SPEECH, LANGUAGE AND HEARING, 2018
https://doi.org/10.1080/2050571X.2018.1458067

Taylor & Francis
Taylor & Francis Group

Check for updates

Delayed silent phoneme monitoring in adults who do and do not stutter

Geoffrey A. Coalson[a] and Courtney T. Byrd[b]

[a]Department of Communication Sciences and Disorders, Louisiana State University, Baton Rouge, LA, USA; [b]Department of Communication Sciences and Disorders, University of Texas at Austin, Austin, TX, USA

ABSTRACT

Previous research employed silent phoneme monitoring tasks to examine differences in phonological encoding in adults who stutter (AWS) compared to adults who do not stutter (AWNS). The primary purpose of this study was to apply a modified version of the task – the delayed silent phoneme monitoring task – to examine the integrity of the phonological speech plan within working memory in AWS and AWNS before and after subvocal rehearsal. The secondary purpose of this study was to examine whether group differences were more apparent when greater phonological demand was placed upon phonological working memory. In Experiment 1, 20 adults (10 AWNS, 10 AWS) identified target phonemes within trochaic nonwords held in memory before the initiation of subvocal rehearsal (1 s) and after subvocal rehearsal (4 s). In Experiment 2, an additional 20 adults (10 AWNS, 10 AWS) monitored identical nonwords with low-frequency iambic stress. Speed and accuracy of manual response was measured, as well as post-trial verbal productions. Both groups identified the initial phoneme of trochaic stimuli fastest, irrespective of stress, and both groups monitored phonemes faster after the 4 s delay. However, AWS identified phonemes within iambic stimuli with less accuracy than AWNS. Group differences in monitoring errors were most evident for phonemes immediately following syllable boundary, and after subvocal rehearsal. Preliminary findings suggest AWS may exhibit distinct difficulties relative to AWNS when accessing segmental information after subvocal rehearsal is required, but only when target words are more phonologically demanding (i.e., low-frequency iambic stress).

ARTICLE HISTORY
Received 17 January 2018
Accepted 22 March 2018

KEYWORDS
Stuttering; phonological working memory; phonological encoding; adults

Introduction

Research has implicated at least two linguistic areas of weakness in adults who stutter (AWS) when compared to adults who do not stutter (AWNS). First, AWS may demonstrate subtle delays or deficiencies during *phonological encoding* (see Byrd, Wolk, & Davis, 2007; Sasisekaran, 2014; cf. Nippold, 2002), the rapid integration of segmental information (i.e., sounds) with metrical information (i.e., syllabic boundaries and syllabic stress) prior to motoric programing (Levelt, Roelofs, & Meyer, 1999). Second, AWS exhibit greater difficulties maintaining information in *phonological working memory* (see Bajaj, 2007; Byrd, Coalson, McGill, & Gkalitsiou, 2016), the system responsible for maintaining the output of phonological encoding prior to production, and/or refreshing the information via subvocal rehearsal as activation approaches temporal decay (Baddeley, 2003). Models of speech production propose that the output of phonological encoding serves as the input for phonological working memory, and suggest that efficient functioning of both systems may reflect the speakers' ability to prepare and maintain the intended utterance prior to fluent speech production (e.g., Baddeley, 2003; Baddeley & Hitch, 1974; Jacquemot & Scott, 2006; Levelt, 1989; Levelt et al., 1999).

The silent phoneme monitoring task has been used by researchers to investigate differences in the time course of phonological encoding in AWS relative to AWNS (e.g., Coalson & Byrd, 2015; Sasisekaran & de Nil, 2006; Sasisekaran, de Nil, Smyth, & Johnson, 2006). Silent phoneme monitoring tasks require nonverbal identification of target sounds within words in the absence of overt speech production. In brief, participants are first trained to silently generate a target word upon presentation of a visual cue, then instructed to respond as quickly as possible, via button press, whether a specified phoneme (presented first) is present or absent in a target word (presented second). The speed and accuracy of response is thought to reflect the time course of phonological encoding in adults. Across these studies, AWS exhibit slower silent monitoring latencies than AWNS, suggesting that phonological encoding processes in AWS are compromised relative to fluent peers.

Researchers also acknowledged that differences in silent phoneme monitoring abilities observed in AWS (or AWNS) may also be related to difficulties storing or rehearsing information in phonological working memory, rather than (or in addition to) phonological encoding difficulties, based on shared architecture of

CONTACT Geoffrey A. Coalson gcoals1@lsu.edu Department of Communication Sciences and Disorders, Louisiana State University, 73 Hatcher Hall, Baton Rouge, LA 70803, USA

both systems and/or cueing methodologies that may have placed greater demand on working memory. That is, deficits in phonological encoding are difficult to isolate from deficits in phonological working memory in AWS, and vice versa. Therefore, the primary purpose of the present study was to examine the integrity of the phonological code during storage and rehearsal in AWNS and AWS by employing a modified paradigm – *delayed silent phoneme monitoring task* – which forces participants to maintain the word in memory before silent phoneme identification. This modification allows assessment of phonological code immediately after initial encoding has been completed (phonological storage), and also after rehearsal is required to maintain the verbatim trace (subvocal rehearsal). It also allows assessment of whether the observed differences in AWS during initial encoding observed in previous studies with no delay in interstimulus interval (i.e., 0 s ISI) are also present during phonological storage (1 s ISI) and/or subvocal rehearsal (4 s ISI), and if so, isolate the stage within the phonological loop more vulnerable to compromise in AWS.

The second purpose of this study was to examine monitor speed and accuracy at each level of processing as phonological demand of the target is increased. Previous studies indicate that even modest shifts in phonological complexity – such as less frequent metrical stress – may be sufficient to disrupt phonological encoding (Coalson & Byrd, 2015) or phonological working memory (Coalson & Byrd, 2017) in AWS. To do so, nonword targets in the present study were presented with low-frequency iambic stress, rather than high-frequency trochaic stress, but identical in phonemic composition. If differences are detected at either stage of phonological working memory (i.e., phonological storage [1 s ISI], subvocal rehearsal [4 s ISI]) for stimuli with more demanding metrical structure, findings will provide novel data of which of these two sub-processes in phonological working memory may be more susceptible to modest increases in complexity in AWS and AWNS.

Phonological encoding in AWNS and AWS

According to Levelt et al. (1999), phonological encoding is defined as the rapid assignment of two independent properties – segmental properties and metrical properties – into an abstract phonological representation that serves as input to the phonetic-articulatory processing system. These abstract speech plans are syllabified with stress demarcation and syllable boundaries assigned based on the properties of the native language. The syllabified representation is then entered into a temporary speech buffer until the entire intended speech utterance has been prepared for overt production or circulation within phonological working memory. Levelt et al. (1999) rely on data from one experimental paradigm – the silent phoneme monitoring task – to estimate the time course of phonological encoding in typically fluent adults.

Wheeldon and Levelt (1995) were the first to employ the silent phoneme monitoring task in AWNS. Twenty Dutch-English bilingual adults were instructed to identify, via manual button press, whether a specified sound was present in the Dutch translation (L1) of a word presented in English (L2). Dutch target words were 20 bisyllabic $C_1VC_2C_3VC_4$(C) stimuli (e.g., '*magnet*;' C = consonant, V = vowel) comprised of trochaic stress ($n = 15$) and iambic stress ($n = 5$). Participants identified each consonant slower than the preceding consonant (i.e., $C_1 < C_2 < C_3 = C_4$), suggesting that segmental information becomes available during phonological encoding in an incremental, left-to-right manner. Wheeldon and Morgan (2002) used the silent phoneme monitoring task with 40 typically fluent monolingual adults. At the start of their experimental session, participants first completed a paired-associate task wherein a semantically-related prompt word (e.g., *frog*) was learned for each $C_1VC_2C_3VC_4$ target word (e.g., *tadpole*). Participants were then instructed to silently monitor for specified phonemes in 18 target words (trochaic: $n = 12$, iambic: $n = 6$) upon auditory presentation of the corresponding prompt word. Overall, data from traditional silent phoneme monitoring tasks in AWNS indicate that segmental information within words are assembled in a relatively distinct temporal pattern:

(1) increasingly longer latencies for each segment from onset to coda ($C_1 < C_2 < C_3 < C_4$),
(2) significant latency differences between C_1 and C_2, presumably due to assignment of the first metrical property (syllabic stress) on the intervening vowel for stimuli with trochaic structure,
(3) significant latency differences between C_2 and C_3, presumably due to assignment of the first metrical property (syllable boundary) for iambic stimuli without initial stress (cf. Jansma & Schiller, 2004; Schiller, 2005), and
(4) a 'leveling-off' of response latencies (i.e., increasingly shorter latency differences) for phonemes following the first metrical property, be it stress or syllable boundary.

Sasisekaran et al. (2006) was the first to compare the silent phoneme monitoring latencies of AWS ($n = 10$) and AWNS ($n = 11$). In their study, participants heard the specified phoneme target (e.g., 'Please respond to the /sə/ sound in the following picture') and then saw one of 14 picture stimuli in random order. All picture names were bisyllabic $C_1VC_2C_3VC_4$, but unlike previous studies in AWNS (Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002), stimuli were restricted to words with trochaic stress. For both groups, segments became available at increasingly slower latencies, C_1 was identified significantly faster than C_2 (as would be expected for trochaic stimuli), and trend analysis confirmed a 'leveling-off' in serial processing speed between phonemes, particularly after the initial stressed syllable. AWS exhibited significantly slower latencies for all consonant positions and supported the notion that individuals who stutter may exhibit an overall slowness in phonological encoding compared to AWNS, although the within-word patterns predicted by Wheeldon and colleagues for words with trochaic stress were present in both groups.

Coalson and Byrd (2015) conducted a silent phoneme monitoring task that directly manipulated syllabic stress of nonword stimuli while controlling for segmental properties. Twenty-two adults (AWNS, $n = 11$; AWS, $n = 11$; Experiment 1) monitored phonemes within trochaic nonwords, and an additional 22 adults (AWNS, $n = 11$; AWS, $n = 11$; Experiment 2) monitored the phonemes within identical nonwords with iambic stress. Both talker groups exhibited similar patterns of encoding for stimuli with trochaic stress, with progressive increases in latencies across positions ($C_1 < C_2 < C_3 < C_4$) and significant C_1–C_2 latency differences. However, the expected within-word latency pattern in AWNS differed from the previous studies that used real word stimuli. Contrary to the patterns observed for real words (Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002), significant C_1–C_2 latency differences were *also* observed in AWNS for iambic nonword stimuli. Further, the predicted C_2–C_3 latency difference thought to accommodate syllable boundary assignment within iambic nonwords (in the absence of initial stress) was detected for AWS, but not AWNS. These data indicate that AWNS exhibited nearly identical within-word latency patterns for both trochaic and iambic nonwords. By comparison, AWS demonstrated unique monitoring patterns compared to AWNS in the presence of lower-frequency iambic stress patterns versus trochaic patterns. Despite the differences between AWNS and AWS, three of the four within-word patterns observed by Wheeldon and colleagues (Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002) and further described by Levelt et al. (1999) were generally maintained within each group: (1) incremental increases in phoneme monitoring latencies from the beginning to end of the word, (2) C_1 phonemes monitored significantly faster than C_2 phonemes (and perhaps irrespective of stress pattern for nonword stimuli), and (3) a 'leveling-off' effect after the first-encountered metrical property (or first syllable for nonwords irrespective of stress).

Across these studies, the tasks have varied from low cognitive demand on phonological working memory (i.e., picture naming task, Sasisekaran et al. [2006]) to higher cognitive-memory load (i.e., L2–L1 translation tasks [Wheeldon & Levelt, 1995]; paired association task with semantic cue [Wheeldon & Morgan, 2002]; nonword recall task with non-semantic visual cues [Coalson & Byrd, 2015]). However, apart from overall differences in response latencies between tasks, the within-word latency patterns reported across studies has been relatively invariant regardless of memory load. This invariance across tasks suggests that the availability of segmental information upon initial encoding may instead reflect the maintained integrity of these segments within phonological working memory during more memory-intensive tasks.

Phonological working memory in AWNS and AWS

Baddeley's (2003) model of working memory incorporates three distinct processes that operate within the central executive system: (1) phonological working memory, which temporarily sustains activation of verbal information, (2) the visuospatial sketchpad, which temporarily stores visual information, and (3) the episodic buffer, which binds visual and verbal information held in working memory into episodic association, which is then compared with information held in the long-term memory. According to Baddeley (2003), novel phonological information is activated for approximately two seconds (i.e., 'phonological storage') before decay occurs. Subvocal rehearsal (i.e., non-vocal repetition of the verbal target) is then required to offset decay and maintain the stored phonological information. Subvocal rehearsal recruits the corresponding motor templates that re-activate the phonological code and sustains phonological targets in working memory.

Phonological working memory is commonly assessed via nonword repetition tasks (i.e., participants repeat novel words of increasing length), serial recall tasks (i.e., participants recall a series of words, nonwords, or digits of increasing length in the correct order), or variations of these tasks designed to assess the temporal limits of working memory processes. In general, accuracy for AWNS during nonword repetition and recall tasks is reduced as the length of stimuli increases. That is, the rate of articulation required to repeat the target silently exceeds the theoretical capacity of the phonological store and, as a consequence, the input used to update the target held in storage within this time frame is incomplete or underspecified. AWS demonstrate reduced accuracy during nonword repetition (Byrd, McGill, & Usler, 2015; Byrd, Vallely, Anderson, & Sussman, 2012) and recall (Coalson & Byrd, 2017; Ludlow, Siren, & Zikria, 1997) compared to AWNS. However, researchers across these studies have acknowledged that the difficulties observed in AWS when reproducing novel phonological sequences is difficult to link to any single subprocess of the phonological loop due to potential deficiencies in phonological encoding in AWS. In addition to this potential confound, and similar to silent phoneme monitoring tasks, deficits in phonological working memory in AWS cannot be attributed to deficits in phonological encoding alone, as verbal reproduction reflects the end product of multiple stages within the phonological loop prior to production.

Phonological working memory and phonological complexity in AWS

One final consideration when interpreting previous investigations of phonological working memory in AWS is the segmental and metrical complexity of the stimuli. Phonological working memory – and, in particular, subvocal rehearsal – in AWS appears vulnerable only if segmental length or complexity is increased (>4 syllables; Byrd, McGill, et al., 2015; Byrd et al., 2012; Ludlow et al., 1997; Sasisekaran & Weisberg, 2014), or metrical structure deviates from high-frequency stress patterns (Coalson & Byrd, 2017). For example, Coalson and Byrd (2017) found AWS ($n = 26$) recalled bisyllabic nonwords with less phonemic accuracy than AWNS ($n = 26$) when targets carried iambic stress, but not trochaic stress. These findings were significant only after participants were forced to produce the nonword in the absence of an auditory cue (i.e., hearing the nonword immediately before verbal response) or orthographic cue (i.e., seeing the nonword in written form immediately before verbal response). These data suggest that subvocal rehearsal may have been necessary to retain the target nonword in the absence of these cues, and if so, this rehearsal may have been less effective for AWS than AWNS. Although the phonemic positions subject to greater error were not reported, findings again indicate that phonological targets held in working memory in AWS are more vulnerable to degradation when segmental properties cannot rely upon language dominant stress-patterns. Findings also correspond with the significantly higher proportion of post-trial production errors for AWS reported by Coalson and Byrd (2015) for iambic targets, but not trochaic targets, a pattern that was not observed for AWNS. Together, these studies suggest that during traditional silent phoneme monitoring tasks, the influence of weaker phonological working memory cannot be completely ruled out, and latency and accuracy patterns observed may reflect inefficiencies in subvocal rehearsal rather than phonological encoding.

In sum, increased phonological complexity may be necessary to assess the limitations of phonological working memory in AWS. Based on the findings of Coalson and Byrd (2015, 2017), AWS and AWNS access segments with trochaic nonwords with similar ease. In contrast, nonwords with less frequent iambic structure may impose greater demand upon phonological encoding in AWS and, in turn, may limit the availability of phonological information in working memory. If this is the case, AWS may exhibit greater difficulty monitoring segmental information held in phonological working memory in the absence of language-dominant stress patterns.

Summary and research questions

The primary aim of this study was to examine the speed and accuracy of AWS and AWNS when identifying individual sounds (C_1, C_2, C_3, C_4) within the speech plan held in working memory before and after subvocal rehearsal. To do so, a delayed silent phoneme monitoring task was administered in which participants received a cue to silently identify the presence or absence of a specified phoneme in nonword held in working memory for either 1 s (i.e., before subvocal rehearsal was required) or 4 s (i.e., after subvocal rehearsal was required. The secondary aim of this study was to examine whether metrical stress affects the ability of each group to retain speech plans before and after subvocal rehearsal. To achieve this, AWS and AWNS completed the delayed silent phoneme monitoring task in two separate experiments that included (a) nonword stimuli with trochaic stress (Experiment 1), or (b) identical nonwords stimuli with iambic stress (Experiment 2). We predicted that AWS would be slower and less accurate when monitoring phonemes after subvocal rehearsal, particularly for stimuli with less common iambic stress, than AWNS. Taken together, present findings will further our understanding of the contribution of phonological encoding and phonological working memory to stuttered speech.

(1) Do AWNS and AWS differ in speed and accuracy when silently identifying phonemes within nonwords with trochaic stress held in working memory before and after subvocal rehearsal? [Experiment 1]
(2) Do AWNS and AWS differ in speed and accuracy when silently identifying phonemes within nonwords with iambic stress held in working memory before and after subvocal rehearsal? [Experiment 2]

Methods

Experiment 1: trochaic stress

Participants

All participants provided oral and written informed consent approved by the first author's university (IRB #3428). Data was collected during two separate sessions, each lasting approximately 90 min. To qualify for inclusion, each participant recruited for participation was required to meet the following criteria: (a) 18 years or older, (b) no current medical, speech, language or hearing difficulties reported by participant or observed by the examiner, (c) no current use of medication that may affect reaction time, (d) native-like proficiency in English (Li, Zhang, Tsai, & Puls, 2014), and (e) pass formal hearing screening (American Speech-Language-Hearing Association [ASHA], 1997) and vision screening (US Dept of Health and Human Services, 1996). A total of 20 adults were included in Experiment 1 (10 AWNS; 5 males, 5 females; $M = 22.80$ years; $SD = 4.23$; 10 AWS; 5 males, 5 females; age range: $M = 22.30$; $SD = 3.02$; $p = .60$).

Based on the correlation between phonological segmentation and working memory with silent phoneme monitoring abilities in previous studies (Coalson & Byrd, 2015; Sasisekaran et al., 2006), participants were also required to score within normal limits (i.e., greater than 2 *SD* below the mean) on a battery of subtests designed to assess phonological processing (Word Segmentation, *Comprehensive Test of Phonological Processing* [CTOPP], Wagner, Torgesen, & Rashotte, 1999, subtest XI; Nonword Segmentation *Comprehensive Test of Phonological Processing – Second Edition* [CTOPP-2], Wagner, Torgesen, Rashotte, & Pearson, 2013; subtest IX) and phonological working memory (Nonword Repetition, CTOPP-2 subtest V; Forward Digit Span, CTOPP-2, Subtest IV; Backward Digit Span, CTOPP-2, Subtest IV). Performance on phonological processing measures was collected to ensure that baseline phonological knowledge was comparable between groups, and that outcomes could be attributed to group classification rather than individual differences in phonological processing abilities. Five independent *t* tests revealed no significant differences between groups on phonological processing subtests (Nonword Repetition: $p = .60$, Word Segmentation: $p = .13$, Nonword Segmentation: $p = .36$, Forward Digit Span: $p = .77$, and Backward Digit Span: $p = .25$).

Talker classification and stuttering severity

A participant was considered an AWS if the following criteria were met: (a) self-identified as a person who stutters with reported onset prior to 7 years of age, (b) prior diagnosis of stuttering by a certified speech-language pathologist, and (c) received a score of 2 or higher on the nine-point stuttering severity scale (O'Brian, Packman, Onslow, & O'Brian, 2004) during an elicited conversational speech sample and reading samples. Stuttering severity for each participant was determined by frequency and severity of stuttering-like disfluencies for both speech samples via the nine-point rating scale by O'Brian et al. (i.e., *1* = no stuttering, *2* = very mild stuttering, *9* = extremely severe stuttering). Audio-video recordings of speech sample from 100% of each talker group (10 AWNS, 10 AWS) were reviewed by one graduate student trained in disfluency measurement and the first author, a certified speech-language pathologist, to determine inter-rater reliability. Inter-rater reliability was sufficiently high for participants in Experiment 1 (conversation: 100.0% agreement; intra-class coefficient [ICC] = .96; reading: 90.0% agreement; ICC = .95; see Table 1 for individual participant scores).

Procedure

Stimuli development. The 12 nonwords used in the present study were identical to those used in Coalson and Byrd (2015, 2017) and developed to control for 10 phonological, phonetic, and linguistic factors thought to influence speech or accuracy of response in AWNS as well as AWS (e.g., word-likeness, segmental and biphone phonotactic probability, phonological neighborhood density and frequency, phonetic complexity, syllable frequency, orthographic transparency, uniqueness point, and syllable boundary clarity). See Appendix A for complete list of nonwords, along with their associated monosyllabic nonword foils. As described in Coalson and Byrd (2015, 2017), all auditory stimuli were recorded by a female native English speaker with North American Western dialect within a sound-treated room using KayPENTAX Computerized Speech Lab at 22050-kHz sampling rate and 16-bit quantization.

Experimental block design and stimuli presentation. The 12 experimental blocks used in the present study were also identical to those used in Coalson and Byrd (2015, 2017). Each of the 12 nonwords, one per experimental block, were presented in a fixed randomized order during the second 90-minute session. Each block contained one bisyllabic nonword target with $C_1VC_2C_3VC_4$ structure, along with three monosyllabic CVC nonword foils. Monosyllabic foils contained all four consonants within the target nonword and were included to reduce anticipation or expectancy between presentation of nonword and manual response during silent phoneme monitoring task.

During the second session, participants were seated comfortably approximately 18 inches away from a computer monitor while wearing headphones and resting fingers on Yes/No response keys. Manual reaction time were recorded during the experimental portion of the study using Superlab Pro (v. 4.5) stimulus presentation software and keypad. All post-trail verbal responses were recorded using a video recorder (Zoom Q2, Tokyo, Japan).

As described in Coalson and Byrd (2015, 2017), each of the 12 experimental blocks were comprised of a three-phase training task, followed by the delayed silent phoneme monitoring task. All participants were exposed to the target nonword a minimum of 12 times during the training task before delayed silent

Table 1. Stuttering severity for participants in Experiment 1 and Experiment 2.

	Reading		Conversation			Reading		Conversation	
	Score	Severity	Score	Severity		Score	Severity	Score	Severity
Experiment 1					*Experiment 2*				
AWS-1	2	VM	1	None	AWS-11	4	M-M	9	ES
AWS-2	5	Mod	6	M-S	AWS-12	8	VS	9	ES
AWS-3	2	VM	3	Mild	AWS-13	2	VM	7	Sev
AWS-4	6	M-S	4	M-M	AWS-14	2	VM	2	VM
AWS-5	2	VM	2	VM	AWS-15	5	Mod	7	Sev
AWS-6	8	VS	7	Sev	AWS-16	2	VM	1	None
AWS-7	4	M-M	2	VM	AWS-17	3	Mild	3	Mild
AWS-8	2	VM	4	M-M	AWS-18	2	VM	2	VM
AWS-9	2	VM	2	VM	AWS-19	1	None	2	VM
AWS-10	1	None	2	VM	AWS-20	1	None	3	Mild
AWNS-1	1	None	1	None	AWNS-11	1	None	1	None
AWNS-2	1	None	1	None	AWNS-12	1	None	1	None
AWNS-3	1	None	1	None	AWNS-13	1	None	1	None
AWNS-4	1	None	1	None	AWNS-14	1	None	1	None
AWNS-5	1	None	1	None	AWNS-15	1	None	1	None
AWNS-6	1	None	1	None	AWNS-16	1	None	1	None
AWNS-7	1	None	1	None	AWNS-17	1	None	1	None
AWNS-8	1	None	1	None	AWNS-18	1	None	1	None
AWNS-9	1	None	1	None	AWNS-19	1	None	1	None
AWNS-10	1	None	1	None	AWNS-20	1	None	1	None

Notes: AWS: adult who stutters; AWNS: adult who does not stutter; Score: score on 9-point scale of stuttering severity (O'Brian et al., 2004); Severity: classification of stuttering severity; None: no stuttering, VM: Very Mild; M-M: Mild-Moderate; Mod: moderate; M-S: Moderate-Severe; SEV: severe; VS: very severe; ES: extremely severe.

phoneme monitoring for that target nonword was initiated. The extensive training sequence was necessary to (a) ensure reliable generation of novel targets in the absence of auditory or orthographic cues (see Gupta, 2003, for accuracy of nonword recall in AWNS after similar number of trials), (b) minimize slowed response times observed for AWS after limited pre-experimental practice (see Smits-Bandstra, 2010 for review), and (c) limit use of visuospatial working memory to retain orthographic representation of nonwords before response (see Hawelka, Huber, & Wimmer, 2006). A detailed account of the training task is provided in Coalson and Byrd (2015), and further outlined in Coalson and Byrd (2017; see https://digitalcommons.lsu.edu/comd_pubs/1 for a brief video demonstration of the training task). A brief summary of the training task is provided below for reference.

Training task. Similar to Coalson and Byrd (2015, 2017), pre-experimental training was comprised of three separate phases: repetition, identification, and generation. This training method was based on the three-phase training paradigm described by Levelt and colleagues (e.g., Cholin, Dell, & Levelt, 2011; Cholin, Levelt, & Schiller, 2006; Levelt & Wheeldon, 1994). During the repetition phase, a target nonword (e.g., 'MAZfoov') and three monosyllabic foils (e.g., 'vef,' 'shoam,' 'zale') were presented in written form in a designated corner of the screen with simultaneous auditory cues. The target and foils were presented, one-by-one, in random order (16 trials total; four times per target, four times per foil). Participants were instructed to repeat each immediately after each presentation.

During the identification phase, the written form of the target nonword was replaced by a 2 × 2 in speaker icon in the corner of the screen established during the repetition phase. The speaker icon was presented simultaneously with the three foils presented in written form in the pre-designated corners. Each target nonword and foil was presented aurally in random order at 2 s intervals, and participants were instructed to point to the corresponding corner of the screen (16 trials total; four times per target, four times per foil). No verbal response was required.

During the generation phase, the speaker icon and three foils were presented, one-by-one, in the same pre-designated corners. No auditory input was provided. Participants were instructed to say the nonword that corresponded with the designated corner of the screen after either the speaker icon or written foil appeared. Each nonword target and foil was presented in randomized order (16 trials total; four times per target, four times per foil). Upon completion, the repetition-identification-generation sequence was repeated with half the number of trials per phase (8 trials per phase; two times per target, two times per foil). If a participant could not say the target with 100% accuracy during the second generation phase, the training sequence was repeated until 100% response accuracy was achieved before advancing to the delayed silent phoneme monitoring task. Initial and secondary rounds of training resulted in 72 total responses per experimental block (18 target nonword, 18 per monosyllabic foil) before proceeding to the delayed silent phoneme monitoring portion of the block.

Delayed silent phoneme monitoring task. After training within the experimental block was complete, the delayed silent phoneme monitoring task was completed as outlined in Table 2 and Figure 1b. The

Table 2. Structure of delayed silent phoneme monitoring task within a single experimental block.

Nonword Stimuli	Visual Cue	Visual Location[†]	ISI	Audio Cue	Participant Response
Target	Neutral Icon	Corner A	1 s or 4 s	[f, z, m, l, v, or ʃ]	1. Press Yes/No 2. Say word
Foil 1	Orthographic	Corner B	1 s or 4 s	[v, m, z, or l]	1. Press Yes/No 2. Say word
Foil 2	Orthographic	Corner C	1 s or 4 s	[z, f, m, or ʃ]	1. Press Yes/No 2. Say word
Foil 3	Orthographic	Corner D	1 s or 4 s	[ʃ, l, v, or f]	1. Press Yes/No 2. Say word

Notes: All nonword stimuli (Target, Foil 1, Foil 2, Foil 3) were presented in fixed randomized order within an experimental block to prevent consecutive presentation of any single item. [†]Corner designation balanced across 12 experimental blocks, with target nonword in each corner four times.

design of each trial was similar to the silent phoneme monitoring task described in Coalson and Byrd (2015; depicted in Figure 1a) with two key modifications. First, the order of presentation of the non-orthographic visual cue and the target sound were reversed. Second, based on the description of phonological working memory provided by Baddeley (2003), the inter-stimulus interval (ISI) between the nonword cue and the target sound was delayed by either (a) 1 s, to capture the initial encoding within the phonological store and before subvocal rehearsal, or (b) 4 s, after subvocal rehearsal had occurred.

The response interval used to capture phonological storage and rehearsal was determined based on theoretical accounts of phonological encoding and working memory, neuroimaging data, and behavioral data. According to Baddeley's (2003) model of the phonological loop, a brief verbatim trace is stored for approximately two seconds prior to decay of activation. Therefore, in theory, all segmental and metrical information should be fully activated, stored, and immediately available after an ISI of one second (1 s). One second ISI is sufficient to completely encode a CVCCVC word based on neurophysiological estimates of phonological encoding (e.g., Hartsuiker, 2007; Indefrey & Levelt, 2004; van Turennout, Hagoort, & Brown, 1997; ~500 ms for six-segment speech plans), and previous silent phoneme monitoring studies that report full encoding of target words (from C_1 to C_4) completed between 124 and 380 ms (e.g., Coalson & Byrd, 2015; Sasisekaran et al., 2006; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002). In contrast, retention of phonological code should theoretically require subvocal rehearsal to update the target and offset temporal decay after two seconds have elapsed. Thus, the amount of time required between presentation of nonword and target sound was set at 1 s ISI to capture storage, and 4 s ISI to capture rehearsal.

The delayed silent phoneme monitoring task within an experimental block contained 12 trials. Six target consonants [f, z, m, l, v, or ʃ] occurred in one position of the $C_1VC_2C_3VC_4$ target and one position of the three CVC foils. Order of the target nonword and foils within a block was controlled to ensure target phonemes and target nonwords were not presented consecutively. After all 12 experimental trials within a block were completed, participants proceeded to the next experimental block with a new target nonword and associated nonword foils. Experimental blocks were presented in a fixed randomized order to ensure visual location and onset phoneme of the target nonwords did not overlap between consecutive blocks. Upon completion of all 12 experimental blocks, a total of 32 true positive responses were collected from each participant (i.e., eight per position, C_1, C_2, C_3, or C_4; four at each position with 1 s ISI, four at each position with 4 s ISI), resulting in 640 true positive responses considered during analyses (i.e., 20 participants × 32 responses per participant) in Experiment 1.

As depicted in Figure 1b, the duration between the onset of the target phoneme and button press was measured to measure manual reaction time. Once either button was pressed, or 3 s had elapsed, the participant was prompted to say the nonword aloud to ensure accurate recall of the nonword during previous response. Participants were then instructed to press the 'Yes' key to continue to the next trial starting at the 'Ready?' screen. Video-audio recording of the entire second session, including the post-trial verbal response for all 12 experimental blocks, were analyzed offline to assess the fluency and accuracy of production.

Data coding, reliability, and token exclusion. Four of the 20 participants (20% sample, two AWS, two AWNS) were randomly selected to determine inter-rater reliability, and an additional four participants (20% sample, two AWS, two AWNS) were randomly selected to determine intra-rater reliability. Intra-rater reliability of post-trial verbal response was found to be sufficiently high (90.3% agreement, Kappa = .84), as was inter-rater reliability (95.5% agreement, Kappa = .93).

Individual responses were removed from reaction time and error analyses if the following four criteria were met: (1) no verbal or manual response [NR], (2) overlapping verbal and manual response [OVR], (3) reaction time outliers characterized by 2 *SD*s above or below participant's position-specific mean reaction time [O], or (4) technical errors that prevented accurate measurement, such as audio-video malfunction or software error [TE]. Of the initial 640 tokens across participants in Experiment 1, 58 (9.1%) were considered unusable prior to all subsequent analyses. A total of

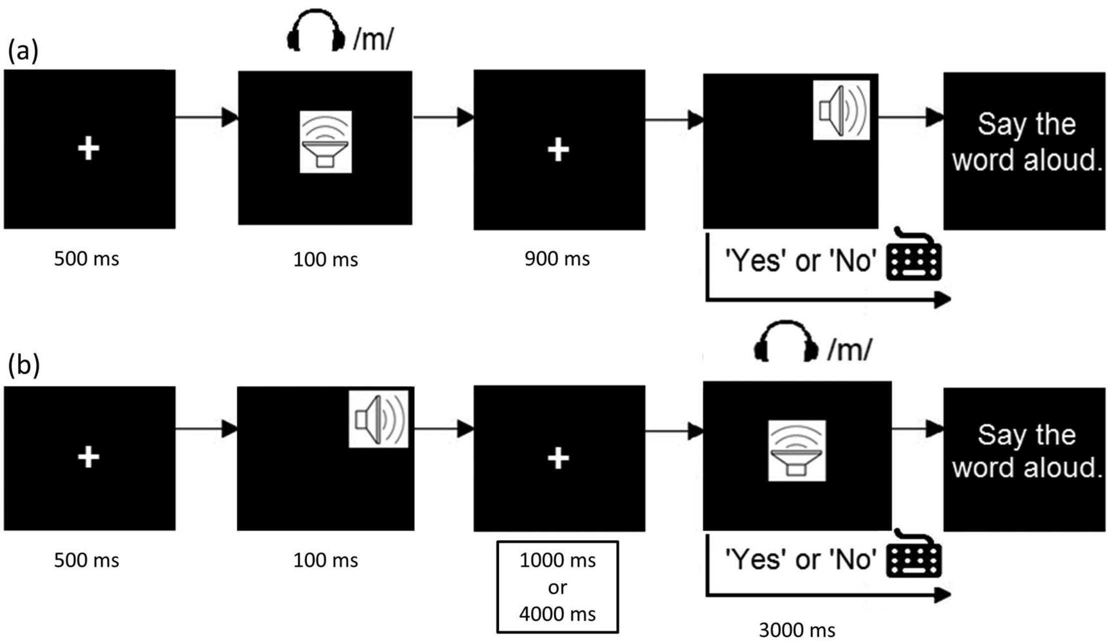

Figure 1. Sequence of events within a single trial during (a) the traditional silent phoneme monitoring task (top; described by Coalson and Byrd [2015]) and (b) the delayed silent phoneme monitoring task (bottom).

582 tokens were considered usable and included in final analyses (AWNS, $n = 288$; AWS, $n = 294$). Table 3 provides detailed breakdown of unusable tokens.

Analyses

Mixed model analyses were conducted using the generalized linear mixed model procedure of SPSS (v. 22), as described by Field (2013) and Heck, Thomas, and Tabata (2012), to assess the relationship between talker group, phoneme position, and inter-stimulus-interval (ISI) upon the speed and accuracy of phoneme identification in nonwords. Two separate mixed models were conducted to assess latency and accuracy data. During analyses of latency, manual reaction time in milliseconds was the continuous response variable. During analyses of accuracy, inaccurate identification of phonemes (i.e., false negative responses) during silent monitoring of the target nonword served as the binomial response variable, with accurate responses coded as the reference category. In both analyses, Group (AWNS, AWS), Position (C_1, C_2, C_3, C_4) and ISI (1 s, 4 s) served as fixed effects. Participants provided multiple responses for each phoneme position (C_1, C_2, C_3, C_4) at each ISI (1 s, 4 s) for each of the 12 nonwords; therefore, Position and ISI served as repeated random effects in both analyses. Pairwise planned comparisons were conducted for all main effects, two-way interactions, and three-way interactions using Least Significant Differences (LSD) adjusted *p*-values. Effect size for main effects and interaction terms was determined using $d = b/(\tau)^{1/2}$, wherein *b* is defined as the coefficient and τ is defined as error variance of the random effects (see Raudenbush & Liu, 2001). Finally, an ANOVA analysis was conducted to examine the frequency of phonemic errors during post–trial verbal production of the target nonword. Mean number of phonemic errors during post-trial response served as the dependent variable and Group (AWS, AWNS) served as the independent variable.

Table 3. Unusable tokens, error tokens, and error combos within data corpus – Experiments 1 and 2.

	Experiment 1: Trochaic Stress			Experiment 2: Iambic Stress		
	AWNS	AWS	*N*	AWNS	AWS	*N*
Initial corpus	320	320	640	320	320	640
Excluded tokens						
NR	18	12	30	13	14	27
OVR	1	6	7	2	4	6
O	10	7	17	8	11	19
TE	3	1	4	4	1	5
Usable *n*	288	294	582	293	290	583
Error tokens						
FN	25	45	70	28	37	65
PE	16	25	41	26	17	43
SE	2	1	3	4	20	24
SLD	2	5	7	0	18	18
nSLD	0	1	1	3	0	3
EC	6	14	20	11	23	34

Notes: AWNS: adults who do not stutter; AWS: adults who stutter; NR: no response; OVR: overlapping verbal response; O: outlier; TE: technical error; FN: false negative response; PE: phonemic error; SE: stress error; SLD: response with stuttering-like disfluency; nSLD: response with non-stuttering-like disfluency; EC: error combination.

Estimation of fit and comparison of models. All fixed effects and interactions of fixed effects up to the third order (Talker Group × Position × ISI) were included in the initial model for each analysis. Third order interaction terms within each analysis were examined first, followed by second-order interaction terms, to determine whether inclusion reduced deviance of chisquared (χ^2) tests and improved the overall fit of each model. Interaction terms were removed if non-significant and change in the −2 log-likelihood value (*−2LL*) between models did not exceed the critical *F*-values that indicate significant change ($p = .01$) reported by Field (2013, p. 898). All final models for latency and accuracy in Experiment 1 satisfied these criteria and, therefore, all interaction terms were included during analyses.

Results

Latencies. To ensure that assessment of latencies were not influenced by inaccurate retrieval and processing of target nonwords or target phonemes, individual tokens were removed from latency analyses, if any one or more of the following six criteria were met: false negative manual response [FN], phonemic error during post-trial production [PE], stress-assignment error during post trial production [SE], stuttering-like disfluencies [SLD], non-stuttering-like disfluencies [nSLD], or any combination of these five errors [EC]. From the usable 582 tokens collected, 142 tokens (24.4%) were excluded from the reaction time analysis based on error response and/or disfluent post-trial verbal response. The final data corpus included 440 fluent, accurate tokens (AWNS, $n = 237$; AWS, $n = 203$; see Table 3 for detailed token exclusion data).

The three-way interaction of Talker Group × Position × ISI was not significant $F < 1$. No significant effects were detected for two-way interactions of Talker Group × Position, ($F < 1$), Talker Group × ISI, ($F < 1$), or ISI × Position $F(3, 424) = 1.16$, $p = .326$. The main effect of Talker Group ($F < 1$) was also not significant. However, significant main effects were detected for Position, $F(3, 424) = 28.82$, $p < .001$ (C_2: $d = .93$, C_3: $d = 1.08$, C_4: $d = 1.40$ [ref. category = C_1]), and ISI, $F(1, 424) = 8.28$, $p = .004$, $d = .47$. Planned comparisons for individual phoneme positions revealed significantly faster phoneme identification in C_1 position ($M =$ 100.37 ms, $SE = 73.51$) compared to each subsequent position (C_2: $M = 1382.31$ ms, $SE = 81.19$, $p < .001$; C_3: $M = 1349.06$ ms, $SE = 84.60$, $p < .001$; C_4: $M =$ 1421.40 ms, $SE = 81.04$; $p < .001$). Planned comparisons for ISI revealed significantly faster overall phoneme monitoring latencies after 4 s delay ($M = 1231.03$ ms, $SE = 73.74$) than 1 s delay ($M = 1355.54$ ms, $SE = 74.58$; see Table 4 and Figure 2).

Table 4. Test of fixed effects for main effects and interactions – Experiment 1.

	Latencies			False Negative Errors		
Factors	*F*	*df*	*p*	*F*	*df*	*p*
Talker Group	0.01	1	.954	2.39	1	.123
ISI	8.28	1	.004**	0.44	1	.510
Position	28.82	3	<.001**	4.88	3	.002**
Talker Group × ISI	0.69	1	.407	0.13	1	.697
Talker Group × Position	0.56	3	.641	0.30	3	.816
ISI × Position	1.16	3	.326	0.34	3	.817
Talker Group × Position × ISI	0.84	3	.473	0.56	3	.651

Notes: ISI = inter-stimulus interval. Reference categories: Talker Group = AWNS; ISI = 1 s ISI; Position = C_1.
*$p < .05$; **$p < .01$.

False negative errors. To ensure that data reflected difficulties in activation and selection of individual phonemes, analysis of false negative errors were based on manual responses provided in the absence of overt post-trial production errors or disfluencies. Of the usable corpus of 582 tokens (AWNS, $n = 288$; AWS, $n = 294$), 72 [12.4%] responses were removed due to the presence of verbal production or fluency error. This resulted in 510 responses considered during analysis (AWNS, $n = 262$, AWS, $n = 248$), including 70 false negative tokens (AWNS, $n = 25$, AWS, $n =$ 45) and 440 accurate manual responses (AWNS, $n =$ 237, AWS, $n = 203$) during analysis.

The three-way interaction of Talker Group × Position × ISI was not significant ($F < 1$). No significant effects were detected for two-way interactions of Talker Group × Position, ($F < 1$), Talker Group × ISI, ($F <$ 1), or ISI × Position ($F < 1$). The main effect of Talker Group $F(1, 494) = 2.39$, $p = .123$, $d = .92$, and ISI ($F < 1$) were also not significant. However, a significant main effect was detected for Position, $F(3, 494) = 5.15$, $p = .002$ (C_2: $d = 1.00$, C_3: $d = .62$, C_4: $d = .84$). Planned comparison of individual phoneme position revealed a significantly lower rate of false negative error for C_1 ($M = 4.3\%$, $SE = 1.8\%$) compared to C_2 ($M = 19.0\%$, $SE = 5.0\%$, $p = .002$) and C_4 ($M = 11.6\%$, $SE = 3.8\%$; $p = .045$), with C_3 approaching but not reaching significance ($M = 10.0\%$, $SE = 3.1\%$, $p = .057$; see Table 4 and Figure 2).

Phonemic errors. To ensure assessment of phonemic errors reflected processes related to segmental accuracy rather than variation related to stress assignment error and/or disfluency, post-trial responses which also included stress or fluency errors were excluded from analysis. Of the usable 582 responses (AWNS, $n = 288$, AWS, $n = 294$), 31 responses were removed. This resulted in 551 responses considered during analysis (AWNS, $n = 278$, AWS, $n = 273$), including 41 phonemic errors (AWNS, $n = 16$, AWS, $n = 25$) and 510 accurate manual responses (AWNS, $n = 262$, AWS, $n =$ 248) during analysis. Results from ANOVA did not reveal a significant difference in mean frequency of phonemic error during post-trial production between AWS and AWNS $F(1, 550) = 2.32$, $p = .129$, $\eta^2 = .001$.

Experiment 2: iambic stress

The purpose of Experiment 2 was to examine whether AWS and AWNS differ in their ability to access phonological information held in working memory as complexity of the stimuli is increased. Intake procedures,

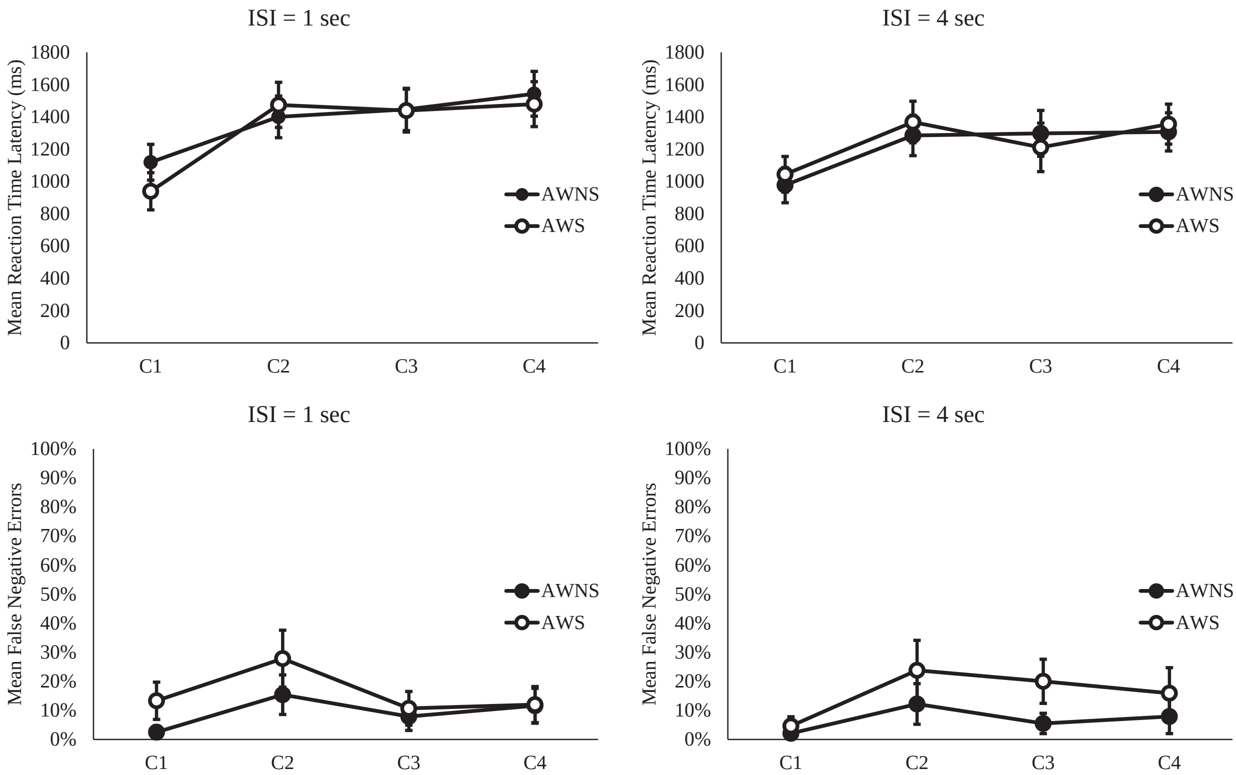

Figure 2. Mean reaction time latencies (top) and mean false negative identification errors (bottom) for adults who stutter (AWS) and adults who do not stutter (AWNS) at 1 s and 4 s inter-stimulus-interval delay (ISI) during delayed silent phoneme monitoring task of $C_1VC_2C_3VC_4$ nonwords with trochaic stress [Experiment 1].

inclusionary and exclusionary criteria, talker group classification criteria, experimental paradigm, stimuli presentation, and analyses in Experiment 2 were identical to those detailed in Experiment 1 with one exception – stimuli were presented with iambic stress rather than trochaic stress.

Participants

A separate cohort of 20 adults who did not participate in Experiment 1 were included in Experiment 2 (10 AWNS; 5 males, 5 females; $M = 20.40$ years; $SD = 1.35$; 10 AWS; 5 males, 5 females; age range: $M = 25.20$; $SD = 4.94$; $p = .31$). No significant differences were detected between groups on phonological processing subtests (Nonword Repetition: $p = .31$, Word Segmentation: $p = .70$, Nonword Segmentation: $p = .16$, Forward Digit Span: $p = .38$, and Backward Digit Span: $p = .51$).

Stuttering severity and reliability

Inter-rater reliability was sufficiently high for 100% of the participants in Experiment 2 (conversation: 75% agreement; ICC = .98; reading: 85% agreement; ICC = .96, see Table 1 for individual stuttering severity scores).

Data coding, reliability, and token exclusion

Similar to Experiment 1, four of the 20 participants (20% sample, two AWS, two AWNS) from Experiment 2 were randomly selected to determine inter-rater reliability, and an additional four participants (20% sample, two AWS, two AWNS) were randomly selected to determine intra-rater reliability. Intra-rater reliability of post-trial verbal response was found to be sufficiently high (87.5% agreement, Kappa = .94), as was inter-rater reliability (88.1% agreement, Kappa = .96). Of the initial 640 tokens collected for Experiment 2, 57 tokens (8.9%) were considered unusable prior to all subsequent analyses. The final data corpus included 583 usable tokens (AWNS, $n = 293$; AWS, $n = 290$; see Table 3).

Estimation of fit and model comparison

For Experiment 2, the final model for latency satisfied goodness-of-fit criteria and all interaction terms were included. However, the final model for phoneme identification accuracy was conducted up to second order interaction terms, as inclusion of third order interaction (Talker Group × Position × ISI) failed to reach significance and did not improve the fit of the model ($-2LL = 2368.29$, $df = 15$) compared to the inclusion fixed effects and two-way interaction terms ($-2LL = 2471.27$, $df = 12$).

Results

Latencies. From the 583 usable tokens, 72 tokens were excluded from AWNS and 115 from AWS participants based on the criteria used in Experiment 1 and described in Table 3. In total, 187 tokens (32.1%) were excluded from reaction time latency analysis on the basis of error response and/or disfluent post-trial verbal response, resulting in a final data corpus that included 396 fluent, accurate tokens (AWNS: $n = 221$; AWS: $n = 175$).

The three-way interaction of Talker Group × Position × ISI was not significant ($F < 1$). No significant effects were detected for two-way interactions of Talker Group × Position ($F < 1$), Talker Group × ISI $F(1, 380) = 2.60$, $p = .108$, $d = .82$ (ref. categories = AWNS, 1 s ISI) and ISI × Position ($F < 1$). The main effect of Talker Group was also not significant ($F < 1$). However, a significant main effects were found for Position, $F(3, 380) = 17.99$, $p < .001$ (C_2: $d = .57$, C_3: $d = .49$, C_4: $d = .78$), and ISI, $F(1, 380) = 4.17$, $p = .042$, $d = .82$. Planned comparisons of individual phoneme positions revealed significantly faster identification of C_1 ($M = 1051.96$ ms, $SE = 74.99$) compared to each subsequent position (C_2: $M = 1309.14$ ms, $SE = 80.55$, $p < .001$; C_3: $M = 1343.17$ ms, $SE = 82.54$, $p < .001$; C_4: $M = 1363.61$ ms, $SE = 80.57$; $p < .001$). Planned comparisons of ISI revealed significantly faster latencies after 4 s ($M = 1225.45$ ms, $SE = 74.54$) than 1 s delay intervals ($M = 1308.37$ ms, $SE = 74.24$; see Table 5 and Figure 3).

False negative errors. Analysis of false negative errors were based on manual responses in the absence of post-trial production errors or disfluencies. Of the usable corpus of 583 tokens (AWNS, $n = 293$; AWS, $n = 290$), 122 (20.9%) responses were removed due to the presence of verbal production or fluency error (see Table 3). This resulted in 461 responses considered during analysis (AWNS, $n = 249$, AWS, $n = 212$), including 65 false negative tokens (AWNS, $n = 28$, AWS, $n = 37$) and 396 accurate manual responses (AWNS, $n = 221$, AWS, $n = 175$).

As depicted in Table 5 and Figure 3, no significant effects were detected for ISI × Position ($F < 1$). The main effects for Talker Group ($F < 1$), ISI $F(1, 448) = 3.10$, $p = .079$, $d = .69$, and Position $F(3, 448) = 1.90$, $p = .128$ (C_2: $d = 0.86$, C_3: $d = .04$, C_4: $d = .43$) were also not significant. However, significant interactions were detected for Talker Group × Position $F(3, 448) = 5.30$, $p < .001$, (C_2: $d = .51$, C_3: $d = 1.08$, C_4: $d = .32$), and Talker Group × ISI $F(1, 448) = 4.45$, $p = .035$, $d = .69$. Decomposition of the Talker Group × Position interaction revealed greater false negative error for AWS at C_3 position ($M = 36.3\%$, $SE = 8.6\%$) than AWNS ($M = 5.4\%$, $SE = 2.8\%$, $p < 0.001$). Further decomposition of the Talker Group × ISI interaction revealed significantly greater errors during silent phoneme identification after 4 s for AWS ($M = 21.2\%$, $SE = 5.8\%$) compared to AWNS ($M = 8.5\%$, $SE = 3.3\%$, $p = .049$). To compare outcomes with and without third-order interactions removed for model fitness, post-hoc analyses were conducted with the inclusion of the highest order interactions (Talker Group × ISI × Position). This interaction was non-significant and did not alter the significance levels reported in the reduced model (i.e., Talker Group × Position, $p = .005$ [C_2: $d = .01$, C_3: $d = .78$, C_4: $d = .12$]; Talker Group × ISI, $p = .034$, $d = .95$).

Table 5. Test of fixed effects for main effects and interactions – Experiment 2.

	Latencies			False Negative Errors		
Factors	F	df	p	F	df	p
Talker Group	0.41	1	.521	0.91	1	.340
ISI	4.17	1	.042*	3.10	1	.079
Position	18.00	3	<.001**	1.90	3	.128
Talker Group × ISI	2.60	1	.108	4.45	1	.035*
Talker Group × Position	0.22	3	.881	5.30	3	<.001**
ISI × Position	0.79	3	.503	0.43	3	.734
Talker Group × Position × ISI	0.30	3	.824	–†		

Notes: ISI = inter-stimulus interval. Reference categories: Talker Group = AWNS; ISI = 1 s ISI; Position = C_1.
†Removed to maintain fit of model.
*$p < .05$; **$p < .01$.

Phonemic errors. Post-trial responses with phonemic errors which also included stress and/or disfluency errors were excluded from analysis. Of the usable 583 responses (AWNS, $n = 293$, AWS, $n = 290$), 79 (13.6%) responses were removed. This resulted in 504 responses considered during analysis (AWNS, $n = 275$, AWS, $n = 229$), including 43 phonemic errors (AWNS, $n = 26$, AWS, $n = 17$) and 461 accurate manual responses (AWNS, $n = 249$, AWS, $n = 212$) during analysis. Results from ANOVA did not revealed no significant difference between groups in mean frequency of phonemic error during post-trial production ($F < 1$).

Discussion

To review, the primary purpose of this study – explored in Experiment 1 – was to examine the integrity of a phonological target held in working memory prior to and after subvocal rehearsal in AWNS and AWS. To do so, a delayed silent phoneme monitoring task was developed wherein participants were required to silently identify target consonants within nonwords (C_1-$VC_2C_3VC_4$) held in the phonological store for 1 s (prior to subvocal rehearsal) and 4 s (after subvocal rehearsal). Participants were significantly faster and more accurate when identifying segments in the initial position, and faster overall after subvocal rehearsal had occurred. No group differences in speed or accuracy were observed between AWNS and AWS.

The second purpose of this study – explored in Experiment 2 – was to examine whether phonological targets held within phonological working memory by AWS and AWNS were more vulnerable to increased phonological demand at either stage. To do so, a second cohort of participants completed the same task described in Experiment 1, but nonword stimuli

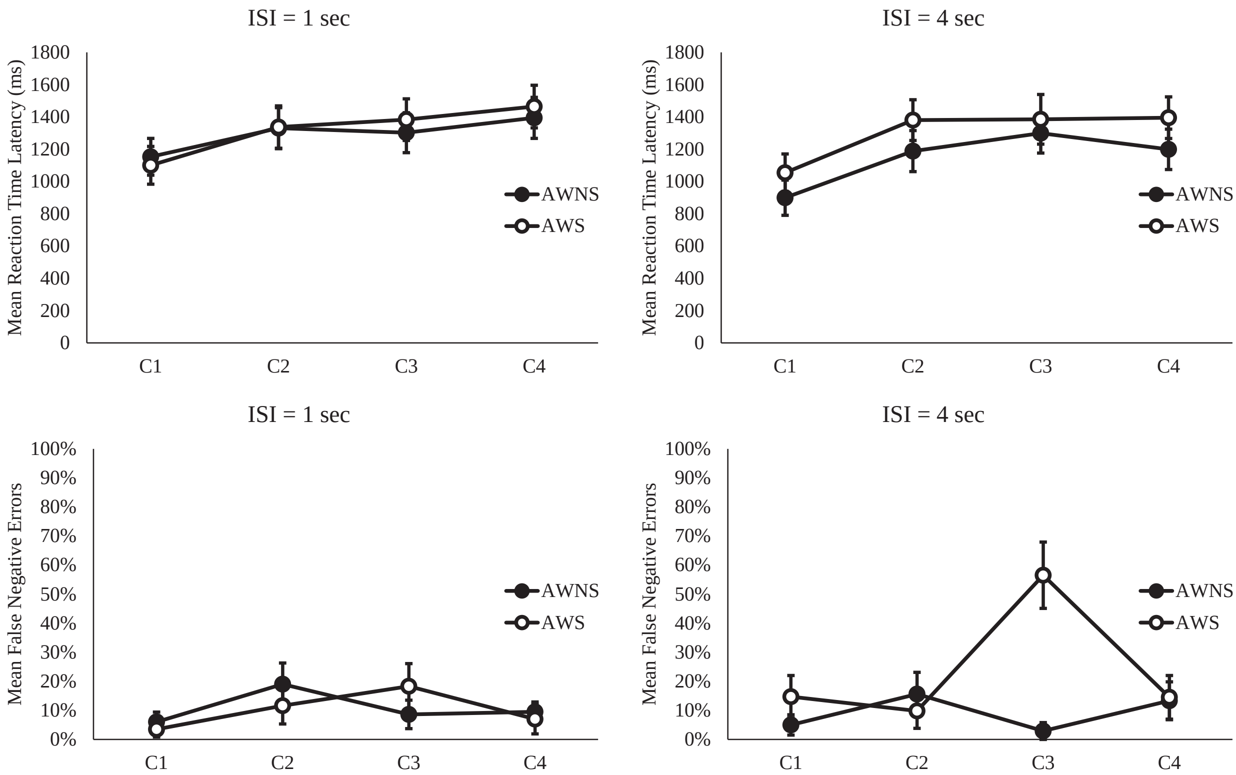

Figure 3. Mean reaction time latencies (top) and mean false negative identification errors (bottom) for adults who stutter (AWS) and adults who do not stutter (AWNS) at 1 s and 4 s inter-stimulus-interval delay (ISI) during delayed silent phoneme monitoring task of $C_1VC_2C_3VC_4$ nonwords with iambic stress [Experiment 2].

were presented with low-frequency stress patterns (i.e., iambic stress) rather than high-frequency stress patterns (i.e., trochaic stress). Similar to Experiment 1, all participants were faster when identifying initial phonemes, and after subvocal rehearsal had commenced. However, in contrast to Experiment 1, AWS were significantly less accurate than AWNS when silently identifying phonemes after subvocal rehearsal had begun. AWS were also significantly less accurate when identifying phonemes that immediately followed the syllable boundary (i.e., C_3). Together, findings from Experiments 1 and 2 suggest that access to phonemes within simple, trochaic sequences is comparable in AWS and AWNS before and after subvocal rehearsal. For more complex, iambic structures, specific phoneme positions were more vulnerable to compromise in AWS after subvocal rehearsal had begun.

Phonological storage and subvocal rehearsal

Based on previous studies implicating subvocal rehearsal as a critical area of compromise for AWS (e.g., see Byrd et al., 2016), we predicted that AWS would differ in speed and accuracy from AWNS when monitoring phonemes in novel phonological sequences that they were forced to hold in memory for 4 s (i.e., beyond the temporal limits of phonological storage) as opposed to 1 s (i.e., within the limits of phonological storage [Baddeley, 2003]). Findings did not support this prediction as both talker groups demonstrated similar overall monitoring speed and accuracy before and after subvocal rehearsal (see Figure 2). These findings are consistent with research that suggests AWS and AWNS perform similarly on phonological working memory tasks when the phonological demand is low. For example, no differences emerge when (a) vocally repeating and/or nonvocally identifying nonwords of shorter lengths (e.g., Byrd et al., 2012; Byrd, McGill, et al., 2015), (b) recalling lists of words that are maximally phonologically dissimilar and maximally semantically similar (e.g., Byrd, Sheng, Gkalitsiou, & Bernstein Ratner, 2015), or (c) the novel phonological sequences have high-frequency trochaic stress assignment (Coalson & Byrd, 2017). In addition, both groups exhibited significantly faster response latencies, on average, when monitoring segmental information after subvocal rehearsal (4 s) than during phonological storage (1 s). Findings also indicate that access to the phonological code for both groups is perhaps stronger after subvocal rehearsal relative to storage.

In addition to comparable group performance, within-group latency and accuracy patterns ran counter to expectations with respect to the availability of segmental information within the phonological loop. We assumed that all segmental information would be fully encoded and immediately available from storage if ample time to complete initial encoding was provided (i.e., 1 s) prior to response. This presumption was based on theoretical models of phonological encoding and working memory (Baddeley, 2003;

Levelt et al., 1999), data from neuroimaging studies (e.g., Hartsuiker, 2007; Indefrey & Levelt, 2004; van Turennout et al., 1997), and time required to fully encode and monitor $C_1VC_2C_3VC_4$ stimuli in previous silent phoneme monitoring studies (Coalson & Byrd, 2015; Sasisekaran et al., 2006; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002). However, even when participants' responses were delayed 1 s (and 4 s) after the target word was cued, participants continued to identify C_1 phonemes faster and with better accuracy than phonemes in C_2–C_4 positions. Findings support that phonemes in initial C_1 position may sustain a stronger level of activation throughout both stages of working memory.

In terms of errors between AWNS and AWS, accuracy of identification was also similar between groups irrespective of ISI delay. Specifically, silent identification for both groups was comparable between storage and rehearsal, although AWS exhibited a slightly higher but non-significant trend toward false negative response. In terms of post-trial errors, AWS exhibited a slight increase in phonemic errors ($n = 25$) versus AWNS ($n = 16$), but this difference did not reach statistical significance. These data, combined with latency data, again suggest that the availability of segments within a nonword does not differ between AWNS and AWS as it transitions from phonological storage to subvocal rehearsal, at least for bisyllabic sequences comprised of high-frequency stress. However, differences observed in AWS when the stimuli to be encoded (Coalson & Byrd, 2015) or held within working memory (Coalson & Byrd, 2017) carry iambic rather than trochaic stress suggest that lack of between-group differences in Experiment 1 may be due to the reduced demand that accompany processing trochaic stress patterns.

Phonological working memory and phonological complexity

Our second aim in Experiment 2 was to examine whether less frequent metrical stress patterns influenced the storage and rehearsal of speech plans in AWS more so than AWNS. Previous studies have indicated that AWS perform similarly during tasks requiring phonological encoding and/or phonological working memory when experimental stimuli were comprised of trochaic stress and did not exceed four syllables (Byrd et al., 2012; Byrd, McGill, et al., 2015; Coalson & Byrd, 2015; Coalson & Byrd, 2017). Iambic stress, on the other hand, has been shown to disrupt both processes to a greater degree in AWS relative to AWNS for stimuli as short as two syllables, and those breakdowns in working memory in AWS are typically observed after subvocal rehearsal is required (Coalson & Byrd, 2015, 2017). Thus, we predicted AWS would demonstrate greater difficulty than AWNS when silently identifying segments within iambic sequences during retention of phonological code in working memory, particularly, when subvocal rehearsal was required (4 s).

In terms of latencies, findings from Experiment 2 do not support these predictions and indicated that speed of identification was similar between groups. In fact, AWNS and AWS exhibited within-word latency patterns for iambic sequences during storage and rehearsal that were remarkably similar to those observed for trochaic sequences in Experiment 1. Data indicated a main effect for ISI and position, suggesting faster access to all segmental positions after subvocal rehearsal for both groups, and stronger activation of C_1 regardless of interval delay. The similarity of within-word latency patterns for trochaic sequences and iambic sequences for AWNS and AWNS may suggest that infrequent metrical properties do not impede how *quickly* phonological information is accessed within working memory for either group. However, significant differences in false negative error between AWS and AWNS that were not observed for identical nonwords with trochaic stress in Experiment 1 indicate that less frequent metrical structure may, instead, compromise the *accuracy* of the phonological code in AWS to a greater extent than AWNS after entry into phonological working memory.

Unlike trochaic stimuli in Experiment 1, AWS exhibited a significantly higher mean proportion of false negative responses when silently monitoring phonemes immediately following the syllable boundary (C_3) within iambic stimuli. AWS also exhibited greater difficulty than AWNS identifying phonemes overall during subvocal rehearsal. As depicted in Figure 3, these two significant interactions were characterized by greater error for C_3 phonemes for AWS ($M = 56.5\%$, $SE = 11.4\%$), if target words were held in memory 4 s prior to response, than AWNS ($M = 2.9\%$, $SE = 2.9\%$). Increased false negative errors were also detected at C_3 for AWS during phonological storage at 1 s ISI ($M = 18.3\%$, $SE = 7.8\%$) compared to AWNS ($M = 8.6\%$, $SE = 4.9\%$), suggesting that difficulties accessing this phonemic position may be present during storage but become more pronounced after rehearsal. These data are consistent with previous reports that retention of more complex phonological code when subvocal rehearsal is required may be uniquely challenging for AWS (e.g., Byrd et al., 2012; Byrd, McGill, et al., 2015; Coalson & Byrd, 2017).

Phonological encoding versus phonological working memory

For the purposes of this paper we considered initial encoding, phonological storage, and subvocal rehearsal as three sequential processes. By allowing 1 s to elapse prior to response, we are assuming that initial encoding has completed and phonological storage has begun (1 s ISI). After 4 s has elapsed prior to response, we are assuming that subvocal rehearsal has begun (4 s ISI). We noted in the introduction that if within-word latency and accuracy patterns during delayed silent phoneme monitoring tasks – even in the absence of between-group differences – were dissimilar from those observed in previous studies which used traditional silent phoneme monitoring tasks with no delay (0 s ISI), this would suggest that traditional and delayed silent phoneme monitoring measure distinct levels of processing (i.e., phonological encoding, phonological working memory, respectively). Across Experiments 1 and 2, and contrary to predictions, within-word latency patterns at 1 s ISI and 4 s ISI appear to resemble those reported in previous studies with 0 s ISI (Coalson & Byrd, 2015; Sasisekaran et al., 2006; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002). That is, AWNS and AWS latency of response were characterized by (a) incrementally slower identification of phonemes (i.e., $C_1 < C_2 < C_3 < C_4$), (b) significantly faster identification of phonemes in C_1 position than all other positions, and (c) an overall 'levelling-off' of latency differences between C_2, C_3, and C_4. Further, given the difficulties AWS demonstrate during initial encoding of nonwords (Coalson & Byrd, 2015) it is possible that 1 s delay was insufficient to fully encode the novel phonological sequences. These patterns may call into question whether data from traditional silent phoneme monitoring tasks in AWNS and AWS may have reflected the contents of the phonological store, or the retention of target words in working memory after subvocal rehearsal had begun, rather than initial phonological encoding.

To explore whether the responses observed during the delayed silent phoneme monitoring task in the present study – thought to reflect two sub-processes of working memory – differ from traditional silent phoneme monitoring task – thought to reflect initial encoding – cross-study comparison is warranted. The most direct comparison of data across studies are provided by Coalson and Byrd (2015), which used identical trochaic and iambic nonword stimuli, procedures, and training protocol as the present study but with no delay between presentation of target nonword and response (i.e., 0 s ISI). Two subtle distinctions in response patterns, at least for trochaic stimuli used in Experiment 1, provide greater confidence that the information monitored by AWNS and AWS after phonological storage (1 ISI) and subvocal rehearsal (4 s ISI) are not identical to initial encoding (0 s ISI; Coalson & Byrd, 2015).

First, C_1–C_2 latency differences are greater in magnitude for both groups after entry into working memory as opposed to initial encoding. That is, the delay of C_2 created C_1–C_2 differences during storage (AWNS: 281 ms; AWS: 535 ms) and rehearsal (AWNS: 308 ms; AWS: 324 ms) were larger than those during initial encoding for AWNS (209 ms) and almost three- to five-times larger for AWS during initial encoding reported at 0 s ISI (107 ms). Notably, these prolonged C_1–C_2 latencies exceed the time required to process the *entire* nonword – C_1 to C_4 – during initial encoding reported by Coalson and Byrd (2015) for both groups (AWNS: 268 ms; AWS: 213 ms). Furthermore, the remaining C_2–C_4 segments in the present study were available almost instantaneously (or faster than the preceding phoneme) during subvocal rehearsal (C_2 to C_4, AWNS: 22 ms, AWS: −13 ms) relative to initial encoding (C_2 to C_4, AWNS: 60 ms, AWS, 106 ms). These patterns suggest a departure from 'incremental' availability of segments and gradual levelling-off that characterize online encoding, and tempers the notion that traditional and delayed silent phoneme monitoring tasks access the same segmental code.

Second, data from trochaic stimuli indicate that both groups identified phonemes in C_1 position with significantly greater accuracy than C_2 whereas Coalson and Byrd (2015) reported no main effect of position for either group. Closer examination of data from each study reveal a greater number of false negative errors for C_2 in the present study during storage (AWS, $M = 27.8\%$, $SE = 9.8\%$; AWNS, $M = 15.4\%$; $SE = 2.2\%$) and rehearsal (AWS, $M = 23.8\%$; $SE = 3.2\%$; AWNS, $M = 12.2\%$; $SE = 7.0\%$) compared to Coalson and Byrd (2015; AWS, $M = 12.0\%$, $SE = 4.0\%$; AWNS, $M = 9.0\%$, $SE = 4.0\%$). Combined with latency data, one may interpret these patterns to suggest that the availability of certain segments differs after entry into the phonological loop, and the speech plan held within working memory is characterized by (a) stable, accurate access to C_1 phonemes, and (b) weaker access to C_2 phonemes. This also suggests that individual segmental positions may be more prone to preservation or degradation once initial encoding has been completed and working memory has been engaged. Taken together, these data provide preliminary support that the delayed silent phoneme monitoring task may reflect fluctuations in the availability of the preverbal speech code as it cycles through each stage of the phonological loop, rather than initial encoding.

Limitations and future studies

Interpretation of the preliminary data in the present study warrant caution for several reasons. First, we acknowledge that the number of participants per experiment was low and less than ideal to assess power. Further, we do not wish to over-interpret null findings derived from data from a limited number of tokens. As is necessary when assessing participant reaction time latencies, false negative manual responses and manual responses followed by inaccurate or disfluent post-trial production were removed from analyses. These criteria allowed for valid comparisons between groups at the expense of data attrition, but we expected attrition to remain relatively similar between groups (with the exception of greater post-trial disfluencies in AWS). This was the case for much of data. Phonemic errors were relatively similar between groups in Experiment 1 (AWNS: $n = 16$ [5%], AWS: $n = 25$ [8%]) and Experiment 2 (AWNS: $n = 26$ [8%], AWS, $n = 17$ [6%]). Post-trial responses that included a disfluency (i.e., SLDs, nSLDs) was higher for AWS than AWNS in Experiment 1 (2% and 0%, respectively) and Experiment 2 (5% and 1%, respectively). False negative errors for both groups in Experiment 1 in the present study across positions and ISI (2% to 27%) were also comparable to Coalson and Byrd (2015 – Experiment 1: 4% to 18%) with nonword stimuli with trochaic stress.

However, two somewhat unexpected patterns emerged related to iambic stimuli in Experiment 2 that reduced the amount of tokens available for latency analyses. First, the proportion of false negative errors for C_3 at 1 s ISI (18%) and 4 s ISI (57%) was greater than expected for AWS compared to Coalson and Byrd (2015 – Experiment 2: 17%). Although position-specific identification errors served as our primary between-group difference, this ultimately decreased the number of valid tokens during latency analysis in Experiment 2. Second, the number of post-trial stress errors when producing iambic tokens in Experiment 2 was notably higher for AWS ($n = 20$, 6%) than AWNS ($n = 4$, 1%), and the combined number of stress-errors in Experiment 1 ($n = 3$, 0.5%). Although these post-trial stress-errors for iambic stimuli remain lower than observed in Coalson and Byrd (2015; 17%), this further reduced the number of valid tokens available for latency analysis. Although no formal analysis of stress-errors could be conducted due to the overall low occurrence, this tendency for persons who stutter to produce greater stress errors for nonwords with lower frequency stress patterns is an interesting trend now observed across three studies (see also Hakim & Bernstein Ratner, 2004). Nonetheless, the combined removal of a high number of false negative errors at C_3 position and post-trial stress errors for AWS decreased the number of valid tokens during reaction time available for analysis at each position and each ISI, particularly in Experiment 2. We recommend that future studies increase the number of tokens per position at each ISI to accommodate these potential trends in AWS response.

Another consideration is the potential influence of individual outliers affecting statistical outcomes. For instance, it is possible that the group differences false negative errors was driven by a single participant outlier due to modest sample size. Post-hoc analysis of false negative error rate for each group across ISI and Position revealed no participant outliers for Experiment 1, and one potential participant outlier for Experiment 2 (AWS-13). Removal of AWS-13 during re-analysis of false negative errors in Experiment 2 did not alter the significance of interactions detected in the original analysis (Talker Group × ISI: $F(1, 430) = 8.30$, $p = .004$, Talker Group × Position: $F(3, 430) = 4.59$, $p = .004$; ISI × Position: $F < 1$), and a significant difference emerged for the main effect of ISI [$F(1, 430) = 5.04$, $p = .025$]. Further breakdown of false negative errors by Talker Group, ISI, and Position in Table 6 did not reveal any notable participant outliers at C_3 at 4 s ISI within AWS, providing greater confidence that group differences were not driven by a single

Table 6. Number of false negative errors out of four trials at each consonant position (C1, C2, C3, C4) and each inter-stimulus-interval (1 s ISI, 4 s ISI) when monitoring iambic nonwords in Experiment 2.

	1 s ISI					4 s ISI				
Participant	C1	C2	C3	C4	*Total*	C1	C2	C3	C4	*Total*
AWNS-11	0	1	0	1	*2*	0	0	1	1	*2*
AWNS-12	1	1	1	0	*3*	0	2	0	0	*2*
AWNS-13	0	2	1	0	*3*	1	2	0	1	*4*
AWNS-14	0	1	0	1	*2*	1	1	0	0	*2*
AWNS-15	0	0	0	0	*0*	0	0	0	0	*0*
AWNS-16	0	0	0	0	*0*	0	0	0	0	*0*
AWNS-17	0	0	0	0	*0*	0	0	0	0	*0*
AWNS-18	0	0	0	1	*1*	0	0	0	1	*1*
AWNS-19	0	1	0	0	*1*	0	0	0	1	*1*
AWNS-20	1	1	1	0	*3*	0	0	0	1	*1*
Total	*2*	*7*	*3*	*3*	*15*	*2*	*5*	*1*	*5*	*13*
AWS-11	0	0	0	0	*0*	0	0	2	0	*2*
AWS-12	0	0	0	1	*1*	0	0	0	0	*0*
AWS-13	1	1	1	1	*4*	0	0	3	1	*4*
AWS-14	0	0	1	0	*1*	1	0	1	2	*4*
AWS-15	0	1	0	0	*1*	0	1	0	0	*1*
AWS-16	0	0	0	0	*0*	1	0	2	1	*4*
AWS-17	0	0	0	0	*0*	0	1	2	0	*3*
AWS-18	0	0	1	0	*1*	0	1	3	0	*4*
AWS-19	0	1	0	0	*1*	2	0	0	0	*2*
AWS-20	0	0	2	0	*2*	0	0	2	0	*2*
Total	*1*	*3*	*5*	*2*	*11*	*4*	*3*	*15*	*4*	*26*

Notes: AWNS: adult who does not stutter; AWS = adult who stutters. Total number of errors for each participant out of four total trials at each Position and ISI.

participant outlier. Nonetheless, the potential influence of individual outliers should not be disregarded in future studies.

The demands placed on the phonological and memory systems in Experiment 2 were achieved by a modest and incremental increase in a single property of speech (i.e., metrical stress). The justification of manipulation of stress alone was based on previous data which indicate that this relatively minor change is sufficient to perturb phonological encoding (Coalson & Byrd, 2015) and phonological working memory (Coalson & Byrd, 2017) in AWS. That said, we do not wish to imply that iambic stress alone is sufficient to provoke moments of stuttered speech, nor should it be considered a distal or proximal cause of stuttering. Instead, the use of lower frequency iambic stress, compared to high-frequency trochaic stress, provided an efficient and uniform method to increase demand of the stimuli without also changing the segmental composition and its phonological-lexical properties (e.g., phonotactic probability, word-likeness, syllable frequency). Additional investigations of the availability of phonological code within working memory with increased segmental demand (e.g., longer nonwords), segmental complexity (e.g., complex combinations of sound segments), or cognitive demand (e.g., increased memory load during ISI) are warranted to further explore differences between AWS and AWNS. We anticipate that as additional demands are imposed on AWNS and AWS in future studies, greater disparity between groups will be observed.

Clinical implications

The primary purpose of this study was to refine our understanding of stuttering and how systems related to speech production and processing – such as phonological working memory – may differ in persons who stutter. Although the present study was conducted in adults, there are sufficient data to suggest that children who stutter exhibit more pronounced difficulties in nonword repetition than fluent peers (Anderson & Wagovich, 2010; Anderson, Wagovich, & Hall, 2006; Hakim & Bernstein Ratner, 2004), and that these difficulties may predict persistence into adulthood (see Spencer & Weber-Fox, 2014). Based on our findings, it is possible that, similar to AWS, children who stutter also demonstrate unique difficulties maintaining word-medial phonemes – such as C_3 – compared to fluent peers or children who recover. Future nonword repetition studies are necessary to determine whether error position analysis, rather than overall accuracy or percent consonant correct, may strengthen the diagnostic and prognostic value of nonword repetition in children.

For many adult clients who stutter, the role of subvocal rehearsal is not trivial. 'Silent' rehearsal of speech prior to production is not an abstract concept, but a deliberate (yet ineffective) act to avoid potential stuttering or feared words (see Jackson, Yaruss, Quesal, Terranova, & Whalen, 2015). During treatment, clinicians may target excessive rehearsal to minimize avoidance behaviors and its potential psychosocial consequences. Based on our preliminary findings, clinicians may have an additional reason to target and perhaps prioritize excessive silent rehearsal. Unlike typically fluent speakers, silent rehearsal may actually degrade the quality of the upcoming speech plan in AWS and further compromise fluent speech. It should be noted that this was observed for AWS even under modest linguistic demand (i.e., less frequent stress patterns). Trace decay during silent rehearsal may be even greater in the presence of more salient demands, such as increased physiological stress (e.g., Bowers, Saltuklaroglu, & Kalinowski, 2012) or oro-motor discoordination (e.g., van Lieshout, Ben-David, Lipski, & Namasivayam, 2014). However, the present study was restricted to non-verbal responses and future studies are warranted to examine the relationship between silent rehearsal, cognitive-linguistic demand, and overt disfluency.

Conclusion

Findings from the present study indicate that AWS do not exhibit greater difficulty than AWNS monitoring phonemes in two-syllable targets with high frequency metrical stress before or after initiation of subvocal rehearsal. However, AWS do exhibit greater difficulty accurately monitoring phonemes when retaining iambic sequences after subvocal rehearsal.

Acknowledgements

We would like to thank the Michael and Tami Lang Stuttering Institute and Dr. Jennifer and Emanuel Bodner for their endowed support of our research efforts.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This project was supported by the LSU-ORED Faculty Research Grant, as well the Louisiana Board of Regents [grant number LEQSF(2015-18)-RD-A-03] and [grant number LEQSF-EPS(2015)-PFUND-392].

References

American Speech-Language-Hearing Association (ASHA). (1997). *Guidelines for audiologic screening* [Guidelines]. Retrieved from www.asha.org/policy.

Anderson, J. D., & Wagovich, S. A. (2010). Relationships among linguistic processing speed, phonological working memory, and attention in children who stutter. *Journal of Fluency Disorders*, *35*, 216–234.

Anderson, J. D., Wagovich, S. A., & Hall, N. E. (2006). Nonword repetition skills in young children who do and do not stutter. *Journal of Fluency Disorders*, *31*, 177–199.

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

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Bajaj, A. (2007). Working memory involvement in stuttering: Exploring the evidence and research implications. *Journal of Fluency Disorders*, *32*, 218–238.

Bowers, A., Saltuklaroglu, T., & Kalinowski, J. (2012). Autonomic arousal in adults who stutter prior to various reading tasks intended to elicit changes in stuttering frequency. *International Journal of Psychophysiology*, *83*, 45–55.

Byrd, C. T., Coalson, G. A., McGill, M., & Gkalitsiou, Z. (2016, June). *Exploring the relationship between phonological working memory and stuttering*. Paper presented at the International Fluency Association 8th World Congress, Lisbon, Portugal.

Byrd, C. T., McGill, M., & Usler, E. (2015). Nonword repetition and phoneme elision in adults who do and do not stutter: Vocal versus nonvocal performance differences. *Journal of Fluency Disorders*, *44*, 17–31.

Byrd, C. T., Sheng, L., Gkalitsiou, Z., & Bernstein Ratner, N. (2015). Veridical and false recall in adults who stutter. *Journal of Speech Language and Hearing Research*, *58*, 28–42.

Byrd, C. T., Vallely, M., Anderson, J., & Sussman, H. (2012). Nonword repetition and phoneme elision in adults who do and do not stutter. *Journal of Fluency Disorders*, *37*, 188–201.

Byrd, C. T., Wolk, L., & Davis, B. L. (2007). Role of phonology in childhood stuttering. In E. Conture, & R. Curlee (Eds.), *Stuttering and related disorders of fluency* (pp. 168–182). New York, NY: Thieme Medical Publishers, Inc.

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 109–122.

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, *99*, 205–235.

Coalson, G. A., & Byrd, C. T. (2015). Metrical encoding in adults who do and do not stutter. *Journal of Speech Language and Hearing Research*, *58*, 601–621.

Coalson, G. A., & Byrd, C. T. (2017). Nonword repetition in adults who stutter: The effects of stimuli stress and auditory-orthographic cues. *PLoS ONE*, *12*, e0188111. doi:10.1371/journal.pone.0188111

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Thousand Oaks, CA: SAGE.

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

Hakim, H. B., & Bernstein Ratner, N. (2004). Nonword repetition abilities of children who stutter: An exploratory study. *Journal of Fluency Disorders*, *29*, 179–199.

Hartsuiker, R. J. (2007). Studies on verbal self-monitoring: The perceptual loop model and beyond. In A. S. Meyer, L. R. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 93–121). Brussels: Psychology Press.

Hawelka, S., Huber, C., & Wimmer, H. (2006). Impaired visual processing of letter and digit strings in adult dyslexic readers. *Vision Research*, *46*, 718–723.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2012). *Multilevel modeling of categorical outcomes using IBM SPSS*. New York, NY: Routledge.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.

Jackson, E. S., Yaruss, J. S., Quesal, R. W., Terranova, V., & Whalen, D. H. (2015). Responses of adults who stutter to the anticipation of stuttering. *Journal of Fluency Disorders*, *45*, 38–51.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *TRENDS in Cognitive Sciences*, *10*, 480–486.

Jansma, B. M., & Schiller, N. O. (2004). Monitoring syllable boundaries during speech production. *Brain and Language*, *90*, 311–317.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. 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*, 1–75.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*, 673–680.

Ludlow, C. L., Siren, K., & Zikria, M. (1997). Speech production learning in adults with chronic developmental stuttering. In W. Hulstijn, H. F. M. Peters, & P. H. H. M. van Lieshout (Eds.), *Speech production: Motor control, brain research, and fluency disorders* (pp. 221–230). Amsterdam: Elsevier Science.

Nippold, M. A. (2002). Stuttering and phonology: Is there an interaction? *American Journal of Speech-Language Pathology*, *11*, 99–110.

O'Brian, S., Packman, A., Onslow, M., & O'Brian, N. (2004). Measurement of stuttering in adults: Comparison of stuttering-rate and severity-scaling methods. *Journal of Speech Language and Hearing Research*, *47*, 1081–1087.

Raudenbush, S. W., & Liu, X.-F. (2001). Effects of study duration, frequency of observation and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, *6*, 387–401.

Sasisekaran, J. (2014). Exploring the link between stuttering and phonology: A review and implications for treatment. *Seminars in Speech and Language*, *35*, 95–113.

Sasisekaran, J., & de Nil, L. F. (2006). Phoneme monitoring in silent naming and perception in adults who stutter. *Journal of Fluency Disorders*, *31*, 284–302.

Sasisekaran, J., de Nil, L. F., Smyth, R., & Johnson, C. (2006). Phonological encoding in the silent speech of persons who stutter. *Journal of Fluency Disorders*, *31*, 1–21.

Sasisekaran, J., & Weisberg, S. (2014). Practice and retention of nonwords in adults who stutter. *Journal of Fluency Disorders*, *41*, 55–71.

Schiller, N. O. (2005). Verbal self-monitoring. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 245–261). Mahwah, NJ: Lawrence Erlbaum.

Smits-Bandstra, S. (2010). Methodological considerations in the measurement of reaction time in persons who stutter. *Journal of Fluency Disorders*, *35*, 19–32.

Spencer, C., & Weber-Fox, C. (2014). Preschool speech articulation and nonword repetition abilities may help predict eventual recovery or persistence of stuttering. *Journal of Fluency Disorders*, *41*, 32–46.

United States Department of Health and Human Services. (1996). The international classification of diseases, 9th revision, clinical modification (ICD-9-CM). 4th edition, volume 1. Retrieved from Centers for Disease Control and Prevention website: http://www.cdc.gov/nchs/icd/icd9cm.html.

van Lieshout, P., Ben-David, B., Lipski, M., & Namasivayam, A. (2014). The impact of threat and cognitive stress on speech motor control in people who stutter. *Journal of Fluency Disorders*, *40*, 93–109.

van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 787–806.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processes – second edition (CTOPP-2)*. Austin, TX: Pro-Ed.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processes (CTOPP)*. Austin, TX: Pro-Ed.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*, 311–334.

Wheeldon, L. R., & Morgan, J. L. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes*, *17*, 503–535.

Appendix A. Target nonword stimuli with associated foils

Block	Target Nonword	Foil 1	Foil 2	Foil 3
1	/viʃ.fuz/	/ʃɛv/	/zom/	/laf/
2	/zæl.ʃov/	/vif/	/miʃ/	/ləz/
3	/ʃiv.lom/	/vuz/	/fəʃ/	/mɛl/
4	/fæz.mul/	/vim/	/zof/	/ʃəl/
5	/lam.vef/	/fɛʃ/	/miv/	/zol/
6	/muf.zoʃ/	/faz/	/vim/	/ʃəl/
7	/foʃ.vul/	/ʃaz/	/zɪf/	/miv/
8	/lev.mof/	/vəl/	/faʃ/	/zim/
9	/mæz.fuv/	/vɛf/	/ʃom/	/zel/
10	/ʃɛm.liz/	/fuʃ/	/zev/	/mæl/
11	/vul.ziʃ/	/ʃaf/	/fɛv/	/lom/
12	/zɪf.ʃom/	/vul/	/feʃ/	/məz/

Note: Lexical, linguistic, and phonotatic properties of each nonword available in Table 2 of Coalson and Byrd (2015, 2017).