Published in final edited form as:

J Exp Child Psychol. 2023 February ; 226: 105567. doi:10.1016/j.jecp.2022.105567.

Infants' short-term memory for consonant–vowel syllables

Rochelle S. Newman[*],

Vidda M. Simpson

University of Maryland, College Park, MD 20742, USA

Abstract

This research examined whether the auditory short-term memory (STM) capacity for speech sounds differs from that for nonlinguistic sounds in 11-month-old infants. Infants were presented with streams composed of repeating sequences of either 2 or 4 syllables, akin to prior work by Ross-Sheehy and Newman (2015) using non-linguistic musical instruments. These syllable sequences either stayed the same for every repetition (constant) or changed by one syllable each time it repeated (varying). Using the head-turn preference procedure, we measured infant listening time to each type of stream (constant vs varying and 2 vs 4 syllables). Longer listening to the varying stream was taken as evidence for STM because this required remembering all syllables in the sequence. We found that infants listened longer to the varying streams for 2-syllable sequences but not for 4-syllable sequences. This capacity limitation is comparable to that found previously for nonlinguistic instrument tones, suggesting that young infants have similar STM limitations for speech and nonspeech stimuli.

Keywords

Short-term memory; Memory; Infant

Introduction

Short-term memory (STM), or working memory, is a capacity-limited system that allows information to be stored temporarily while undergoing processing (Baddeley & Hitch, 1974). This system is a critical underpinning for language because phonological STM is related to children's language acquisition and vocabulary development (Baddeley et al., 1998; Gathercole & Baddeley, 1989; Gathercole et al., 1997; Hoff et al., 2008).

STM has a very constrained capacity; individuals can hold only a limited amount of information at once. Ross-Sheehy and Newman (2015; henceforth R-S&N) reported the first study examining infants' auditory STM capacity. They tested 10.5-month-olds' ability to recognize when a sequence of instruments changed, arguing that discrimination of changing versus repeating streams implied that the length of the stream was within STM capacity limits. Specifically, they found that infants listened longer to sequences of varying tones compared with sequences that repeated identically but did so only when sequence durations

[*]Corresponding author. rnewman1@umd.edu (R.S. Newman).

were 700 ms, not 1400 ms. They argued that this represented the capacity limit for infants at this age; thus, infants had an auditory STM capacity of approximately 1 second's worth of information (or, more precisely, of less than 1.4 s). They also suggested that this capacity could place limits on infants' learning of the language spoken around them.

Studies of adult STM suggest that capacity estimates differ across verbal and nonverbal tasks (Li et al., 2013) and that the systems that underlie different types of auditory memory appear to be dissociable (Schulze & Tillmann, 2013). There is also evidence that working memory for phonological versus nonlinguistic auditory information may be served by distinct cortical networks (Stevens, 2004). Given these findings with adults, R-S&N (2015) noted that it was unclear whether their findings from young infants using nonlinguistic auditory sounds would extend to verbal stimuli. If adults use specialized mechanisms for chunking linguistic information into units, this might take time to develop; in that case, young infants would show similar capacities for linguistic and nonlinguistic stimuli at early ages, but these capacities would diverge over time. In contrast, if adults have entirely separate stores for auditory and phonological information, this distinction could be present even in very young infants.

STM for verbal stimuli is critical for young infants who are in the earliest stages of learning their native language. To acquire language, infants need to store examples of the input they receive in long-term memory; presumably, STM is required as a temporary holding place for this information as it is processed for longer-term storage. Thus, the limits on infants' STM for speech information are likely to play a critical role in their acquisition of linguistic representations. R-S&N (2015) pointed out this fact when they suggested, "If the findings of limited capacity in infant auditory STM were to extend to phonological STM, it would have implications for a wide range of theories of verbal language." They raised two particular examples: the nature of infant-directed speech (IDS) and statistical learning of syntax. Parental utterances to infants are typically short, often only single words; some have suggested that the purpose of such short sentences is to aid word segmentation (Brent & Siskind, 2001). Yet infants do not appear to need isolated words to recognize them (Fernald & Hurtado, 2006). A limited-capacity STM could be an alternative explanation for this aspect of IDS; perhaps parents provide short utterances to ensure that they fall within children's limited memory store.

Researchers studying statistical learning have questioned how infants identify which of the myriad potential relationships among words are most important to track (Newport, 1990). The "less is more" hypothesis (Newport, 1990) suggests that a limited STM capacity might help young learners to focus on the most important relationships in the linguistic signal (see also Elman, 1993). These theories would garner tremendous support if infants' STM capacity was indeed limited in this way. Thus, theories of both syntactic acquisition and parental speech would be affected by findings that infants' verbal STM has similar capacity limits to their nonverbal STM.

In summary, the input children hear serves as the basis for language learning, which is a fundamental aspect of children's early development. But this input is filtered by the children's own intake systems, including STM. Understanding capacity limits for language input therefore is critical for understanding how children learn language and why parents speak to infants in the way that they do. In this study, we employed a method similar to that of R-S&N (2015) but used speech syllables rather than musical tones to assess capacity limitations for verbal stimuli at this age.

Method

Participants

Participants were 32 healthy infants (20 male) aged approximately 11 months (M_{age} = 336 days, range = 311–371) from predominantly English-speaking homes (minimum 80 % English); the approximate age and number of participants was based on R-S&N (2015). Data from an additional 11 infants[1] were excluded from analyses due to failure to meet the age requirement (history of prematurity or above age limit; $n = 3$), not meeting the language requirement ($n = 2$), fussiness ($n = 5$), or experimenter error ($n = 1$). Racial distribution was 72 % Caucasian, 6 % African American, 3 % Asian, and 19 % mixed race. Regarding parents' education, 4 of the infants' mothers had some college education, 9 had a 4-year degree, 12 had a master's degree, and 7 had a doctoral degree.

Stimuli

The stimuli consisted of multi-syllable sequences. Each syllable within a sequence consisted of a consonant–vowel (CV) combination that does not make a known word. The eight consonants were the voiced stop consonants *d* and *g*, the voiceless stop consonant *k*, the voiceless fricative *sh*, the voiced fricative *v*, the voiceless affricate *ch*, the voiced glide *w*, and the voiced liquid *l*. The eight vowels were those in b*ea*d, b*oa*t, b*ai*t, b*oo*t, b*u*t, b*ir*d, b*o*g, and b*a*g. Vowels and consonants were pseudo-randomly combined such that we avoided combinations that would have resulted in real words (e.g., *lay*, *show*); the final eight syllables were *ka*, *dir* (as in "dirt"), *gee*, *shay*, *wuh*, *voo*, *lae* (as in "lap"), and *cho*. The CV combinations were recorded at a 44.1-kHz sampling rate, 16 bits quantization, by a female native speaker of English (the first author) using a Shure SM81 microphone (Shure, Niles, IL, USA) and Mackie 1202 VLZ mixer/amplifier (Mackie, Woodinville, WA, USA). The syllables were edited so that they were each 350 ms in length, the same duration as the musical tones in R-S&N (2015).

The syllables were combined into multisyllabic sequences and presented in lists. There was a total of four different sequence types, consisting of a crossing between the number of syllables in a sequence (2 vs 4) and whether the sequence repeated identically each time (constant) or one syllable would vary from repetition to repetition (varying). For example, the 2-syllable constant sequence consisted of a 2-CV combination that repeated over and over, with the repetitions separated by 350 ms of silence, such as "shay-cho … shay-cho … shay-cho." In the 2-syllable varying sequences, however, 1 CV syllable changed in either the first or second position each time the item occurred (e.g., "shay-cho … shay-gee … ka-gee"). The 4-syllable constant sequence consisted of a combination of 4 CV syllables

[1]Part-way through the study, we identified a minor error in the stimulus list for some orders; to be safe, we replaced all children who had been run in those orders. Children who were replaced for this reason are not counted in the 32 good infants or the 11 drops.

that repeated over and over, such as “lae-cho-shay-gee … lae-cho-shay-gee … lae-cho-shay-gee,” and the 4-syllable varying series likewise had 4 syllables but 1 syllable would vary from repetition to repetition (“lae-cho-shay-gee … lae-cho-ka-gee … dir-cho-ka-gee”). Because each syllable was 350 ms in length, both the 2-syllable sequences were 700 ms and the 4-syllable sequences were 1400 ms.

Changes occurred pseudo-randomly in either the first or second position for 2-syllable items and either the first, second, third, or fourth position for 4-syllable items (e.g., ABCD, <pause>, ABED, <pause>, FBED, <pause>, FBEG, <pause>, FBHG). There were limits imposed on the randomization. First, the same position could not change more than three times in a row. Second, a sequence could not immediately follow an identical sequence in the list (i.e., the change could not be back to itself), although syllable sequences could repeat within the list (ABCD, ABED, ABCD). Third, sequences were eliminated if they sounded like an English word because the goal of the investigation was to have nonsense syllable sequences. In particular, sequences such as “wah-dir” and “lae-dir” were avoided because they sounded too similar to the real words “water” and “ladder.”

These combinations were arranged into 16 different sequences. There were eight 2-syllable pairs (4 constant and 4 varying sequences) and eight 4-syllable pairs (4 constant and 4 varying sequences).

Procedure

Parents first completed questionnaires on their child’s demographic, language, and developmental background. Then the infant was seated on the parent’s lap in a three-sided test booth. The head turn preference paradigm (Kemler Nelson et al., 1995) was used to assess the child’s behavior, and the procedure mirrored that of R-S&N (2015). A researcher (previously tested for reliability) monitored the session and coded the infant’s head turns in real time by pressing buttons on a computer-controlled response box.

To acquaint infants with the task, infants heard a 4-trial pretrial phase of musical passages. The test phase consisted of 16 trials that were blocked into groups of 4. Each block included a 2-syllable constant passage, a 2-syllable varying passage, a 4-syllable constant passage, and a 4-syllable varying passage presented in random order.

Trials began with the center yellow light blinking. Once the infant oriented toward the light, it was turned off and one of the red lights on the side panels began to flash. Once the child looked in that direction, the recording would play from the loudspeaker on the same side of the booth as the light so that it seemed as if the flashing light was the source of the sound. As is standard practice, in the pretrial phase the red blinking light turned off as soon as the child oriented because there was no advantage to maintaining the infant’s attention on the musical pretrial stimulus. In the test phase, the red light continued to flash (encouraging further attention) until the recording ended or the child looked away for a minimum of 2 s.

The experimenter sat behind the center panel and pressed a button on a response box every time the child looked at or away from the flashing lights. Listening time was assessed by the amount of time the infant spent looking at the flashing light (the “source” of the sound);

any time the infant spent looking away was excluded. To minimize bias, both the parent and experimenter wore Peltor aviation headphones (3 M, St. Paul, MN, USA) playing masking music.

Analysis

The analysis mirrored that from R-S&N (2015). Looking time for the four different types of passages (2-syllable constant, 2-syllable varying, 4-syllable constant, and 4-syllable varying) were recorded and averaged by the computer. A 2 (Series Length: 2 or 4 syllables) × (Series Type: constant or varying) repeated-measures analysis of variance (ANOVA) was conducted to examine the effect of sequence length and sequence type. Given the results from R-S&N (2015), this was followed by planned analyses examining the two stimulus lengths separately.

Results

The analysis revealed a marginal main effect of sequence type, $F(1, 31) = 3.50$, $p = .071$, $\eta_p^2 = .101$, indicating that, overall, infants tended to listen longer to the varying series (10.50 s) than to the constant series (9.36 s), in line with prior results from R-S&N (2015). The main effect of series length was significant, $F(1, 31) = 5.06$, $p = .032$, $\eta_p^2 = .140$, indicating that infants listened longer in general to the 4-syllable sequences (10.52 s) than to the 2-syllable sequences (9.34 s). The length by series type interaction was not significant, $F(1, 31) = 1.26$, $p = .27$, $\eta_p^2 = .039$, but planned comparisons (see Table 1) showed a reliable difference in listening times to the varying and constant sequences (Fig. 1), as in the prior nonlinguistic study. Infants listened longer to the varying items than to the constant items in the 2-syllable series, but not in the 4-syllable condition. The lack of preference for the 4-syllable condition suggested that 4 items might be beyond infants' STM capacity.

These findings suggest that infants' memory for syllables is quite short. Infants seemed to be able to remember 2 syllables (700 ms), but they did not remember series that were 4 syllables long (1400 ms). These results match the memory duration found in R-S&N (2015) and suggest that infants' STM for speech sounds appears to be equivalent to that for other auditory sounds. That said, the increased variability found in this study resulted in the interaction between factors not being significant.

Discussion

This study extends findings from R-S&N (2015) to establish that by 11 months of age infants can encode and remember sequences of syllables, much like they can encode and remember sequences of musical tones. Moreover, infant STM capacity for speech is capacity-limited, and this capacity appears (at least superficially) to be very similar to that shown previously for nonlinguistic stimuli among infants of the same age. Whereas R-S&N (2015) reported that infants could remember 2, but not 4, tones of 350 ms duration each, the current results suggest that infants can likewise remember 2, but not 4, nonsense syllables of 350 ms duration each—or roughly 1 second's worth of auditory information. This finding places clear limits on the duration of a given speech utterance that infants are likely to process and remember. It also provides an explanation for findings from Brent and Siskind (2001) that the frequency with which children hear a word in isolation predicts the likelihood of that word being learned; words in isolation are highly likely to fall within infants' STM capacity limits, whereas longer sentences will not. When hearing a longer utterance, infants may resort to processing only a portion of that utterance, reducing the likelihood that any given word within the utterance will be learned.

The finding of a 2-syllable capacity limit may appear to be somewhat surprising given that infants of the same age or younger have been shown to learn statistical patterns in 3-syllable sequences (e.g., Saffran et al., 1996). However, these findings do not necessarily conflict. First, the finding that infants can succeed with 2, but not 4, syllables does not indicate whether they would succeed with the intervening number of 3 syllables. Moreover, success in many statistical learning studies, including Saffran et al. (1996), depends on transitional probabilities, which can be acquired from pairs of syllables only. That is, whereas the "words" presented were trisyllabic, infants needed only to remember the relationships between pairs of syllables to succeed.

However, there have been studies that required infants of this age to remember longer sequences that cannot be reduced to transitional probabilities (Gerken, 2004; Marcus et al., 1999). Some have tested older infants, who might be expected to have slightly longer capacities, or have included strings of varying lengths, making it unclear whether infants could have learned the pattern from only the shorter presentations (for some examples, see Gómez & Gerken, 1999). But Gerken (2004) found that 9-month-old infants could learn stress patterns from 3- to 5-syllable words, which could be difficult to reconcile with the syllable capacity limit found here. On the other hand, Santelmann and Jusczyk (1998) reported that 18-month-old infants recognized discontinuous dependencies between words when they had only 1 to 3 syllables intervening, but not 4 or 5 syllables intervening, in keeping with the current findings.

But perhaps more critically, whereas we refer here to capacity in terms of the number of syllables (2 but not 4), R-S&N (2015) suggested that the true bottleneck might be best measured in terms of *duration* rather than *set size*. By such accounting, we have shown that infants can easily remember 700 ms of speech information but do not show memory for 1400 ms worth of information. Most statistical learning studies with young infants have used syllables that were presented at an extremely rapid rate. For example, the stimuli in Saffran et al. (1996) were presented at 270 syllables per minute, or an average of 222 ms per syllable. Even 3 syllables at that rate (666 ms) would fall well within the capacity limitations shown here. Gerken (2004) did not report a syllable rate, but based on the stimulus duration information (90-s list including 59 words with 500-ms intervening breaks), words averaged approximately 1 s each, which could fall within this durational limit. Unfortunately, it is difficult to separate durational limits from set-size limits in verbal STM without also varying other aspects of the stimuli (e.g., speaking rate, syllable complexity, phonological neighborhood properties), and work with adults suggests that both duration and set-size factors may have an impact on capacity (Cowan et al., 1997).

The limited capacity we found here does not match that found in the adult literature. Moreover, as noted earlier, adult STM studies have found different capacity limitations in verbal tasks than in nonverbal tasks, whereas the current findings show similar limits in infants. One possibility is that verbal STM and nonverbal STM rely on different systems, which just happen to have similar capacities early on but develop differently over time as a result of experience or maturation. The similarity in capacity early in development may be a coincidence rather than an indication of a shared store.

Another possibility has to do with how listeners "chunk" information into units. Work with adults suggests that extended experience with a particular type of stimulus can lead to advanced chunking mechanisms that greatly extend the apparent STM (Chase & Ericsson, 1982). This is an important means of surpassing what would otherwise be fundamental limitations of our capacity. Young infants are born with limited auditory experience and have likely not yet learned such strategies. As they gain more experience, they may develop such heuristics, which would lead to larger apparent capacity during adulthood and to differential capacity limitations for stimuli that better match this prior experience. Given humans' extensive language experience, it is reasonable that verbal and nonverbal capacity appears to be different in adults but not in children. Testing children across a range of ages would allow us to map such strategy development. Such results would also have important implications for theories of linguistic development, particularly in the domain of statistical learning; as children learn to chunk larger groups of speech input, they could presumably track longer-distance dependencies (Elman, 1993; Santelmann & Jusczyk, 1998).

One question is the extent to which the current stimulus sequences were treated as independent syllables as compared with holistic words (e.g., were infants remembering "shay" followed by "cho" or learning a new word "shaycho"). The fact that sequences were bounded by silence ("shay-cho" <pause> "shay-cho" <pause>) likely made them cohere as utterances, and this type of coherence has been generally treated as making an item a "word" (as in the statistical learning literature; see Saffran et al., 1996). Thus, infants may have been remembering bisyllabic words rather than sequences of two distinct syllables. Studies with adults have demonstrated that they show serial position effects for syllables within a novel nonword, much like they do for lists of digits, and this has led some to argue that a novel word, even for adults, is "literally processed like a list (i.e., a sequence of sounds) when it is first encountered" (Gupta et al., 2005, p. 142). If so, what differentiates a list of syllables from a word may simply be whether it has been previously stored in long-term memory; there may be little or no functional difference between treating items as a series of syllables versus a holistic word.

Conclusions

The current study examined whether linguistic stimuli have different capacity limitations than nonlinguistic stimuli for young infants; such could have been the case because of a specialized phonological/verbal memory store, via more sophisticated chunking mechanisms that influence what constitutes a "unit" for linguistic stimuli, or due to differences in neural circuitry. Our findings mirror patterns in a prior study with nonlinguistic stimuli (R-S&N, 2015). Although a similar capacity size does not necessitate having a shared STM system, it does suggest that for young infants the STM system used for speech might not differ from that used in other auditory tasks. The separability of systems found in adults, then, may be the result of extended maturational or experiential change or may be the result of changes in chunking. Future work could examine when during development this dissociation occurs.

Acknowledgments

This project represents the honors thesis of the second author; we thank her committee member Art Popper, the 2010 University of Maryland Senior Summer Scholars Award, and the Howard Hughes Medical Institute Undergraduate Research Fellowship. We thank Lloyd Frei, Keith Miller, Don Mathis, Jim Sawusch, and Jim Garmon for design and implementation of the test booth and software, and we thank the following students for assistance in scheduling and testing infants: Candance Ali, Amelie Bail, Catherine Bender, Jennifer Coon, Alyssa Cooke, Justine Dombroski, Sarah Edelberg, Daniel Eisenberg, Lauren Evans, Lauren Fischer, Whitney Goodrich-Smith, Sean Hendricks, Laura Horowitz, Megan Janssen, Mina Javid, Esther Kim, Rachel Lieberman, Eileen McLaughlin, Kelly McPherson, Giovanna Morini, Courtenay O'Connor, Sabrina Panza, Amanda Pasquarella, Elise Perkins, Rachel Rhodes, Allie Rodriguez, Maria Rodriguez, Judith Segal, Mara Steinberg, Lisa Tuit, Chelsea Vogel, Amanda Wildman, and Cavena Williams. Thanks also go to Shannon Ross-Sheehy and Amritha Mallikarjun for helpful discussions. This work was supported by a National Institutes of Health (NIH) grant (R01HD081127) and a National Science Foundation (NSF) grant (BCS0642294) to the University of Maryland.

Data availability

Data will be made available on request.

References

Baddeley AD, Gathercole SE, & Papagno C (1998). The phonological loop as a language learning device. Psychological Review, 105(1), 158–173. [PubMed: 9450375]

Baddeley AD, & Hitch GJ (1974). Working memory. In Bower G (Ed.). Recent advances in learning and motivation (Vol. 8, pp. 47–89). Academic Press.

Brent MR, & Siskind JM (2001). The role of exposure to isolated words in early vocabulary development. Cognition, 81, B33–B44. [PubMed: 11376642]

Chase WG, & Ericsson KA (1982). Skill and working memory. Psychology of Learning and Motivation, 16, 1–58.

Cowan N, Wood NL, Nugent LD, & Treisman M (1997). There are two word-length effects in verbal short-term memory: Opposed effects of duration and complexity. Psychological Science, 8(4), 290–295.

Elman JL (1993). Learning and development in neural networks: The importance of starting small. Cognition, 48(1), 71–99. [PubMed: 8403835]

Fernald A, & Hurtado N (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. Developmental Science, 9(3), F33–F40. [PubMed: 16669790]

Gathercole SE, & Baddeley AD (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. Journal of Memory & Language, 28, 200–213.

Gathercole SE, Hitch GJ, Service E, & Martin AJ (1997). Phonological short-term memory and new word learning in children. Developmental Psychology, 33(6), 966–979. [PubMed: 9383619]

Gerken LA (2004). Nine-month-olds extract structural principles required for natural language. Cognition, 93, B89–B96. [PubMed: 15178379]

Gómez RL, & Gerken L (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. Cognition, 70, 109–135. [PubMed: 10349760]

Gupta P, Lipinski J, Abbs B, & Lin P-H (2005). Serial position effects in nonword repetition. Journal of Memory and Language, 53, 141–162.

Hoff E, Core C, & Bridges K (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. Journal of Child Language, 35, 903–916. [PubMed: 18838017]

Kemler Nelson DG, Jusczyk PW, Mandel DR, Myers J, Turk A, & Gerken LA (1995). The headturn preference procedure for testing auditory perception. Infant Behavior and Development, 18, 111–116.

Li D, Cowan N, & Saults JS (2013). Estimating working memory capacity for lists of nonverbal sounds. Attention, Perception, & Psychophysics, 75, 145–160.

Marcus GF, Vijayan S, Bandi Rao S, & Vishton PM (1999). Rule learning by seven-month-old infants. Science, 283, 77–80. [PubMed: 9872745]

Newport EL (1990). Maturational constraints on language learning. Cognitive Science, 14, 11–28.

Ross-Sheehy S, & Newman RS (2015). Infant auditory short-term memory for non-linguistic sounds. Journal of Experimental Child Psychology, 132, 51–64. [PubMed: 25590900]

Saffran JR, Aslin RN, & Newport EL (1996). Statistical learning by 8-month-old infants. Science, 274(5294), 1926–1928. [PubMed: 8943209]

Santelmann LM, & Jusczyk PW (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. Cognition, 69(2), 105–134. [PubMed: 9894402]

Schulze K, & Tillmann B (2013). Working memory for pitch, timbre, and words. Memory, 21(3), 377–395. [PubMed: 23116413]

Stevens AA (2004). Dissociating the cortical basis of memory for voices, words and timbres. Cognitive Brain Research, 18, 162–171. [PubMed: 14736575]

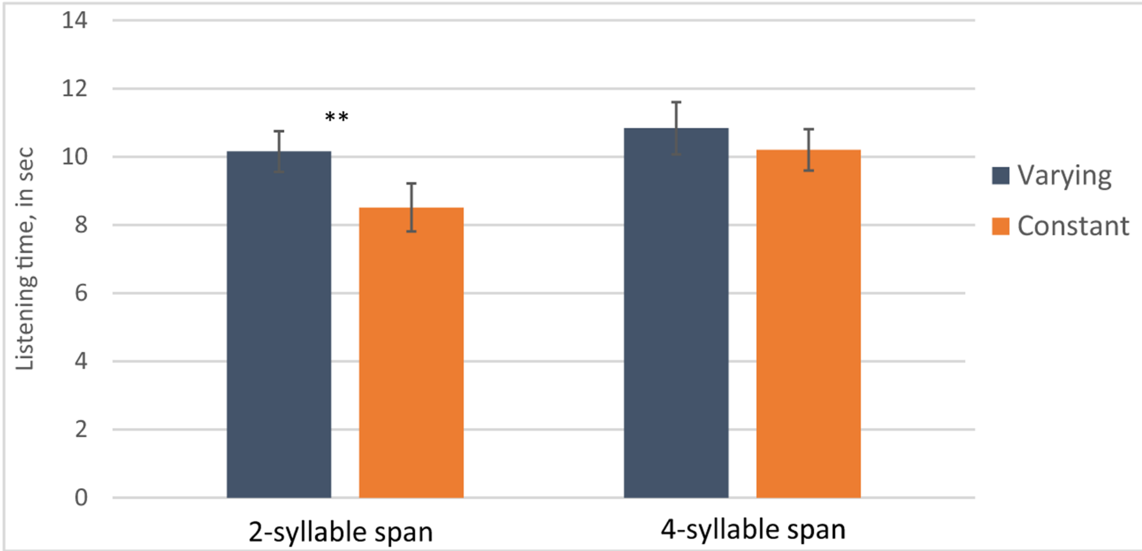

Fig. 1.

Listening time to repeated sequences of either 2 or 4 syllables that either repeat identically (constant) or change (varying). Error bars show standard errors. $^{**}p < .01$.

Table 1

Listening times and planned comparisons.

Stimulus type		Listening time in seconds (*SD*)	Planned comparison
2-Syllable			
	Constant	8.52 (3.99)	$t(31) = 1.22$, $p > .20$ Cohen's $d = 0.162$
	Varying	10.16 (3.37)	
4-Syllable			
	Constant	10.20 (3.43)	$t(31) = 1.22$, $p > .20$ Cohen's $d = 0.162$
	Varying	10.84 (4.31)	