

# Monkeys have a limited form of short-term memory in audition

Brian H. Scott<sup>a</sup>, Mortimer Mishkin<sup>a,1</sup>, and Pingbo Yin<sup>a,b,1</sup>

<sup>a</sup>Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892; and <sup>b</sup>Neural Systems Laboratory, Institute for Systems Research, University of Maryland, College Park, MD 20742

Contributed by Mortimer Mishkin, June 7, 2012 (sent for review January 29, 2012)

A stimulus trace may be temporarily retained either actively [i.e., in working memory (WM)] or by the weaker mnemonic process we will call passive short-term memory, in which a given stimulus trace is highly susceptible to “overwriting” by a subsequent stimulus. It has been suggested that WM is the more robust process because it exploits long-term memory (i.e., a current stimulus activates a stored representation of that stimulus, which can then be actively maintained). Recent studies have suggested that monkeys may be unable to store acoustic signals in long-term memory, raising the possibility that they may therefore also lack auditory WM. To explore this possibility, we tested rhesus monkeys on a serial delayed match-to-sample (DMS) task using a small set of sounds presented with ~1-s interstimulus delays. Performance was accurate whenever a match or a nonmatch stimulus followed the sample directly, but it fell precipitously if a single nonmatch stimulus intervened between sample and match. The steep drop in accuracy was found to be due not to passive decay of the sample’s trace, but to retroactive interference from the intervening nonmatch stimulus. This “overwriting” effect was far greater than that observed previously in serial DMS with visual stimuli. The results, which accord with the notion that WM relies on long-term memory, indicate that monkeys perform serial DMS in audition remarkably poorly and that whatever success they had on this task depended largely, if not entirely, on the retention of stimulus traces in the passive form of short-term memory.

macaque | primate | vocalization

**W**orking memory (WM) is a system that enables the temporary maintenance and manipulation of information necessary to guide behavior (1, 2). The term “working memory” has sometimes been applied to parametric sensory discriminations (3) [e.g., comparing the acoustic frequency of two successive tones, or the visual contrast of two successive images, separated by a short interstimulus interval (ISI)]. However, in the absence of the need for maintaining and manipulating the stimuli, such discriminations may be more properly described as tests of a type of short-term memory (STM) that we will call *passive* short-term memory (pSTM) rather than WM.

Definitions and models of WM vary (4), but the concepts of STM (particularly pSTM) and WM differ along a dimension of increasing attention to the stimulus item and greater reliance on its stored representation. Indeed, WM has been posited to differ from other forms of STM by operating not on a recently presented item, per se, but on the activation of a representation of that item stored in long-term memory (LTM) (4–7). This distinction is related to another, viz. the distinction between categorical perception and continuous, noncategorical perception, the former term implying that perception of some stimuli activates their previously stored representations sorted into categories on the basis of either their physical similarity or some more abstract factor. The capacity of WM in vision has been estimated at four to seven items (2, 4), but if the stimuli cannot be distinguished categorically the capacity is much smaller, perhaps as small as a single item (8, 9). The capacity of WM thus reflects processes beyond a passively retained sensory trace (10).

A closely related benefit of a system that actively maintains stimuli in memory by activating their stored representations is greater attentional control (1). Specifically, the stored representation of an item can be reactivated continuously or repeatedly (i.e., rehearsed), and thus the maintained memory of the item can be more readily protected not only from passive, temporal decay but also from the retroactive interference produced by incoming stimuli (10, 11). In fact, increased resistance to interference may well be the basis for any WM capacity greater than one.

The results of auditory studies in dogs and monkeys (12, 13) have raised the possibility that these animals are unable to store the representations of acoustic stimuli in LTM. However, these animals do have the ability to retain acoustic stimuli within the period of WM (i.e., on the order of tens of seconds), and in fact, it was proposed that this duration of retention in audition was served by WM. However, if WM, itself, is dependent on LTM, and if the animals do not have LTM, then their short-term retention cannot be attributed to WM. For the purpose of this study, we will posit reliance on stored representations in LTM\* as a heuristic definition of WM, as distinct from a passively retained sensory trace, or pSTM. In this regard, pSTM may be equivalent to the “long auditory store” described by Cowan (5, 17), which persists on the order of 10–30 s.

In an attempt to determine whether WM and categorical perception play any role in the auditory memory of the monkey, we tested rhesus monkeys on delayed matching-to-sample (DMS; Fig. 1) using a small set of acoustic items representing several different putative auditory categories (pure tones, environmental sounds, monkey calls, etc.; Fig. S1). Because of the known difficulty monkeys have in acquiring and applying the rule for auditory DMS (12, 18–22), we simplified task requirements by requiring the animal to (i) remember a single sample stimulus for 1 s, and (ii) respond at test to an identical match stimulus and withhold responding to a nonmatch that, importantly, always belonged to a stimulus category other than the sample’s category. The only

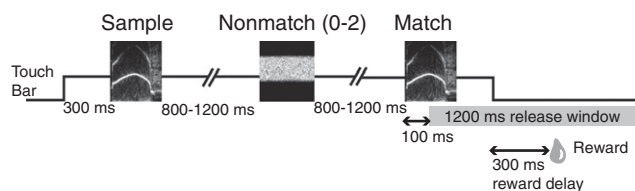
Author contributions: M.M. and P.Y. designed research; B.H.S. and P.Y. performed research; B.H.S. and P.Y. analyzed data; and B.H.S. and M.M. wrote the paper.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence may be addressed. E-mail: pyin@umd.edu or mishkinm@mail.nih.gov.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1209685109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1209685109/-DCSupplemental).

\*By our definition, the LTM store on which auditory WM depends must allow for reactivation of previously experienced sounds to mediate an identity comparison (i.e., recognition memory). Thus, according to the definition, sound–response learning (habit formation) and sound–picture learning (cross-modal association) demonstrate other long-term retention abilities, not auditory LTM. The reverse form of cross-modal association, viz. picture–sound learning, would fall within our definition of auditory LTM, because the stored associate triggered by the object would be the neural representation of the sound and would therefore be expected to be located in the cortical auditory system. However, to our knowledge, picture–sound association has not been demonstrated in the monkey; only sound–picture association has been shown (14, 15). In the latter case, the stored associate triggered by the sound would be the representation of the picture located in the cortical visual system, a supposition supported by the finding that auditory interference during an interstimulus delay period of a cross-modal association task had little effect on performance, whereas visual interference during the delay disrupted performance severely (16).



**Fig. 1.** Schematic diagram of the timing of a DMS trial. The animal initiated a trial by holding a contact bar for 300 ms. A sample stimulus (~300 ms in duration) was presented, followed by one to three test sounds with a varied ISI of 800–1200 ms. When the test sound was the same as the sample (a match), the animal was required to release the bar within a 1,200-ms response window beginning 100 ms after match onset. A correct response (a “hit”) earned a liquid reward 300 ms after bar release. A response within the first 100 ms following match onset was considered an early release error. Failure to release by the end of the response window was counted as a “miss” error. If the stimulus following the sample was a nonmatch, the animal was required to hold the bar (a “correct rejection”) until the match stimulus was presented. Release to the nonmatch was counted as a FA error. Any type of error aborted the trial and was penalized by a 3-s time out in addition to the standard 3-s intertrial interval, to discourage animals from aborting trials with multiple nonmatches. Each trial ended after release of the bar, but if the bar was released during stimulus presentation, the full stimulus played out before the trial was reset. Trials with zero, one, or two nonmatch sounds were randomly generated with equal probability. In an attempt to aid the animal’s performance, the task was designed such that the nonmatch stimuli, which were selected pseudorandomly on each trial, did not belong to the same stimulus category as the sample. Trials were organized in blocks such that each stimulus in the set served as the sample in a pseudorandom order before the same stimulus was used again.

task complication was that if the animal correctly withheld responding to the nonmatch, a second test stimulus—either a match or a different nonmatch—was presented 1 s later, thereby requiring that the monkey maintain the sample in memory despite the potential interference from one or two intervening nonmatch stimuli. However, the monkey could readily overcome this complication simply by matching to category or, if the experimenter’s categories differed from its own, by activating a stored representation of each sound.

Successful performance on our DMS task would suggest that monkeys have an auditory system capable of supporting WM, which is to say they have categorical auditory perception (i.e., stored representations of auditory stimuli sorted into categories) or at least stored but unsorted representations that can be activated by presentation of the DMS stimuli. Conversely, failure to master the DMS task would suggest that monkeys do not have auditory WM and, consequently, are limited mnemonically to auditory pSTM.

## Results

**DMS Performance.** After pretraining two monkeys to release a touch-bar at the immediate repetition of a sound, they were trained on the DMS rule (release to same, hold to different) until they attained a stable level of performance significantly above chance, a stage that required ~200 daily sessions (details in *SI Methods, Training and Acquisition of the DMS Rule* and Fig. S2). Performance data on the final version of the DMS task were then collected across many additional daily sessions (monkey F: 360 sessions, >250,000 trials; monkey S: 116 sessions, >82,000 trials).

The two monkeys performed similarly. Averaged across all three trial types (zero, one, or two nonmatches), both animals performed at a mean of 67% correct (Fig. 2A), but there was a strong effect of trial type: scores were high on trials with no nonmatch stimulus (93% and 89% correct for monkeys F and S, respectively) but dropped steeply when nonmatch stimuli were included in the sequence (73% for each animal with one nonmatch; 38% and 40%, respectively, with two nonmatches). To

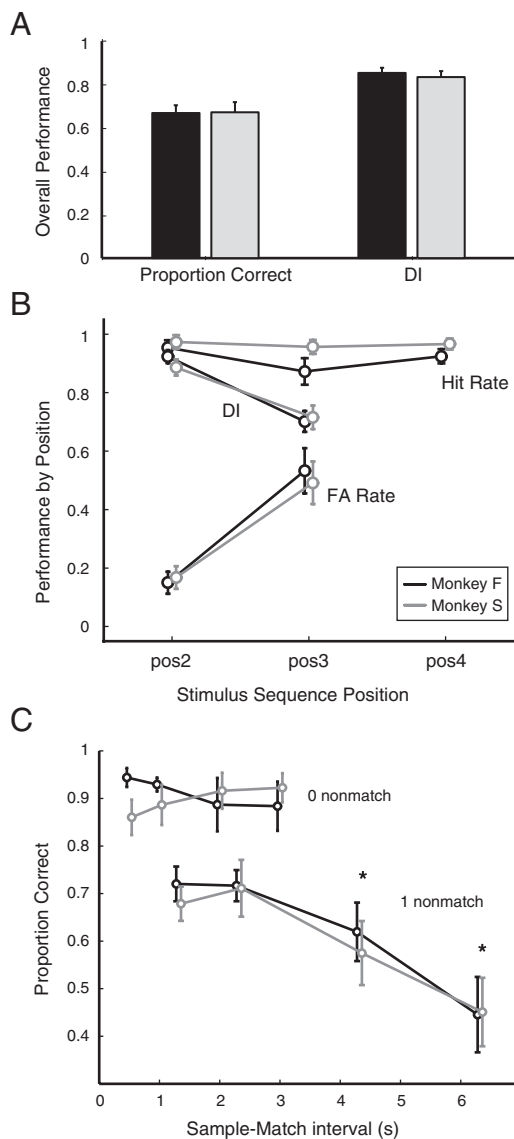
clarify this trend in performance, the hit rate, false alarm (FA) rate, and Discrimination Index (DI) were calculated for each position in the task (Fig. 2B). DI is a measure based on signal detection theory, and it takes a value between 1 for perfect performance and 0.5 at chance (*SI Methods* and Fig. S3). The hit rate was uniformly high at all stimulus positions, indicating that the monkeys rarely made “miss” errors ( $\leq 5\%$  of all trials in each animal). By contrast, the FA rate increased from ~0.15 at stimulus position 2 to ~0.5 at stimulus position 3, and the DI decreased from ~0.9 to 0.7. The cumulative effect of this tendency toward FA errors was a very low success rate for the two-nonmatch trial type. However, performance at stimulus position 3 was still significantly better than chance: DI was >0.7 in both animals, compared with the threshold (*SI Methods*) of 0.56. All three metrics (hit rate, FA rate, and DI) were subjected to an ANOVA in each animal, and the results confirmed the strong effect of sequence position (all  $F$  values >84.2, all  $P$  values <  $10^{-4}$  for both animals).

**Delay Duration.** Because the number of nonmatch stimuli is confounded with the elapsed time between sample and match presentation in a sequential DMS task, the decline in performance across stimulus position could have been attributable to interference by the nonmatch stimulus, to a decay in memory of the sample, or to both. To determine which was the case, we varied ISI in a control task, using either zero nonmatch stimuli or one. Within each session, ISI was fixed at 0.5, 1, 2, or 3 s. As shown in Fig. 2C, there was no significant decrement in performance as the sample–match interval increased from 0.5 to 3 s for trials without a nonmatch stimulus, but performance in trials with one nonmatch was significantly lower at all ISIs. This strongly suggests that the decline in performance at later stimulus positions on the standard testing schedule was likely due to interference from the nonmatch sound rather than to decay of the memory for the sample over the short delay periods used here.

Whereas performance was unaffected by an increase in delay duration on trials with zero nonmatches, performance at the longer delays declined sharply on trials with one nonmatch [Fig. 2C; one-way ANOVA on proportion correct, monkey F:  $F(3,36) = 53.1, P < 10^{-4}$ ; monkey S:  $F(3,39) = 40.7, P < 10^{-4}$ ; accounting for multiple comparisons,  $1.3 = 2.3 < 4.3 < 6.3$ ]. The performance decrement at longer delays in the presence of a nonmatch stimulus suggests an interaction between delay duration and the interference effects of the nonmatch stimulus. However, for durations up to 2 to 3 s (the longest at which a direct comparison can be made), the limiting factor was clearly the presence of a nonmatch stimulus, not delay duration.

**Sound Category.** The 21 stimuli used for the DMS task were drawn from seven diverse sound categories (Fig. S1), with the expectation that monkeys would make use of these categories in detecting a match. Naturalistic stimuli, particularly conspecific monkey vocalizations (Mvocs), could have had a privileged representation owing to their ethologic significance, whereas tones and noise would not. Surprisingly, however, not only was there no advantage for Mvoc over other sample categories, there was a counterintuitive trend toward better performance for temporally simple synthetic stimuli [band-pass noise (BPN), pure tones (PTs), and frequency-modulated tone sweeps (FMs)] over temporally complex stimuli, including Mvocs (Fig. 3). Performance (DI) varied as a function of category and between animals, with a significant interaction effect [two-way ANOVA, sound category,  $F(6,3318) = 86.9, P < 10^{-4}$ ; animal,  $F(1,3318) = 85.7, P < 10^{-4}$ ; category  $\times$  animal,  $F(6,3318) = 32.6, P < 10^{-4}$ ].

Performance for modulated noise (temporally orthogonal ripple complexes, or TORCs) was clearly different between animals, being among the best stimuli for monkey S, but the worst for monkey F. A one-way ANOVA on the remaining six categories, accounting for multiple comparisons, distinguished two distinct groups:



**Fig. 2.** (A) Overall performance (mean  $\pm$ SD) as measured by proportion of correct trials (Left) and DI (Right) for all standard testing sessions (black bars, monkey F,  $n = 360$  sessions,  $>250,000$  trials; gray bars, monkey S,  $n = 116$  sessions,  $>82,000$  trials). (B) Hit rate, FA rate, and DI computed separately at each position within the trial sequence for monkey F (black) and monkey S (gray). FA rate and DI are not computed for the fourth stimulus, because it is always a match, and so no FA can occur (details in *SI Methods, Analysis of Behavioral Performance* and Fig. S3). (C) Control task 1 with variable ISIs between sample offset and match onset. Performance is plotted as a function of the total sample-match interval (e.g., a one-nonmatch trial at an ISI of 0.5 s would include two 0.5-s delay intervals plus the 0.3-s duration of the intervening nonmatch stimulus, for a total of 1.3 s) and is shown separately for trials with (i) no nonmatch stimulus (upper curves) and (ii) one intervening nonmatch stimulus (lower curves). In the absence of a nonmatch stimulus, DMS performance in both animals is stable up to 3 s. With a single intervening nonmatch stimulus, scores at 1- and 2-s intervals are significantly below those without an intervening nonmatch, and there is a strong effect of delays  $\geq 4$  s. Asterisks mark scores that are significantly different from the scores at 1 and 2 s (which are equivalent) according to ANOVA followed by multiple comparisons. Performance between trial types was compared by proportion correct, because DI is not applicable to trials with no nonmatch stimuli (and thus no possible FA response). In this control task, the ISI (0.5, 1, 2, or 3 s) was fixed within daily sessions and randomly interleaved across 40 sessions, 10 at each ISI. Median number of trials per ISI was 919 for monkey F (black) and 727 for monkey S (gray). Each point on the curves for one intervening nonmatch is offset from a whole-number delay by 0.3 s to account for the added delay due to the presentation of that 0.3-s nonmatch stimulus.

performance for BPN, PT, and FM was equivalent, and significantly better than performance for Mvoc, other vocalizations, and environmental noise. We will refer to these sets of categories as “temporally simple” and “temporally complex,” respectively. Performance on the three stimuli within a category was generally consistent, and hierarchical clustering of performance for all 21 stimuli verified the distinction between temporally simple and temporally complex sounds (*SI Results, Performance by Stimulus* and Fig. S4).

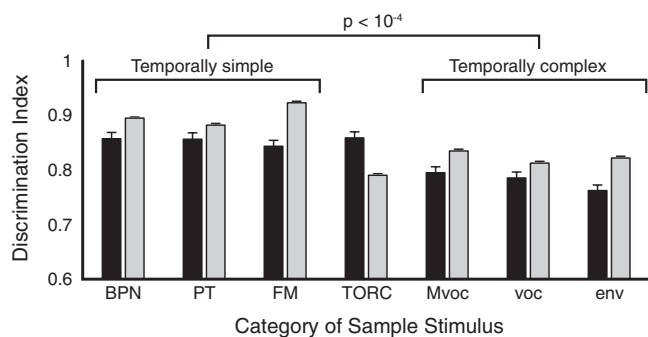
As noted earlier, the DMS task design of always presenting sample and nonmatch stimuli from different categories was intended to aid the animals both in acquiring the DMS rule and in applying it. In a second control experiment ( $n = 29$  sessions for each animal), that category restriction was lifted, but overall performance [DI (mean  $\pm$  SD)] remained the same:  $0.84 \pm 0.01$  and  $0.84 \pm 0.01$  for with and without category restriction, respectively, for monkey F; and  $0.85 \pm 0.02$  and  $0.84 \pm 0.02$  for with and without category restriction, respectively, for monkey S. Because the probability of the sample and nonmatch being of the same category was still only 10% (i.e., after a sample is selected, only 2 of 20 remaining stimuli are of the same category), there are relatively few trials on which such pairings occurred ( $<300$  for each animal). Performance on these within-category trials using Mvocs did not differ from that on within-category trials using other categories (ANOVA on DI across sessions, controlling for multiple comparisons). In particular, performance in the DMS task was no better on Mvocs relative to other stimuli, whether a Mvoc served as sample, nonmatch, or both.

## Discussion

We have demonstrated that, although monkeys can perform a serial DMS task with a small set of auditory stimuli, their accuracy quickly degrades across serial position despite an ISI of only 1 s. This rapid decay of performance is not due simply to a decay of the memory over time: presentation of an intervening nonmatching sound after the sample sound is a far more important factor (Fig. 2C). The results indicate that, under our task conditions, auditory memory in nonhuman primates is surprisingly poor, perhaps limited to a single item, largely because of retroactive interference with the sample sound by subsequently presented items. This retroactive interference effect in audition seems to be extremely powerful. Prior studies of auditory memory in monkeys tested with nonserial DMS have shown forgetting thresholds (75% correct responses) of up to 30 s in the absence of intervening stimuli, whether the studies used trial-unique stimuli (12) or a set of only two tones (20).

**Effect of Small Stimulus Sets.** Compared with the present results on serial DMS in audition, the performance of monkeys on serial DMS in vision was far more robust in the face of intervening nonmatching stimuli. Scores on the original, visual version of the task declined from  $\sim 98\%$  correct with no intervening nonmatching stimulus to  $\sim 82\%$  correct after three such intervening stimuli (23). By contrast, performance on our auditory task declined from  $\sim 90\%$  correct with no intervening stimulus to  $\sim 40\%$  correct after just two intervening nonmatch stimuli. Importantly, the visual study also used a restricted stimulus set (six images within each session), and data were not collected until the animals were highly familiar with the stimulus set for that session. Serial DMS and a small stimulus set combined to produce high rates of FAs for two different response/reward-related reasons: (i) only an active response (bar release in both studies) was rewarded, and (ii) a nonmatch stimulus tended to elicit an active response if an active response to that same stimulus was recently rewarded (this cross-trial interference effect is described in *SI Results* and Fig. S5). These two response/reward-related effects are greatly exaggerated in serial auditory compared with serial visual DMS owing, presumably, to the greater susceptibility of acoustic sample stimuli to





**Fig. 3.** Performance across sample categories (DI, mean + SE across sessions) for monkeys S and F (black and gray bars, respectively). Performance varied significantly as a function of category and animal, with a significant interaction effect (two-way ANOVA, all effects  $P < 10^{-4}$ ). Sound categories are BPN, PT, FM, TORC, Mvoc, other species' vocalizations (voc), and environmental sounds (env). Categories have been sorted left to right by average performance across the two animals. Multiple-comparisons ANOVA identified performance for BPN, PT, and FM (grouped by the left bracket) as statistically equivalent, and better than performance for all other categories (grouped by the right bracket). We have distinguished these groups as "temporally simple" and "temporally complex." TORCs were excluded from the one-way ANOVA because performance for this category clearly differed between subjects: TORCs were remembered well by monkey S but poorly by monkey F. The entire set of stimuli is illustrated in Fig. S1.

retroactive interference. A possible explanation for this difference in outcome in the two modalities is considered below (*Discussion, Role of Sound Category*).

An earlier study reported a less powerful effect of retroactive interference in monkeys, even at delays approaching the limit of auditory STM [28 s (18)]. In that study, however, the interfering sounds (monkey calls or music) filled the delay intervals instead of serving as potential targets that required the monkey to listen attentively to determine whether to respond for a reward. Rather, the monkeys had been trained to ignore these nontest sounds, because they were irrelevant. This may be the reason that, although the delay-interval, nontest sounds produced some retroactive interference (a drop from control levels of approximately 15%), they did not produce nearly as much retroactive interference as our nonmatch test sounds did (up to 50% with two nonmatch sounds). Although this account provides a plausible explanation of the discrepancy, the effects of the two different types of interference need to be compared directly in future experiments.

Comparison of auditory memory in humans with that in other animals is complicated by the human ability to tag most sounds with verbal labels. Pitch, however, is a continuous sound quality, such that a particular pitch cannot easily be labeled or categorized, so comparison between PTs must depend on memory for this acoustic feature alone. This pure sensory memory is independent of auditory verbal memory in humans (24) and thus is tractable to study in an animal model. Macaques have been trained to make frequency discriminations across a delay, and their performance is stable up to ISIs of at least 1 s (25, 26). In human listeners, memory for pitch is also subject to rapid decay over time (27) and to interference from intervening stimuli (24, 28, 29). When the reference tone is varied from trial to trial, thresholds for distinguishing changes in tone frequency are stable for intertone intervals up to 1 s, and steadily worsen for intervals  $\geq 3$  s (27). Our monkeys' performance on trials with a nonmatch stimulus also began to drop at delays  $> 3$  s, suggesting that the temporal dynamics of their sensory memory are similar to those of humans.

**Role of Sound Category.** A curious aspect of the preceding result is that the animals did not appear to exploit sound category in performing the DMS task. Because nonmatch stimuli were always

selected to be of a different category from the sample, matching to category alone would have been sufficient for perfect performance. However, these categories were defined by the experimenter, and although some may not have had any a priori significance for the animal (e.g., PT, BPN, FM), confusion of conspecific vocalizations with environmental sounds or modulated noise was surprisingly common. In the context of our task, it seems that vocalizations were not retained with any more accuracy than stimuli of lesser ethological significance.

A recent study of *nonserial auditory DMS* in rhesus monkeys by Ng et al. (30) reported better performance and faster reaction times for conspecific vocalizations relative to other sound categories (similar to the categories used here) at delays of 5 s. In contrast, as shown in Fig. 3, our results indicate slightly better performance for tonal stimuli compared with all others, including monkey vocalizations. There are several methodological differences between the two studies, but the most likely source of the difference in category effect is the stimulus set size. Whereas the stimuli used in the nonserial DMS task were nearly trial-unique, our stimuli were repeated many times as both sample and nonmatch stimuli within each session, thus requiring contradictory responses across trials (Fig. S5). When Ng and colleagues used a restricted set of eight stimuli (one from each of their categories), their monkeys too performed slightly better for PTs than for all other categories, including vocalizations (31). Apparently, interference among items in a small stimulus set influences the monkeys' strategy, in this case favoring the retention in memory of simple stimuli compared with spectrotemporally richer ones.

Within both audition and vision, the capacity of working memory has been estimated using stimulus items that are categorically different (e.g., spoken words or visual objects), whose representations have been stored in LTM. Contrary to those earlier estimates, Olsson and Poom (8) estimated the capacity of short-term visual memory to be only a single item when those items cannot be separated by category. When challenged with stimuli that are equivalently difficult to discriminate and label, visual STM is no better than auditory (9). If monkeys do not bring the concept of "category" to bear in serial auditory DMS, perhaps their poor performance is not surprising.

**Implications Regarding Underlying Neural Mechanisms.** Temporally simple synthetic stimuli (BPN, PT, and FM) were the more accurately discriminated items (i.e., error rate was relatively low when they served as sample) but also led to high miss rates subsequent to their presentation as a nonmatch. Apparently, a temporally simple stimulus, which can be described using a single parameter (the frequency of a tone, or center frequency of a BPN, or a regularly changing tone frequency), is not only comparatively resistant to interference from subsequent distracter stimuli but is itself a potent distracter. This salience may be related to the physiological representation of these stimuli. In a tonotopically organized brain structure, such as the auditory cortex, a tone or BPN will activate a discrete population of neurons throughout its duration (32, 33). By contrast, a more complex stimulus such as a vocalization will likely evoke a distributed and asynchronous activation across the auditory cortex. If a subsequent test stimulus is also complex, it may well engage some or all of this same population of neurons, albeit in a different temporal order. Thus, the comparison of serially presented complex stimuli, whose activation patterns overlap, requires comparison of population responses through time. By contrast, discrimination of simple static stimuli can be accomplished by a place code, without regard to activation through time. In short, in the absence of categorical perception in audition, the tonotopic organization of the system may favor parametric comparisons of frequency.

## Conclusions

Long-term auditory memory in nonhuman primates has yet to be demonstrated in the laboratory. Fritz et al. (12) measured a decay or forgetting threshold of ~30 s for trial-unique sounds, far shorter than the monkey's forgetting threshold for trial-unique pictures [15–20 min (12, 34)]. As indicated at the outset, retention for 30 s is considered to be within the capacity of WM, and it was originally proposed that the auditory memory the monkeys had displayed was mediated by WM. However, the results of the present experiment—particularly the evidence that the monkeys (*i*) failed to make use of auditory categories, (*ii*) were highly susceptible to retroactive interference, and (*iii*) could retain in memory only a single item—all weigh against the notion that monkeys possess a mechanism for auditory WM. The present results suggest instead that, in performing an auditory DMS task, monkeys rely on a passively retained sensory trace in STM, which we have labeled pSTM. Further, if WM does indeed depend on reactivation of a stimulus representation stored in LTM, then the absence of auditory WM could well be due to the monkey's inability to achieve long-term storage in the auditory modality.

Although the above interpretation provides a cohesive account of our findings, with the apparent absence of auditory WM in the present study and of auditory LTM in the earlier one (12) reinforcing each other, there are other possible explanations of the results that need to be considered. For example, monkeys might actually have both forms of auditory memory, but their WM system may not be capable of maintaining LTM traces in an active state in the face of interference. Or perhaps the monkeys can maintain long-term traces in an active state but are unable to determine which long-term trace—that of the sample or that of the nonmatch—is the target of the reactivation. Although neither possibility has any independent support, they (and related

variants) cannot logically be ruled out, and so our current interpretation must be considered a tentative one.

Whatever the final outcome, however, the present results indicate that the monkey's strikingly poorer mnemonic ability in audition than in vision observed earlier in LTM extends equally to working memory. This sharp mnemonic difference between sensory modalities calls out for further study.

## Methods

All procedures were conducted in accordance with the National Institutes of Health's *Guide for the Care and Use of Laboratory Animals* and were approved by the Animal Care and Use Committee of the National Institute of Mental Health. Two adult male rhesus monkeys (*Macaca mulatta*) were trained on a sequential DMS task. Animals were kept on a controlled-access schedule for water. Details of the apparatus and training procedure are provided in *SI Methods*. The animal was seated in a primate chair in a soundproof booth, facing a speaker ~1 m directly in front at eye level. The task timing and trial structure are described in Fig. 1, with additional details of the apparatus and training procedure provided in *SI Methods* and Fig. S2.

The stimulus set used during testing consisted of 21 sounds (Fig. S1), including three exemplars from each of seven experimenter-defined categories. Among these categories, four were synthetic: PT, BPN, FM, and modulated noise (TORCs). The three remaining categories were all recorded natural sounds, including Mvoc, other species' vocalizations, and environmental sounds. Synthetic sounds were 300 ms in duration; natural sounds ranged from 195 to 300 ms in duration. All sounds were presented at 60–70 dB sound pressure level.

**ACKNOWLEDGMENTS.** We thank Helen Tak, Kathleen Moorhead, Peter Sergo, and Holly Vinal for assistance with animal training and data collection, and Alexander Kloth for programming assistance as well as data collection. We also thank Michael Colombo, Nelson Cowan, and Gregg Recanzone for providing valuable constructive comments on the original manuscript. This work was supported by the Division of Intramural Research Programs, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services.

- Baddeley A (2003) Working memory: Looking back and looking forward. *Nat Rev Neurosci* 4:829–839.
- Miller GA, Galanter E, Pribram KH (1960) *Plans and the Structure of Behavior* (Holt, Rinehart & Winston, New York).
- Pasternak T, Greenlee MW (2005) Working memory in primate sensory systems. *Nat Rev Neurosci* 6:97–107.
- Cowan N (2008) What are the differences between long-term, short-term, and working memory? *Prog Brain Res* 169:323–338.
- Cowan N (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol Bull* 104:163–191.
- Fuster JM (1995) *Memory in the Cerebral Cortex* (MIT Press, Cambridge, MA).
- Norman DA (1968) Toward a theory of memory and attention. *Psychol Rev* 75:522.
- Olsson H, Poom L (2005) Visual memory needs categories. *Proc Natl Acad Sci USA* 102:8776–8780.
- Visscher KM, Kaplan E, Kahana MJ, Sekuler R (2007) Auditory short-term memory behaves like visual short-term memory. *PLoS Biol* 5:e56.
- Saults JS, Cowan N (2007) A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *J Exp Psychol Gen* 136:663–684.
- Lewandowsky S, Oberauer K, Brown GD (2009) No temporal decay in verbal short-term memory. *Trends Cogn Sci* 13:120–126.
- Fritz J, Mishkin M, Saunders RC (2005) In search of an auditory engram. *Proc Natl Acad Sci USA* 102:9359–9364.
- Kowalska DM, Kuśmierk P, Kosmal A, Mishkin M (2001) Neither perirhinal/entorhinal nor hippocampal lesions impair short-term auditory recognition memory in dogs. *Neuroscience* 104:965–978.
- Colombo M, Gross CG (1994) Responses of inferior temporal cortex and hippocampal neurons during delayed matching to sample in monkeys (*Macaca fascicularis*). *Behav Neurosci* 108:443–455.
- Gaffan D, Harrison S (1991) Auditory-visual associations, hemispheric specialization and temporal-frontal interaction in the rhesus monkey. *Brain* 114:2133–2144.
- Colombo M, Graziano M (1994) Effects of auditory and visual interference on auditory-visual delayed matching to sample in monkeys (*Macaca fascicularis*). *Behav Neurosci* 108:636–639.
- Cowan N (1984) On short and long auditory stores. *Psychol Bull* 96:341–370.
- Colombo M, D'Amato MR (1986) A comparison of visual and auditory short-term memory in monkeys (*Cebus apella*). *Q J Exp Psychol B* 38:425–448.
- Colombo M, D'Amato MR, Rodman HR, Gross CG (1990) Auditory association cortex lesions impair auditory short-term memory in monkeys. *Science* 247:336–338.
- Colombo M, Rodman HR, Gross CG (1996) The effects of superior temporal cortex lesions on the processing and retention of auditory information in monkeys (*Cebus apella*). *J Neurosci* 16:4501–4517.
- D'Amato MR, Salmon DP (1984) Processing and retention of complex auditory stimuli in monkeys (*Cebus apella*). *Can J Psychol* 38:237–255.
- Stepien LS, Cordeau JP, Rasmussen T (1960) The effect of temporal lobe and hippocampal lesions on auditory and visual recent memory in monkeys. *Brain* 83:470.
- Miller EK, Li L, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J Neurosci* 13:1460–1478.
- Deutsch D (1970) Tones and numbers: Specificity of interference in immediate memory. *Science* 168:1604–1605.
- Brosch M, Oshurkova E, Bucks C, Scheich H (2006) Influence of tone duration and intertone interval on the discrimination of frequency contours in a macaque monkey. *Neurosci Lett* 406:97–101.
- Brosch M, Selezneva E, Bucks C, Scheich H (2004) Macaque monkeys discriminate pitch relationships. *Cognition* 91:259–272.
- Harris JD (1952) The decline of pitch discrimination with time. *J Exp Psychol* 43:96–99.
- Deutsch D (1972) Mapping of interactions in the pitch memory store. *Science* 175:1020–1022.
- Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.
- Ng CW, Plakke B, Poremba A (2009) Primate auditory recognition memory performance varies with sound type. *Hear Res* 256:64–74.
- Ng CW (2011) Behavioral and neural correlates of auditory encoding and memory functions in rhesus macaques. PhD thesis (Univ of Iowa, Iowa City, IA).
- Phillips DP, Semple MN, Calford MB, Kitzes LM (1994) Level-dependent representation of stimulus frequency in cat primary auditory cortex. *Exp Brain Res* 102:210–226.
- Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* 435:341–346.
- Murray EA, Mishkin M (1998) Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *J Neurosci* 18:6568–6582.