



Published in final edited form as:

Curr Biol. 2014 December 1; 24(23): 2767–2775. doi:10.1016/j.cub.2014.10.004.

Neural correlates of auditory short-term memory in rostral superior temporal cortex

Brian H. Scott^a, Mortimer Mishkin^a, and Pingbo Yin^{a,b}

^a Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892 USA

^b Neural Systems Laboratory, Institute for Systems Research, University of Maryland, College Park, MD 20742 USA

Summary

Background—Auditory short-term memory (STM) in the monkey is less robust than visual STM and may depend on a retained sensory trace, which is likely to reside in the higher-order cortical areas of the auditory ventral stream.

Results—We recorded from the rostral superior temporal cortex as monkeys performed serial auditory delayed-match-to-sample (DMS). A subset of neurons exhibited modulations of their firing rate during the delay between sounds, during the sensory response, or both. This distributed subpopulation carried a predominantly sensory signal modulated by the mnemonic context of the stimulus. Excitatory and suppressive effects on match responses were dissociable in their timing, and in their resistance to sounds intervening between the sample and match.

Conclusions—Like the monkeys' behavioral performance, these neuronal effects differ from those reported in the same species during visual DMS, suggesting different neural mechanisms for retaining dynamic sounds and static images in STM.

Introduction

Auditory perception and language depend on linking sounds through time [1, 2]. In vision and touch, short-term memory (STM) is thought to rely on the same regions of secondary sensory and association cortex that support perception [3], such as the inferotemporal (IT) visual cortex [4]. The rostral superior temporal cortex (rSTC), including the rostral supratemporal plane and superior temporal gyrus, occupies a position in the auditory processing hierarchy similar to that of IT in the visual processing hierarchy [5, 6], and may play an analogous functional role.

Corresponding Author: Brian Scott 49 Convent Drive, Room 1B80 Bethesda, MD 20892 USA brianscott@mail.nih.gov Phone: 1 (301) 443-7867 Fax: 1 (301) 402-0046.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Neurons in rSTC show long response latency and a preference for complex stimuli [7, 8]; ablation of rSTC disrupts auditory pattern discrimination and delayed-match-to-sample (DMS) performance [9, 10]; and rSTC affords a bridge to the prefrontal cortex (PFC; [11]), known to function in concert with IT during visual DMS [12], and implicated in auditory DMS as well [13-16].

Despite these commonalities between the visual and auditory systems, recent behavioral studies indicate that auditory DMS performance in the monkey is less robust than that for visual DMS, and is likely to depend on a retained sensory trace [17, 18]. To test the hypothesis that the rSTC supports this trace, we recorded neurons throughout rSTC while rhesus monkeys performed auditory DMS (Fig. 1). A substantial population of neurons exhibited sustained modulation of their firing rate during the delay interval, as well as task-related modulation of their sensory responses, as observed in IT during visual DMS [19-22]. Our findings confirm the engagement of these areas during auditory DMS, and suggest that the disparity between modalities evident in behavior [17] is rooted in concomitant neurophysiological differences.

Results

Three monkeys (F, S, and K) were trained to perform auditory serial DMS (Fig. 1A). Sequences of two to four sounds (~300 ms in duration) were presented at an interstimulus interval of ~1 s. Monkeys released a touch bar to indicate the repetition of the first sound (sample) as a match, and withheld response to any intervening nonmatch sounds. Stimuli were drawn from a set of 21 exemplars including both synthetic and natural sounds. Behavioral performance declined markedly as the number of nonmatch stimuli in the trial increased [17, 18]. Performance of monkeys F and S was quite similar, but Monkey K could not be trained to criterion with >1 nonmatch stimulus (data from this animal are included where appropriate).

Recording sites spanned the rostral auditory cortical areas, including auditory core (R and RT), the adjacent medial and lateral belt, rostral parabelt, and tissue extending rostrally to the dorsal temporal pole (Fig. 1B, Supplementary Table S1). Auditory responses were obtained at 36% of 640 sites, yielding 280 responsive units (37% of 749 units tested; 85 from monkey F [all in left hemisphere], 148 from monkey S [117 right, 31 left], and 47 from monkey K [all in left]). The median number of effective stimuli was 6, and responses were predominantly excitatory (80%). Of the auditory units, 13% also responded at the time of reward delivery, but this epoch of the trial is excluded in later analysis.

Modulation of delay-period activity

In about one third of the units, a sustained modulation of firing rate during at least one of the delay epochs in the trial was observed (98/280, 35%). Activity was measured over the last 600 ms of each delay, and compared to the 600 ms pre-trial baseline (Wilcoxon rank-sum, $p < 0.008$ correcting for multiple comparisons). As shown in Figure 2, this modulation could take the form of delay suppression (DS) or delay enhancement (DE), which occurred in roughly equal proportion (50/280, 18%, and 48/280, 17%, respectively). Delay enhancement diminished across the three epochs within the trial, but the sensory responses evoked by

match and nonmatch stimuli did not differ in magnitude relative to the sample (Fig. 2C). By contrast, DS was sustained across all three epochs of the trial, and the responses evoked by match and nonmatch sounds were suppressed relative to that for the sample (Fig. 2D).

Firing rate during the first delay was dependent on the identity of the preceding sample stimulus in 10 units (3.6%; Kruskal-Wallis test using sample identity [1-21] as the single factor, $p < 0.008$ correcting for multiple comparisons). No unit showed a significant effect of sample identity in the second delay. For comparison to previous studies that used smaller stimulus sets (e.g., [23]), stimuli were ranked by the magnitude of the sample response, and trials were grouped between the top and bottom half of the stimuli. By this analysis, firing rate during the first delay carried information about sample identity in 16 units (16/280 = 6%), including 8 DE units (8/42 = 19%) and 1 DS unit (1/36 = 2.8%; K-W test, $p < 0.008$). Even among the sub-population showing elevated activity during the delay, that activity was selective for the prior sample sound in $< 20\%$ of units.

Among units also recorded during passive presentation, 15% (20/133) showed 'delay' modulation during the inter-stimulus interval (see Supplemental Experimental Procedures), a proportion lower than that observed during behavior (χ^2 , $p = 0.004$). Tested separately, this distinction was stronger for DE (7% passive, 15% behaving, $p = 0.016$) than for DS (8%, 13%, $p = 0.17$). However, passive effects were observed in only a small minority of those units that showed DE (12%) and DS (24%) during behavior, suggesting that these phenomena were largely specific to the DMS task. As task engagement has been shown to induce both phasic and tonic shifts in firing rate in auditory cortex [23-27], delay modulation (particularly suppression) may reflect a passive process that is strengthened or recruited during DMS performance.

Modulation of the match response

The response of a single unit could be influenced not only by stimulus selectivity, but also by the context in which that sound appeared in the DMS task. The unit in Figure 3 exhibited match suppression (MS), a reduction in response magnitude for match presentations, relative to those for the same sounds presented as samples (Fig. 3 A-D; Supplemental Fig. S1). This effect persisted through at least two intervening nonmatch stimuli (Fig. 3 B-D), spanning a total interval of >3 s. This MS appears to result from stimulus-specific repetition suppression of the sample sound, because no such effect was seen when the same sounds were presented as a nonmatch (Fig. 3, G-H). To isolate the effect of repetition we compared responses to match and nonmatch presentations within each trial position (Fig. 3 E-F), revealing a significant effect only at position 2.

An overall match suppression was evident in the averaged population response, and persisted through the full trial duration (Fig. 4A-C; all $p < 10^{-8}$, Wilcoxon sign-rank [WSR] test on firing rates from 25-200 ms). Match responses were also suppressed relative to the nonmatch response at position 2 (Fig. 4D, $p = 0.001$, WSR), but not position 3 ($p = 0.14$). Suppressive effects were not entirely stimulus-specific across the full population of neurons, as revealed by a generalized suppression of nonmatch responses relative to the preceding sample at positions 2 (Fig. 4E) and 3 (both $p < 0.0004$, WSR). The magnitude of the match/sample suppression at position 2 was greater than that of either the match/nonmatch or

sample/nonmatch comparison ($p < 0.001$, $p < 0.0009$ respectively, WSR). Nonmatch suppression may stem from partial adaptation to shared features between sample and nonmatch sounds, which we have previously shown to predict matching errors during DMS [18].

Timing of match response modulation

The proportion of units showing a significant difference in firing rate between stimulus contexts was calculated in a sliding 100-ms window, and overlaid on each panel of Figure 4. The match effect (Fig. 4A-C) showed a biphasic time course, which appears to reflect the sum of two underlying processes (Fig 4F): a transient effect peaking at ~100 ms, and a steady buildup during and beyond the stimulus presentation that is also evident in the match/nonmatch comparison (Fig. 4D). The early component could reflect recognition of the match, though it was also seen to a lesser degree in the sample/nonmatch comparison (Fig. 4E), suggesting it may result from shared features between the sample and nonmatch sounds that were not sufficient to trigger a behavioral ‘match’ response. The latter component could reflect an accumulating decision process, preparation of the motor response, or anticipation of reward. To control for motor and reward effects, we compared activity between nonmatch presentations that did or did not lead to an erroneous response (Supplemental Experimental Procedures; Fig. S2), and confirmed that bar release and reward anticipation had no effect during the stimulus period in the vast majority of neurons.

Averaging across the population obscures the heterogeneity of response modulations in individual units. In AA trials, modulation of the match response relative to the sample was observed in 19% of units (53/280), but these effects were not universally suppressive: 12% showed MS (33/280; Figs. 3, S1), but 7% exhibited the opposite effect, match enhancement (ME; 20/280; Fig S3). Averaged responses of these subpopulations are presented in Figure 5 (for the proportion of units showing effects in the match/nonmatch and sample/nonmatch comparisons see Fig. S4). Whereas MS was evident throughout the first 200 ms of the response (peaking at ~100 ms), ME peaked later in the response (~180 ms after stimulus onset; compare Figs. 5A and 5E). After correction for the onset latency of each unit, ME effects lagged MS by a mean of > 50 ms ($p = 0.003$, Wilcoxon rank-sum). A contingency analysis (Table S2) revealed a tendency for ME and DE, or MS and DS, to co-occur within the same units at both trial positions (binominal test, $p < 0.003$).

Delay and match effects diminish selectively across the trial

To control for differences in statistical power and anticipatory effects across trial positions, a subset of trials from monkeys F and S ($N = 233$ units) was re-analyzed as described above. The proportion of units exhibiting significant DS (9%) was unchanged between delay 1 and delay 2 (χ^2 , $p = 1$; Fig. 6A), but the proportion of units showing DE declined from 13% to 6% (χ^2 , $p = 0.008$). Similarly, the proportion of units showing MS was equivalent at positions 2 and 3 (11% and 9%; χ^2 , $p = 0.43$), whereas ME was observed in 5% of units at position 2, but was nearly absent at position 3 (1% of units; χ^2 , $p = 0.03$). Thus, whereas suppressive effects persisted across the duration of the trial, excitatory effects were apparently ‘reset’ by the intervening nonmatch stimulus. (Despite changes in prevalence of the effects, the average magnitude of MS and ME was equivalent across trial lengths; Fig.

S5.) Coincident with this shift in the physiological phenomena associated with the DMS task, the behavioral accuracy of the animals declined sharply after the first nonmatch (Fig. 6B), indicating that DMS performance was related, not to the degree of suppression, but to the degree of enhancement in the delay activity and match response.

Time course of stimulus encoding and retention

For the first epoch of the trial, firing rate during the sample presentation and subsequent delay was analyzed with a single factor ANOVA (sample identity, 1-21) in a 100-ms sliding window. An example unit in STGr (Fig. 7A) exhibited a slow sustained response that was selective for the preceding sample stimulus well into the delay period (Fig. 7B). However, this unit was an exception among the population: the average variance explained by sample identity decayed to zero ~300 ms after stimulus offset (Fig 7C), well before the presentation of the sound at position 2. Only 5/280 units, two in STGr and three in rostral belt, exhibited persistent selectivity for > 500 ms after the sample presentation (inset in Fig. 7C).

Relative strength of sensory and mnemonic signals

To capture the relative weight of sensory and mnemonic influences in the second epoch of the trial, the ANOVA model was expanded to include three factors at position 2. The first was the identity of the preceding sample (an integer from 1 to 21), which seldom showed a significant effect. Second, the match/nonmatch condition at position 2 (a value of zero or one) was taken to represent mnemonic information in the response. Third, the identity of the stimulus at position 2 (1-21) was nested within the match/nonmatch factor, and taken to represent purely sensory information. The unit in Figure 7D showed strong ME, particularly in the latter half of the sensory response. As revealed by the ANOVA model (Fig. 7E), sensory selectivity of the response reached its maximum 100 ms after sound onset, and persisted for 100 ms after sound offset; by contrast, the influence of match/nonmatch status was maximal between 200-300 ms. Although this unit showed clear sensory and mnemonic selectivity, the population as a whole conveyed primarily sensory information (Fig. 7F), with relatively little influence of the abstract match/nonmatch distinction. At position 2, 41/280 units (15%) showed an effect of the match/nonmatch factor (criterion: > 1 significant time bin between 0-300 ms), and among those units the mean variance explained was 5.4%. By contrast, stimulus identity was a significant factor in 103/280 units (37%), with a mean explained variance of 19%.

Anatomical distribution of memory effects

Our recording sites spanned cortical areas across four hierarchical levels, from core and belt regions to parabelt and STGr (Fig. 1B). To quantify whether the prevalence of memory effects differed across levels, the population was split into two groups: Group 1 (N=101), comprising rostral core and belt, and Group 2 (N=167), comprising parabelt, RTp, dorsal temporal pole, and upper bank of the STS (Supplementary Table S1). The proportion of units exhibiting match or delay effects did not differ between the two populations (χ^2 test, $p = 0.4$ for DS/DE, $p = 0.1$ for MS/ME). By the ANOVA analysis (Fig. 7E), match/nonmatch status significantly affected firing rate in 16% of Group 1 units and 14% of Group 2 units.

Discussion

Delay-period effects

Firing rate during the memory delay was modulated in 35% of units (Fig. 2), but delay activity was seldom selective for the preceding stimulus. Similar results have been reported in the caudal belt [23], dorsal temporal pole (our TGd; [28]), and recently in AI [29]. Our serial DMS paradigm revealed that while DS persisted throughout the trial, DE was not robust to interference, and diminished in tandem with behavioral accuracy (Fig. 6). In this regard, DE seems more closely tied to the sensory trace, whereas DS may represent a more general attentional effect that may be necessary, but not sufficient, to support STM.

Delay suppression is seldom mentioned in the visual DMS literature, but common in primate studies of auditory STM [13, 23, 28, 29]. Inhibitory projections from PFC to the STGr [30] provide a possible substrate for the robust DS observed in our unit data (Fig. 2C) and in human fMRI [31-33]. Whereas visual STM in humans has been reported to rely on active maintenance of information in early sensory cortex [34], auditory STM elicits delay-period suppression [33] that may protect the STM trace from interfering sounds [32]. Of particular relevance to the present study, Linke and colleagues [32] describe delay suppression that was strongest in subjects who relied on passive, echoic memory, as opposed to active rehearsal. We believe nonhuman primates are limited to this type of auditory memory [17], as they lack the ‘phonological loop’ [35] necessary for rehearsal.

Match Suppression and Enhancement

If the sensory trace is not evident as sustained stimulus-specific delay activity, a tenable alternative is a sub-threshold mechanism such as synaptic plasticity [36], which would affect subsequent responses. Responses to match stimuli were modulated relative to responses to sample stimuli in 19% of units, with roughly two-thirds exhibiting MS, and one third ME. The ME that we observed, which appeared 80-180 ms after stimulus onset, has not been described previously in auditory cortex; ME 300 ms after sound onset has been reported in AI and TGd [28, 29], but likely represents response selection and/or feedback from PFC [13]. By contrast, short-latency MS has been reported in AI (23% of units; [29]), caudal auditory belt (22%; [23]), and TGd (9%; [28]). Collectively these data argue against a specialization for STM at the temporal pole, and in favor of a more distributed representation that includes core and belt. Consistent with this, the ‘rSTG’ lesion of Fritz et al. [10] comprised the higher-level areas we designated as ‘Group 2’ (Table S1), yet those animals did not show a deficit in auditory DMS at a 5-s delay.

Comparisons to visual DMS

The prevalence of match and delay effects we observed is similar to that reported in some studies of visual DMS in IT cortex, which described excitatory and inhibitory delay activity that carried little information about the preceding stimulus [19, 21], and a relatively weak influence of match/nonmatch status on sensory responses [21]. Those recordings covered a broad area of IT cortex, as did ours in the rSTC, using tasks that required only sensory memory for simple colors or patterns.

Delay activity and match effects were observed to a greater degree by Miller and colleagues [20, 37], who recorded from a restricted IT region in or near the perirhinal cortex, which is strongly associated with visual recognition memory [38]. Our DMS paradigm is modeled after theirs, which required the animals to overcome multiple nonmatch items in a series of complex images. Responses to match stimuli were more suppressed than responses to nonmatch stimuli in rSTC and in IT [20], indicating that MS is stimulus-specific. In parallel with our findings, MS appeared at the same latency as the response, suggesting it originated at or before the level of IT [20]. However, ME in IT cortex appeared at the same latency as MS and survived intervening nonmatch stimuli [22], unlike the ME we observed in auditory cortex, which occurred ~50 ms later than MS and did not survive intervening distractors. The time lag suggests that the ME we observed could have arisen via a top-down signal; Plakke et al. [13] recently found that the population response in lateral PFC shows ME within the first 100 ms after cue onset [13], a latency short enough to potentially drive ME in the rSTC. Alternatively, the lag may reflect temporal integration of the dynamic auditory signal within the rSTC itself, as required for recognition of sounds that evolve over time, but not for recognition of static images.

Adaptation and context effects

The latency of the MS effect in rSTC suggests it is a local or bottom-up process, possibly an outgrowth of adaptive processes evident in the primary auditory cortex (A1). Although the duration of forward masking or enhancement in A1 would be insufficient to span the 1-s delay in our task [39, 40], context effects lasting up to 1 s have been reported in A1 of the awake primate [41-43]. The time course of adaptation has not been systematically studied in the fields downstream from A1, but evidence from human electrophysiology suggests that the decay of the activation trace is slower in auditory association cortex than in A1 [44], consistent with the long-lasting MS we observed.

Conclusions

Despite ethological evidence for long-term learning and storage of sounds by monkeys (e.g., [45]), their auditory memory falls short of visual and tactile memory when tested by DMS [9, 10], a discrepancy across modalities that may extend to humans as well [46]. Visual and tactile memory appear to tap the same cortical system [4], and tactile objects may be encoded as visual images or shapes regardless of the modality of input. Auditory ‘objects’, by contrast, are more likely to refer to transient events that unfold over time, complicating their storage and retrieval.

Miller and Desimone [22] proposed two parallel mechanisms for visual STM in the temporal lobe: match suppression, representing a passive memory trace, and match enhancement, representing an actively retained memory, in a distinct population of neurons. In rSTC, match enhancement was neither widespread nor robust to interference, bolstering prior behavioral evidence implying that monkeys may depend primarily upon the passive sensory trace [17, 18]. Alternatively, the active mechanism in audition may emerge in PFC, though whether the match enhancement recently described in lateral PFC [13] is robust to interference remains unknown.

Experimental Procedures

All procedures accorded with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Animal Care and Use Committee of the NIMH. Subjects were three adult male rhesus monkeys (*Macaca mulatta*). Details of the task, stimuli, training, and behavioral performance were published previously [17, 18]. Detailed methods are available as Supplemental Experimental Procedures. Briefly, animals sat in a primate chair within a sound-attenuating booth. Head position was fixed, and a sipper tube was positioned for delivery of water rewards. The trial sequence is shown in Figure 1A. The standard stimulus set included three exemplars from each of seven categories: modulated noise; band-pass noises; pure tones; frequency-modulated sweeps; rhesus monkey vocalizations; other species' vocalizations; and environmental sounds. Synthetic sounds were 300 ms in duration, whereas the duration of the natural sounds varied slightly (195-282 ms). Stimuli were presented at 60-70 dB SPL via a loudspeaker (Ohm Acoustics, NY) located 1 m directly in front of the animal.

An MRI-compatible recording chamber was implanted to allow a vertical approach to the rSTG (Fig. 1B). Electrode tracks were guided by alignment to an MRI acquired after implantation of the chamber. Most sites (81%) yielded one or two simultaneously recorded units; the 280 units in this report derive from 114 sites yielding one unit, 57 yielding two, 16 yielding three, and one site that yielded four separable units. Spike sorting was verified offline by principal components analysis (Spike 2, CED), and spike and event times were exported to MATLAB (Mathworks) for analysis.

After a unit was isolated, sounds were presented in pseudorandom order 8-10 times with an inter-stimulus interval (ISI) of 2.5 s as the animal sat passively. If a unit evinced an auditory-evoked response, then the animal was presented with the DMS paradigm. After completion of the recordings, all sites from each hemisphere were aligned to the left hemisphere of an averaged MRI template (Fig. 1B; [47]), registered to a combined MRI and histology atlas [48].

Memory effects were investigated in 280 units that were responsive to at least one stimulus. To identify match suppression or enhancement (Figs. 3-5), responses from correct trials were segregated by stimulus context (sample, match, or nonmatch) and sequential position within the trial. In all statistical comparisons, responses were pooled across stimuli, and the number of trials per stimulus was equated between contexts. For each trial type, spike counts during sample and match presentations were compared by a Wilcoxon rank-sum test in a 100-ms sliding window moved in 20-ms steps. A unit was classified as showing an effect if two adjacent bins between 0 and 300 ms were significantly different between contexts ($p < 0.01$, Bonferroni corrected for overlap of time bins).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank H Tak, M Muñoz-Lopez, K Moorhead, P Sergio, A Kloth, and H Vinal for assistance with animal training and data collection, and RC Saunders and M Malloy for providing technical expertise. This research was supported by the Intramural Research Program of the NIMH/NIH/DHHS.

References

1. Cowan N. On short and long auditory stores. *Psychol Bull.* 1984; 96:341–370. [PubMed: 6385047]
2. Demany, L.; Semal, C. The Role of Memory in Auditory Perception.. In: Yost, WA.; Popper, AN.; Fay, RR., editors. *Auditory Perception of Sound Sources.* Vol. 29. Springer US; New York: 2007. p. 77-113.
3. Pasternak T, Greenlee MW. Working memory in primate sensory systems. *Nat Rev Neurosci.* 2005; 6:97–107. [PubMed: 15654324]
4. Mishkin M. A memory system in the monkey. *Philos Trans R Soc Lond B Biol Sci.* 1982; 298:83–95. [PubMed: 6125978]
5. Hackett TA. Information flow in the auditory cortical network. *Hear Res.* 2011; 271:133–146. [PubMed: 20116421]
6. Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci.* 2013; 17:26–49. [PubMed: 23265839]
7. Kikuchi Y, Horwitz B, Mishkin M. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J Neurosci.* 2010; 30:13021–13030. [PubMed: 20881120]
8. Perrodin C, Kayser C, Logothetis NK, Petkov CI. Voice cells in the primate temporal lobe. *Current biology : CB.* 2011; 21:1408–1415. [PubMed: 21835625]
9. Colombo M, Rodman HR, Gross CG. The effects of superior temporal cortex lesions on the processing and retention of auditory information in monkeys (*Cebus apella*). *J Neurosci.* 1996; 16:4501–4517. [PubMed: 8699260]
10. Fritz J, Mishkin M, Saunders RC. In search of an auditory engram. *Proc Natl Acad Sci U S A.* 2005; 102:9359–9364. [PubMed: 15967995]
11. Romanski LM, Bates JF, Goldman-Rakic PS. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol.* 1999; 403:141–157. [PubMed: 9886040]
12. Fuster JM, Bauer RH, Jervey JP. Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Res.* 1985; 330:299–307. [PubMed: 3986545]
13. Plakke B, Ng CW, Poremba A. Neural correlates of auditory recognition memory in primate lateral prefrontal cortex. *Neuroscience.* 2013; 244:62–76. [PubMed: 23583656]
14. Chao LL, Knight RT. Contribution of human prefrontal cortex to delay performance. *J Cogn Neurosci.* 1998; 10:167–177. [PubMed: 9555105]
15. Zatorre RJ, Evans AC, Meyer E. Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci.* 1994; 14:1908–1919. [PubMed: 8158246]
16. Zatorre RJ, Samson S. Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain.* 1991; 114(Pt 6):2403–2417. [PubMed: 1782523]
17. Scott BH, Mishkin M, Yin P. Monkeys have a limited form of short-term memory in audition. *Proc Natl Acad Sci U S A.* 2012; 109:12237–12241. [PubMed: 22778411]
18. Scott BH, Mishkin M, Yin P. Effect of acoustic similarity on short-term auditory memory in the monkey. *Hear Res.* 2013; 298:36–48. [PubMed: 23376550]
19. Fuster JM, Jervey JP. Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *J Neurosci.* 1982; 2:361–375. [PubMed: 7062115]
20. Miller EK, Li L, Desimone R. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J Neurosci.* 1993; 13:1460–1478. [PubMed: 8463829]
21. Eskandar EN, Richmond BJ, Optican LM. Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioral context. *J Neurophysiol.* 1992; 68:1277–1295. [PubMed: 1432084]

22. Miller EK, Desimone R. Parallel neuronal mechanisms for short-term memory. *Science*. 1994; 263:520–522. [PubMed: 8290960]
23. Gottlieb Y, Vaadia E, Abeles M. Single unit activity in the auditory cortex of a monkey performing a short term memory task. *Exp Brain Res*. 1989; 74:139–148. [PubMed: 2924831]
24. Benson DA, Hienz RD, Goldstein MH Jr. Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: spatial tuning and behavioral dependency. *Brain Res*. 1981; 219:249–267. [PubMed: 7260632]
25. Brosch M, Selezneva E, Scheich H. Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J Neurosci*. 2005; 25:6797–6806. [PubMed: 16033889]
26. Brosch M, Selezneva E, Scheich H. Representation of reward feedback in primate auditory cortex. *Frontiers in systems neuroscience*. 2011; 5:5. [PubMed: 21369350]
27. Miller JM, Sutton D, Pfingst B, Ryan A, Beaton R, Gourevitch G. Single cell activity in the auditory cortex of Rhesus monkeys: behavioral dependency. *Science*. 1972; 177:449–451. [PubMed: 4625421]
28. Ng CW, Plakke B, Poremba A. Neural correlates of auditory recognition memory in the primate dorsal temporal pole. *J Neurophysiol*. 2013
29. Bigelow J, Rossi B, Poremba A. Neural correlates of short-term memory in primate auditory cortex. *Frontiers in neuroscience*. 2014; 8:250. [PubMed: 25177266]
30. Barbas H, Medalla M, Alade O, Suski J, Zikopoulos B, Lera P. Relationship of prefrontal connections to inhibitory systems in superior temporal areas in the rhesus monkey. *Cereb Cortex*. 2005; 15:1356–1370. [PubMed: 15635060]
31. Huang S, Seidman LJ, Rossi S, Ahveninen J. Distinct cortical networks activated by auditory attention and working memory load. *Neuroimage*. 2013
32. Linke AC, Vicente-Grabovetsky A, Cusack R. Stimulus-specific suppression preserves information in auditory short-term memory. *Proc Natl Acad Sci U S A*. 2011; 108:12961–12966. [PubMed: 21768383]
33. Rinne T, Koistinen S, Salonen O, Alho K. Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. *J Neurosci*. 2009; 29:13338–13343. [PubMed: 19846721]
34. Serences JT, Ester EF, Vogel EK, Awh E. Stimulus-specific delay activity in human primary visual cortex. *Psychological science*. 2009; 20:207–214. [PubMed: 19170936]
35. Baddeley A. Working memory: looking back and looking forward. *Nat Rev Neurosci*. 2003; 4:829–839. [PubMed: 14523382]
36. Sugase-Miyamoto Y, Liu Z, Wiener MC, Optican LM, Richmond BJ. Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS computational biology*. 2008; 4:e1000073. [PubMed: 18464917]
37. Miller EK, Li L, Desimone R. A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*. 1991; 254:1377–1379. [PubMed: 1962197]
38. Meunier M, Bachevalier J, Mishkin M, Murray EA. Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *J Neurosci*. 1993; 13:5418–5432. [PubMed: 8254384]
39. Brosch M, Scheich H. Tone-sequence analysis in the auditory cortex of awake macaque monkeys. *Exp Brain Res*. 2008; 184:349–361. [PubMed: 17851656]
40. Brosch M, Schulz A, Scheich H. Processing of sound sequences in macaque auditory cortex: response enhancement. *J Neurophysiol*. 1999; 82:1542–1559. [PubMed: 10482768]
41. Bartlett EL, Wang X. Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol*. 2005; 94:83–104. [PubMed: 15772236]
42. Fishman YI, Steinschneider M. Searching for the mismatch negativity in primary auditory cortex of the awake monkey: deviance detection or stimulus specific adaptation? *J Neurosci*. 2012; 32:15747–15758. [PubMed: 23136414]
43. Werner-Reiss U, Porter KK, Underhill AM, Groh JM. Long lasting attenuation by prior sounds in auditory cortex of awake primates. *Exp Brain Res*. 2006; 168:272–276. [PubMed: 16328295]

44. Lu ZL, Williamson SJ, Kaufman L. Human auditory primary and association cortex have differing lifetimes for activation traces. *Brain Res.* 1992; 572:236–241. [PubMed: 1611518]
45. Seyfarth RM, Cheney DL, Marler P. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science.* 1980; 210:801–803. [PubMed: 7433999]
46. Bigelow J, Poremba A. Achilles' ear? Inferior human short-term and recognition memory in the auditory modality. *PLoS One.* 2014; 9:e89914. [PubMed: 24587119]
47. McLaren DG, Kosmatka KJ, Oakes TR, Kroenke CD, Kohama SG, Matochik JA, Ingram DK, Johnson SC. A population-average MRI-based atlas collection of the rhesus macaque. *Neuroimage.* 2009; 45:52–59. [PubMed: 19059346]
48. Saleem, KS.; Logothetis, NK. Coronal and Sagittal series. 2nd edition with Horizontal. Elsevier/ Academic press; San Diego, CA: 2012. A combined MRI and histology atlas of the rhesus monkey brain in stereotaxic coordinates..

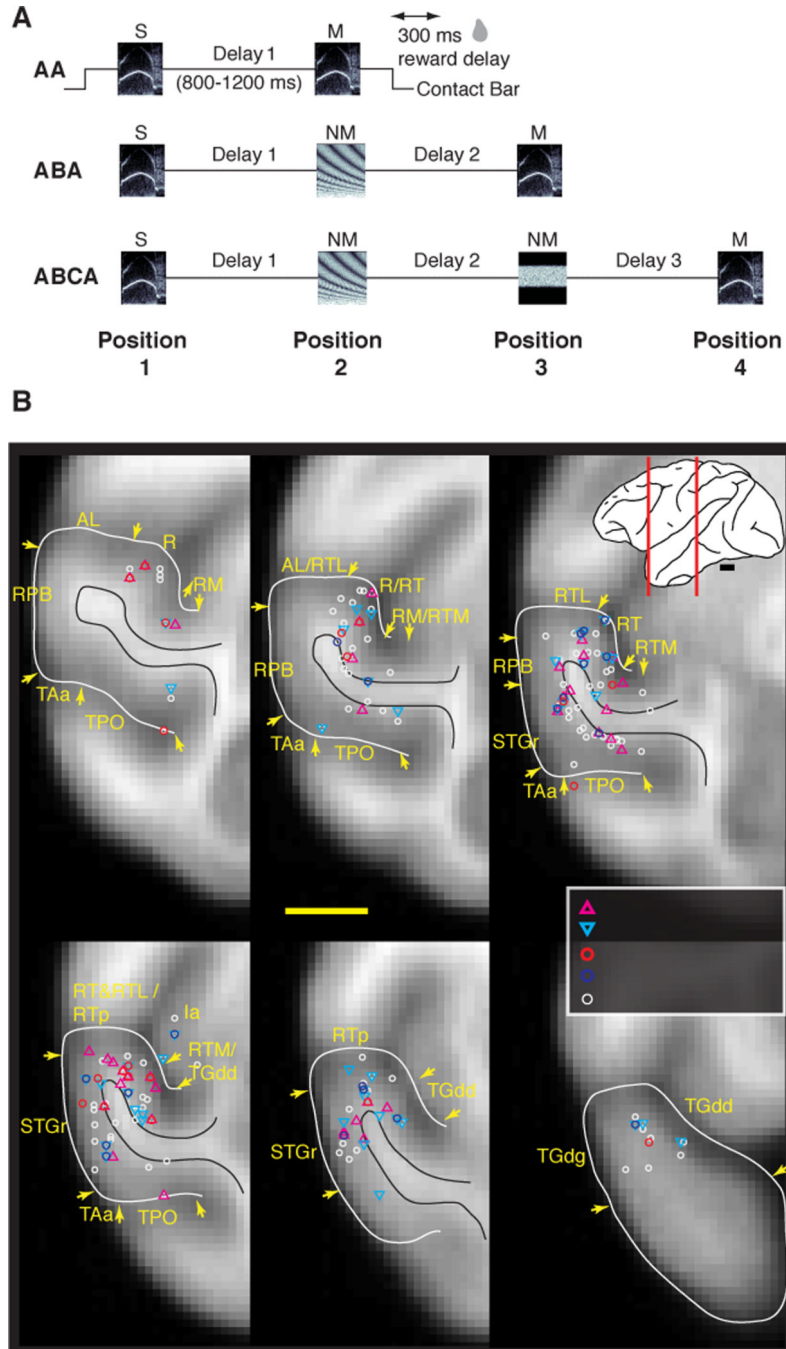


Figure 1. Monkeys performed an auditory short-term memory task while activity was recorded from single cortical neurons in the rostral STG. (A) Schematic diagram of the three trial types in the auditory DMS task. Sounds were ~300 ms in duration, here represented by frequency-time spectrograms. The monkey initiated a trial by holding a contact bar for 300 ms, after which a sample sound was presented, followed by 1-3 test sounds at a randomized interstimulus delay of 800-1200 ms. When the test sound was identical to the sample (i.e., a match), the monkey could release the bar within a 1200-ms window beginning 100 ms after

match onset to earn a reward delivered 300 ms after bar release. If the stimulus was a nonmatch, the animal was required to continue holding the bar until the match appeared. Release following a nonmatch or failure to release after the match was counted as an error and punished by an extended inter-trial interval. Note that the stimulus at position 1 was always a sample; at position 2 and 3, a match or nonmatch could be presented; the stimulus at position 4 was always a match. Abbreviations and example stimuli for an ABCA trial: S, sample; NM1, nonmatch 1; NM2, nonmatch 2; M, match. **(B)** Recording sites from four hemispheres aligned to an averaged MRI volume for rhesus macaques [47]. Recordings spanned 18 mm, from 11-28 mm rostral to ear bar zero (EBZ), collapsed here onto 6 representative coronal sections at the level of fields R, RT, RTp, and the temporal pole. White lines outline the STG from the fundus of the lower limb of the circular sulcus to the fundus of the STS; black lines mark the border of the white matter. Inset at top right: lateral view of a macaque brain with red lines indicating the caudorostral extent of the recordings sites. Field abbreviations: AL, anterolateral (belt); Ia, agranular insula; ls, lateral sulcus; R, rostral (core); RM, rostromedial (belt); RPB, rostral parabelt; RT, rostrot temporal (core); RTL, rostrot temporal-lateral (belt); RTM, rostrot temporal-medial (belt); RTp, rostrot temporal-polar; STGr, rostral superior temporal gyrus; sts, superior temporal sulcus; TAa and TPO, sts dorsal bank areas; TGdd/g, area TG dorsal, dysgranular/granular. Scale bar = 5 mm. Unit counts by field are in Supplemental Table S1; organization of cortical fields is reviewed in [5].

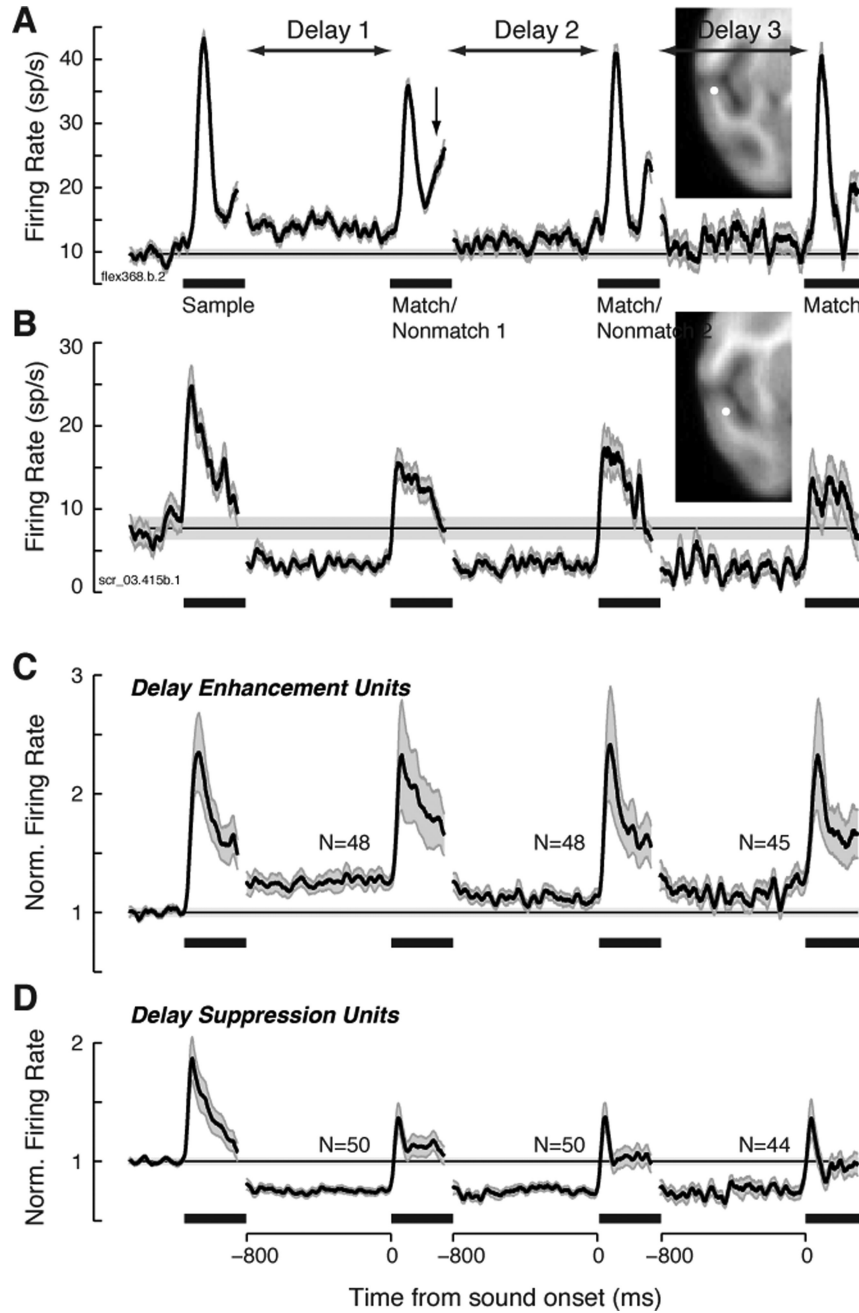


Figure 2. Firing rate during the interstimulus delay periods was suppressed or elevated relative to baseline in 35% of units. (A) Example unit showing delay enhancement (DE), recorded in field RTp of the rostral supratemporal plane (see inset). The late component of the auditory response to Match 1 (arrow) shows evidence of match enhancement (ME), as illustrated for this same unit in Supplemental Fig. S3. Black traces plot mean firing rate across all correct trials, the horizontal line marks baseline firing rate, and gray shading indicates ± 1 SEM across trials; black bars indicate time of stimulus presentation. Noisier traces at later delays are attributable to averaging fewer correct trials, owing to the sequential nature of the task,

and to the higher error rate on long trials than on short trials. Traces are discontinuous because delay duration varied from 800 to 1200 ms; for simplicity, activity is plotted for the 800 ms preceding the next stimulus onset. **(B)** Example unit showing delay suppression (DS), recorded at the medial edge of RTp (see inset). **(C)** Mean normalized firing rate for the subset of units exhibiting DE (48/280, 17%). Traces from delay 3 include fewer units than traces from delays 1 and 2 because one of the three subjects was not tested with the longest trial type. Firing rate was normalized within each unit by dividing by its baseline rate, before averaging across units (shading indicates ± 1 SEM across units). **(D)** Mean normalized firing rate for the subset of units exhibiting DS (50/280, 18%).

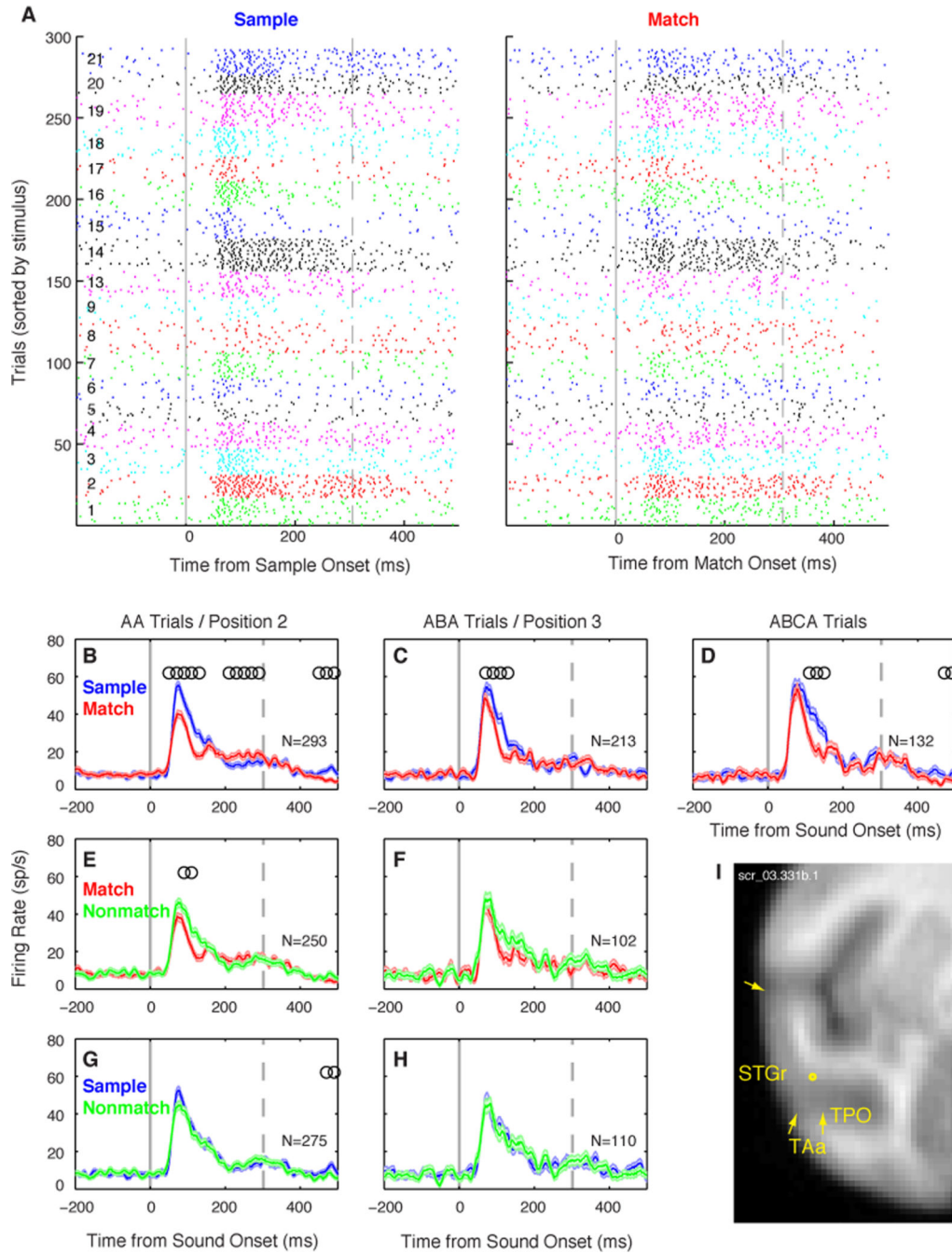


Figure 3. Match suppression (MS) in a well-isolated single unit recorded in ventral STGr. (A) Spike-time rasters of sample and match responses across 293 correct DMS trials (AA trials only), sorted by stimulus (numbered 1-21 at left, and indicated by tick colors). Solid gray line marks stimulus onset; dashed gray line marks offset of the longest stimuli. This unit responded vigorously to complex stimuli like rippled noise (1-3), a rhesus bark vocalization (14), and environmental sounds (19-21), with an onset latency of 55 ms. (B-D) Overlay of firing rate (mean \pm SEM across trials) to sample (blue) and match (red) presentations in all

correct trials (N indicated in each panel). In all panels, responses to different stimuli are pooled, and the number of trials per stimulus is equal across conditions. Open circles mark the centers of 100-ms time bins with a significant difference in firing rate between conditions (Wilcoxon rank-sum, $p < 0.01$ corrected). Match responses were suppressed relative to responses to the sample in all trial types, though the onset component ‘recovers’ on ABCA trials. **(E,F)** Overlay of match (red) and nonmatch (green) responses at position 2, the first stimulus to follow the sample, and at position 3, after an intervening nonmatch. This unit shows a significant match/nonmatch effect at position 2, but not at position 3. **(G,H)** Overlay of responses to the nonmatch (green) and the sample (blue), for a nonmatch at positions 2 and 3. No significant difference was seen, implying that suppression was specific to the match, and was not driven by a generalized suppression of responses later in the trial. **(I)** Recording location aligned to the MRI atlas, with the rostro-caudal position indicated in mm relative to the interaural axis. See Figure S1 for additional MS example units.

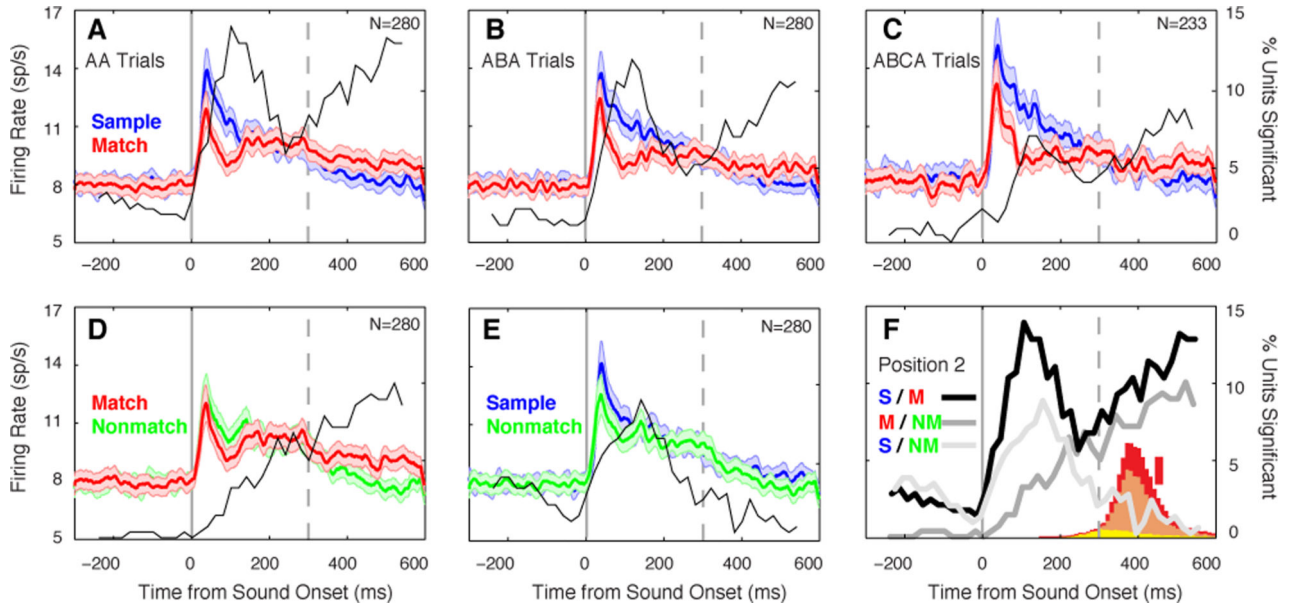
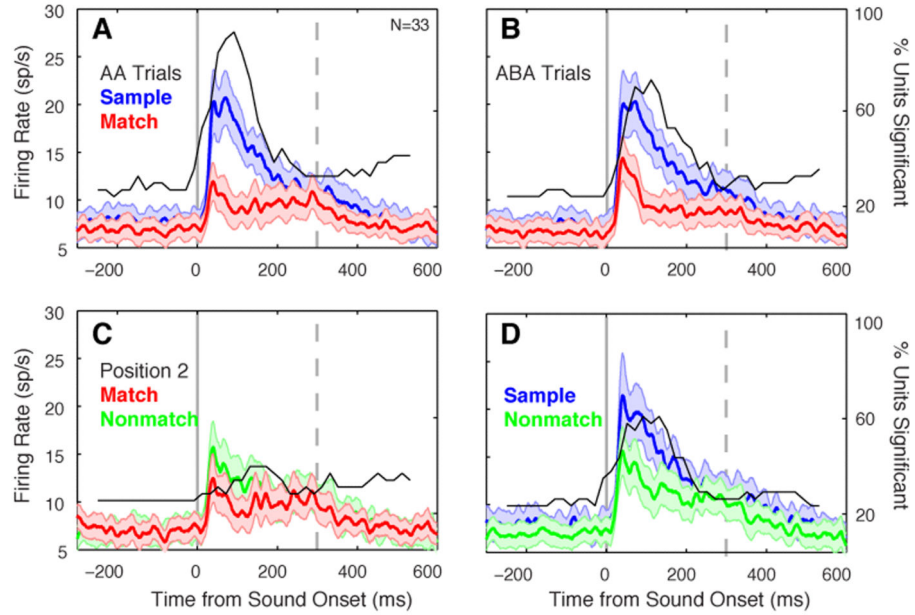


Figure 4.

The mean population response to match presentations was suppressed relative to that for sample and nonmatch presentations. (A-C) Firing rate (mean \pm SEM across units) to sample and match presentations on all correct trials. Suppression of the match response is strongest \sim 100 ms after stimulus onset, and persists across zero, one, or two intervening nonmatch stimuli (A-C, respectively). Fine black line indicates the proportion of units showing a significant difference in firing rate in each 100-ms time bin (see axis on the far right). (D) Match and nonmatch responses to the stimulus at position 2, the first stimulus to follow the sample. (E) Responses to the sample and to the nonmatch stimulus at position 2. (F) Overlay of curves tracking the prevalence of significant effects through time, for sample/match (AA trials, from panel A; black line), match/nonmatch at position 2 (from panel D; dark gray line), and sample/nonmatch (from panel E; light gray line). Inset histogram shows distributions of reaction times on match trials for all three subjects (red, orange, and yellow represent monkeys F, S, and K, respectively; vertical red scale bar indicates 500 trials). See Figs. S1 and S3 for single-unit examples.

Match Suppression Units



Match Enhancement Units

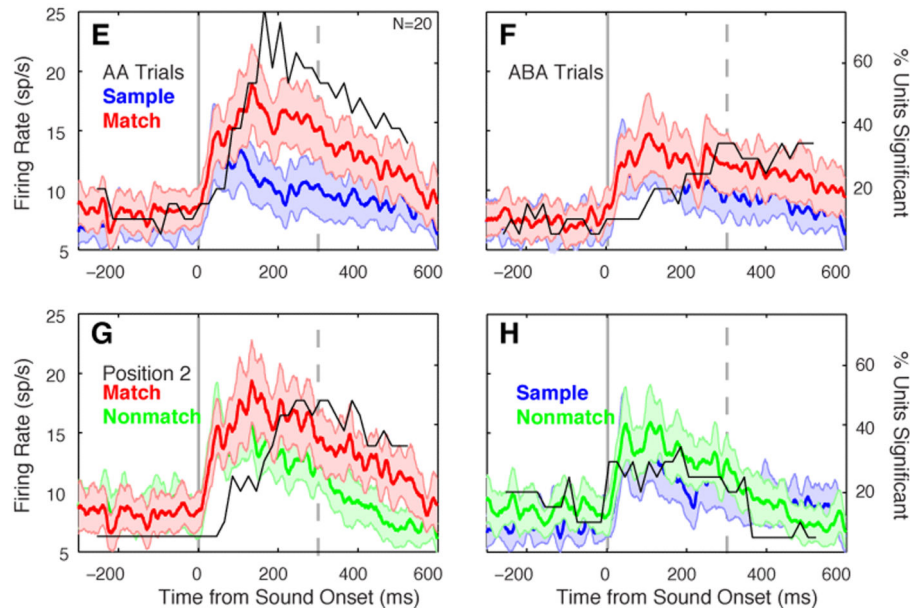


Figure 5. Modulation of auditory response magnitude by task context in the subsets of units showing MS (33/280, 12%; panels A-D), and ME (20/280, 7%; panels E-H). Conventions as in Figure 4. Note that MS affects the response over its full duration, including the onset, whereas ME occurs later (compare panels A and E). See Figs. S1 and S3 for single-unit examples.

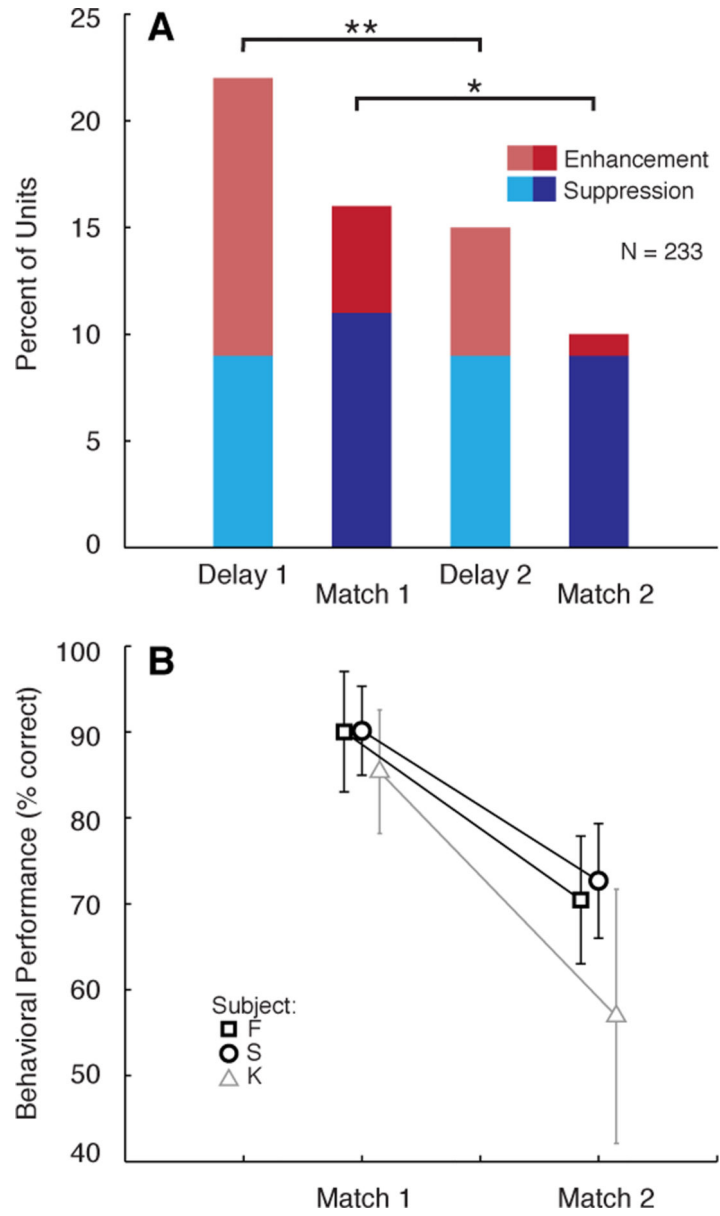


Figure 6. The prevalence of enhancement and suppression decrease differentially over the course of the trial. **(A)** Bars plot percentage of units (N = 233) showing significant DS (light blue) or DE (light red), and significant MS (dark blue) or ME (dark red), at two different points in the trial. The proportion of units showing DE decreases between delays 1 and 2, but the proportion showing DS is unchanged. Similarly, the proportion of units showing ME decreases between match positions 1 and 2, but the proportion showing MS is unchanged. **p = 0.008, *p = 0.03. **(B)** Behavioral performance on the DMS task for the corresponding AA (match at position 2) and ABA (match at position 3) trial types, for monkeys F (squares) and S (circles); physiological data from monkey K (triangles) are not included (see text), but this animal’s performance is shown in gray for comparison with that of the others (percent correct, mean ± SD across sessions). Percent correct for monkeys F and S, respectively was:

93% and 89% on AA trials; 73% and 73% on ABA trials; 38% and 40% on ABCA trials. Monkey K performed at 85% correct on AA trials, but only at 57% on ABA trials. The false alarm rate for F and S at position 2 was 14% and 18%, respectively, and at position 3 was 48% and 47%, respectively; the false alarm rate for K at position 2 was 39%.

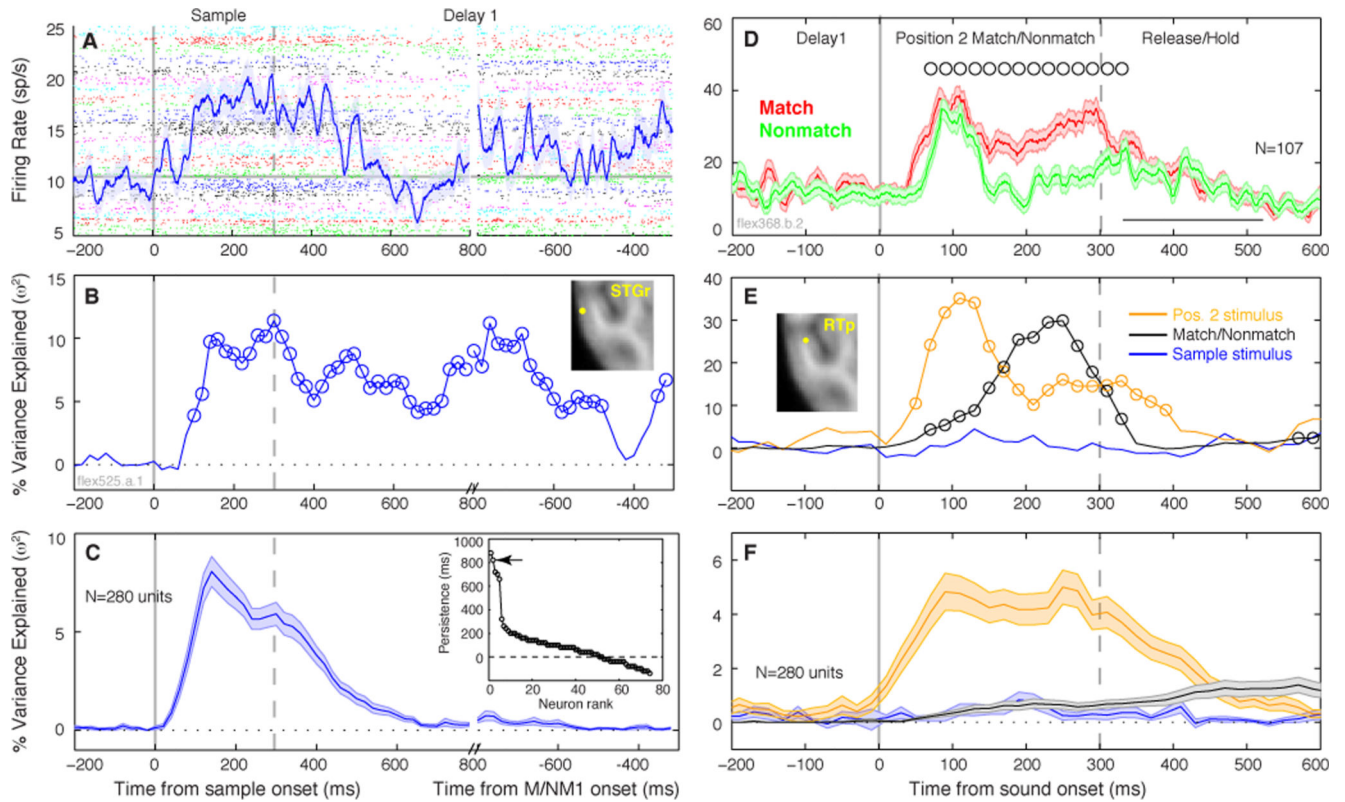


Figure 7.

Firing rate across time may be influenced by sensory and mnemonic factors within individual units, with a predominance of sensory encoding at the population level. **(A)** Sliding-window ANOVA describes the rise and decay of sensory encoding through time. Firing rate (mean \pm SEM across trials) of an example STGr unit (see inset in panel B) during both sample presentation and the ensuing delay period, overlaid on spike-time rasters (colors designate 21 different stimuli). Firing rate exceeds baseline + 3SD at 100 ms latency, and remains elevated for several hundred ms after sound offset. The discontinuity in the time axis is placed at 800 ms after onset of the sample and 800 ms before onset of the following sound to accommodate the variable delay durations. Gray horizontal line marks pre-trial baseline firing rate. **(B)** Single-factor ANOVA in a 100-ms sliding window revealed that the variance in firing rate during the delay was influenced by the preceding sample stimulus, even when mean firing rate during the delay (panel A) dropped below the pre-stimulus baseline (600-700 ms). Open circles in panel B mark bins with a significant F-value ($p < 0.05$ after FDR correction). **(C)** The mean variance explained (\pm SEM across units) by sample identity across the full population peaked between 100-200 ms after sound onset, and faded to zero by \sim 300 ms after sound offset, well before the end of the delay interval. Sustained encoding like that seen for the unit in (B) was rarely observed. Inset in panel C: Units are sorted by ‘persistence’, i.e. the last time bin to evince significant stimulus encoding, relative to sound offset. Only \sim 50 units showed sustained selectivity after sound offset (dashed horizontal line at zero); arrow marks the example unit in B. **(D)** Firing rate of a unit in field RTp (see inset in panel E) for match and nonmatch presentations at position 2 (the same ME unit depicted in Fig. S3, upper panels). Open circles mark centers of 100-ms

bins with significantly different spike counts for match and nonmatch stimuli; open square ~420 ms marks mean \pm SD bar-release time on match trials. **(E)** Corresponding ANOVA result from the unit in **(D)**, showing the proportion of variance in a 100-ms sliding window explained by three factors: the match/nonmatch status of the position 2 stimulus (black curve); nested within that, the identity of the position 2 stimulus (yellow curve); and the identity of the preceding sample stimulus (blue curve). Open circles mark bins with a significant F-value ($p < 0.05$ after FDR correction). **(F)** Mean explained variance (\pm SEM) for the population, same conventions as in **(E)**.