Published in final edited form as:
Hear Res. 2014 March ; 309: 17–25. doi:10.1016/j.heares.2013.11.001.

Behavioral correlates of auditory streaming in rhesus macaques

Kate L. Christison-Lagay[1] and **Yale E. Cohen**[2]

[1]Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, PA 19104

[2]Dept. Otorhinolaryngology, Perelman School of Medicine, U. Pennsylvania, Philadelphia, PA 19104

Abstract

Perceptual representations of auditory stimuli (i.e., sounds) are derived from the auditory system's ability to segregate and group the spectral, temporal, and spatial features of auditory stimuli—a process called "auditory scene analysis". Psychophysical studies have identified several of the principles and mechanisms that underlie a listener's ability to segregate and group acoustic stimuli. One important psychophysical task that has illuminated many of these principles and mechanisms is the "streaming" task. Despite the wide use of this task to study psychophysical mechanisms of human audition, no studies have explicitly tested the streaming abilities of non-human animals using the standard methodologies employed in human-audition studies. Here, we trained rhesus macaques to participate in the streaming task using methodologies and controls similar to those presented in previous human studies. Overall, we found that the monkeys' behavioral reports were qualitatively consistent with those of human listeners, thus suggesting that this task may be a valuable tool for future neurophysiological studies.

1. Introduction

One of the fundamental tasks of the auditory system is to transform low-level sensory representations of acoustic stimuli into perceptual representations (i.e., sounds) that can guide behavior (Bizley et al., In Press; Griffiths et al., 2004; Shamma et al., 2010; Zatorre et al., 2004). These perceptual representations form the core building blocks of our hearing experience (Bregman, 1990; Griffiths et al., 2004; Shamma, 2008) and are derived from the auditory system's ability to segregate and group the spectral, temporal, and spatial features of auditory stimuli—a process called "auditory scene analysis" (Bregman, 1990; McDermott, 2009; Winkler et al., 2009). Auditory scene analysis enables a listener to follow, for example, the melody that is carried by a banjo in a band or to track a friend's voice in a noisy restaurant (McDermott, 2009; Shinn- Cunningham, 2008).

Corresponding author: Kate Christison-Lagay, Department of Otorhinolaryngology, 3400 Spruce St – 5 Ravdin, Philadelphia, PA 19104, katechri@mail.med.upenn.edu phone: 1 215 898 7504, fax: 1 215 898 9994.

Psychophysical studies have identified several of the principles and mechanisms that underlie a listener's ability to segregate and group acoustic stimuli (Horvath et al., 2001; Rahne et al., 2009; Sussman, 2005; Sussman et al., 2007). One important psychophysical task that has illuminated many of these principles and mechanisms is the "streaming" task (Bregman, 1990; Carlyon et al., 2001; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007). Typically, the streaming task is a one-interval, two-alternative forced choice task in which an auditory stimulus—composed of an interleaved sequence of tone bursts (Fig. 1)—is presented and a listener reports whether she heard one or two streams. By varying the spectral, temporal, and other properties of this sequence, the probability that a listener reports one or two streams is systematically altered. For example, when the frequency difference between the tone bursts in the two sequences is small (e.g., ≤1 semitone difference), listeners systematically report hearing one stream. On the other hand, when the frequency difference between these tone-burst sequences is large (e.g., ≥10 semitones), listeners systematically report hearing two separate streams. When the frequency difference is intermediate between these two extremes, the reports become less reliable: on alternating trials, listeners report hearing one or two streams.

Despite the wide use of this task (and variants of it) to study psychophysical mechanisms of human audition (Shamma et al., 2011), no studies have explicitly tested the streaming abilities of non-human animals using the standard methodologies employed in human-audition studies. Instead, previous studies have indirectly tested streaming (Izumi, 2002; Ma et al., 2010; Moerel et al., 2012). For example, in Ma et al. (2010), ferrets reported hearing a "target" tone that was embedded in a tone-burst sequence. This experimental strategy to test streaming is reasonable because many non-human animals process auditory stimuli and hear sounds in a manner similar to that of human listeners (Izumi, 2002; Kuhl et al., 1975; Kuhl et al., 1982; Miller et al., 2001; Petkov et al., 2003; Petkov et al., 2007; Recanzone et al., 2008). Consequently, it was assumed that, like humans (Elhilali et al., 2009), these ferret listeners could only detect the target tone when the auditory stimulus was segregated into two streams.

However, if the goal of testing the auditory perceptual abilities of non-human animals is to develop them as models of human-brain function, it is imperative to use methodologies and controls that are comparable to those used with human listeners so that valid inferences can be made regarding human audition and cognition. Here, we trained rhesus macaques to participate in a streaming task using methodologies and controls similar to those presented in previous human studies. Overall, we found that the monkeys' behavioral reports were consistent with those of human listeners, thus suggesting that this task may be a valuable tool for future neurophysiological studies.

2. Methods

The University of Pennsylvania Institutional Animal Care and Use Committee approved the experimental protocols.

2.1 Experimental Chamber

Psychophysical sessions were conducted in a darkened room with sound-attenuating walls. A monkey (*Macaca mulatta*; Monkey H [male, 12 years old] or Monkey S [male, 8 years old]) was seated in a primate chair in the center of the room. A touch-sensitive joystick was attached to the chair. The monkey moved the joystick during the behavioral task to indicate his behavioral report.

2.2 Auditory Stimulus

The auditory stimulus was a sequence tone bursts (40-ms duration with a 5-ms $\cos^2$ ramp at a sound level of 65 dB SPL) that alternated between two types of tone bursts, called here "tone A" and "tone B". The inter-tone-burst interval was 13 Hz. Auditory stimuli were generated using the RX6 digital-signal-processing platform (TDT Inc.) and were presented by a studio-monitor speaker (Yamaha MSP7).

2.3 Behavioral Task

The streaming task was a single-interval, two-alternative-forced-choice discrimination task that required the monkey to report whether he heard one or two streams (Fig. 2). A trial began with the presentation of the auditory sequence (Fig. 1). Following offset of the auditory stimulus, an LED was illuminated, and the monkey had 3000 ms to move the joystick (a) to the right to report one stream or (b) to the left to report two streams.

2.4 Training Procedure and Reward Structure

During the initial training sessions, tones A and B were presented at frequency differences that, in humans (Cusack, 2005; Micheyl et al., 2005), elicit reliable reports of one or two streams (i.e., ≤1.0 or ≥10 semitones, respectively); this strategy was appropriate because macaque monkeys and humans have similar hearing (Brown et al., 1980; Kuhl et al., 1983; Petkov et al., 2003; Pfingst et al., 1978; Recanzone et al., 2008; Recanzone et al., 2000; Serafin et al., 1982; Sinnott et al., 1976; Tsunada et al., 2011). On these trials, the monkey received consistent feedback: he was only rewarded for reporting a "correct" response. Specifically, when the frequency difference between tone A and tone B was ≤1.0 semitone, the monkey was rewarded when he moved the joystick to the right. When the frequency difference was ≥10 semitones, the monkey was rewarded when he moved the joystick to the left.

After the monkey's performance stabilized (i.e., they were performing significantly above chance during entire behavioral sessions for 3 or more days a week; monkeys were trained 5 days a week with ~200 trials a day for 2 years before their performance stabilized), we presented auditory sequences that contained both the "extreme" frequency differences (≤1.0 or ≥10 semitones) as well as frequency differences that were "intermediate" between these two extremes (i.e., >1 and <10 semitones). Because stimuli with these intermediate frequency differences do not elicit reliable reports of one or two streams in human listeners (Bregman, 1990; Bregman et al., 2000; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007), there was not a "correct" answer. Consequently, on these trials, the monkeys did not

receive consistent feedback: they received rewards on 50% of randomly selected trials; the decision to reward was independent of their behavioral report.

2.5 Behavioral-testing Strategy

We manipulated four parameters of the tone-sequence: the frequency difference between tones A and B; the duration of the auditory sequence; the temporal relationship between tones A and B; and the frequency of tone A. These first three parameters manipulations tested whether the monkeys' reports were modulated in a manner consistent with human listeners' reports (Bregman, 1990; Elhilali et al., 2009; Micheyl et al., 2007). The last parameter manipulation controlled for the possibility that the monkeys were not actually reporting the number of heard streams but, instead, reported two streams whenever they heard a stimulus that contained high frequencies.

Next, we describe the details of these manipulations. First, on a trial-by-trial basis, we randomly varied the frequency difference between tones A and B. During ~93% of these trials, we presented those frequency differences that provided consistent feedback (i.e., ≤1.0 or ≥10 semitones). For the remaining trials (~7% of the trials or ~44 trials/day), we presented those frequency differences that did not provide consistent feedback (i.e., >1 and <10 semitones). Second, on a trial-by-trial basis, we randomly varied tone A's frequency (range: 865–2226 Hz; mean: 1500 Hz). Third, on a trial-by-trial basis, we varied the sequence duration (i.e., "listening duration"; 180–2022 ms; mean: 778 ms). Fourth, on a subset of days, we manipulated the temporal relationship between tones A and B. On most days, tones A and B were presented in their standard asynchronous format; see Figure 1. However, on select days, tones A and B were presented simultaneously on a randomly subset of trials (~27%); see Figure 1 inset. The time between the onsets of the simultaneous was 13 Hz, the same as the asynchronous timing. When tones A and B were presented simultaneously, their frequency difference was always 10 semitones. For the simultaneous trials, the monkeys received rewards independent of their behavioral response.

2.6 Data Analyses

We quantified the monkeys' performance by calculating the probability of the monkey reporting two streams (i.e., the monkey moved the joystick to the left). This analysis was conducted as a function of the (a) the frequency difference between tones A and B, (b) the frequency of tone A, (c) listening duration, and (d) the temporal relationship between tones A and B. The 95%-confidence interval on each of these probability values was calculated using the following formula: $1.96*(p*(1-p)/n)^{0.5}$ (Zar, 1996). p was the probability (i.e., the proportion of trials when the monkey reported two streams), and n was the number of trials. The monkeys' performance was considered reliable when the 95%-confidence interval did not overlap with chance performance (i.e., 0.5). A Wilcoxon test was also used to determine whether a probability value differed from chance; the *p*-values that are reported in the text reflect the results of this test. Probability values that were generated from different stimulus-parameter manipulations (e.g., for the upper half of listening durations and the lower half of listening durations) were considered to be significantly ($p<0.05$) different when the 95%-confidence intervals for the two conditions did not overlap.

In a second set of analyses, we conducted two different bootstrap procedures. These bootstrap procedures were conducted to establish performance thresholds, which were then used to identify runs of trials that exceeded these thresholds. The first bootstrap procedure generated a "null" distribution. This null distribution reflected the probability that the monkeys responded randomly: that is, their responses were independent of the stimulus. To generate this distribution, we first identified those trials in which the frequency difference between tones A and B was 0.5, 1, 10, or 12 semitones and then shuffled the relationship between these frequency differences and the monkeys' reports. Since these frequency differences generate consistent reports in human listeners (Bregman, 1990; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007), when we shuffled the relationship, we hypothesized that we could systematically divorce the stimulus from the response. In contrast, because other frequency differences (i.e., 3 and 5 semitones) do not generate consistent reports in human listeners, there is no "incorrect" answer and the stimulus cannot be divorced from the response. Therefore, we did not include these trials within our shuffling procedure. Next, we selected, with replacement, N of these shuffled stimulus-report pairings; N was the number of trials/day. We then determined whether a shuffled pair was "correct" (e.g., frequency difference was ≤1 semitones and the report was "one stream") or "incorrect" (e.g., the frequency difference was ≤1 semitones and the report was "two streams"). Third, to simulate the temporal dynamics of a behavioral session, we treated these shuffled pairs as if they consecutive trials of a behavioral testing session. We then analyzed performance as a function of different running-average window sizes (i.e., 10, 20 or 50 consecutive shuffled stimulus-response pairings). This procedure was repeated 1000 times for each behavioral session. From this procedure, we generated, as a function of each window size, a distribution of running averages. Finally, we calculated the "running-average window (RAW) threshold". In one variant, we calculated the RAW threshold from each session's running-average distribution: the RAW threshold was defined as the upper boundary of each distribution's 95% confidence interval. In a second variant, all of the individual session distributions were pooled together (as a function of window size), and the "population" RAW threshold was defined as the upper boundary of this pooled distribution's 95% confidence interval.

The second bootstrap procedure generated a distribution of simulated data that, unlike the first bootstrap procedure, maintained the relationship between the auditory stimulus and the monkeys' responses. This bootstrap procedure tested whether, within an experimental session(s), there were temporal epochs or "runs" of performance that were above chance. First, for each experimental session, we identified those trials in which the frequency difference between tones A and B was 0.5, 1, 10, or 12 semitones; analogous to the logic described above, we did not use the other frequency-difference values. Next, while maintaining the relationship between the stimuli and response, we shuffled the order of the trials. This procedure maintained the relationship between the stimulus and response but disrupted the temporal order of these stimulus-response pairings. Finally, to simulate the temporal dynamics of a behavioral session, we analyzed performance as a function of different running-average window sizes (i.e., 10, 20 or 50 consecutive shuffled stimulus-response pairings). This procedure was repeated 1000 times for each behavioral session.

Like with the first bootstrap procedure, we calculated the RAW threshold using the session-by-session-data or the pooled data.

To compare the monkeys' performance with the bootstrapped performance, we extracted consecutive blocks of data that contained 10, 20, or 50 trials in which the frequency difference was 0.5, 1, 10, or 12 semitones. However, because the actual dataset contained trials from all of the tested-frequency differences, the actual length of the data block could be longer than the window size. For example, if the window size was 20 trials, the data block might contain 25 trials: 20 trials in which the frequency difference was 0.5, 1, 10, or 12 semitones and 5 trials in which the frequency difference was 3 or 5 semitones. When the monkey's performance on the 0.5, 1, 10, and 12 semitone trials exceeded the RAW threshold, the entire trial block (including trials in which the frequency difference was 3 or 5 semitones) was considered "suprathreshold". To be clear, the determination of "suprathreshold" was only based on the 0.5-, 1-, 10-, and 12-semitone trials because only these trial types were used in the bootstrap procedure. Using the suprathreshold data, we calculated, as a function of each window size and each frequency difference, the probability that the monkey reported two streams. These values were generated from individual behavioral sessions or from the dataset that was generated when the individual sessions were pooled together, analogous to that done with the bootstrap procedures. Finally, this analysis was conducted independently for each of the RAW thresholds that were calculated from each of the two bootstrap procedures.

3. Results

3.1 Monkeys' reports are modulated by the frequency difference between tones A and B

The results from 388 behavioral sessions (Monkey S: 227 sessions; Monkey H: 161 sessions) are shown in Figure 3; since monkeys S and H had comparable behavior, we pooled their behavioral data. Figure 3 plots the probability (i.e., the proportion of trials) that the monkeys reported two auditory streams as a function of frequency difference between tones A and B. When the frequency difference was ≤1 semitone, the probability that the monkeys reported two streams was less than chance. That is, the probability plus/minus its 95%-confidence interval was less than and did not include 0.5 (i.e., chance performance): 0.5-semitone difference: p=0.454±0.005, $p<0.05$; 1-semitone difference: p=0.465±0.006, $p<0.05$. The interpretation of this result is that the monkeys reliably reported one stream. When the frequency difference was ≥10 semitones, a different pattern emerged: the probability that the monkeys reported two streams exceeded chance: 10-semitone difference: p=0.550±0.007, $p<0.05$; 12-semitone difference: p=0.551±0.005, $p<0.05$. The monkeys' reports for the intermediate frequency differences (3 and 5 semitones) were between the reports for the other frequency differences; however, only the 5-semitone difference did not differ from chance (3- semitone difference: p=0.464±0.018, $p<0.05$; 5-semitone difference: p=0.487±0.020, $p>0.05$).

Although our behavioral data were reliable, the monkeys' behavior clearly did not differ substantially from 0.5 and was poor relative to human performance (Bregman, 1990; Cusack, 2005; Micheyl et al., 2007). However, during the behavioral sessions, we observed short periods (i.e., 10–50 consecutive trials) of high performance. To gain further insight

into this observation, we conducted further analyses of their behavior using two different bootstrap procedures.

In the first bootstrap procedure, we shuffled the relationship between the auditory stimulus and the monkeys' responses to generate a null distribution. This distribution tested the hypothesis that, over short windows of trials, the monkeys performed better than chance and were using the stimulus to guide their choices. Panels A and B in Figure 4 show the RAW thresholds that were generated from this procedure and the respective suprathreshold subset of behavioral data (see **Methods**). Figure 4A shows the monkeys' performance when the RAW thresholds were calculated from the population data. This threshold calculation is a reflection of performance for a given running average relative to the monkeys' general behavior. Figure 4B shows the monkey's performance using the session-by-session RAW thresholds. These thresholds provide a measure of performance relative to a particular day's behavior. We again found that monkeys (a) significantly reported one stream at the smallest frequency differences; (b) significantly reported two streams at the largest frequency differences; and (c) for intermediate frequency differences, behavior did not differ from chance (i.e., it fell below the running-average threshold). More specifically, for running windows of 10 trials (green data), we found that the monkeys' behavior was ~30% better than their overall behavior that was shown in Figure 3. The monkeys' performance improved modestly for larger running-average windows (blue and red data): for windows of 50 trials (red data), behavior improved by ~10%. Like the data in Figure 3, this bootstrap analysis indicated that the monkeys' behavior was guided by the stimuli. However, unlike the data shown in Figure 3, this bootstrap analysis indicated that for runs of 10 to 50 trials, the monkeys' performance can closely approximate the performance of human listeners.

To further evaluate these windows of high performance, we performed a second bootstrap procedure. In this procedure (and unlike the first one), we maintained the integrity of the stimulus-response pairings but shuffled the temporal order of these pairings. This procedure tested explicitly the reliability of the running-average windows; that is, this procedure tested whether there were short "runs" of performance that were above chance. Figures 4C and 4D show the monkeys' performance for those runs of trials that exceeded the bootstrap's performance at each of the RAW thresholds. Once again, we identified runs of trials in which the monkeys' behavior exceeded the RAW thresholds. We again found that short running-average windows of 10 trials (green data) were ~30% than the overall data in Figure 3; with more modest gains of ~10% over the overall data for windows of 50 trials (red data).

Together, all three analyses indicate that the monkeys successfully learned the streaming task. Using all of the data (Fig. 3), we found that their performance was reliable, and the pattern of their behavior was consistent—albeit poorer—than human performance. However, importantly, we found periods of high performance, defined as having a running average that fell above the RAW thresholds. These periods of high performance, which more closely approximated human performance, were found in windows of 10–50 trials (Fig. 4).

3.2 The monkeys' behavior was independent of tone A's frequency

Next, we tested whether the trial-by-trial variability in the frequency of tone A (range: 865–2226 Hz) affected the monkeys' behavioral reports. As a reminder, because the frequency of tone B was based on tone A's frequency, when we changed tone A's frequency, we changed the frequency content of the auditory sequence. This analysis is important because if the monkeys were using a strategy of reporting "two streams" whenever they heard a high-frequency stimulus, then changing the frequency of tone A should affect their behavior. However, if the monkeys were simply reporting the number of heard streams, their reports should be independent of tone A's frequency. The results of this analysis are shown in Figure 5. In this Figure, we again plot the probability that the monkeys' reported two streams as a function of the frequency difference between tones A and B. However, here, we subdivided the data: the "low-frequency" data contained the monkeys' reports when the tone A's frequency was between 865–1500 Hz (the lower half of the distribution of tone A frequencies), whereas the "high-frequency" data contained reports when tone A's frequency was 1501–2226 Hz (the upper half of the distribution of tone A frequencies). Using the two bootstrap procedures (see **Methods**), we calculated the running-average thresholds independently for both the low-frequency and high-frequency data groups; because data for all RAW thresholds followed the same pattern, Figure 5 only shows the data relative to the 20-trial RAW threshold. As can be seen, for each of those frequency differences that exceeded the bootstrap threshold (i.e., 0.5, 1, 10 and 12 semitones), in most cases, the confidence intervals on the monkeys' reports for the low-frequency data overlapped with those of the high-frequency data. That is, the frequency of tone A did not significantly ($p>0.05$) affect the monkeys' reports. When the confidence intervals did not overlap, we could not identify any consistent trend between the frequency of tone A and the monkeys' reports. These results are consistent with the hypothesis that the monkeys' reports were independent of tone A's frequency.

3.3 Longer stimulus durations biased the monkeys to report two streams

Next, we tested how the trial-by-trial variability in the amount of time that the monkeys' listened to the auditory sequence time (listening duration; 180–2022 ms) affected their behavior. We divided the behavior into trials when the listening duration was 180–770 ms (the lower half of the distribution of listening durations) and into trials when the listening duration was 771–2022 ms (the upper half of the distribution of listening durations). The results of this analysis are shown in Figure 6; because data for all RAW thresholds followed the same pattern, Figure 6 only shows the data relative to the 20-trial RAW threshold. As can be seen, for each of those frequency differences that exceeded the bootstrap threshold (i.e., 0.5, 1, 10 and 12 semitones), the confidence intervals on the monkeys' reports for the longer-duration data never overlap with, and are always higher than, those of the shorter-duration sequences. Like human listeners (Micheyl et al., 2007), longer-duration sequences biased the monkeys to report "two streams" more often than shorter-duration sequences.

3.4 Simultaneous presentation of tones A and B biases the monkeys to report one stream

Finally, we tested whether the temporal relationship of tone A and tone B affected the monkeys' behavioral reports. If, as discussed above, the monkeys were simply reporting

"two streams" whenever they perceived a high-frequency stimulus, their reports should not depend on the tones A and Bs' temporal relationship. However, if the monkeys were reporting the number of heard streams, then, like human listeners (Elhilali et al., 2009), their reports should be biased toward reporting one stream when tone A and B were presented simultaneously and even when the frequency difference between tones A and B is large (e.g., ≥ 10 semitones).

Because the simultaneous presentation of tones A and B sounded different than the normal asynchronous presentation, we limited its presentation to a small subset of behavioral sessions (N = 18). Consequently, this data set was not large enough for our bootstrap procedure. Finally, to maximize the informative trials with the least exposure to the simultaneous trials as possible, we limited this presentation to a 10-semitone frequency difference.

Figure 7 shows the results of this analysis. As noted above, when the tones were asynchronous and the frequency difference was 10 semitones, the probability that the monkeys reported two streams was significant (p=0.524±0.007; $p<0.05$; this proportion represents the monkeys' behavior during those sessions when simultaneous tones were also presented). However, when tones A and B were presented simultaneously, the monkeys were more likely to report one stream (10 frequency semitones: p=0.459±0.051). This proportion of trials was significantly ($p<0.05$) smaller than the one when tones A and B were presented asynchronously. However, it is not different than chance performance (0.5; $p>0.05$). Nonetheless, this result is consistent with the hypothesis that the simultaneous presentation of tones A and B biased the monkeys toward reports of "one stream".

4. Discussion

The streaming task has been used extensively to test auditory perception in humans. Here, we demonstrated for the first time that rhesus macaques' behavioral reports were qualitatively consistent with those of human listeners. We found that monkeys reported small frequency differences as one stream, large ones as two streams, and intermediate ones as either one or two streams. We further found that the monkeys' reports were independent of the absolute frequency content of the stimulus but that longer listening durations biased the monkeys toward reporting two streams, suggestive of a buildup of streaming (Bregman, 1978). Moreover, simultaneous presentation of tones A and B biased the monkey toward reporting one stream. Below, we discuss the interpretation of our findings, as well as caveats regarding performance and implications for auditory processing across species.

Although our current findings are consistent with human studies, training monkeys on the streaming task presented challenges that are not faced in training humans on this task. Namely, monkeys could not be explicitly told to report one or two streams. Therefore, without controls, our results could have been interpreted as the monkeys merely reporting any stimulus with a high frequency as two streams and anything else as one stream. However, three controls support the hypothesis that the monkeys were reporting the number of heard streams. First, by presenting tone A across a range of frequencies that spanned nearly 2.5 octaves—considerably larger than the frequency difference between tones A and

B—we demonstrated that the monkeys' reports were independent of the frequency of tone A (Fig. 5). Second, like human listeners (Micheyl et al., 2007), longer stimulus durations biased the monkeys to report two streams. This result is consistent with findings that the perception of two streams "builds up" over time (Elhilali et al., 2009; Micheyl et al., 2007) and is inconsistent with a hypothesis of simply reporting frequency differences. Finally, similar to human listeners (Elhilali et al., 2009), when the tone bursts were presented simultaneously and the frequency difference was large (which normally elicits reports of "two streams), the monkeys' reports were biased toward those of "one stream" (Fig. 7). Overall, these controls are consistent with the hypothesis that the monkeys reported the number of heard streams.

Although the monkeys' performance was reliable and the three stimulus controls yielded results qualitatively similar to those of humans, the monkeys overall performance (Fig. 3) indicated that this task was difficult. However, in observing the monkeys' performance, it was apparent that there were times when the monkeys had short runs of good performance. Indeed, our two bootstrap procedures indicated that the monkeys used the stimulus to guide their behavior and had high levels of performance over windows of 10–50 trials (Fig. 4) that more closely mirrored that of human-performance levels (Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007). Importantly, since trials with a given frequency difference were randomly distributed within a session, these periods of high performance did not represent runs of "easy" trials (e.g., blocks when the same frequency difference was presented multiple times in succession).

How do our results fit into the general comparative psychophysical literature? Our findings support this literature, much of which has found that humans and non-human animals similarly process auditory stimuli. For example, several sets of studies have found that humans, monkeys, quail and chinchillas have similar categorical boundaries for human phonemes (Kuhl et al., 1975; Kuhl et al., 1982). Similarly, monkeys exhibit amodal completion in a manner similar to humans (Miller et al., 2001; Petkov et al., 2003; Petkov et al., 2007) and group sounds in a manner similar to humans (Izumi, 2002). Other studies have demonstrated that non-human animals parse the auditory scene like human listeners (Aulanko et al., 1993; Coath et al., 2005; DeWitt et al., 2012; Moerel et al., 2012; Narayan et al., 2007). Finally, our data are consistent with those studies that used indirect assays of streaming (Izumi, 2002; Ma et al., 2010; Moerel et al., 2012).

Where in the brain is this information being processed? Several studies have recorded from the monkey primary auditory cortex while monkeys were listening passively to auditory sequences similar to those used in our study (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005). Although the monkeys were not actively engaged in a streaming task during these studies, the pattern of neural activity indicated that this cortical region may be involved in the grouping and segregation of auditory stimuli into auditory streams. Indeed, other sets of findings in the core and belt regions of the auditory cortex have also hinted at a role for these brain regions in auditory scene analysis (Bendor et al., 2006; Bizley et al., 2013; Fishman et al., 2004; Fishman et al., 2000; Fishman et al., 2001a; Fishman et al., 2001b; Micheyl et al., 2007; Niwa et al., 2012; Wang et al., 2008). More generally, the ventral auditory pathway, which is specialized for mediating auditory perception (Bizley et

al., In Press; Cohen, 2012; Kaas et al., 1999; Rauschecker et al., 2009; Romanski et al., 2009), likely plays a role in the neural computations that allow a listener to segregate or group an auditory stimulus into one or more auditory streams.

Finally, this task will provide a powerful tool to disassociate brain activity that is related to the features of the auditory stimulus from activity that is related to a listeners' behavioral report. In particular, since listeners reports vary, on a trial-by-trial basis, for sequences with intermediate frequency differences (>1 semitone and <10 semitones), this stimulus can be considered akin to a "bistable percept" (Andersen et al., 1996; Bregman, 1990; Logothetis et al., 1989; Parker et al., 1998). In other words, by holding the stimulus constant and analyzing neural responses as a function of the listener's behavioral report, we can identify and differentiate between the brain regions and the computations that underlie auditory scene analysis, auditory perception and decision-making.

5. Conclusion

In conclusion, we have shown that monkeys can be trained to perform the streaming task. Moreover, their behavioral reports are consistent with human reports across a variety of experimental manipulations. These findings add further evidence that monkeys group and segregate acoustic stimuli similarly to humans. Therefore, they provide an excellent model to study the neural coding that underlies this behavior, and more generally, auditory perception.

Acknowledgments

We thank Joji Tsunada, Steven Eliades, and Heather Hersh for helpful comments on the preparation of this manuscript. We also thank Harry Shirley for outstanding veterinary support. KLCL and YEC were supported by grants from NIDCD-NIH and the Boucai Hearing Restoration Fund.

Abbreviations

RAW threshold	running-average window

References

Andersen RA, Bradley DC, Shenoy KV. Neural mechanisms for heading and structure-from-motion perception. Cold Spring Harb Symp Quant Biol. 1996; 61:15–25. [PubMed: 9246431]

Aulanko R, Hari R, Lounasmaa OV, Naatanen R, Sams M. Phonetic invariance in the human auditory cortex. Neuroreport. 1993; 4:1356–8. [PubMed: 8260620]

Bendor D, Wang X. Cortical representations of pitch in monkeys and humans. Curr Opin Neurobiol. 2006; 16:391–9. [PubMed: 16842992]

Bizley JK, Cohen YE. The what, where, and how of auditory-object perception. Nat Rev Neurosci. In Press.

Bizley JK, Walker KM, Nodal FR, King AJ, Schnupp JW. Auditory Cortex Represents Both Pitch Judgments and the Corresponding Acoustic Cues. Curr Biol. 2013

Bregman AS. Auditory streaming is cumulative. J Exp Psychol Hum Percept Perform. 1978; 4:380–387. [PubMed: 681887]

Bregman, AS. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press; Cambridge, MA: 1990.

Bregman AS, Ahad PA, Crum PAC, O'Reilly J. Effects of time intervals and tone durations on auditory stream segregation. Percept Psychophys. 2000; 62:626–636. [PubMed: 10909253]

Brown CH, Beecher MD, Moody DB, Stebbins WC. L ocalization of noise bands by Old World monkeys. J Acoust Soc Am. 1980; 68:127–32. [PubMed: 6771312]

Carlyon RP, Cusack R, Foxton JM, Robertson IH. Effects of attention and unilateral neglect on auditory stream segregation. J Exp Psychol. 2001; 27:115–127.

Coath M, Brader JM, Fusi S, Denham SL. Multiple views of the response of an ensemble of spectro-temporal features support concurrent classification of utterance, prosody, sex and speaker identity. Network. 2005; 16:285–300. [PubMed: 16411500]

Cohen, YE. Auditory Cognition: The Integration of Psychophysics with Neurophysiology. In: Cohen, YE.; Popper, AN.; Fay, RR., editors. Neural Correlates of Auditory Cognition, Vol. Springer Handbook of Auditory Research. Springer-Verlag; New York: 2012. p. 1-6.

Cusack R. The Intraparietal Sulcus and Perceptual Organization. J Cogn Neurosci. 2005; 17:641–651. [PubMed: 15829084]

DeWitt I, Rauschecker JP. Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci USA. 2012; 109:E505–14. [PubMed: 22308358]

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron. 2009; 61:317–29. [PubMed: 19186172]

Fishman YI, Arezzo JC, Steinschneider M. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am. 2004; 116:1656–70. [PubMed: 15478432]

Fishman YI, Reser DH, Arezzo JC, Steinschneider M. Complex tone processing in primary auditory cortex of the awake monkey. I Neural ensemble correlates of roughness. J Acoust Soc Am. 2000; 108:235–46. [PubMed: 10923888]

Fishman YI, Reser DH, Arezzo JC, Steinschneider M. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res. 2001a; 151:167–187. [PubMed: 11124464]

Fishman YI, Volkov IO, Noh MD, Garell PC, Bakken H, Arezzo JC, Howard MA, Steinschneider M. Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. J Neurophysiol. 2001b; 86:2761–88. [PubMed: 11731536]

Griffiths TD, Warren JD. What is an auditory object? Nat Rev Neurosci. 2004; 5:887–892. [PubMed: 15496866]

Horvath J, Czigler I, Sussman E, Winkler I. Simultaneously active pre- attentive representations of local and global rules for sound sequences in the human brain. Cognitive Brain Research. 2001; 12:131–144. [PubMed: 11489616]

Izumi A. Auditory stream segregation in Japanese monkeys. Cogn. 2002; 82:B113–B122.

Kaas JH, Hackett TA. 'What' and 'where' processing in auditory cortex. Nat Neurosci. 1999; 2:1045–1047. [PubMed: 10570476]

Kuhl PK, Miller JD. Speech-perception by chinchilla - phonetic boundaries for synthetic vot stimuli. J Acoust Soc Am. 1975; 57:S49–S50.

Kuhl PK, Padden DM. Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. Perception & Psychophysics. 1982; 32:542–550. [PubMed: 7167352]

Kuhl PK, Padden DM. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. J Acoust Soc Am. 1983; 73:1003–1010. [PubMed: 6221040]

Logothetis NK, Schall JD. Neuronal correlates of subjective visual perception. Science. 1989; 245:761–3. [PubMed: 2772635]

Ma L, Micheyl C, Yin P, Oxenham AJ, Shamma SA. Behavioral measures of auditory streaming in ferrets (Mustela putorius). J Comp Psychol. 2010; 124:317–30. [PubMed: 20695663]

McDermott J. The cocktail party problem. Curr Biol. 2009; 19:R1024–R1027. [PubMed: 19948136]

Micheyl C, Tian B, Carlyon RP, Rauschecker JP. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron. 2005; 48:139–48. [PubMed: 16202714]

Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Courtenay Wilson E. The role of auditory cortex in the formation of auditory streams. Hear Res. 2007; 229:116–31. [PubMed: 17307315]

Miller CT, Dibble E, Hauser MD. Amodal completion of acoustic signals by a nonhuman primate. Nat Neurosci. 2001; 4:783–4. [PubMed: 11477422]

Moerel M, De Martino F, Formisano E. Processing of Natural Sounds in Human Auditory Cortex: Tonotopy, Spectral Tuning, and Relation to Voice Sensitivity. J Neurosci. 2012; 32:14205–14216. [PubMed: 23055490]

Narayan R, Best V, Ozmeral E, McClaine E, Dent M, Shinn-Cunningham B, Sen K. Cortical interference effects in the cocktail party problem. Nat Neurosci. 2007; 10:1601–7. [PubMed: 17994016]

Niwa M, Johnson JS, O'Connor KN, Sutter ML. Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci. 2012; 32:3193–210. [PubMed: 22378891]

Parker AJ, Newsome WT. Sense and the single neuron: probing the physiology of perception. Annu Rev Neurosci. 1998; 21:227–77. [PubMed: 9530497]

Petkov CI, O'Connor KN, Sutter ML. Illusory sound perception in macaque monkeys. J Neurosci. 2003; 23:9155–61. [PubMed: 14534249]

Petkov CI, O'Connor KN, Sutter ML. Encoding of illusory continuity in primary auditory cortex. Neuron. 2007; 54:153–65. [PubMed: 17408584]

Pfingst BE, Laycock J, Flammino F, Lonsbury-Martin B, Martin G. Pure tone thresholds for the rhesus monkey. Hear Res. 1978; 1:43–47. [PubMed: 118150]

Rahne T, Sussman E. Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. European Journal of Neuroscience. 2009; 29:205–211. [PubMed: 19087164]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 2009; 12:718–724. [PubMed: 19471271]

Recanzone GH, Sutter ML. The biological basis of audition. Annu Rev Psychol. 2008; 59:119–42. [PubMed: 17678445]

Recanzone GH, Guard DC, Phan ML. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol. 2000; 83:2315–2331. [PubMed: 10758136]

Romanski LM, Averbeck BB. The Primate Cortical Auditory System and Neural Representation of Conspecific Vocalizations. Ann Rev Neurosci. 2009; 32:315–346. [PubMed: 19400713]

Serafin JV, Moody DB, Stebbins WC. Frequency selectivity of the monkey's auditory system: psychophysical tuning curves. J Acoust Soc Am. 1982; 71:1513–8. [PubMed: 7108026]

Shamma S. On the Emergence and Awareness of Auditory Objects. PLoS Biol. 2008; 6:e155. [PubMed: 18578570]

Shamma SA, Micheyl C. Beyond the scenes of auditory perception. Curr Opin Neurobiol. 2010; 20:361–6. [PubMed: 20456940]

Shamma SA, Elhilali M, Micheyl C. Temporal coherence and attention in auditory scene analysis. Trends Neurosci. 2011; 34:114–23. [PubMed: 21196054]

Shinn-Cunningham BG. Object-based auditory and visual attention. Trends Cogn Sci. 2008; 12:182–6. [PubMed: 18396091]

Sinnott JM, Beecher MD, Moody DB, Stebbins WC. Speech sound discrimination by monkeys and humans. J Acoust Soc Am. 1976; 60:687–95. [PubMed: 824334]

Sussman ES. Integration and segregation in auditory scene analysis. The Journal of the Acoustical Society of America. 2005; 117:1285–1298. [PubMed: 15807017]

Sussman ES, Horvath J, Winkler I, Orr M. The role of attention in the formation of auditory streams. Perception & Psychophysics. 2007; 69:136–152. [PubMed: 17515223]

Tsunada J, Lee JH, Cohen YE. Representation of speech categories in the primate auditory cortex. J Neurophysiol. 2011

Wang X, Lu T, Bendor D, Bartlett E. Neural coding of temporal information in auditory thalamus and cortex. Neuroscience. 2008; 154:294–303. [PubMed: 18555164]

Winkler I, Denham SL, Nelken I. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci. 2009; 13:532–40. [PubMed: 19828357]

Zar, J. Biostatistical Analysis. Prentice Hall; Upper Saddler River, NJ: 1996.

Zatorre RJ, Bouffard M, Belin P. Sensitivity to auditory object features in human temporal neocortex. J Neurosci. 2004; 24:3637–42. [PubMed: 15071112]

1. We test rhesus monkeys in an auditory streaming task
2. We find that rhesus behavior is consistent with that of human listeners
3. These findings validate this model to study neural correlates of auditory scene analysis

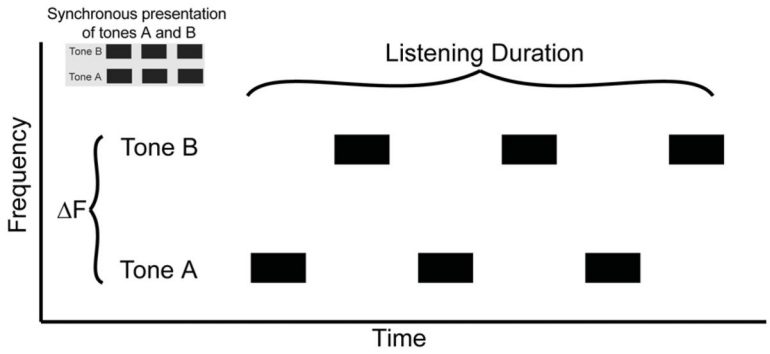

Figure 1. Schematic of the auditory stimulus to test auditory streaming

The auditory stimulus was an asynchronous sequence of two types of tone bursts: tone A and tone B. Typically, tones A and B were presented asynchronously but were at times presented simultaneously (see inset at upper left). The frequency of tone A, the frequency difference between the tones A and B (ΔF), and the listening duration (i.e., the duration of the auditory sequence) varied on a trial-by-trial basis. The units on the x- and y-axes are arbitrary.

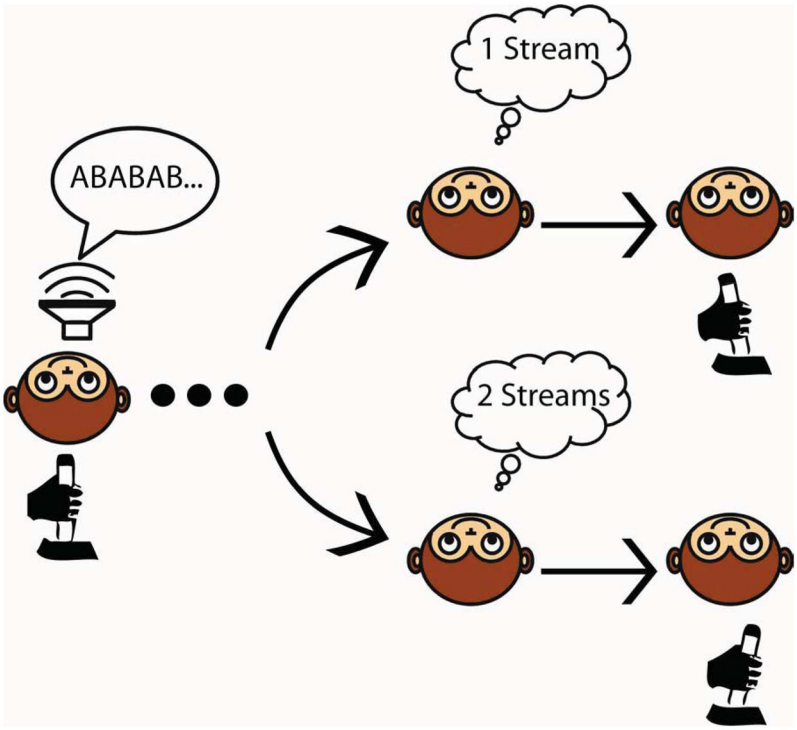

Figure 2. Schematic of the streaming task

The streaming task is a one-interval, two- alternative, forced-choice task requiring a monkey to report whether he heard one or two auditory streams by moving a joystick to the right (one stream) or left (two streams). When the frequency difference between tones A and B was ≤1 semitone or ≥10 semitones, the monkeys received a juice reward for reporting the correct answer. For all other frequency differences, the monkeys received a reward on 50% of randomly selected trials; the decision to reward was made independent of their behavioral report.

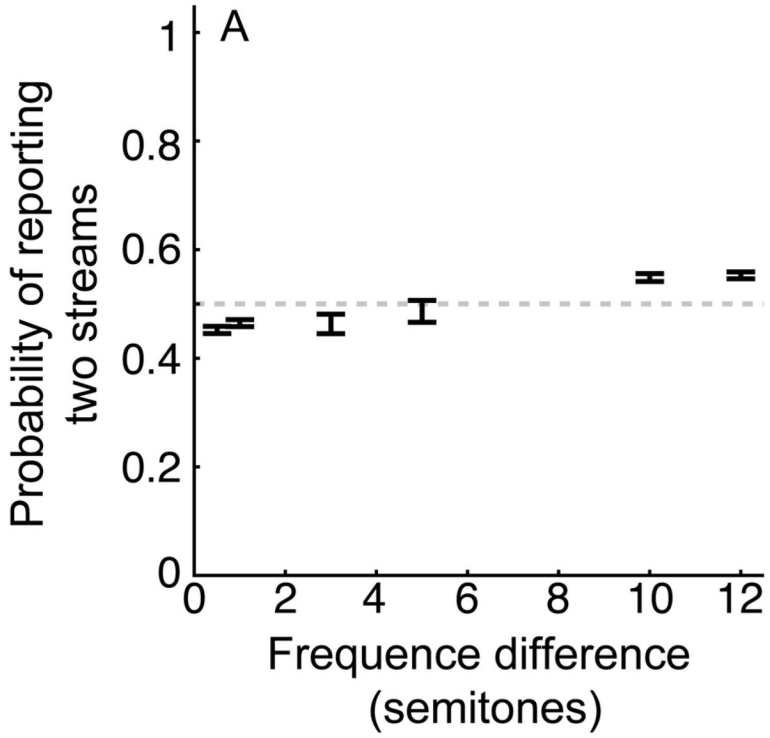

Figure 3. Behavioral performance: all data and all sessions

The average performance of both monkeys from all of the behavioral sessions reported in this manuscript (except for those trials when tone A and B were presented simultaneously; see Fig. 7). The center of each bar indicates the average probability (i.e., the proportion of trials) that the monkeys reported two streams; the length of the bars indicates the 95% confidence interval. The gray dashed line represents chance performance (0.5) of answering one or two streams.

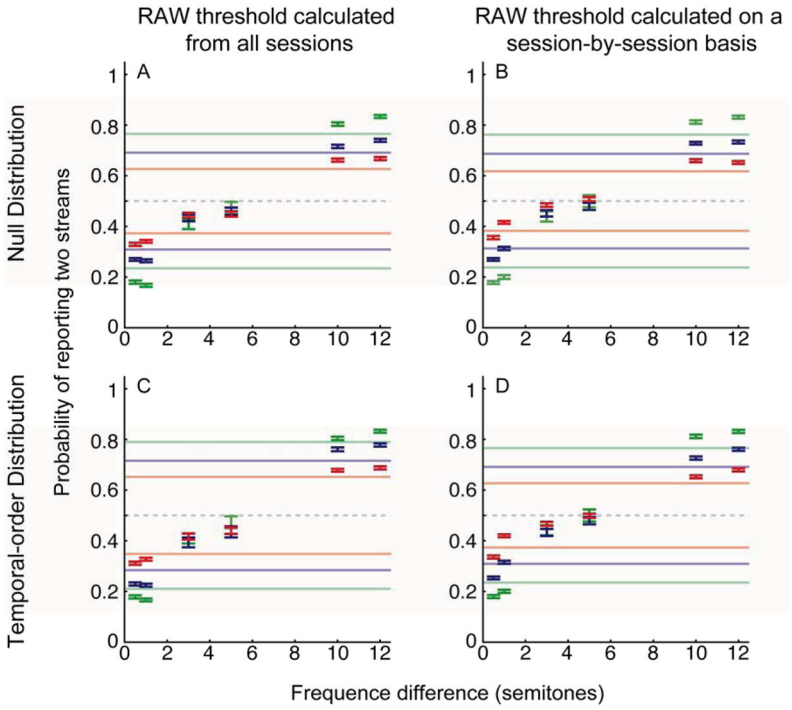

Figure 4. Behavioral performance: behavior relative to the bootstrapped RAW thresholds

The data on the top row show the monkeys' behavior relative to a bootstrapped null distribution (i.e., one in which there is no relationship between the stimulus and the monkeys' responses). The data on the bottom row show the monkeys' behavior relative a second bootstrap distribution that maintained the integrity between the stimulus and the monkeys' responses but shuffled the temporal order. This bootstrap procedure tested explicitly whether there were significant temporal runs of performance. For data in the left column, the RAW thresholds were calculated from data that was pooled across all behavioral sessions. For data in the right column, the RAW thresholds were calculated on a session-by-session basis. The color of each of the solid lines illustrates the upper and lower boundaries of the different RAW thresholds: green is 10 trials, blue is 20 trials, and red is 50 trials. The center of each bar indicates average suprathreshold performance; the color of the data points is consistent with the color of the threshold values. The length of the bars indicates the 95% confidence interval. If error bars from one color are not visible, it is because the confidence intervals for multiple conditions overlap completely. The gray dashed line represents chance performance (0.5) of answering one or two streams.

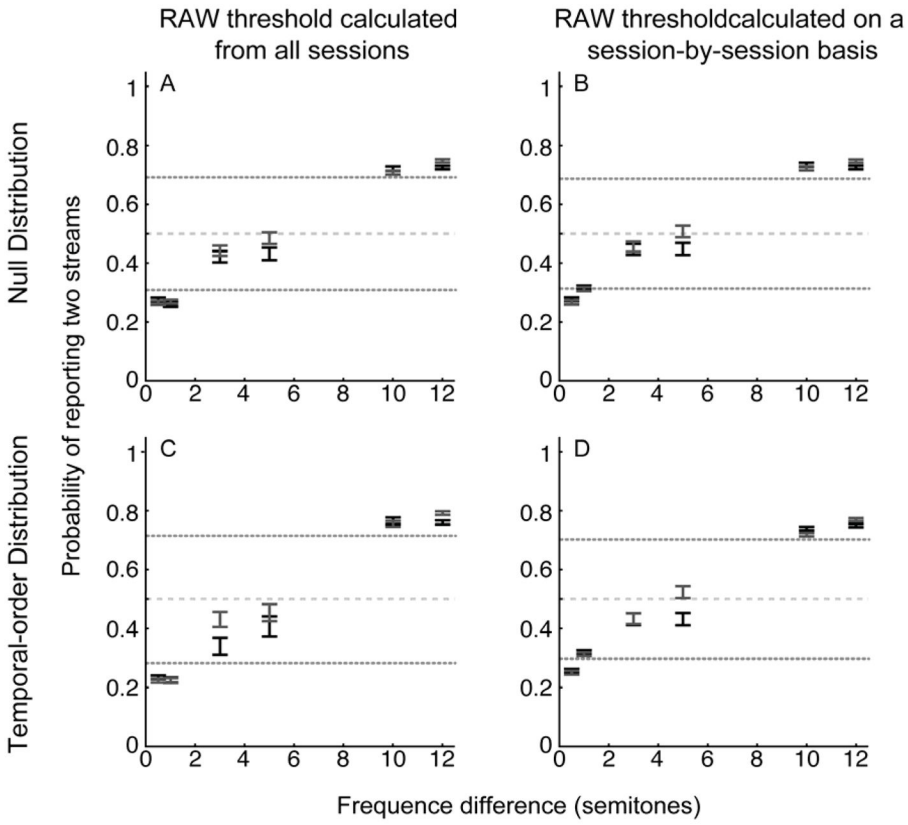

Figure 5. Behavioral performance: dependence on the frequency of tone A

The data in each row and column are organized analogous to that in Figure 4. The dotted lines illustrate the upper and lower boundaries of the 20-trial RAW threshold; the other thresholds are not shown. The data in black indicate average suprathreshold performance when the frequency of tone A was relatively low (865–1500 Hz). The data in gray indicate average suprathreshold performance when the frequency of tone A was relatively high (1501–2226 Hz). The center of each bar indicates average suprathreshold performance; the length of the bars indicates the 95% confidence interval. If error bars from one color are not visible, it is because the confidence intervals for multiple conditions overlap completely. The gray dashed line represents chance performance (0.5) of answering one or two streams.

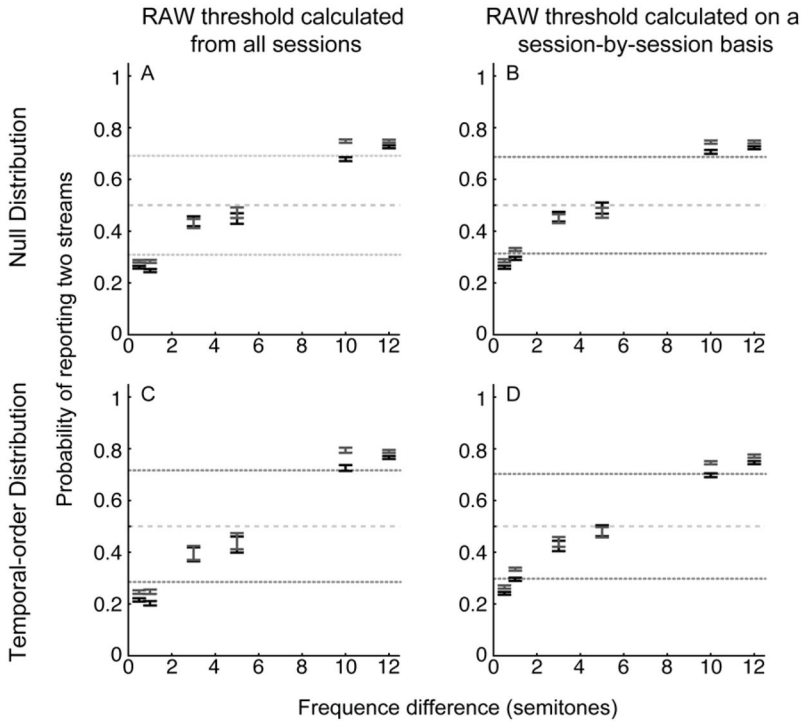

Figure 6. Behavioral performance: dependence on listening duration

The data in each row and column are organized analogous to that in Figure 4. The dotted lines illustrate the upper and lower boundaries of the 20-trial RAW threshold; the other thresholds are not shown. The data in black indicate average suprathreshold performance when the listening duration was short (180–770 ms). The data in gray indicate average suprathreshold performance when the listening duration was long (771–2022 ms). The center of each bar indicates average suprathreshold performance; the length of the bars indicates the 95% confidence interval. If error bars from one color are not visible, it is because the confidence intervals for multiple conditions overlap completely. The gray dashed line represents chance performance (0.5) of answering one or two streams.

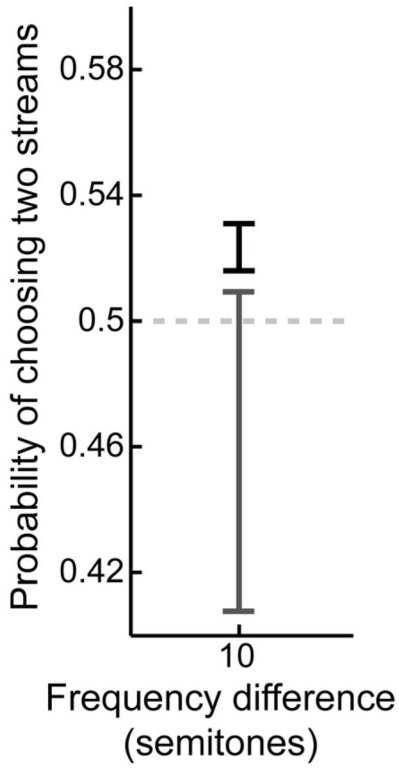

Figure 7. Behavioral performance: dependence on the temporal structure of tones A and B

The black bar indicates average performance for trials when tones A and B were presented asynchronously. The gray bar indicates average performance for trials when tones A and B were presented simultaneously. The center of each bar indicates the average probability (i.e., the proportion of trials) that the monkeys reported two streams; the length of the bars indicates the 95% confidence interval.