Sensory and Motor Systems

Temporal Integration of Auditory Information Is Invariant to Temporal Grouping Cues[1,2,3]

Andrew S.K. Liu,[1] Joji Tsunada,[2] Joshua I. Gold,[3,*] and Yale E. Cohen[2,3,*]

DOI:http://dx.doi.org/10.1523/ENEURO.0077-14.2015

[1]Bioengineering Graduate Group, [2]Department of Otorhinolaryngology, Perelman School of Medicine, and [3]Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Abstract

Auditory perception depends on the temporal structure of incoming acoustic stimuli. Here, we examined whether a temporal manipulation that affects the perceptual grouping also affects the time dependence of decisions regarding those stimuli. We designed a novel discrimination task that required human listeners to decide whether a sequence of tone bursts was increasing or decreasing in frequency. We manipulated temporal perceptual-grouping cues by changing the time interval between the tone bursts, which led to listeners hearing the sequences as a single sound for short intervals or discrete sounds for longer intervals. Despite these strong perceptual differences, this manipulation did not affect the efficiency of how auditory information was integrated over time to form a decision. Instead, the grouping manipulation affected subjects' speed−accuracy trade-offs. These results indicate that the temporal dynamics of evidence accumulation for auditory perceptual decisions can be invariant to manipulations that affect the perceptual grouping of the evidence.

Key words: auditory cortex; decision-making; scene analysis; speed-accuracy tradeoff; streaming

---

Significance Statement

Auditory perception depends on perceptual grouping cues, which relate to how the brain parses the auditory scene into distinct perceptual units, and auditory decisions, which relate to how the brain identifies a sound. These two processes are not independent; both rely on the temporal structure of the acoustic stimulus. However, the effects of this temporal structure on perceptual grouping and decision-making are not known. Here, we combined psychophysical testing with computational modeling to test the interaction of temporal perceptual grouping cues with the temporal processes that underlie perceptual decision-making. We found that temporal grouping cues do not affect the efficiency by which sensory evidence is accumulated to form a decision. Instead, the grouping cues modulate a subject's speed−accuracy trade-off.

Introduction

Auditory perception depends on both perceptual grouping and decision-making. Perceptual grouping is a form of feature-based stimulus segmentation that determines whether acoustic events are grouped into a single sound or segregated into distinct sounds (Bregman, 1990). Auditory decision-making involves the brain's interpretation

of information within and across discrete stimuli to detect, discriminate, or identify their source or content.

Auditory perceptual grouping and decision-making each depend critically on the temporal structure of incoming acoustic events. For instance, when a person is walking, each step is a unique acoustic event, but our auditory system groups these events together to form a stream of footsteps. However, if the time between events is long, the auditory system segregates these events into unique, discrete sounds. Decision-making can also depend on the temporal structure of a sound because decision-making is a deliberative process in which listeners often accumulate and interpret auditory information over time to form categorical judgments (Green et al., 2010; Brunton et al., 2013; Mulder et al., 2013).

Although we know that perceptual grouping can affect some forms of decision-making (Bey and McAdams, 2002; Roberts et al., 2002; Micheyl and Oxenham, 2010; Borchert et al., 2011; Thompson et al., 2011), the interplay between the temporal properties of an auditory stimulus, perceptual grouping, and decision-making is not known. We cannot infer this interplay from visual studies because analogous manipulations in the visual domain (Kiani et al., 2013) do not relate directly to auditory perceptual grouping. Further, because temporal processing is fundamentally different for audition than for vision (Bregman, 1990; Griffiths and Warren, 2004; Shinn-Cunningham, 2008; Shamma et al., 2011; Bizley and Cohen, 2013), it is reasonable to hypothesize that this interplay may be different in these two sensory systems. Thus, it remains an open and fundamental question whether and how grouping temporal information interacts with auditory grouping and decision-making.

To examine this question, we measured the performance of human subjects who participated in a series of auditory tasks in which they reported whether a sequence of tone bursts was increasing or decreasing in frequency. Temporal information was manipulated by changing the interval between the onsets of consecutive tone bursts. This manipulation affected the subjects' perceptual grouping of the tone-burst sequence: they heard one sound when the interval was short but a series of discrete sounds when it was long. The quality of the sensory evidence was manipulated by changing the proportion of tone bursts that linearly increased or decreased in frequency. We found that subjects accumulated sensory evidence over time to form their decisions. However, the time interval between consecutive tone bursts did not affect how this incoming stimulus was accumulated to form the decision about the change in frequency. Instead, the time between the tone bursts affected how the subjects balanced the speed and accuracy of their decisions,

which fundamentally trade-off for certain decisions like this one that require incoming, noisy information to be accumulated over time (Gold and Shadlen, 2007; Bogacz et al., 2010). Specifically, for our task, longer time intervals between tone bursts (i.e., slower rates of incoming sensory information) led to a higher premium on speed at the expense of accuracy. Overall, these findings indicate that temporal manipulations that affect the perceptual grouping of sounds do not necessarily affect how information from those sounds are accumulated over time to form a decision, even when the temporal manipulations have clear effects on the trade-off between the speed and accuracy of the decision.

Materials and Methods

Prior to their participation, subjects provided informed consent. Human subjects were recruited at the University of Pennsylvania. All subjects (age range: 25-48 years) reported normal hearing; three of the subjects were authors on the study.

Experimental setup

All experimental sessions took place in a single-walled acoustic chamber (Industrial Acoustics) that was lined with echo-absorbing foam. Each subject was seated with his or her chin in a chin rest that was approximately 2 ft from a calibrated Yamaha (model MSP7) speaker. Auditory stimuli were generated using the RX6 digital-signal-processing platform (TDT Inc.). The task structure was controlled through the Snow Dots toolbox (http://code.google.com/p/snow-dots) that ran in the Matlab (Mathworks) programming environment. Subjects indicated their responses by pressing a button on a gamepad (Microsoft Sidewinder). Tasks instructions and feedback were presented on a LCD flat-panel monitor (Dell E171FP) that was placed above the speaker.

Auditory stimuli

The auditory stimulus was a sequence of tone bursts (duration: 30 ms with a 5 ms $\cos^2$ gate; level: 65 dB SPL). The interburst interval (IBI) was the time between the offset of one tone burst and the onset of the next tone burst (range: 10-150 ms).

At the beginning of each trial, the frequency of a sequence's first tone burst was randomly sampled from a uniform distribution (500-3500 Hz). Next, we generated a monotonically increasing or decreasing sequence by adding or subtracting a 7.5 Hz to the previous tone-burst frequency. Finally, on a trial-by-trial basis, we perturbed the temporal order of the tone bursts. This perturbation changed the quality of the sensory evidence. In particular, for each trial, we defined a sequence's coherence, which was the proportion of tone bursts in a sequence whose frequency value changed by a fixed increment relative to the previous tone burst. If the coherence was 100%, all of the tone bursts monotonically increased or decreased. If the coherence was <100%, the temporal order of a subset of tone bursts was randomly shuffled. For example, if the coherence was 50%, half of the tone bursts were shuffled. If the coherence was 0%, all of the tone bursts in the sequence were shuffled.

Correspondence should be addressed to Dr. Yale E. Cohen, Department of Otorhinolaryngology, 3400 Spruce St-5 Ravdin, Philadelphia, PA 19104. E-mail: ycohen@mail.med.upenn.edu.
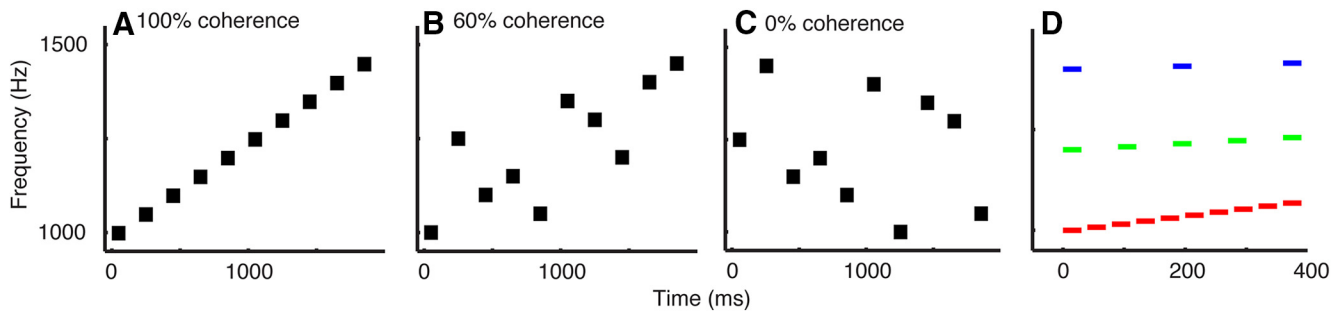
Figure 1 Example auditory sequences characterized in terms of coherence and different interburst intervals. *A*, Frequency direction (i.e., increasing) is easiest to discriminate when the stimulus is 100% coherent. As coherence gets smaller (60% in *B* and 0% in *C*), frequency direction becomes more ambiguous and, thus, can lead to more errors. In *A−C*, we show increasing auditory sequences. Decreasing sequences are analogous but with negative coherence values. *D* shows a 100% coherence auditory sequence at three different IBIs: 10 ms (red), 60 ms (green), and 150 ms (blue).

This stimulus design ensured that the frequency range of the sequences overlapped and that, on average, each sequence contained the same frequency values (but in a different temporal order). This procedure also minimized the possibility that subjects based their decisions on the specific frequency content of a sequence rather than its sequence direction.

Overall, each tone-burst sequence could be characterized by three parameters (Fig. 1): (1) direction, which indicated whether a sequence of tone bursts increased or decreased in frequency; (2) IBI; and (3) coherence. The values of each of these three parameters were determined on a trial-by-trial basis depending on the constraints of each auditory task; see Behavioral tasks, below.

Behavioral tasks
Subjects participated in three versions of the discrimination task. For all three versions, subjects reported the direction (increasing or decreasing) of the tone-burst sequence and received visual feedback, via the LCD monitor, regarding their report on every trial. The time between trials was ~2 s and was independent of the subjects' report.

Response-time task
For this task, subjects reported their perceptual decision at any time following sequence onset (Fig. 2A). They were instructed to respond as quickly as possible but not to sacrifice accuracy. We tested six subjects (5 male and 1 female) in four weekly 1.5 h sessions. Each session contained four blocks of trials; a short break was provided between blocks. In each block, we varied sequence direction (increasing or decreasing), IBI (10, 60, or 150 ms), and coherence (0, 10, 25, 50, or 100%) on a trial-by-trial basis. Each combination of these sequence parameters was presented five times within a block for a total of 150 trials/block. The maximum response-time (RT) (and, hence, maximum sequence duration) was 5000 ms; a trial was aborted if a subject did not respond by the end of the sequence. The stimulus was terminated as soon as the subject reported the decision.

Four subjects participated in additional sessions (four blocks of trials in each of four sessions per subject) of the

RT task to test for effects of changes in the speed−accuracy trade-off. For these sessions, the IBI was held constant at 60 ms and subjects were instructed to "emphasize speed" or "emphasize accuracy" in alternating blocks. Stimulus tone-sequence direction (increasing or
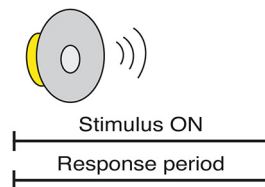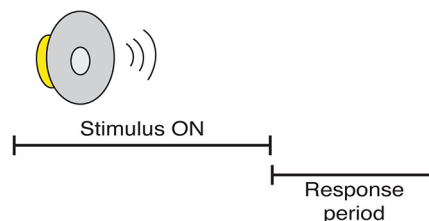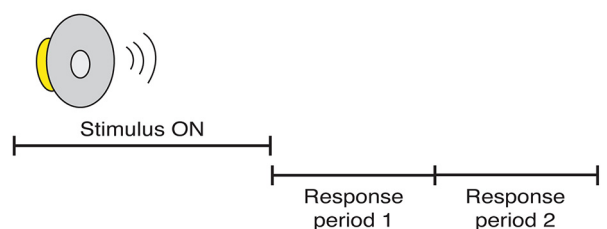
Figure 2 Task design. *A*, For the RT task, subjects indicated their choice (i.e., sequence direction) any time after onset of the sequence. *B*, For the variable-duration task, subjects indicated their choice after sequence offset. *C*, For the hybrid task, subjects indicated sequence direction and whether they heard the sequence as one sound or discrete sounds after sequence offset in separate response periods. For all three tasks, subjects were provided feedback regarding frequency direction at the end of each trial (not shown).

decreasing) and coherence (0, 10, 25, 50, or 100%) were varied on a trial-by-trial basis.

Variable-duration task

For this task, the experimenter controlled listening duration: subjects reported their direction decision following offset of the auditory sequence (Fig. 2B). Five of the six subjects tested in the RT task (4 male and 1 female) also participated in six weekly 1.5 h sessions of the variable-duration task. On each trial, we chose the duration of the sequence by sampling from a truncated exponential distribution (rate parameter = 2000 ms for all IBIs); choosing the sequence duration from this distribution minimized the possibility that subjects could anticipate the end of the sequence (Gold and Shadlen, 2003). The upper- and lower-stimulus durations were a function of IBI: for IBI = 10 ms, min = 160, max = 1400 ms; for IBI = 60 ms, min = 360, max = 3150 ms; and for IBI = 150 ms, min = 720, max = 6300 ms. These limits were chosen so that each sequence, independent of IBI, contained 4-35 tone bursts. At the end of each sequence, a response cue was flashed on the LCD monitor and subjects had 800 ms to respond. Each session contained four blocks of trials; a short break was provided between blocks. In each block, we varied sequence direction (increasing or decreasing), IBI (10, 60, or 150 ms), and coherence (0, 10, 25, 50, or 100%) on a trial-by-trial basis. Each combination of these sequence parameters was presented five times within a block, for a total of 150 trials per block.

Hybrid task

For this task, subjects participated in a version of the variable-duration task that also required them to report, on each trial, whether they heard the sequence as one sound or as a series of discrete sounds (Fig. 2C). Four of the six subjects tested in the RT task plus one new subject (3 male and 2 female) participated in four weekly 1.25 h sessions of the hybrid task. Sequence duration was sampled from a truncated exponential distribution and the limits of the stimulus durations were set to ensure that each sequence had 4-35 tone bursts. Subjects reported their two decisions during two separate 800 ms response periods. Prior to each sequence's onset, a colored cue, which was presented on the LCD monitor, indicated the temporal order in which subjects were to report their decisions; this order alternated on a block-by-block basis.

For this task, we set both IBI and coherence to values that were centered on each subject's psychophysical threshold. IBI threshold was the IBI value rated as one sound 50% of the time. Because the subjects' 50%-IBI threshold varied on a day-by-day basis, we measured this threshold daily, prior to their participation in the hybrid task. IBI threshold was measured using a one-up/one-down adaptive procedure (Treutwein, 1995; García-Pérez, 1998). The sequence always had 16 tone bursts and used a 50% coherence stimulus.

Coherence threshold was defined to be 70.7% correct performance, which corresponds to a d′ of 0.77. Because preliminary experiments indicated that coherence threshold was constant across experimental sessions, it was measured once for each subject. Coherence threshold was calculated using a two-up/one-down adaptive procedure (Treutwein, 1995; García-Pérez, 1998). During this procedure, the sequence's IBI was set to the IBI threshold.

For the hybrid task, each session contained four blocks of trials; a short break was provided between blocks. We varied IBI, on a trial-by-trial basis, between three different values: (1) IBI threshold minus 15 ms (20 trials per block), (2) IBI threshold (80 trials per block), and (3) IBI threshold plus 15 ms (20 trials per block). Coherence was set to each subject's coherence threshold.

Fitting of behavioral data to sequential-sampling models

Behavioral data were fit to variants of sequential-sampling models related to the drift diffusion model (DDM) (Ratcliff et al., 2004; Smith and Ratcliff, 2004; Gold and Shadlen, 2007; Green et al., 2010; Brunton et al., 2013; Mulder et al., 2013) to quantify the effects of sequence coherence and IBI on the decision-making process. These models describe the process of converting incoming sensory evidence, which is represented in the brain as the noisy spiking activity of populations of relevant sensory neurons, into a decision variable that can guide behavior.

RT task

For the RT task, a key benefit of these sequential-sampling models is that they make quantitative predictions about both choice and RT as a function of the coherence of the auditory sequence. In other words, these models simultaneously fit: (1) the psychometric function, which describes accuracy versus sequence coherence; and (2) the chronometric function, which describes RT versus sequence coherence. We used several model variants.

Model variant #1

The first variant was a standard, symmetric DDM in which a perceptual decision is based on an accumulation over time of noisy evidence to a fixed bound, a process that is mathematically equivalent to the one-dimensional movement of a particle undergoing Brownian motion to a boundary (Ratcliff et al., 1999; Gold and Shadlen, 2002; Shadlen et al., 2006; Eckhoff et al., 2008; Green et al., 2010; Ding and Gold, 2012; Brunton et al., 2013; Mulder et al., 2013). In brief, this version had seven free parameters: one drift rate (k) per IBI; one symmetric bound for either increasing or decreasing ($+A$ or $-A$; i.e., the height of the bound was the same for both choices but with an opposite sign) choices per IBI; and a single non-decision time (T_{ND}) that accounts for sensory-processing and motor-preparation time. Drift rate governs sensitivity and is implemented in terms of the moment-by-moment sensory evidence, which has a Gaussian distribution $N(\mu,1)$ with a mean μ that scales with sequence coherence (C): $\mu = kC$. The decision variable is the temporal accumulation of this momentary sensory evidence. A decision (i.e., the subject reports that the sequence is increasing or decreasing) occurs when this decision variable reaches a decision bound ($+A$ or $-A$, respectively). The decision time is operationally defined as the time between

auditory-sequence onset and the cross of either bound. RT is the sum of the decision time and the associated non-decision time. The probability that the decision variable first crosses the $+A$ bound is $\frac{e^{2\mu A} - 1}{e^{2\mu A} - e^{-2\mu A}}$. The mean decision time is $\frac{2A}{\mu}\coth(2\mu A) - \frac{A}{\mu}\coth(\mu A)$ for increasing choices and $\frac{2A}{\mu}\coth(2\mu A) - \frac{A}{\mu}\coth(\mu A)$ for decreasing choices.

Model variant #2
The second variant was a leaky accumulator, in the form of a stable Ornstein−Uhlenbeck process (Busemeyer and Townsend, 1993; Bogacz et al., 2006), in which positive values of the leak term imply that a given decision is influenced most strongly by the most recent samples of sensory evidence. Because these processes do not have simple, exact analytic solutions for both the psychometric and chronometric functions, we conducted simulations to fit the data. For each fit, we simulated 40,000 trials per iteration, with the decision variable computed in 1 ms steps.

Model variant #3
The final variant was an accumulate-to-bound model with non-leaky drift and bounds that can collapse (decrease towards zero) as a function of time within a trial (Ditterich, 2006b; Drugowitsch et al., 2012; Thura et al., 2012; Hawkins et al., 2015). The dynamics of the collapsing process were governed by a two-parameter Weibull function (Hawkins et al., 2015). These fits were also obtained using simulations like the leaky-accumulator model, above.

All model fits were computed by first using a pattern-search algorithm to find suitable initial conditions ("patternsearch" in the Matlab programming environment) and then a gradient descent to find the best-fitting parameters ("fmincon" in Matlab) that minimized the negative log-likelihood of the data, given the model fits (Palmer et al., 2005). The likelihood function for subjects' choices was modeled as binomial errors and the mean response times were modeled as Gaussian errors.

Variable-duration and hybrid tasks
For these tasks, the experimenter, not the subject, controlled listening duration. Therefore, we used models that had a different stopping rule than those used for the RT tasks. Specifically, these models assumed that the decision was made based on the sign of the accumulated evidence at the end of the stimulus presentation. These models included two basic parameters: (1) a coherence-scaling term (k), which governed the relationship between sequence coherence and the strength of the evidence to accumulate; and (2) accumulation leak (λ), which governed the efficiency of the accumulation process. We computed the probability (p) of a correct response as a function of sequence coherence (C) and listening duration (D) as: $p(C, D) = 0.5 * \left(1 - (1 - L) * \varphi\left(- kC\sqrt{2(e^{\lambda D} - 1) / \lambda(e^{\lambda D} + 1)}\right)\right)$, where φ is the normal standard cumulative distribution function. The lapse rate (L) was set to a small value (0.01) to provide better fits (Klein, 2001; Wichmann and Hill, 2001). The models were fit using Matlab's "fmincon" function to minimize the log-likelihood of the data, given the parameters, and assuming binomial errors.

Fitting of RT task data to the LATER model
RT distributions from the RT task were also analyzed using the LATER (Linear Approach to Threshold in Ergodic Rate) model. This model assumes that RT distributions are distributed as an inverse Gaussian because they result from a process with a linear rate of rise, which are distributed across trials as a Gaussian, that trigger a movement when reaching a fixed threshold (Carpenter and Williams, 1995; Reddi et al., 2003). We used maximum-likelihood methods to fit RT distributions to a model with two free parameters that represented the mean rate-of-rise and the threshold. We tested if and how each parameter varied with IBI for each subject, direction, and stimulus coherence.

Calculating psychophysical kernels from the RT data
Finally, we calculated the subjects' psychophysical kernels from the RT task to further support the idea that the subjects were using a process akin to bounded accumulation. The kernels were computed directly from the data by taking, from each 0% coherence trial, the mean-subtracted stimulus sequence and then computing the mean (and standard error) value of these time-dependent sequences separately for trials leading to increasing or decreasing choices. We fit these kernels to a model that sorted the 0%-coherence trials into two categories based on the sign of the slope of a linear regression of frequency versus burst number. We used a grid search to find the values of two parameters—one governing the number of tone bursts from stimulus onset that is used to compute the linear regression and a second that scales the stimulus-frequency values used to compute the kernel—that maximized the likelihood of obtaining the IBI-specific kernels measured from the data given the model.

Results
We used human psychophysics to test if and how the time course of evidence accumulation for an auditory-discrimination task was affected by a temporal manipulation that modified the perceptual grouping of the sensory evidence. The task required subjects to report whether the frequency direction of a tone-burst sequence was increasing or decreasing (Fig. 1A). Task difficulty was manipulated by controlling the coherence of the sequence, which corresponded to the fraction of tone bursts whose frequencies increased or decreased systematically (Fig. 1A−C).

The temporal manipulation was a change in the IBI of the stimulus sequence, which affected perceptual grouping. For short IBIs ($<\sim$30 ms), subjects tended to report that the sequence was one sound. For medium IBIs ($\sim$30-100 ms), subjects alternated trial-by-trial between reports that the sequence was one sound or a series of discrete sounds. For long IBIs ($>\sim$100 ms), subjects reliably reported that the sequence was a series of discrete sounds (Fig. 3). To test how this grouping manipulation affected the temporal dynamics of the perceptual decision (i.e., the frequency direction of the tone-burst sequence), we used
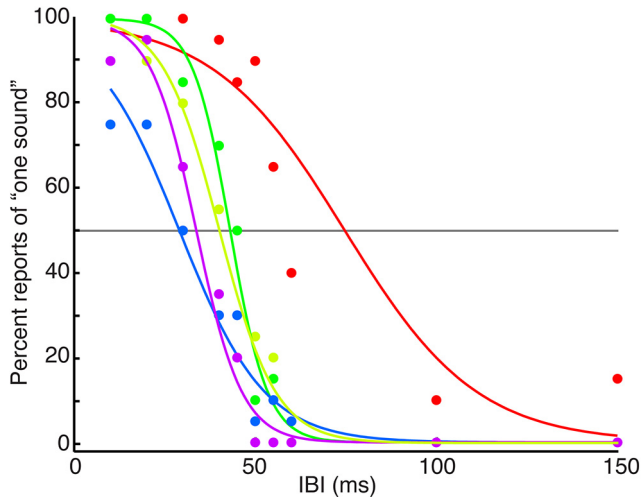
Figure 3 Influence of IBI on reports of perceived grouping. Subjects reported whether they perceived the stimulus as one sound or discrete sounds. The graph shows proportion of trials in which each subject chose one sound as a function of IBI. The points indicate each subject's performance. Each curve represents a logistic function that was fit to each subject's reports across four sessions. The gray line indicates 50% of trials reported as "one sound", which was IBI threshold. Colors represent different subjects.

three versions of a discrimination task: an RT task in which subjects controlled listening duration; a variable-duration task in which the experimenter controlled listening duration; and a hybrid task that combined the variable-duration task with an explicit grouping judgment about whether the sound was one sound or a series of discrete sounds. Results from each task are presented below.

Response time task
Both accuracy and RT depended systematically on coherence and IBI. Figure 4 summarizes the behavioral per-

formance of all six subjects. Choices tended to be more accurate and faster when stimulus coherence was high than when it was low.

Subjects' choices were clearly based on the direction of frequency change and not simply the absolute frequency value of the sequence. Specifically, when we analyzed behavior relative to the beginning, mean, and ending frequency value of the sequence, performance was near chance: the median (full range) percent correct of all six subjects' choices, when collapsed across coherences and IBIs, was 56.9% (55.3-58.0), 50.6% (50.2-51.9), and 55.4% (54.6-56.6), respectively.

We also found that the relationships between choice, RT, and coherence were modulated by IBI. Longer IBIs led to shallower psychometric functions (i.e., lower sensitivity; Fig. 4A) and longer RTs (Fig. 4B). When we considered the portion of the RT that only included presentation of the tone bursts and not the silent periods by subtracting out the cumulative IBIs (which we refer to as "signal RT" and which treats the cumulative IBIs, like sensory and motor processing, as part of the non-decision time on a given trial), the effects of IBI on RT were reversed: the longest signal RTs corresponded to the shortest IBIs (Fig. 4C).

We quantified the effects of coherence and IBI on the time-dependent decision process by fitting the choice and RT data for each subject and from pooled data across subjects to several variants of models that are related to the DDM. All of the models had the same basic form. They assumed that the decision was based on the temporal accumulation of noisy evidence until reaching one of two prespecified boundaries. The signal-to-noise ratio of this decision variable was governed by a drift rate, which was proportional to stimulus coherence and, in some cases, was also subject to leaky accumulation (Busemeyer and Townsend, 1993; Usher and McClelland, 2001; Tsetsos et al., 2012). Choice was governed by the identity of the reached boundary, which, in some cases, could change as a function of time within a trial to reflect an increasing
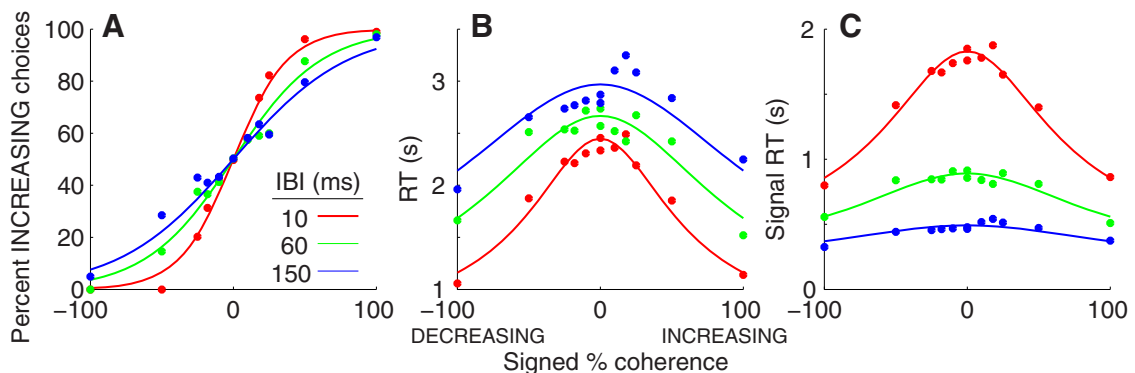
Figure 4 Performance on the RT task. **A,** The fraction of trials in which subjects reported a sequence of tone bursts was increasing in frequency as a function of signed coherence and IBI. Positive coherence values indicate that the sequence was increasing in frequency; negative indicates decreasing. **B,** The mean RT (i.e., the time between sequence onset and button press) as a function of signed coherence and IBI on correct trials only (and all 0% coherence trials). **C,** As in **B,** but using signal RT (i.e., RT not including cumulative IBI). The solid curves are simultaneous fits of psychometric and chronometric data to a DDM (see Results). The psychometric data (**A**) only show the fit to the signal-RT data (**C**). In each panel, the points indicate performance data that was pooled across subjects. Colors represent different IBIs, as indicated in **A.**

Table 1 Statistical table

	Data structure	Statistical test	Power
a	Normal distribution	Pearson Correlation	$p < 0.01$
b	Normal distribution	Likelihood-ratio test; Bonferroni-corrected for three parameters	$p < 0.001$
c	Normal distribution	Likelihood-ratio test; Bonferroni-corrected for three parameters	$p < 0.01$
d	Normal distribution	Likelihood-ratio test	$p > 0.24$
e	Normal distribution	Likelihood-ratio test	$p > 0.1$
f	Normality not assumed	Mann−Whitney test	$p < 0.01$
g	Normality not assumed	Kruskal−Wallis test	$p < 0.001$
h	Normal distribution	Likelihood-ratio test, Bonferroni-corrected for two parameters	$p < 0.01$
i	Normality not assumed	Kruskal−Wallis test	$p > 0.05$
j	Normal distribution	Likelihood-ratio test, Bonferroni-corrected for two parameters	$p < 0.01$
k	Normal distribution	Likelihood-ratio test	$p = 0.2838$
l	Normality not assumed	Kruskal−Wallis test	$p = 0.011$
m	Normality not assumed	Kruskal−Wallis test	$p = 0.125$
n	Normal distribution	Likelihood-ratio test	$p > 0.05$
o	Normality not assumed	Kruskal−Wallis test	$p > 0.05$
p	Normality not assumed	Mann−Whitney test	$p < 0.05$
q	Normal distribution	Likelihood-ratio test, Bonferroni-corrected for two parameters	$p < 0.01$
r	Normal distribution	Likelihood-ratio test, Bonferroni-corrected for two parameters	$p < 0.01$

urgency to respond (Ditterich, 2006b; Drugowitsch et al., 2012; Thura et al., 2012; Hawkins et al., 2015). RT was governed by the time to reach the boundary plus extra non-decision time. The height of the boundary governed the trade-off between speed and accuracy: higher boundaries provided longer decision times and higher accuracy, whereas lower boundaries increased speed at the expense of accuracy (Gold and Shadlen, 2007). Because performance tended to be symmetric for the two choices [across all subjects and IBIs and using balanced stimulus presentations, the median (interquartile range) absolute value of the difference in the fraction of increasing vs decreasing choices was only 0.001 (0.000–0.004)], we used an unbiased model with seven parameters: one non-decision time, three parameters representing drift rate per each of the three IBI conditions, and three parameters representing the bound height per IBI.

Fits of this model to choice data and either raw or signal RTs were consistent with a decision variable that was based on the signal portion of the stimulus sequence but not the time between bursts (the IBI). Drift rate has units of change of standardized evidence per unit time. Therefore, converting from raw to signal RT affects the (linear) scaling of this value, which is governed by the duty cycle associated with the given IBI: $b/(\text{IBI} + b)$, where b is burst duration of 30 ms. Accordingly, best-fitting values of the DDM parameters that were fit to choices and raw RT were scaled versions of those fit to choices and signal RTs, with the scale factors approximately equal to the IBI-specific duty cycles [slopes of linear regressions of subject-specific signal vs raw drift rates = 0.80 (duty cycle = 0.75), 0.43 (0.33), and 0.23 (0.17) for IBI = 10, 60, and 150 ms, respectively]. Furthermore, best-fitting signal drift rates that were rescaled and expressed in units of the change in evidence per unit of raw time (i.e., multiplied by the duty cycle) were strongly correlated with the associated, best-fitting raw drift rates across IBIs and subjects ($r = 0.98$, $p < 0.01^a$; Table 1). These results imply that the IBI manipulation affected only the duty-cycle-dependent

scaling of drift rates. Therefore, we used signal RTs for the model fits, which corresponded to drift rates that had the same temporal scaling and thus could be compared directly across IBI conditions.

We found that the effects of IBI on choice and signal RT primarily reflected changes in the decision boundary but not the drift rate (Fig. 5). The height of a symmetric, fixed decision boundary (i.e., the same height for both increasing and decreasing choices) declined systematically with increasing IBI for all six subjects and for data combined across subjects (likelihood-ratio test comparing a seven-parameter model with separate values of drift rate and bound height per IBI plus a non-decision time to a five-parameter model with a single value of drift rate shared across IBIs, $p < 0.001^b$ in all cases; Bonferroni-corrected for three parameters; Fig. 5A). In contrast, drift rate depended on IBI for only one of the six subjects ($p < 0.01^c$; Bonferroni-corrected for three parameters) and not for the other subjects or combined data (Fig. 5B). These model fits were not improved by adding to the model either leaky accumulation (likelihood-ratio test, $p > 0.24^d$ across subjects and for all data combined) or collapsing bounds (likelihood-ratio test, $p > 0.1^e$ for five of the six subjects and for all data combined). There also was little evidence for slow errors that can be expected in models with collapsing bounds, with only eight of 216 conditions separated by subject/coherence/IBI showing such an effect (Mann−Whitney test comparing median correct vs error RTs, $p < 0.01$; Ditterich, 2006a)[f]. Likewise, there was little evidence for fast errors that can be expected in models with variable bounds, with only three conditions showing such an effect (Ratcliff and Rouder, 1998). Thus, changes in IBI, which affected perceptual grouping and the rate of arrival of decision-relevant signals, caused systematic, robust changes in the speed−accuracy trade-off governed by a fixed, time-independent bound. In contrast, the changes in IBI did not cause systematic changes in the efficiency with which sensory evidence was accumulated over time to form the decision.
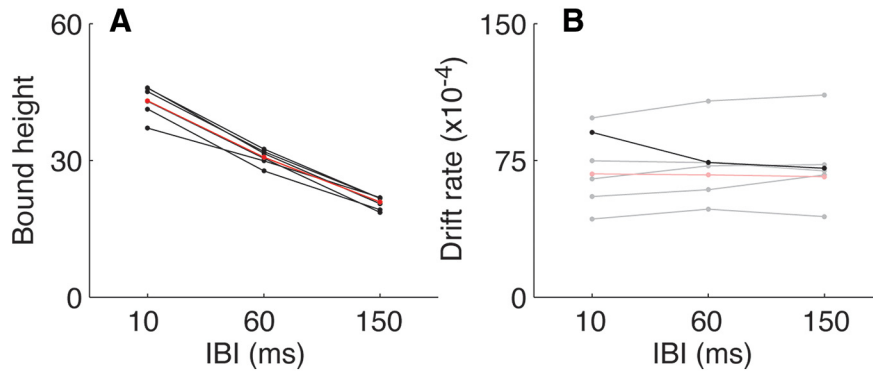
Figure 5 Parameter values from fits of the basic DDM to the RT-task data. Each panel shows best-fitting values of bound height (**A**) or drift rate (**B**) plotted as a function of IBI for fits to data from individual subjects (black) or combined across all subjects (red). Dark lines/symbols indicate that the model fits were improved significantly by fitting the given parameter separately for each IBI condition (likelihood-ratio test, $p < 0.01$, Bonferroni-corrected for three parameters).

These results were supported by independent analyses of the signal-RT distributions. A useful way to assess possible changes in drift rate and/or bound height in a simple accumulate-to-bound framework is to use the LATER model (Carpenter and Williams, 1995; Reddi et al., 2003). According to this model, a decision variable rises linearly to a threshold (bound) in order to trigger a motor response. Assuming a fixed bound, but a noisy decision variable with a rate of rise that is normally distributed across trials, RT is distributed as an inverse Gaussian. This distribution can be plotted as a straight line on "reciprobit" axes (i.e., percent cumulative frequency on a probit scale vs the reciprocal of RT from 0%-coherence trials; Fig. 6A). Horizontal shifts of these lines imply changes in the mean rate-of-rise of the decision variable, whereas swivels about a fixed point at infinite RT imply changes in the bound height (Reddi et al., 2003). When we fit the LATER model to signal-RT data separately for each subject, coherence, and IBI (correct trials only), we found that increasing IBI caused systematic decreases in the bound (Kruskal−Wallis test for H_0: equal median values per IBI, across subjects and coherences, $p < 0.001^g$; 34 of 36 individual subject−coherence pairs had a significant dependence of bound height on IBI, all of which had a

lower bound for the longest vs the shortest IBI, $p < 0.01^h$, likelihood-ratio test, Bonferroni-corrected for two parameters; Fig. 6B). In contrast, increasing IBI did not cause a systematic change in the best-fitting mean rate-of-rise (Kruskal−Wallis test, $p > 0.05^i$; 13 of 36 individual subject−coherence pairs had a significant dependence of rate-of-rise on IBI, of which seven showed an increasing rate-of-rise and six showed a decreasing rate-of-rise; $p < 0.01^j$, likelihood-ratio test, Bonferroni-corrected for two parameters; Fig. 6C).

These results were also supported by correlation analyses that related choices to the noisy auditory stimulus (Knoblauch and Maloney, 2008; Murray, 2012). We computed two kernels per IBI condition, one for increasing choices and the other for decreasing choices, from the 0%-coherence trials from all six subjects. Each kernel represented the mean, within-trial time course of the mean-subtracted, stochastic auditory sequence that led to the given choice (Fig. 7). Subjects made increasing choices when the frequency tended to increase throughout most of the trial, with average kernels that started below the within-trial mean, then increased steadily to a peak value above the within-trial mean around the time of the median RT, then reverted back towards the mean.
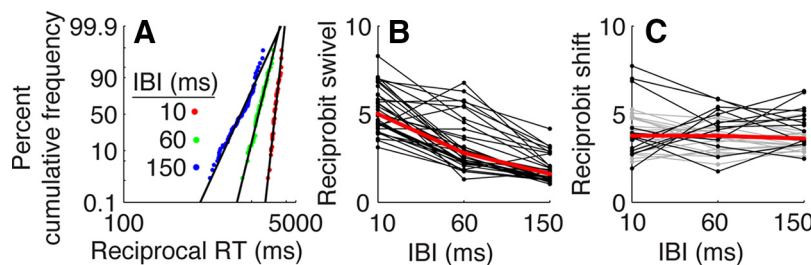
Figure 6 LATER model fits to signal RT data. **A**, Distributions of signal RT from 0%-coherence trials for one subject are plotted on a reciprobit plot: reciprocal RT versus percentage of cumulative frequency on a probit scale (Reddi et al., 2003), separately per IBI. Best-fitting values of the bound height (**B**) and mean rate-of-rise (**C**) of the LATER model (see Results and Materials and Methods for details) are plotted as a function of IBI for each subject and coherence (black/gray lines and data points). The data in black indicate that the model fits were improved significantly by fitting the given parameter separately for each IBI condition (likelihood-ratio test, $p < 0.01$, Bonferroni-corrected for two parameters). Shaded lines/symbols indicate that the model fits were not improved significantly. Red data points/lines represent the median values across all conditions.
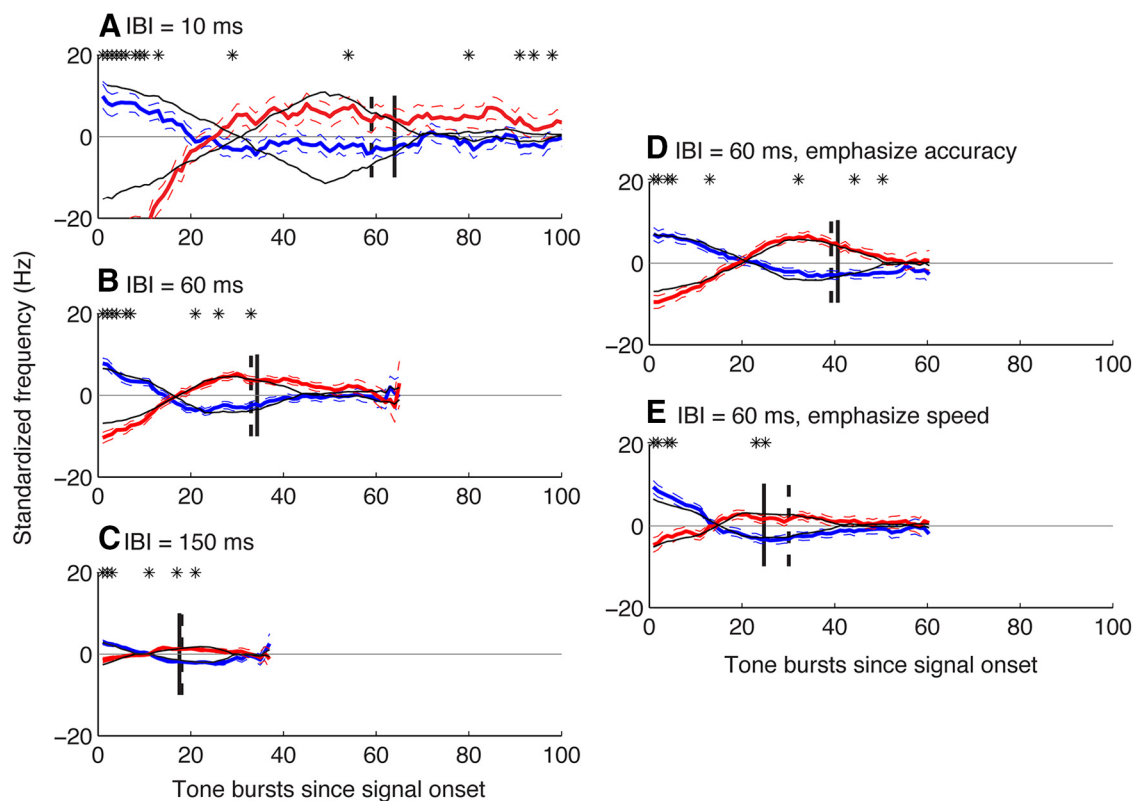
Figure 7 Psychophysical kernels. Kernels represent the average of the stimulus sequence (mean subtracted on each trial) presented on all 0%-coherence trials across subjects for increasing (red) and decreasing (blue) choices. *A−C*, Kernels computed per IBI condition, as indicated, and smoothed using a 21-sample moving mean. *D*, *E*, Kernels for the 60 ms IBI condition when subjects were told to emphasize accuracy or speed, respectively. Thick and broken lines are mean and standard error, respectively. Data are aligned relative to onset of the sequence. Asterisks indicate that the increasing and decreasing kernels were significantly different from one another for a given time bin (Mann−Whitney test, $p < 0.05^p$, using the raw, unsmoothed kernels). Black lines indicate best-fitting simulated kernels (see text for details). Solid vertical lines indicate median RT. Dashed vertical lines indicate the end of the integration time from the best-fitting simulated kernels.

Likewise, subjects made decreasing choices on trials in which the frequency progression of the stochastic stimulus moved in the opposite direction, starting relatively high and then decreasing for much of the trial.

These kernels were consistent with a DDM-like decision process that had lower bounds for longer IBIs, corresponding to an increasing emphasis on speed at the expense of accuracy. Specifically, the choice selectivity of these kernels (i.e., the time bins in which increasing and decreasing kernels differed from each other), measured in terms of tone bursts within a stimulus sequence and thus consistent with the signal RT analyses described above, was longest for the shortest IBI and shortest for the longest IBI (compare asterisks in Fig. 7*A−C*). We found a similar effect when using just the 60 ms IBI but providing explicit instructions to the subjects about the speed−accuracy trade-off, with relatively longer choice selectivity under an "emphasize accuracy" condition (Fig. 7*D*) and relatively shorter choice selectivity under an "emphasize speed" condition (Fig. 7*E*). These kernels were also qualitatively consistent with an evidence-accumulation process with little or no leak because choice selectivity was strongest at the beginning of a trial; if leak was present, it

would tend to show up as less choice selectivity at the beginning of a trial.

To more quantitatively relate these kernels to the underlying decision process, we fit them to a simple, two-parameter model that assumed that choices were based on particular frequency progressions within the given stimulus. One parameter governed the time course of the relevant progression, which could range from just the first two bursts to the full sequence; i.e., simulated increasing or decreasing choices occurred when the slope of a linear regression of frequency versus burst number for the first n bursts in a sequence was >0 or <0, respectively. The second parameter scaled the contribution of each stimulus sequence to the final kernel, akin to the signal-to-noise ratio (SNR) of the internal stimulus representation. We found that the best-fitting value of the integration time decreased systematically with increasing IBI (and for the "emphasize speed" relative to the "emphasize accuracy" instruction), in each case closely matching the IBI-specific median RTs (compare solid and dashed vertical lines in Fig. 7). In contrast, the best-fitting scale factor did not differ as a function of IBI (likelihood-ratio test, $p = 0.28^k$), implying a consistent SNR across conditions (like the
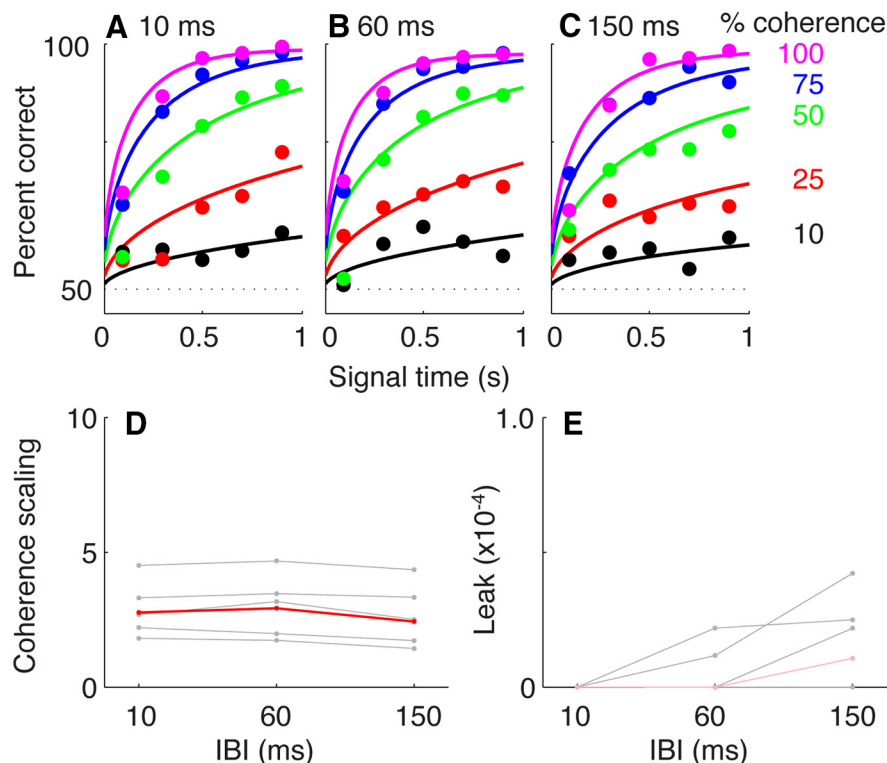
Figure 8 Performance on the variable-duration task. **A–C**, Psychometric data are plotted as a function of listening duration for different coherences and IBIs, as indicated. Each data point reflects mean performance for all five subjects as a function of coherence and signal time (plotted in 0.2 s bins, up to 1.0 s, but fit using unbinned data). The solid curves are fits from the best-fitting model with two parameters: drift rate and accumulation leak. **D, E**, Best-fitting values of drift rate (**D**) and accumulation leak (**E**) plotted as a function of IBI for fits to data from individual subjects (black) or combined across all subjects (red). Dark lines and symbols indicate that the model fits were improved significantly by fitting the given parameter separately for each IBI condition (likelihood-ratio test, $p < 0.01^q$, Bonferroni-corrected for two parameters). Shaded lines and symbols indicate that the model fits were not improved significantly.

IBI-independent drift rate in Fig. 5). Thus, like the DDM- and LATER-based analyses described above, these kernel analyses implied that the IBI manipulation affected the speed−accuracy trade-off on the RT task but not how information was accumulated over time to form the decision.

Variable-duration task
Because the psychometric and chronometric data from the RT task depended on each subject's individual speed-−accuracy tradeoff, it is possible that the DDM-model fits reflected complex interactions between the rate of sensory-evidence accumulation and bound heights across IBI conditions (Ratcliff and Tuerlinckx, 2002). To directly test the relationship between IBI and the rate of evidence accumulation, subjects participated in the variable-duration task. During this task, we experimentally controlled the duration of the auditory sequence and, hence, the amount of sensory evidence (Fig. 2B). Analogous to the RT task, we analyzed performance as a function of signal time to standardize the sequence duration with respect to the rate of tone-burst presentation.

Mean performance accuracy for all five subjects improved systematically as a function of both coherence and signal time, in a manner that was qualitatively similar

for all three IBI conditions (Fig. 8A−C). For each condition, accuracy tended to reach an upper asymptote of >99% correct in <1000 ms of signal time for the highest coherences; accuracy rose steadily at longer listening times for lower coherences.

We again quantified these effects by fitting the choice data to DDM-like models (Fig. 8D,E). Like the DDM described in the Response time task section (above), all of these models assumed that the decision was based on the value of a decision variable, which represented the accumulation of noisy sensory evidence over time. Like the RT fits, we assumed a drift rate that scaled linearly with coherence and an accumulation process that might include a leak. However, unlike bounded diffusion in the DDM when it was applied to RT data, these models assumed that the accumulation process continued until the stimulus was turned off, at which point the decision was based on the current sign of the decision variable (Gold and Shadlen, 2003; Bogacz et al., 2006). Therefore, these models did not have parameters representing bounds or non-decision times and were fit to psychometric data only (percent correct as a function of both coherence and listening duration).

The model-fitting results suggested only modest, if any, dependence of drift rate or leak on IBI. Drift rate was
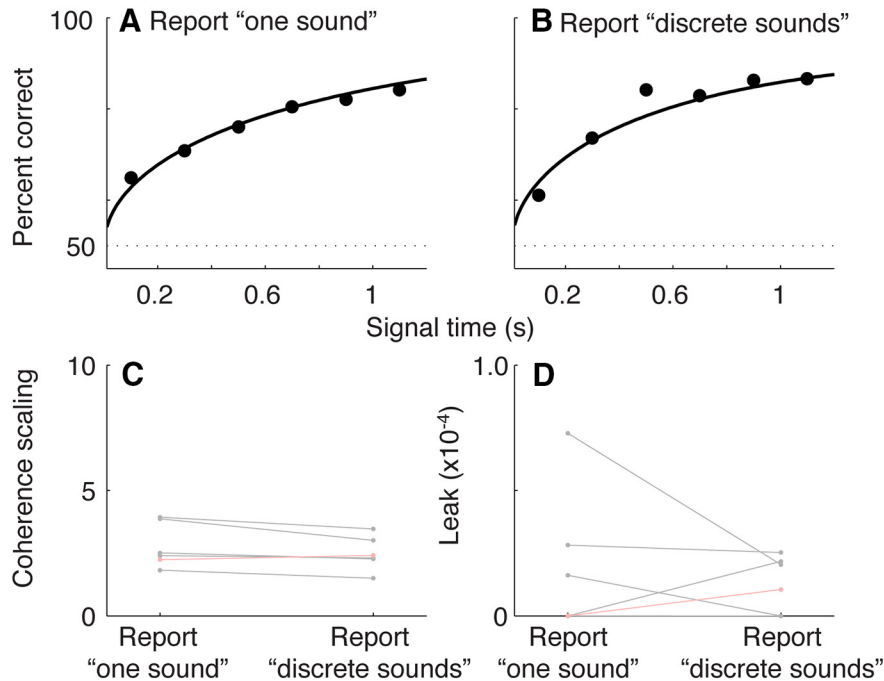
Figure 9 Performance on the hybrid task. **A**, **B**, Psychometric data are plotted as a function of listening duration for trials in which the subject reported that the sequence was one sound (**A**) or a series of discrete sounds (**B**). The solid curves are fits from the best-fitting model with two parameters: drift rate and accumulation leak. **C, D,** Best-fitting values of drift rate (**C**) and accumulation leak (**D**) plotted as a function of perceptual grouping for fits to data from individual subjects (black) or combined across all subjects (red). The model fits were not improved by fitting the given parameter separately for each IBI condition (likelihood-ratio test, $p > 0.1$[r] in all cases, Bonferroni-corrected for two parameters).

independent of IBI for all five subjects and depended non-monotonically on IBI for data combined across subjects (Fig. 8*D*). There was also a slight trend for the best-fitting values of drift rate to depend systematically on IBI across subjects (Kruskal−Wallis test for H_0: equal median values per IBI, across subjects, $p = 0.011$[l]). Accumulation leak was not significantly affected by IBI for any of the individual subjects or the data combined across subjects (Kruskal−Wallis test for H_0: equal median values per IBI, across subjects $p = 0.125$[m]; Fig. 8*E*). Thus, the subjects' decisions improved systematically as a function of the number of tone bursts but were largely independent of the time between bursts.

Hybrid task

To test directly the relationship between perceptual grouping and the decision process, subjects participated in the hybrid task, which is a variant of the variable-duration task (Fig. 2*C*). In this task, on each trial, tone-burst sequence was set to each subject's coherence threshold and IBI threshold (see Materials and Methods, above). At the end of each trial, the subject gave two sequential responses to indicate: (1) perceptual grouping (i.e., one sound or a series of discrete sounds) and (2) sequence direction (i.e., increasing or decreasing). Figure 9 shows psychometric data pooled across all of the subjects, separated into trials in which the subject reported perceiving the sequence as one sound (Fig. 9*A*) or a series of discrete sounds (Fig. 9*B*). Accuracy tended to increase

steadily as a function of listening duration in a similar manner for the RT and variable-duration tasks.

To quantify these effects, we fit the data to the same models as those described for the variable-duration task but applied to one sound versus discrete conditions instead of different IBIs (Fig. 9*C,D*). These fits indicted that the grouping report did not have any effect on the perceptual decision. For both the drift-rate and accumulation-leak parameters, the data from individual subjects and across subjects were better fit by models that used a single parameter for all trials, as opposed to separate parameters for one sound and discrete reports (likelihood-ratio test, $p > 0.05$[n]). Moreover, when the data were fit separately for the two grouping reports, the resulting best-fitting values did not differ from each other across subjects (Kruskal−Wallis test for H_0: equal median values per grouping judgment, across subjects, $p > 0.05$[o] for both drift and leak). Thus, the perceptual-grouping judgment did not appear to have a substantial effect on the accumulation efficiency of the sensory evidence.

Discussion

We examined the relationship between auditory perceptual grouping and decision-making. Our focus was on the role of time in both processes. Specifically, does the temporal manipulation of a stream of acoustic events, which affects its perceptual grouping (Bregman, 1990), affect decisions about its identity? We found that the time interval between sequentially presented tone bursts had

strong effects on whether the tone bursts were perceptually grouped as a single sound or heard as discrete sounds (Fig. 3). In contrast, this manipulation did not systematically affect how subjects accumulated information to form a decision about whether frequencies of the tone bursts were increasing or decreasing (Figs. 4-9). Thus, for this task and stimulus, the temporal accumulation of sensory evidence is invariant to the temporal intervals (gaps) between pieces of evidence that affects perceptual grouping.

The effect of time gaps on the time-course of evidence accumulation has also been studied in the visual system (Kiani et al., 2013). In that study, the accumulation of visual evidence was invariant to the temporal gap between pulses of motion evidence for a visual motion direction-discrimination task. We extended those findings by demonstrating that this accumulation invariance is accompanied by a change in the speed−accuracy trade-off that accounts for the different time intervals between the tone bursts. Our results also show that similar principles may govern the temporal dynamics of auditory and visual decisions, despite differences in how the underlying sensory mechanisms process temporal information (Carr and Friedman, 1999; Schnupp and Carr, 2009; Raposo et al., 2012). Below, we discuss the relationships between auditory perceptual grouping and decision-making and then discuss potential neural bases for our findings.

Temporal dynamics of auditory perceptual grouping and decision-making

Perceptual-grouping cues can affect auditory judgments (Bey and McAdams, 2002; Roberts et al., 2002; Micheyl and Oxenham, 2010; Borchert et al., 2011; Thompson et al., 2011). For example, judgments about the timing differences between auditory stimuli are more accurate when stimuli are grouped into the same auditory stream versus when they are segregated into different auditory streams (Roberts et al., 2002). Similarly, a listener's ability to detect a deviant tone burst improves when the tone burst is segregated (e.g., by frequency) into a separate auditory stream (Rahne and Sussman, 2009; Sussman and Steinschneider, 2009). In other situations, stream segregation enhances a listener's ability to identify a tone sequence (Bey and McAdams, 2002).

However, despite evidence for the roles of temporal cues in both auditory grouping and decision-making, little is known about how those roles interact. We addressed this issue by building upon on the rich history of auditory psychophysics and quantitative modeling (Green et al., 1957; Green, 1960; Greenwood, 1961; Luce and Green, 1972; Green and Luce, 1973). Specifically, we applied sequential-sampling models, in particular the DDM, to assess how auditory information presented sequentially over time was used to form a decision about the direction of change of the stimulus frequency.

We used a series of complementary approaches to demonstrate that subjects' decisions were consistent with a DDM-like process that accumulates sensory evidence over time. First, we fit choice and mean RT data from the RT task directly to several variants of the DDM,

all of which effectively described the relationships between stimulus coherence (strength), IBI, and the subjects' speed−accuracy trade-offs (Figs. 4, 5) (Green and Luce, 1973; Wickelgren, 1977; Palmer et al., 2005; Ratcliff and McKoon, 2008). Second, the subjects' full RT distributions were also consistent with a rise-to-bound process (Fig. 6), which we fit using a simplified version of DDM models (i.e., the LATER model) that assumes that the rising process is stochastic across trials (as opposed to within trials, for the DDM) and is effective at describing RT distributions across a range of conditions (Carpenter and Williams, 1995; Reddi et al., 2003). Third, analyses of our noisy auditory stimulus indicated that, at least on average, subjects were using information that extended from the beginning of the trial until around the time of the response (Fig. 7). That is, choice selectivity was strongest at the beginning of a trial. Fourth, performance on the variable-duration task increased systematically as a function of listening duration, in a manner consistent with the evidence-accumulation process described by the DDM (Figs. 8, 9).

Our primary result from these analyses was that the accumulation process was invariant to the IBI manipulation. In particular, we found that there was no systematic leak associated with the accumulation process in any of the tested task conditions. Further, the psychophysical kernels were consistent with an evidence-accumulation process with little leak: choice selectivity was strongest at the beginning of a trial and not at the end (which would be expected of a leaky process; Fig. 7). Thus, decisions were consistent with a lossless form of information accumulation (Brunton et al., 2013; Kiani et al., 2013). Moreover, the rate of information accumulation—which was measured in our models as a drift rate and represents the average rate of change of the underlying decision variable—depended only on the coherence of the tone bursts and not the temporal gaps between them. The temporal gaps, instead, affected the speed−accuracy trade-off of the decision (Figs. 5–7). Thus, the evidence-accumulation process was able to use the signals as they arrived without losing information in the intervening gaps, regardless of their duration.

This result was particularly striking in light of the fact that the temporal gaps had a strong effect on the subjects' percept of the tone-burst stimulus: short gaps gave rise to the percept of a single, grouped sound, whereas longer gaps gave rise to the percept of discrete sounds (Fig. 3). Our findings are, therefore, somewhat surprising, given that previous work has noted an interaction between perceptual grouping and auditory judgments (Bey and McAdams, 2002; Roberts et al., 2002; Micheyl and Oxenham, 2010; Borchert et al., 2011; Thompson et al., 2011). Further work is needed to fully explore these interactions and their relationship to temporal processing. For example, it might be interesting to add nonuniform temporal manipulations to each stimulus sequence (e.g., variable IBIs or tone-burst durations) to get a better sense of how specific timing cues presented at specific times in the stimulus sequence affect both grouping and decision-making.

Neural basis

Much of auditory perceptual grouping is pre-attentive and has substantive neural signatures in the auditory midbrain and auditory cortex (Carlyon et al., 2001; Fishman et al., 2004; Micheyl et al., 2007; Sussman et al., 2007; Pressnitzer et al., 2008; Shinn-Cunningham, 2008; Winkler et al., 2009). In contrast, auditory decision-making is generally associated with neural mechanisms that are found in the ventral auditory pathway of cortex (Romanski et al., 1999; Romanski and Averbeck, 2009; Rauschecker, 2012; Bizley and Cohen, 2013). In primates, this pathway includes the core and belt fields of the auditory cortex, which project directly and indirectly to regions of the frontal lobe. Because neural activity in these regions, particularly in the belt fields, is not modulated by subjects' choices, it is thought that this activity represents the sensory evidence used to form an auditory decision but not the decision itself (Tsunada et al., 2011; Tsunada et al., 2012; Bizley and Cohen, 2013; but see Niwa et al., 2012; Bizley et al., 2013). In contrast, frontal lobe activity is modulated by subjects' choices, consistent with the notion that neural activity in this part of the brain reflects a transition from a representation of sensory evidence to a representation of choice (Binder et al., 2004; Kaiser et al., 2007; Russ et al., 2008; Lee et al., 2009). This hierarchy of information processing is qualitatively similar to that seen in the visual and somatosensory systems (Shadlen and Newsome, 1996; Parker and Newsome, 1998; Gold and Shadlen, 2001; Romo and Salinas, 2001; Romo et al., 2002; Gold and Shadlen, 2007; Hernández et al., 2010).

Thus, our results imply that frontal-mediated decision-making can temporally accumulate evidence from the auditory cortex, independent of how that evidence has been parsed into temporally continuous or distinct groups earlier in the auditory pathway. One possible explanation for our invariance to grouping is that the decision computations use information that is processed separately from the grouping percept. Unfortunately, whereas several studies have reported signatures of grouping in the core auditory cortex (Fishman et al., 2004; Micheyl et al., 2005; Brosch et al., 2006; Selezneva et al., 2006; Elhilali et al., 2009; Shamma and Micheyl, 2010; Fishman et al., 2013; Noda et al., 2013) and representations of the grouping percept in non-core regions of the human auditory cortex (Gutschalk et al., 2005; Gutschalk et al., 2008), neurophysiological studies elucidating where and how perceptual-grouping cues interact with decisions have yet to be conducted. Taken together, however, these aforementioned studies predict that the grouping percept may be mediated in the auditory cortex and frontal activity represents the decision process. A second, alternative explanation is that the decision process may be very flexible and able to efficiently accumulate different forms of noisy evidence under different conditions (Brunton et al., 2013; Kaufman and Churchland, 2013). The degree of this flexibility might depend on the type and quality of the sensory information, memory load, the nature of the environment in which the subject is making the decisions, and other task demands.

References

Bey C, McAdams S (2002) Schema-based processing in auditory scene analysis. Percept Psychophys 64:844-854. Medline

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci 7:295-301. CrossRef Medline

Bizley JK, Cohen YE (2013) The what, where, and how of auditory-object perception. Nat Rev Neurosci 14:693-707. CrossRef Medline

Bizley JK, Walker KM, Nodal FR, King AJ, Schnupp JW (2013) Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr Biol 23:620-625. CrossRef Medline

Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol Rev 113:700-765. CrossRef Medline

Bogacz R, Wagenmakers EJ, Forstmann BU, Nieuwenhuis S (2010) The neural basis of the speed-accuracy tradeoff. Trends Neurosci 33:10-16. CrossRef Medline

Borchert EM, Micheyl C, Oxenham AJ (2011) Perceptual grouping affects pitch judgments across time and frequency. J Exp Psychol Hum Percept Perform 37:257-269. CrossRef Medline

Bregman AS (1990) Auditory scene analysis. Boston, MA: MIT.

Brosch M, Oshurkova E, Bucks C, Scheich H (2006) Influence of tone duration and intertone interval on the discrimination of frequency contours in a macaque monkey. Neurosci Lett 406:97-101. CrossRef Medline

Brunton BW, Botvinick MM, Brody CD (2013) Rats and humans can optimally accumulate evidence for decision-making. Science 340:95-98. CrossRef Medline

Busemeyer JR, Townsend JT (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psych Rev 100:432-459. CrossRef

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

Carpenter RH, Williams ML (1995) Neural computation of log likelihood in control of saccadic eye movements. Nature 377:59-62. CrossRef Medline

Carr CE, Friedman MA (1999) Evolution of time coding systems. Neural Comput 11:1-20. Medline

Ding L, Gold JI (2012) Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. Neuron 865-874. CrossRef

Ditterich J (2006a) Stochastic models of decisions about motion direction: behavior and physiology. Neural Netw 8:981-1012.

Ditterich J (2006b) Evidence for time-variant decision making. Eur J Neurosci 24:3628-3641. CrossRef Medline

Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The cost of accumulating evidence in perceptual decision making. J Neurosci 32:3612-3628. CrossRef Medline

Eckhoff P, Holmes P, Law C, Connolly PM, Gold JI (2008) On diffusion processes with variable drift rates as models for decision making during learning. New J Phys 10:nihpa49499. CrossRef Medline

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA (2009) Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61:317-329. CrossRef Medline

Fishman YI, Arezzo JC, Steinschneider M (2004) Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116:1656-1670. Medline

Fishman YI, Micheyl C, Steinschneider M (2013) Neural representation of harmonic complex tones in primary auditory cortex of the awake monkey. J Neurosci 33:10312-10323. CrossRef Medline

García-Pérez MA (1998) Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. Vision Res 38:1861-1881. CrossRef

Gold JI, Shadlen MN (2001) Neural computations that underlie decisions about sensory stimuli. Trends Cogn Sci 5:10-16. Medline

Gold JI, Shadlen MN (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36:299-308. Medline

Gold JI, Shadlen MN (2003) The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. J Neurosci 23:632-651. Medline

Gold JI, Shadlen MN (2007) The neural basis of decision making. Annu Rev Neurosci 30:535-574. CrossRef Medline

Green CS, Pouget A, Bavelier D (2010) Improved probabilistic inference as a general learning mechanism with action video games. Curr Biol 20:1573-1579. CrossRef Medline

Green DM (1960) Auditory detection of a noise signal. J Acoust Soc Am 32:121-131. CrossRef

Green DM, Luce RD (1973) Speed-accuracy trade off in auditory detection. In: Attention and performance IV (Kornblum S, ed), pp 547–570. New York: Academic.

Green DM, Birdsall TG, Tanner WP Jr (1957) Signal detection as a function of signal intensity and duration. J Acoust Soc Am 29:523-531. CrossRef

Greenwood DD (1961) Auditory masking and the critical band. J Acoust Soc Am 33:484-502. CrossRef

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5:887-892. CrossRef Medline

Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005) Neuromagnetic correlates of streaming in human auditory cortex. J Neurosci 25:5382-5388. CrossRef Medline

Gutschalk A, Micheyl C, Oxenham AJ (2008) Neural correlates of auditory perceptual awareness under informational masking. PLoS Biol 6:e138. CrossRef Medline

Hawkins GE, Forstmann BU, Wagenmakers EJ, Ratcliff R, Brown SD (2015) Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. J Neurosci 35: 2476–2484.

Hernández A, Nácher V, Luna R, Zainos A, Lemus L, Alvarez M, Vázquez Y, Camarillo L, Romo R (2010) Decoding a perceptual decision process across cortex. Neuron 29:300-314. CrossRef

Kaiser J, Lennert T, Lutzenberger W (2007) Dynamics of oscillatory activity during auditory decision making. Cereb Cortex 17:2258-2267. CrossRef Medline

Kaufman MT, Churchland AK (2013) Cognitive neuroscience: sensory noise drives bad decisions. Nature 496:172-173. CrossRef Medline

Kiani R, Churchland AK, Shadlen MN (2013) Integration of direction cues is invariant to the temporal gap between them. J Neurosci 33:16483-16489. CrossRef Medline

Klein SA (2001) Measuring, estimating, and understanding the psychometric function: a commentary. Percept Psychophys 63:1421-1455. Medline

Knoblauch K, Maloney LT (2008) Estimating classification images with generalized linear and additive models. J Vis 8(16):10.1-10.19. CrossRef

Lee JH, Russ BE, Orr LE, Cohen YE (2009) Prefrontal activity predicts monkeys' decisions during an auditory category task. Fron Integr Neurosci 3:16. CrossRef

Luce RD, Green DM (1972) A neural timing theory for response times and the psychophysics of intensity. Psych Rev 79:14-57. Medline

Micheyl C, Oxenham AJ (2010) Objective and subjective psychophysical measures of auditory stream integration and segregation. J Assoc Res Otolaryngol 11:709-724. CrossRef Medline

Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48:139-148. CrossRef Medline

Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Courtenay Wilson E (2007) The role of auditory cortex in the formation of auditory streams. Hear Res 229:116-131. CrossRef Medline

Mulder MJ, Keuken MC, van Maanen L, Boekel W, Forstmann BU, Wagenmakers EJ (2013) The speed and accuracy of perceptual decisions in a random-tone pitch task. Atten Percept Psychophys 75:1048-1058. CrossRef Medline

Murray RF (2012) Classification images and bubbles images in the generalized linear model. J Vis 12:2. CrossRef Medline

Niwa M, Johnson JS, O'Connor KN, Sutter ML (2012) Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci 32:3193-3210. CrossRef Medline

Noda T, Kanzaki R, Takahashi H (2013) Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. PLoS One 8:e83544. CrossRef Medline

Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. J Vis 5:376-404. CrossRef Medline

Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. Annu Rev Neurosci 21:227-277. CrossRef Medline

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Curr Biol 18:1124-1128. CrossRef Medline

Rahne T, Sussman E (2009) Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. Eur J Neurosci 29:205-211. CrossRef Medline

Raposo D, Sheppard JP, Schrater PR, Churchland AK (2012) Multisensory decision-making in rats and humans. J Neurosci 32:3726-3735. CrossRef Medline

Ratcliff R, McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput 20:873-922. CrossRef Medline

Ratcliff R, Rouder JN (1998) Modeling response times for two-choice decisions. Psychol Sci 9:347-356. CrossRef

Ratcliff R, Tuerlinckx F (2002) Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. Psychon Bull Rev 9:438-481. Medline

Ratcliff R, Van Zandt T, McKoon G (1999) Connectionist and diffusion models of reaction time. Psychol Rev 106:261-300. Medline

Ratcliff R, Gomez P, McKoon G (2004) A diffusion model account of the lexical decision task. Psychol Rev 111:159-182. CrossRef Medline

Rauschecker JP (2012) Ventral and dorsal streams in the evolution of speech and language. Front Evol Neurosci 4:7. CrossRef

Reddi BA, Asrress KN, Carpenter RH (2003) Accuracy, information, and response time in a saccadic decision task. J Neurophysiol 90:3538-3546. CrossRef Medline

Roberts B, Glasberg BR, Moore BC (2002) Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. J Acoust Soc Am 112:2074-2085. CrossRef

Romanski LM, Averbeck BB (2009) The primate cortical auditory system and neural representation of conspecific vocalizations. Annu Rev Neurosci 32:315-346. CrossRef Medline

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2:1131-1136. CrossRef Medline

Romo R, Salinas E (2001) Touch and go: decision-making mechanisms in somatosensation. Annu Rev Neurosci 24:107-137. CrossRef Medline

Romo R, Hernández A, Salinas E, Brody CD, Zainos A, Lemus L, de Lafuente V, Luna R (2002) From sensation to action. Behav Brain Res 135:105-118. Medline

Russ BE, Orr LE, Cohen YE (2008) Prefrontal neurons predict choices during an auditory same-different task. Curr Biol 18:1483-1488. CrossRef Medline

Schnupp J, Carr CE (2009) On hearing with more than one ear: lesions from evolution. Nat Neurosci 12:692-697. CrossRef Medline

Selezneva E, Scheich H, Brosch M (2006) Dual time scales for categorical decision making in auditory cortex. Curr Biol 16:2428-2433. CrossRef Medline

Shadlen MN, Newsome WT (1996) Motion perception: seeing and deciding. Proc Natl Acad Sci U S A 93:628-633. Medline

Shadlen MN, Hanks TD, Churchland A, Kiani R, Yang T (2006) The speed and accuracy of a simple perceptual decision: a mathematical primer. In: Bayesian brain (Rao R, Pouget A, eds). Cambridge, MA: MIT.

Shamma S, Micheyl C (2010) Behind the scenes of auditory perception. Curr Opin Neurobiol 20:361-366. CrossRef

Shamma S, Elhilali M, Micheyl C (2011) Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34:114-123. CrossRef Medline

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends Cogn Sci 12:182-186. CrossRef Medline

Smith PL, Ratcliff R (2004) Psychology and neurobiology of simple decisions. Trends Neurosci 27:161-168. CrossRef Medline

Sussman E, Steinschneider M (2009) Attention effects on auditory scene analysis in children. Neuropsychologia 47:771-785. CrossRef Medline

Sussman ES, Horváth J, Winkler I, Orr M (2007) The role of attention in the formation of auditory streams. Percept Psychophys 69:136-152. Medline

Thompson SK, Carlyon RP, Cusack R (2011) An objective measurement of the build-up of auditory streaming and of its modulation by attention. J Exp Psychol Hum Percept Perform 37:1253-1262. CrossRef Medline

Thura D, Beauregard-Racine J, Fradet CW, Cisek P (2012) Decision making by urgency gating: theory and experimental support. J Neurophysiol 108:2912-2930. CrossRef Medline

Treutwein B (1995) Adaptive psychophysical procedures. Vision Res 35:2503-2522. Medline

Tsetsos K, Chater N, Usher M (2012) Salience driven value integration explains decision biases and preference reversal. Proc Natl Acad Sci U S A 109:9659-9664. CrossRef Medline

Tsunada J, Lee JH, Cohen YE (2011) Representation of speech categories in the primate auditory cortex. J Neurophysiol 105:2634-2646. CrossRef Medline

Tsunada J, Lee JH, Cohen YE (2012) Differential representation of auditory categories between cell classes in primate auditory cortex. J Physiol 590:3129-3139. CrossRef Medline

Usher M, McClelland JL (2001) The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 108:550-592. Medline

Wichmann FA, Hill NJ (2001) The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys 63:1293-1313. Medline

Wickelgren WA (1977) Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica 41:67-85. CrossRef

Winkler I, Denham SL, Nelken I (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci 13:532-540. CrossRef Medline