2913–2922.

de Gelder, B., & Vroomen, J. (1992). Abstract versus modality-specific memory representations in processing auditory and visual speech. *Memory & Cognition, 20*, 533–538.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*, 223–234.

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development, 5*, 261–276.

Engle, R. W. (1977). A developmental study of the prelinguistic auditory store (PAS). *Intelligence, 1*, 358–368.

Engle, R. W., Fidler, D. S., & Reynolds, L. H. (1981). Does echoic memory develop? *Journal of Experimental Child Psychology, 32*, 459–473.

Fassbender, C. (1993). *Auditory grouping and segregation processes in infancy*. Norderstedt, Germany: Kaste Verlag.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America, 99*, 1730–1741.

Frank, H. S., & Rabinovitch, M. S. (1974). Auditory short-term memory: Developmental changes in precategorical acoustic storage. *Child Development, 45*, 522–526.

Gillam, R. B., Cowan, N., & Day, L. S. (1995). Sequential memory in children with and without language impairment. *Journal of Speech and Hearing Research, 38*, 393–402.

Goldman, R., & Fristoe, M. (2015). *Goldman-Fristoe test of articulation* (3rd ed.). Bloomington, MN: NCS Pearson.

Greenberg, S. N., & Engle, R. W. (1983). Voice change in the stimulus suffix effect: Are the effects structural or strategic? *Memory & Cognition, 11*, 551–556.

Hitch, G. J. (1975). The role of attention in visual and auditory suffix effects. *Memory & Cognition, 3*, 501–505.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1141.

Lee, J. H., & Humes, L. E. (2012). Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background. *Journal of the Acoustical Society of America, 132*, 1700–1717.

Leibold, L. J., Buss, E., & Calandruccio, L. (2018). Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch. *Ear and Hearing, 39*, 935–945.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science, 219*, 1347–1349.

Macnamara, B. N., Moore, A. B., & Conway, A. R. (2011). Phonological similarity effects in simple and complex span tasks. *Memory & Cognition, 39*, 1174–1186.

Maidment, D. W., Macken, B., & Jones, D. M. (2013). Modalities of memory: Is reading lips like hearing voices? *Cognition, 129*, 471–493.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Morra, S. (2015). How do subvocal rehearsal and general attentional resources contribute to verbal short-term memory span? *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00145.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*, 482–488.

Nairne, J. S., & Crowder, R. G. (1982). On the locus of the stimulus suffix effect. *Memory & Cognition, 10*, 350–357.

Nairne, J. S., & Walters, V. L. (1983). Silent mouthing produces modality- and suffix-like effects. *Journal of Verbal Learning and Verbal Behavior, 22*, 475–483.

Nicholls, A. P., & Jones, D. M. (2002). Capturing the suffix: Cognitive streaming in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 12–28.

Nittrouer, S., Caldwell-Tarr, A., Low, K. E., & Lowenstein, J. H. (2017). Verbal working memory in children with cochlear implants. *Journal of Speech Language and Hearing Research, 60*, 3342–3364.

Nittrouer, S., Caldwell-Tarr, A., & Lowenstein, J. H. (2013). Working memory in children with cochlear implants: Problems are in storage, not processing. *International Journal of Pediatric Otorhinolaryngology, 77*, 1886–1898.

Nittrouer, S., & Lowenstein, J. H. (2014). Separating the effects of acoustic and phonetic factors in linguistic processing with impoverished signals by adults and children. *Applied Psycholinguistics, 35*, 333–370.

Nittrouer, S., & Tarr, E. (2011). Coherence masking protection for speech in children and adults. *Attention, Perception, & Psychophysics, 73*, 2606–2623.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17*, 398–422.

Rowe, E. J., & Rowe, W. G. (1976). Stimulus suffix effects with speech and nonspeech sounds. *Memory & Cognition, 4*, 128–131.

Salame, P., & Baddeley, A. D. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society, 24*, 263–265.

Salter, D., & Colley, J. G. (1977). The stimulus suffix: A paradoxical effect. *Memory & Cognition, 5*, 257–262.

Spoehr, K. T., & Corin, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. *Memory & Cognition, 6*, 583–589.

Summers, V., & Leek, M. R. (1998). FO processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech Language and Hearing Research, 41*, 1294–1306.

Sussman, E., Wong, R., Horvath, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research, 225*, 117–127.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1256–1271.

Wilkinson, G. S., & Robertson, G. J. (2006). *The Wide Range Achievement Test (WRAT)* (4th ed.). Lutz, FL: Psychological Assessment Resources.