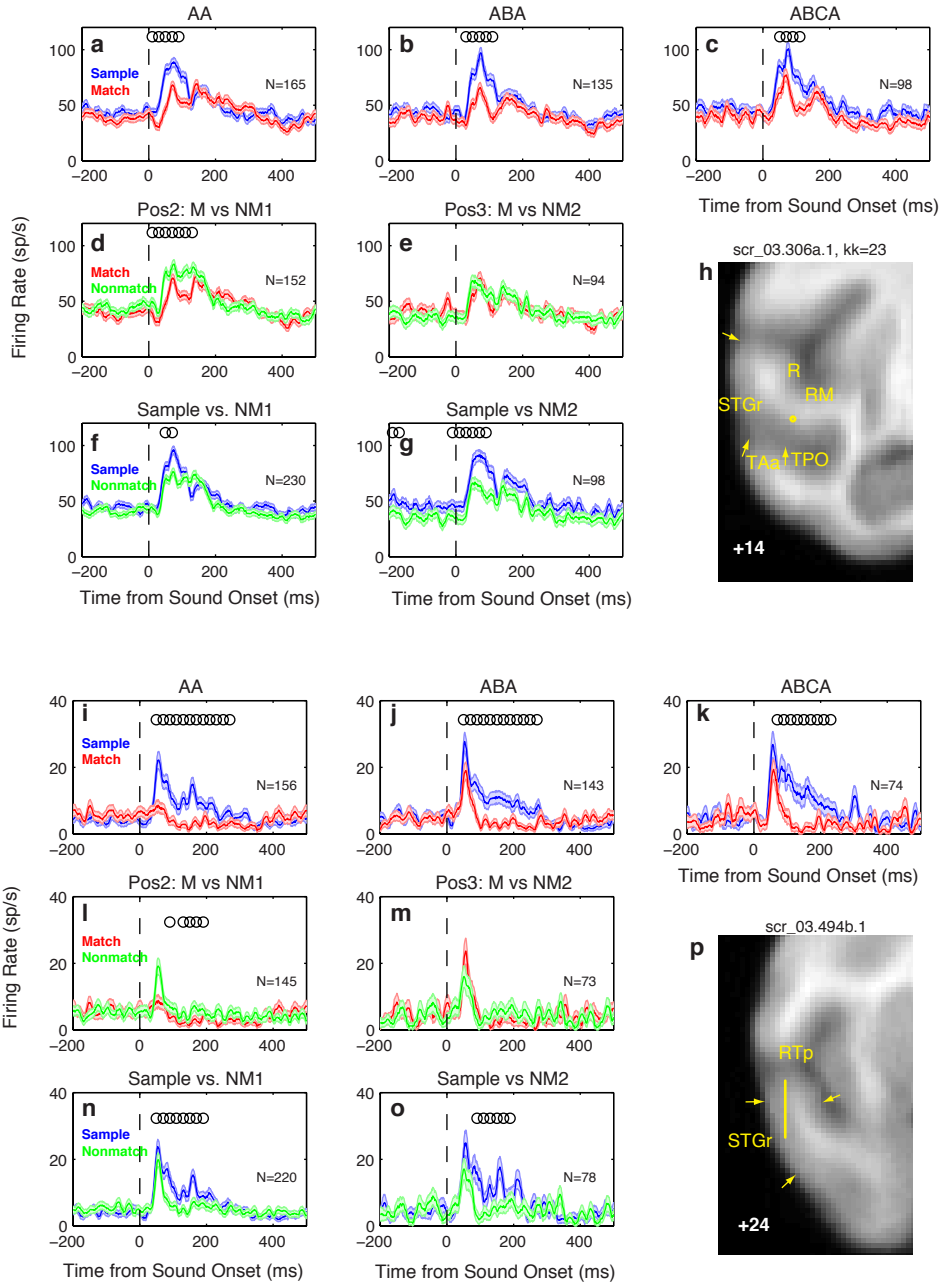

Figure S1: (related to Figures 4 & 5) Two units with robust MS for all three trial types. **(a-c)** Overlay of firing rate (mean ± SEM) to sample (blue) and match (red) presentations from all responsive units on all correct trials (N indicated in each panel). In all panels, responses to different stimuli 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. **(d,e)** 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. **(f,g)** Overlay of responses to the nonmatch (green) and the sample (blue), for a nonmatch at positions 2 (f) and 3 (g). **(h)** Recording location aligned to the averaged MRI atlas of McLaren et al. (2009), with the rostrocaudal position indicated in mm relative to the interaural axis. The unit in a-h was recorded from area TPO, in the upper bank of the STS, ventral to the core field R. **(i-p)** A unit from RTp or the STGr; the exact recording depth could not be determined, but the electrode track is indicated by the yellow line in (p). Conventions as in panels a-h. Note that both units evince significant match/nonmatch suppression at position 2, but this is not observed at position 3 after the intervening nonmatch (panels d and l vs. e and m). By contrast, generalized suppression of responses following the sample is evident at positions 2 and 3 (panels f-g, n-o). This suppression may be driven by adaptation that is not specific for feature identity, but sensitive to shared features between sample and nonmatch. The match response in the second unit is entirely suppressed in AA trials (i), but the onset component of the response recovers after one or two intervening nonmatch stimuli (panels j, k). This effect on the onset component appears to be stimulus-specific, in that it is seen for the match/nonmatch comparison (l), but not for the sample/nonmatch comparison (n,o). The sustained component of the response, from 100-300 ms, is present in responses to the sample only and is suppressed in all subsequent responses (match or nonmatch).

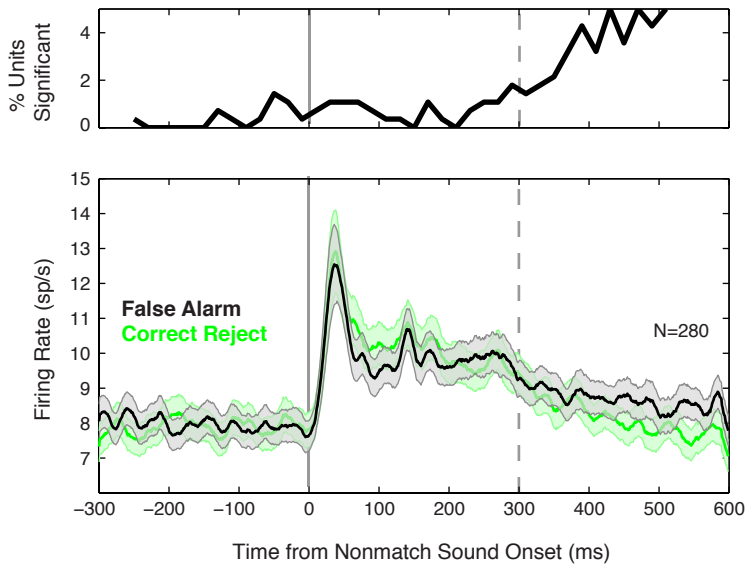

Figure S2: (related to Figure 4F) By the same criterion used to define match suppression or match enhancement (see Experimental Procedures), firing rates were compared between responses to nonmatch stimuli that elicited an incorrect bar release (false alarm), and those that did not (correct reject). Because no match has occurred in either condition, a difference in firing rate may be taken to reflect motor preparation, but not a memory or match effect *per se*. Lower panel: mean firing rates (±SEM) among all units in the population (N=280) on nonmatch presentations that were followed by bar release (black trace) or not (green trace). The number of trials per stimulus is equated between conditions within each unit. Upper panel: percentage of units in the population showing an effect, which did not exceed 2% during the epoch used to define MS and ME (0-300 ms). Among the 53 units showing MS or ME, only one (an MS unit in field TGd, from monkey K) showed a significant difference between false alarm and correct reject trials, and an increase in firing rate when spikes were aligned to the time of the bar relase. That unit also showed an effect of reward expectancy, responding only to unexpected rewards during passive listening, demonstrating that several task-relevant variables may be represented in the firing rates of neurons in higher auditory cortex.

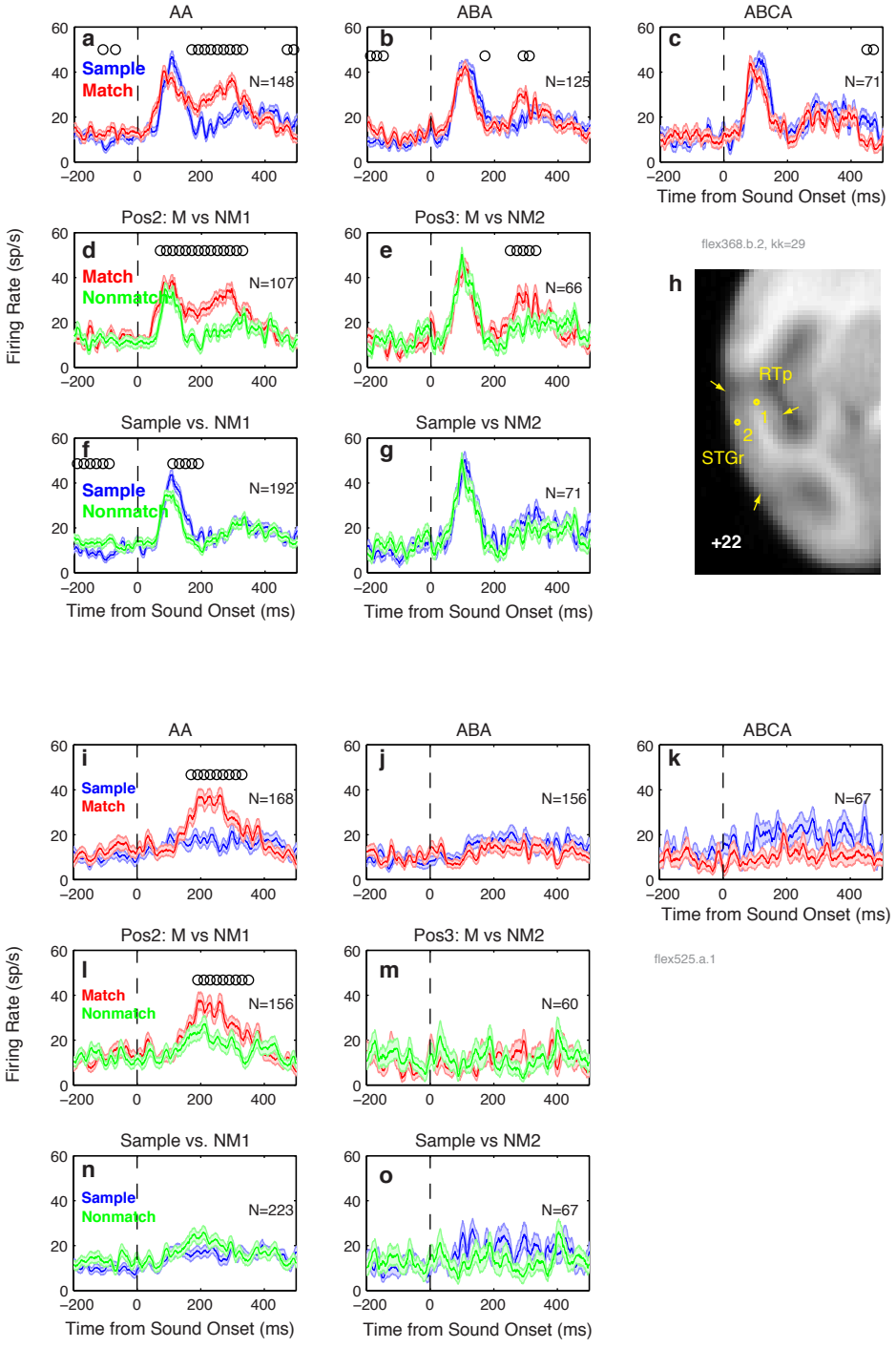

Figure S3: (related to Figures 4 & 5) Two units showing ME, which in both cases appeared only > 100 ms after the onset of the match sound, and was observed only on trials with no intervening nonmatch. Figure conventions as in Supplementary Figure S1. **(a-h)** This unit in field RTp (unit 1 in panel h) evinced reliable ME only on AA trials and only after the initial onset transient, which was identical between sample and match (a). This enhancement was clearly match-specific, as it was also observed in the match/nonmatch comparison at position 2 (d), whereas nonmatch responses were actually suppressed relative to sample (f). This same unit is depicted in Figure 7A and B of the main text. **(i-o)** This unit in the STGr (unit 2 in panel h) shows a clear enhancement of the match response relative to the sample response (i) or to the nonmatch response (l), beginning ~100 ms after sound onset, and is not seen later in the trial. Although the sample and nonmatch responses appear ineffective, this is an artifact of scaling the axes to fit the match response; the sample response for this unit is better illustrated in Figure 7D of the main text.

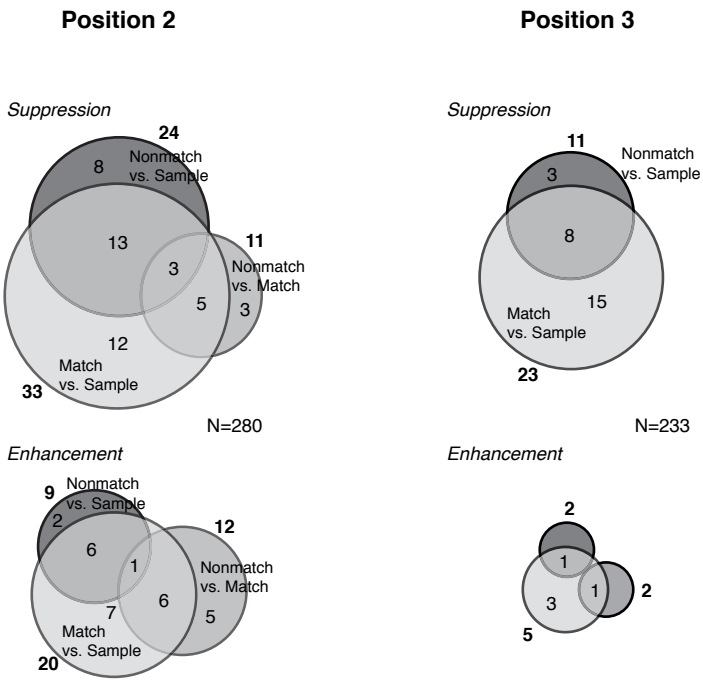

Figure S4: (related to Figures 5 & 6) Venn diagrams summarize the number of units showing suppression or enhancement effects, for three comparisons. Left column: Sample and match responses are taken from AA trials, and nonmatch responses are from position 2 (the same trial position as the match). Right column: Sample and match responses are taken from ABA trials, and nonmatch responses are from position 3 (the same trial position as the match). The MS and ME populations (see Fig. 5) were defined by the match/sample comparison (lightest gray), of which a subset also showed an effect in the match/nonmatch comparison (medium gray), and/or the nonmatch/sample comparison (dark gray). Due to the limited number of correct trials with nonmatch presentations, these two control comparisons are of lower statistical power than the sample/match comparison, but they serve to reveal the underlying processes driving match suppression. Note that the number of trials was not equated between position 2 and position 3, as was done to compare effects across trial position in Figure 6A of the main text.

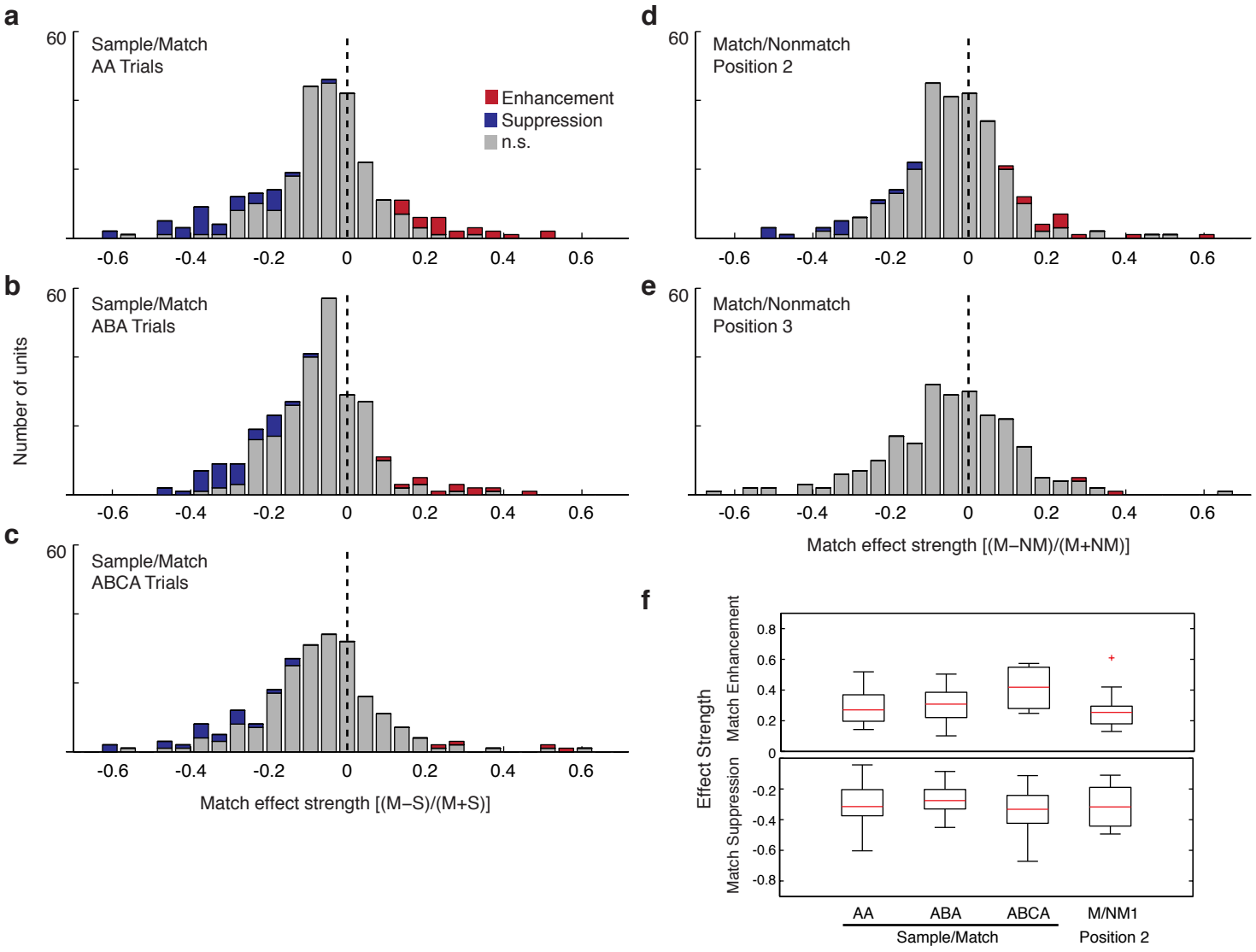

Figure S5: (related to Figures 5 & 6) Although the number of units evincing ME was reduced on longer trials (see Fig. 6A of the main text), the relative strength of MS and ME was consistent across trial type. Histograms plot the magnitude of match suppression and enhancement effects in individual units for AA, ABA, and ABCA trials (**panels a-c**, respectively), and for the match/nonmatch comparison at position 2 (**panel d**) and at position 3 (**panel e**). Effect magnitude was measured as the ratio of the difference in firing rates between conditions to the sum of firing rates between conditions (e.g., [FRmatch – FRsample]/[FRmatch + FRsample]). Negative values indicate match suppression, and positive values indicate match enhancement; there was an overall skew toward negative values, as predicted from the population firing rates in Figure 4. Blue and red bars count units with significant MS and ME, respectively; for these units, firing rates were measured over the range of contiguous 100-ms time bins showing a significant difference in firing rate ($p < 0.001$ by Wilcoxon rank-sum test; see Methods). Gray bars count units for which the firing rate difference was not significant; for these units, firing rates were measured over 0-300 ms from stimulus onset. **(f)** Boxplots of effect strength across trial type, for MS (lower plot) and ME (upper plot). The median size of a significant effect was ~0.3 in all comparisons, and did not differ across the four comparisons either for MS (ANOVA, $F = 1.45$, $p = 0.23$) or for ME ($F = 1.2$, $p = 0.32$). There is a trend toward a stronger effect in ABCA trials than in AA or ABA trials, but this may result from the limited number of correct ABCA trials available for analysis; due to limited statistical power, a larger effect size would be necessary to detect a significant difference.

Table S1: Distribution of DMS effects by cortical field (related to Fig. 1B)

		Unit Counts by hemisphere					*DMS Effects*									
	Field	F left	K left	S right	S left	**Total**	DS	DS %	DE	DE %	MS	MS %	ME	ME %	Any fx	Any %
Group 1	R	9	0	17	0	26	4	15%	4	15%	2	8%	5	19%	11	42%
	RT	9	3	29	4	45	7	16%	5	11%	4	9%	3	7%	14	31%
	RM	0	0	4	0	4	1	25%	1	25%	0	0%	1	25%	2	50%
	AL	0	0	0	0	0	0	-	0	-	0	-	0	-	0	-
	RTM	2	1	12	2	17	1	6%	5	29%	0	0%	1	6%	6	35%
	RTL	2	0	4	3	9	1	11%	3	33%	2	22%	0	0%	5	56%
Group 2	RTp	25	0	25	6	56	13	23%	10	18%	10	18%	4	7%	29	52%
	RPB	2	5	0	3	10	2	20%	0	0%	0	0%	0	0%	2	20%
	STGr	15	24	10	8	57	10	18%	14	25%	11	19%	3	5%	27	47%
	TGd	9	3	0	4	16	6	38%	0	0%	2	13%	1	6%	7	44%
	TAa/TPO	10	4	14	0	28	3	11%	5	18%	1	4%	0	0%	8	29%
	insula	0	6	0	0	6	2	33%	0	0%	1	17%	0	0%	2	33%
	other	2	1	2	1	6	0	0%	1	17%	0	0%	2	33%	3	50%
	Total:	**85**	**47**	**117**	**31**	**280**	**50**	18%	**48**	17%	**33**	12%	**20**	7%	**116**	41%
	Group 1	22	4	66	9	101	14	14%	18	18%	8	8%	10	10%	38	38%
	Group 2	61	36	49	21	167	34	20%	29	17%	24	14%	8	5%	73	44%

Table S1 presents the locations of all recorded units in each hemisphere (left columns) and all task-related effects (right columns). Unit counts for delay suppression, delay enhancement, match suppression, and match enhancement are presented as raw counts, and as a percentage of total units in each field. Effect prevalence was compared between Group one (core and belt) and group 2 (higher-order regions), though no strong difference was seen (see main text). Field locations are shown in coronal MRI sections in Figure 1B, and in a schematic diagram of the supratemporal plane and superior tempral gyrus at right; color roughly indicates hierarchical level: core (dark blue), belt (medium blue), parabelt, STGr, and dorsal temporal pole (TGd; light blue), field RTp (purple), and the dorsal bank of the STS (green). Nomenclature follows Hackett (2010) and Saleem and Logothetis 2012 (for abbreviations see Figure 1B).

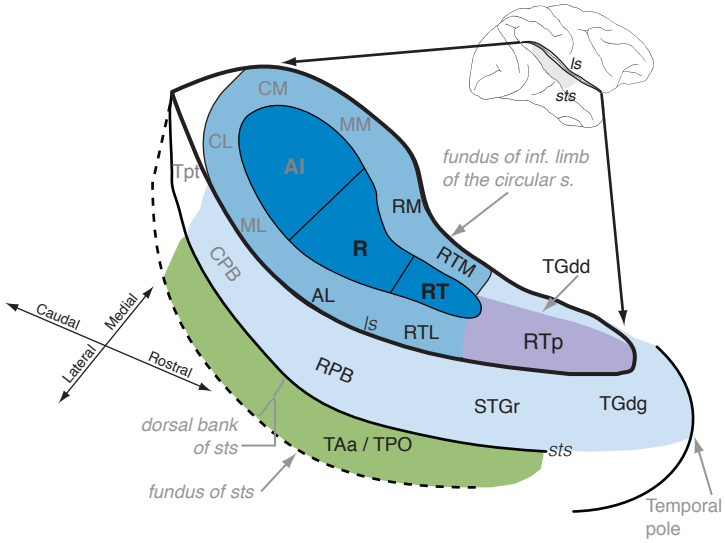

Table S2: Co-occurrence of match and delay effects by unit (related to Figure 5)

Delay 1/ Match 1				
	Delay:			
Match:	DS	DE	No effect	Total
MS	12**	5	16	33
ME	2	10**	8	20
No effect	22	27	178	227
Total	36	42	202	280
Delay 2/ Match 2				
	Delay:			
Match:	DS	DE	No effect	Total
MS	8**	0	25	33
ME	0	5**	5	10
No effect	17	11	209	237
Total	25	16	239	280

The hypothesis that delay modulation and match modulation at the subsequent stimulus position are independent can be rejected for both delay 1/match 1, and delay 2/match 2 (χ^2, $p < 10^{-7}$). Dependency between DS and MS, and between DE and ME, could arise if match effects were caused, at least in part, by the ongoing delay modulation. That is, because responses were measured as absolute firing rates, a match response 'riding' atop elevated delay activity may appear to be ME, though in fact both sample and match elicit responses of similar magnitude (relative to their respective pre-stimulus baselines). However, subtracting the pre-stimulus baseline firing rate did not eliminate the dependence between match and delay effects at either trial position (χ^2, $p < 0.002$). Furthermore, among units in which subtracting baseline did eliminate the match effect, there was no clear tendency for such units to show the corresponding delay effect. Thus, although the effects are not independent, modulation of the match response cannot be explained merely as a by-product of ongoing delay-period modulation. MS, match suppression; ME, match enhancement; DS, delay suppression; DE, delay enhancement. **significantly different from chance by binomial test, $p \leq 0.003$.

Supplemental Experimental Procedures

Behavioral task

The trial sequence is shown in Figure 1A. The monkey initiated a trial by holding a contact bar for 300 ms, after which a sample stimulus (~300 ms duration) was presented, followed by 1-3 test sounds at an interstimulus delay of ~1 s. The delay was randomized over a range of 800-1200 ms so that the exact time of sound onset could not be predicted; this forced the animals to rely on the actual onset of the sound to guide their response timing and may have lessened repetition-suppression effects due to temporal predictability [1]. When the test sound was identical to the sample (a match), the monkey could release the bar within a 1200-ms window beginning 100 ms after match onset to earn a liquid reward delivered 300 ms after bar release. If the stimulus was a nonmatch, the animal was required to continue holding until the match appeared. Release following a nonmatch or failure to release after the match was counted as an error and initiated a punitive 3-s timeout in addition to the standard 3-s intertrial interval. The number of nonmatch stimuli was 0, 1, or 2, presented pseudorandomly with equal probability (referred to as AA, ABA, and ABCA trials, respectively). 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. Despite extensive training, monkey K could not reach 70% overall performance when all three trial types were used and so was tested on AA and ABA trials only; thus, for this animal, only trial position 2 presented a match/nonmatch choice. Behavioral data in this report are from the same sessions during which the unit activity was recorded. Our previously published behavioral data [2, 3] were collected during interleaved sessions when recording was not performed, and thus reflect an independent data sample.

The standard stimulus set consisted of 21 sounds, each ~300 ms in duration. The set included three exemplars for each of seven categories: (1) temporally orthogonal ripple complexes (TORCs); (2) 1/3-octave band-pass noises (BPN) at center frequencies of 512, 2048, and 8192 Hz; (3) pure tones (PT) at the same frequencies; (4) frequency-modulated sweeps (FM) − upward, downward, and bi-directional − between 0.25 and 16 kHz; (5) rhesus monkey vocalizations (Mvoc) − archscream, bark, and coo; (6) other species' vocalizations (voc) – dog mbark, bird song, and female human vowel /a/; (7) environmental sounds (env) − cage door

closing, click of water solenoid opening, and metallic noise. All synthetic sounds were 300 ms in duration, whereas the duration of the natural sounds varied slightly (Mvocs tended to be shorter than the other categories: 282 ms, 246 ms, and 195 ms). A variant of this stimulus set was presented to monkey F during recording from a subset of units (N = 30/280); these consisted of five exemplars each from four categories: TORC, BPN, PT, and Mvoc (each category included the same three sounds described above, plus two additional exemplars). All stimuli were equalized in root-mean-square amplitude to have approximately equal loudness and were presented at 60 or 70 dB SPL via a loudspeaker (Ohm Acoustics, NY) located 1 m directly in front of the animal. Sample sounds were chosen pseudorandomly in blocks, such that each served as sample (and match) in one correct trial before the order was reshuffled for the next block.

Physiological recording

The behavioral task was controlled by NIMH-Cortex software (dally.nimh.nih.gov), which triggered sound playback via a custom-built interface with a second computer running SIGNAL software (Engineering Design, www.engdes.com/). For recordings in the left hemisphere of monkey S, the Cortex/SIGNAL system was replaced by Presentation software (Neurobehavioral Systems, www.neurobs.com). The output of the SIGNAL buffers (or Presentation PC sound card) was flattened across frequency (Rane RPM 26v parametric equalizer, Mukilteo, WA), attenuated (Agilent HP 355C and 355D), amplified (NAD, Pickering, Ontario), and delivered via a loudspeaker (Ohm Acoustics, NY) located 1 m directly in front of the animal's head. Sound level was calibrated with a Brüel and Kjær 2237 sound-level meter using A-weighting.

After behavioral training was complete, an MRI-compatible recording chamber (19-mm diameter, 6-ICO-J2A, Crist Instruments) was implanted under aseptic surgical conditions to allow a vertical approach to the supratemporal plane (STP) and lateral surface of the superior temporal gyrus (STG) rostral to the primary auditory cortex (Fig. 1B). The center of the chamber was placed 20-24 mm anterior to the interaural axis (ear bar zero), and as lateral as possible to allow access to the STG. In monkeys F and K, recordings were made in the left hemisphere, whereas in monkey S, recordings were made first in the right hemisphere, at two different

chamber placements, and then in the left. Perhaps owing to earlier recordings from caudal auditory cortex in monkey S's left hemisphere [4], the yield of responsive neurons in this animal's left rostral auditory cortex was somewhat lower than that in the other hemispheres studied.

In daily sessions, epoxy-insulated tungsten electrodes (1-3 MΩ, FHC) were lowered through a 1-mm grid (6- YGD-D2, Crist) using sharpened stainless steel guide tubes. Electrodes were lowered under computer control, and tracks were guided by alignment to an MR image acquired after implantation of the chamber. Recordings in three hemispheres were carried out using a single-electrode hydraulic microdrive (Narishige, Japan), but up to three electrodes were used simultaneously in the left hemisphere of monkey S (NAN system, Plexon). Neuronal activity was amplified and filtered (TDT Bioamp, or RZ2) into a high-frequency signal (~500-8000 Hz, sampled at 25 kHz) from which spike waveforms were extracted. Physiological data and task events were acquired and saved to disk through a Power-1401 acquisition system controlled by Spike2 software (Cambridge Electronic Design). 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 stable unit(s) was isolated, the 21 sounds were presented in pseudorandom order at least 8-10 times with an inter-stimulus interval (ISI) of 2.5 s as the animal sat passively. A liquid reward was delivered randomly between stimuli with a probability of 0.1 per interval. If a unit evinced an auditory-evoked response, then the animal was presented with the DMS paradigm. (During recordings from the left hemisphere of monkey S, the DMS paradigm was presented first, but this did not appear to increase the yield of auditory-responsive neurons.) For most units (245/280) the full stimulus set was used; in the remainder, a subset (median, N = 9) of the most effective stimuli was selected to serve as the samples (although nonmatch sounds were still drawn from the full set). As long as the unit waveform was stable, the monkey was allowed to work to satiety, or until at least ten correct trials were completed for each sample stimulus.

In order to assign recording sites to cortical fields (Fig. 1B, Supplementary Table S1), sites were aligned to the left hemisphere of an averaged MRI template for macaque [5], registered to a

combined MRI and histology atlas [6]. The field boundaries corresponding to tonotopic reversals (R/RT, and RT/RTp) were modified from the atlas to accord with observed frequency tuning to pure tones or band-pass noise (presented under passive listening conditions), and the dimensions of these fields in the published physiology and imaging literature (e.g.,[4, 7, 8]).

Data analysis

Auditory responses were identified by two methods. First, for each stimulus, spike counts during the stimulus period (from 0-300 ms after stimulus onset, combining sample, nonmatch, and match sounds) were compared to the baseline spike counts (starting 300 ms before stimulus onset) by a Wilcoxon rank-sum test ($p < 0.05$, Bonferroni corrected for the number of stimuli in the set). Second, to identify transient responses, spike times were binned at 1-ms resolution and convolved with an exponential kernel to generate a spike-density function; a response was considered significant if the spike rate exceeded 2.8 standard deviations (SD; $p < 0.005$) above the baseline rate for 10 consecutive bins. Memory effects were investigated in the subset of units that was responsive to at least one stimulus by either test, and was recorded during valid DMS behavior (> 100 trials, performance > 2 SD above chance). Response latency was taken as the time at which the spike density function exceeded 3 SD above the mean baseline firing rate. Latency was computed using all stimuli and the subset of effective stimuli (as indicated by the tests above), and the shorter value was chosen. If both excitatory and inhibitory deviations were identified, again the shorter value was taken.

Delay intervals were defined as the 800 ms preceding the onset of the match or nonmatch stimulus at position 2, 3, or 4 (Fig. 2; because the actual ISI varied from 800-1200 ms, only the final 800 ms was extracted to average across trials). To identify units with delay activity, spike counts were measured over the last 600 ms of each delay period, and compared to spike counts during the 600 ms preceding the sample (Wilcoxon rank-sum, $p < 0.05$ corrected for multiple comparisons). For display purposes only, smoothed spike density functions were generated by binning spike times at 1-ms resolution, convolving with a Gaussian kernel (σ=20 ms), and normalizing to the pre-trial baseline firing rate (Fig. 2B,C).

Responses to acoustically identical sample, match, and nonmatch presentations were compared to identify modulation of the sensory response by task context (Figs. 3-5). For each unit, responses from correct trials were segregated by condition (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 conditions. 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, generating a vector of 40 p-values spanning -300 to 480 ms relative to stimulus onset. A unit was classified as showing an effect if two adjacent bins between 0 and 300 ms were significantly different between conditions ($p < 0.01$, Bonferroni corrected for overlap of time bins). Effects after 300 ms were excluded due to potential confounding with the motor response and reward delivery. The same method and criterion were applied to compare responses to match and nonmatch responses at position 2 (Fig. 3D), and, separately, at position 3 (Fig. 3E); finally, sample responses were compared to nonmatch responses at position 2 (Fig. 3G) and at position 3 (Fig. 3H).

The analyses above included data from all correct trials, which precludes a fair comparison of memory effects across the different epochs of the trial. Because each trial is sequential, and error responses on long trials were common, there is more statistical power at delay 1, or the position 2 match/nonmatch, than at later trial positions. In addition, reaction times for match stimuli at position 4 were significantly shorter than those at positions 2 and 3, suggesting that monkeys learned that the fourth stimulus was always a match and anticipated their response; these anticipatory effects may be unrelated to the memory of the sample stimulus itself. To eliminate these confounds, the analysis in Figure 6A systematically excluded a subset of the data. (1) Delay activity was sampled only from correct ABA and ABCA trials, so each trial contributed a firing rate value for delay 1 and delay 2 (this excluded units from monkey K, who did not perform ABCA trials). (2) Match effects were examined at positions 2 and 3 only, selecting an equal number of trials, and an identical distribution of stimuli, within each unit. For these reasons, the number of included units (233/280) and overall prevalence of effects is slightly lower in Figure 6A than in prior figures.

To assess the relative influence of sensory and task factors on firing rate across time, an ANOVA model was applied in a 100-ms sliding window. Firing rate during the sample presentation and subsequent delay were analyzed with a single factor (sample identity, 1-21; Fig. 7 B,C). For the trial epoch surrounding the position 2 stimulus (Fig. 7 E,F) factors included: the identity of the preceding sample (1-21), the match/nonmatch condition at position 2 (a dummy variable of 0 or 1), and nested within that factor, the identity of the stimulus at position 2 (1-21). The single-factor model was also applied to a subset of 120 units recorded during both passive listening and DMS performance, to determine whether the time course of stimulus encoding differed between task conditions. For each unit, an equivalent number of trials per stimulus was drawn from each condition, using only sample presentations during behavior. The ANOVA models were applied to spike counts measured in a 100-ms sliding window moved in 20-ms steps (Matlab Statistics Toolbox 'anovan', using Type III sum of squares). To control for false discovery rate (FDR), p-values from all time bins, factors, and units within each model were sorted, and the 95th percentile taken to indicate an alpha of 0.05 (in practice ~0.005). Variance explained by each factor was measured by partial ω^2, an unbiased alternative to eta-squared:

$$\omega^2_{partial} = (SSQeffect - df_{effect} * MS_{error}) / (SSQ_{effect} + [N - df_{effect}] * MS_{error})$$

where SSQ is the sum of squares, df the degrees of freedom, N the sample size, and MS the mean-squared error. For averaging across the population, values of ω^2 below the FDR-corrected threshold were set to zero, and standard error of the mean was measured across units. The proportion of units showing a significant effect was reliably near zero for time bins preceding the stimulus onset, verifying that this correction was sufficient to quash spurious false positives.

Control analyses: motor and reward effects

We attribute the suppression or enhancement of responses to match stimuli as a potential signal for the repetition of the sample stimulus in the context of an auditory memory task. However, in addition to the repetition itself, the match presentation also triggers a motor response, i.e. the release of a touch bar, in correct trials. In turn, this motor action is associated with the delivery of a fluid reward 300 ms later. Previous reports have identified spiking activity related to bar release in the auditory cortex of monkeys over-trained on an auditory task [9], which could present a confound with the MS and ME we report. To identify the effect of motor preparation and reward expectation on firing rate, we compared responses to identical distributions of

nonmatch stimuli sorted by the ensuing behavioral response: a correct reject (holding the bar), or a false alarm (incorrect bar release). This procedure was identical to that used to identify MS and ME above: firing rates 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 conditions ($p < 0.01$, Bonferroni corrected). Because no match occurred in either condition, a difference in firing rate may be interpreted to reflect motor preparation, and/or reward expectation, but not a memory or match effect per se. As seen in Figure S5, the difference in the population firing rate between correct reject and false alarm trials was negligible, and no more than 2% of 280 units showed a significant effect. As noted in the main text, of the 53 units that exhibited MS or ME, only one showed a significant effect in this comparison.

Control analyses: modulations during passive listening

If the modulations of delay and match activity we report were exclusively related to the performance of the auditory DMS task, we should not expect to observe similar effects during passive stimulus presentation. Under passive conditions, sounds were presented in pseudorandom order with an ISI of 2.5 s, with a 10% chance of reward delivery after each sound. These conditions did not replicate the sequence and timing used during DMS testing, but these differences reduce the likelihood that the subjects were "covertly" performing the task during passive trials.

Delay activity was examined by comparing 'baseline' firing rate during the 600 ms preceding sound onset to a 'delay' epoch spanning 500-1100 ms after sound onset, using the same criteria applied during behavioral performance (Wilcoxon rank-sum, $p < 0.05$ corrected for multiple comparisons; rewarded trials were excluded from analysis). An effect was observed in 15% of units (20/133), lower than the 28% that showed an effect during the first delay epoch of the DMS task (χ^2, $p = 0.004$). However, the percentages being compared encompass largely independent populations of neurons. Of the 38 units that showed delay modulation during the task and were also recorded during passive listening, only 12% of DE units, and 24% of DS units, showed the corresponding effect during passive listening. Furthermore, if the departure from baseline activity during the delay were merely an aftereffect of the sensory response, it should be present

within every delay in the trial sequence. This was not the case for DE, which was significantly reduced after delay 1 (Fig. 6). Taken together, these results suggest that the delay modulation we observed – particularly delay enhancement – is related to performance of the DMS task, as it is not consistent with the firing rate modulations observed during passive listening.

The shifts in firing rate during passive listening may result from the intermittent, probabilistic reward schedule (in this sense, the animal may not have been truly 'passive'). A similar proportion (~ 15%) of units in AI has been reported to show tonic shifts in firing rate related to the omission of an expected reward, or expectation of reward size [10]. A subset of these neurons appears to be recruited in performance of DMS, as may be expected for any reward-driven auditory/cognitive task. However, as shown above, 88% of DE units and 76% of DS units showed these effects only during performance of the DMS task.

To simulate "match suppression" in the passively collected data, trials were split into two arbitrary groups by segregating the even- and odd-numbered presentations of each stimulus. These groups were subjected to the same statistical procedure used to identify MS or ME in the data collected during DMS, above: firing rates 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 conditions ($p < 0.01$, Bonferroni corrected). For comparison, an identical number of trials was selected from the behaving condition and re-analyzed (typically, the trial count per stimulus was lower in the split passive data set than the full behaving data set). In the subset of behavior trials, 9% showed MS or ME (vs. 19% when all trials were used), but no units showed any significant effect in the passive control condition.

Supplemental References

1. Costa-Faidella, J., Baldeweg, T., Grimm, S., and Escera, C. (2011). Interactions between "what" and "when" in the auditory system: temporal predictability enhances repetition suppression. J Neurosci *31*, 18590-18597.
2. Scott, B.H., Mishkin, M., and Yin, P. (2012). Monkeys have a limited form of short-term memory in audition. Proc Natl Acad Sci U S A *109*, 12237-12241.

3. Scott, B.H., Mishkin, M., and Yin, P. (2013). Effect of acoustic similarity on short-term auditory memory in the monkey. Hear Res *298*, 36-48.
4. Yin, P., Mishkin, M., Sutter, M., and Fritz, J.B. (2008). Early stages of melody processing: stimulus-sequence and task-dependent neuronal activity in monkey auditory cortical fields A1 and R. J Neurophysiol *100*, 3009-3029.
5. McLaren, D.G., Kosmatka, K.J., Oakes, T.R., Kroenke, C.D., Kohama, S.G., Matochik, J.A., Ingram, D.K., and Johnson, S.C. (2009). A population-average MRI-based atlas collection of the rhesus macaque. Neuroimage *45*, 52-59.
6. Saleem, K.S., and Logothetis, N.K. (2012). A combined MRI and histology atlas of the rhesus monkey brain in stereotaxic coordinates. 2nd edition with Horizontal, Coronal and Sagittal series, (San Diego, CA: Elsevier/Academic press).
7. Fukushima, M., Saunders, R.C., Leopold, D.A., Mishkin, M., and Averbeck, B.B. (2012). Spontaneous high-gamma band activity reflects functional organization of auditory cortex in the awake macaque. Neuron *74*, 899-910.
8. Petkov, C.I., Kayser, C., Augath, M., and Logothetis, N.K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. PLoS Biol *4*, e215.
9. Brosch, M., Selezneva, E., and Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J Neurosci *25*, 6797-6806.
10. Brosch, M., Selezneva, E., and Scheich, H. (2011). Representation of reward feedback in primate auditory cortex. Frontiers in systems neuroscience *5*, 5.