

TEMPORAL PROCESSING IN AUDITORY PERCEPTUAL GROUPING AND DECISION-  
MAKING

Andrew S. K. Liu

A DISSERTATION

In

Bioengineering

Presented to the Faculties of the University of Pennsylvania  
in  
Partial Fulfillment of the Requirements for the  
Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

---

Yale Cohen  
Associate Professor of Otorhinolaryngology

Graduate Group Chairperson

---

Jason Burdick, Professor of Bioengineering

Dissertation Committee

Diego Contreras, Professor of Neuroscience (Chair)  
Josh I. Gold, Associate Professor of Neuroscience  
Alan A. Stocker, Assistant Professor of Psychology

TEMPORAL PROCESSING IN AUDITORY PERCEPTUAL GROUPING AND DECISION-  
MAKING

COPYRIGHT

2014

Andrew Shih Kang Liu

# Acknowledgements

I want to thank my mentors, Yale Cohen and Josh Gold, and the rest of my thesis committee, Diego Contreras and Alan Stocker for their guidance. I want to thank the past and present members of the Cohen Lab, especially Adam Gifford, Kate Christison-Lagay, Joji Tsunada, Sharath Bennur, and Harry Shirley for their advice and support.

I wish to express my heartfelt gratitude to my wife, Wei, who has been tremendously supportive and understanding during the writing process, to my parents, Jung-Dee and Jaung-Shen, and my sister, Betsy, for their guidance and support through all of the years.

## ABSTRACT

### TEMPORAL PROCESSING IN AUDITORY PERCEPTUAL GROUPING AND DECISION- MAKING

Andrew S. K. Liu

Yale E. Cohen

How do perceptual decision-making and auditory perceptual grouping interact on a perceptual, computational, and neural level? The work in this dissertation lays the groundwork to investigate the neural basis for auditory perceptual decisions by examining the perceptual and computational effects of the temporal structure of an auditory stimulus. I examined the role of auditory perceptual grouping on auditory perceptual judgments by asking whether the presentation rate of a stimulus sequence, which can affect the perceptual grouping of the stimulus, affects how sensory evidence converted into a decision. I devised a task that allows us to test, under different grouping conditions, whether the observed performance was consistent with changes in the representation of sensory evidence used to make the perceptual judgment or in the process by which the sensory evidence is converted into the decision. Subjects made a judgment on the frequency changes over time of a tone sequences while the interburst interval (IBI), or the time between tones of the stimulus, was varied across trials. I examined how subjects processed the sensory evidence to form their decisions as well as modeled the effect of IBI on their decision-making process. The results show that subjects accumulated sensory evidence over time to form their judgment and while IBI and perceived grouping did not affect the accumulation rate, subjects accumulated less

total sensory evidence for long IBIs consistent with a collapsing decision boundary. By understanding how the brain converts sensory stimuli into a perceptual decision with our task, we can better understand the computational principles and the neural implementation of how auditory percepts are formed.

# Table of Contents

	Acknowledgements.....	iii
	Abstract.....	iv
	List of Illustrations .....	vii
1	Introduction .....	1
2	Integration of Auditory Contour Information is Invariant to Time Gaps and Perceptual Grouping of Stimulus.....	14
3	Discussion .....	51
	Bibliography.....	60

# List of Illustrations

Figure 1 – Stimulus	42
Figure 2 – IBI vs. Perceptual Grouping	43
Figure 3 – Task	44
Figure 4 – Response Time Psychometric & Chronometric Data	45
Figure 5 – Response Time Model Fits	46
Figure 6 – Absolute Frequency Decision Model Comparison	47
Figure 7 – Psychophysical Kernel	48
Figure 8 – Interrogation Task Data & Fits	49
Figure 9 – Hybrid Task Data & Fits	50

# Chapter 1

## Introduction

We inform our decisions by inferring the state of the world, which is a process called perceptual decision-making. This process uses information from a stimulus to make a categorical judgment about the stimulus, which can then guide a behavior, such as a motor action. The stimulus is initially transformed by the peripheral sensory organs, such as the retina in the eye or the cochlea in the inner ear, into a neural representation of the physical variables of the stimulus. The cochlea, for example, transforms the acoustic stimulus, a time-varying pressure waveform, into a neural-code representing the physical attributes of the stimulus, such as the frequency content of the acoustic waveform. As sensory information is processed further downstream, through the midbrain and then to the cerebral cortex, the information represented by neurons is transformed from a representation primarily encoding the physical attributes of the stimulus to a representation that is more closely related to the perceptual report. This transformation is typically thought to have two parts: first, the stimulus features that correspond to the relevant perceptual dimensions for the judgment are extracted from early sensory representations by the downstream neurons; second,

this higher-level representation of the relevant sensory evidence for the decision is converted by neurons further downstream into a perceptual decision, which typically correlates well with the behavioral report (Shadlen & Newsome 2001). In this dissertation, I explore the behavioral and computational basis for the transformation of an auditory stimulus into a perceptual decision. By understanding how the brain represents sensory stimuli and how sensory information is transformed into a perceptual decision, we can gain insight into the neural mechanisms that allow us to adaptively interact with the environment.

Towards the goal of understanding the neural representation of auditory stimuli and the conversion from sensory stimuli into auditory percepts, in this dissertation I am investigating how certain perceptual decisions are affected by the processing of the temporal structure of an auditory stimulus. Auditory perception depends on the brain extracting and analyzing the temporal structure of a stimulus. For example, the temporal structure of a Bach concerto for two violins is one factor that allows a listener to separate out the different parts corresponding to each violin. To separate out the different parts of the concerto for two violins, the auditory system uses the temporal structure of the auditory stimulus to separate the incoming stimulus into two perceptual representations, corresponding to each of the two violin parts (Bregman 1994). This process is called auditory perceptual grouping.

Auditory perceptual grouping converts the sensory representation of an auditory stimulus into the perceptual representation of an auditory stimulus. This

process, which itself depends on the temporal structure of the stimulus, may affect the processing of the stimulus to extract perceptually relevant information from the stimulus as well as how this information is converted into a categorical perceptual decision. For many perceptual tasks, this transformation of the sensory information into a perceptual decision depends on combining multiple samples of sensory information over time in order to deal with unreliable sensory information (Mauzrek et al 2003, Gold & Shadlen 2007, Shadlen & Kiani 2013). While some recent studies have explored the effect of the stimulus temporal structure on the perceptual decision-making process (Kiani et al 2013, Brunton & Brody 2013), whether auditory perceptual grouping affects the process of combining multiple samples of sensory information over time to make a decision remains to be explored.

In this dissertation, I used human auditory psychophysics to examine the role of temporal processing in auditory perceptual decision-making. Specifically, my goal was to establish a novel auditory task in humans to test how the presentation rate of a tone sequence affected how subjects converted the stimulus into a perceptual judgment on how the frequencies of the tone sequence change over time. To understand how auditory grouping based on the temporal structure of a stimulus affects the transformation of the stimulus into a perceptual decision, subjects' accuracy as well as the time needed to make a perceptual judgment was measured while performing a perceptual task and across trials, the stimulus temporal structure was varied. I used the drift diffusion model, a computational model of

decision-making process, to relate the stimulus information on a given trial to both the accuracy and the time needed to make a judgment. This model simultaneously accounts for both measures of subject performance on the perceptual task. By examining how the model differed across the different stimulus temporal structure conditions, my aim was to test whether the temporal structure of the stimulus, and specifically auditory grouping, had an effect on the decision-making process. Here, I will first review the background for how the temporal structure of an auditory stimulus affects how sounds are perceived. Then, I will review the approach I used to study the effect of the stimulus temporal structure on auditory perceptual decisions. Finally, I will conclude with a summary of the aims of this dissertation.

### **Stimulus Temporal Structure and Auditory Perception**

To hear sounds and to use sound information to guide behavior, our auditory system transforms acoustic stimuli into a perceptual representation (Griffiths & Warren 2004; Bizley & Cohen 2013). These perceptual representations, also referred to as auditory objects, form the basis of our experience of the auditory environment. Auditory objects are the perceptual entity representing an experienced sound, corresponding to a particular sound source or sound event. Forming these perceptual representations depend on analyzing and extracting behaviorally relevant information from the acoustic stimulus that may evolve through time and may even be separated in time. For example, the brain is able to selectively process the pitches of musical notes separated in time to form a

perceptual representation of the part of a melody being played by an instrument as part of a trio. This process begins in the cochlea, where the time-varying pressure waveform of the acoustic stimulus is converted into neural activity. Downstream areas in the brain convert this sensory representation based on the physical attributes of the stimulus into a perceptual representation that we experience as sound.

Several experimental paradigms have been used to explore how perceptual representations are formed from the sensory information extracted from the stimulus (Van Noorden 1975, Kidd et al 1994, Kidd et al 2003, Bregman 1994, Bregman et al 2000, Nelken et al 1999), but one common process that has been implicated in the formation of the perceptual representations is auditory perceptual grouping (Bregman 1990, Bizley & Cohen 2013, Shamma et al 2011). Auditory perceptual grouping tries to explain how the low-level, physical attributes of a stimulus are converted into a perceptual representation. The auditory perceptual grouping process extracts perceptual representations for the different auditory objects present in the stimulus and is based on the principles of Gestalt psychology. An acoustic scene is a mixture of acoustic waves from different sound sources that can either be perceived as a whole, or as the elements that comprise the scene. For example, a jazz trio can be perceived as a unified whole, or as the individual components, as the pianist, bass, or vocalist parts of the trio. According to Gestalt principles, the perception of the unified scene, like the trio playing a jazz piece, is

not just the combination of the percepts of the elements of the scene, such as the different parts of the trio, but an entirely distinct perceptual entity.

While substantial progress has been made to understand the stimulus parameters that promote auditory perceptual grouping, how cortical auditory neurons implement the auditory perceptual grouping process to transform the sensory input into a high-level perceptual representation remains to be elucidated. While top-down factors such as attention certainly can play a role in the auditory perceptual grouping process (Lakatos et al 2013, Bey & McAdams 2002), the grouping process is typically driven by bottom-up factors derived from the physical attributes of a stimulus (Bregman 1994, Hartman & Johnson 1991, Beauvois 1998, Bregman et al 2000, Bizley & Cohen 2013). Bottom-up, or stimulus-driven, perceptual grouping relies on regularities of the spectral and temporal structure of a stimulus to allow different stimulus components to be grouped together. Acoustic features such as spectral and temporal proximity of the different stimulus components as well as similar spectral and amplitude modulation over time contribute to the bottom-up processing. For example, the frequency separation and the presentation rate of an 'ABA' tone triplet (where A and B represent tones with different frequencies), can determine whether subjects perceive a sequence of the tone triplets as a "gallop" or trill versus as two "streams" or two separate, but continuously playing tones.

In this dissertation, I am interested in how auditory perceptual grouping, which affects whether a stimulus is grouped into a unified perceptual

representation or separated into distinct perceptual representations and is based on the temporal structure of a stimulus, affects the process that converts the sensory information into a judgment. The perceptual representation of a stimulus may affect how sensory information is extracted and converted into a perceptual decision. The hypothesized effect is that when the stimulus is represented by a unified, grouped, perceptual representation, sensory information is more efficiently extracted from the stimulus than if the stimulus were not grouped and was separated into distinct perceptual representations. Thus, by having subjects make a judgment on how the frequency of a tone sequence changes over time while varying the presentation rate of the tone sequence across trials, the effects of auditory grouping on the overall perceptual decision-making process should affect subjects' accuracy as well as the amount of time needed to make their judgment. In the next section, I will explain the basis for the methods that quantify the effect of auditory grouping based on the temporal structure of the stimulus on how sensory information is processed and how this information is used to make a perceptual decision.

### **Modeling perceptual decisions to understand auditory perceptual processing**

In this dissertation, subjects discriminated whether the frequency of a tone sequence was increasing or decreasing over time while the presentation rate of the tone sequence varied across trials. I used computational modeling to test whether subjects' behavioral performance on an auditory task was affected by auditory

grouping. The approach I took infers the nature of the neural representation of sensory information and the decision-making process from subjects' performance on a perceptual task. By modeling subjects' performance for the different stimulus temporal structure conditions, I could test whether auditory grouping, based on the temporal structure of the stimulus, affected the extraction of sensory information from the stimulus or the conversion of the sensory information into a perceptual decision. This approach uses the drift diffusion model (DDM) of decision-making that relates the physical parameters of a stimulus to subjects' accuracy as well as the time they took to make a decision for a perceptual task (Gold & Shadlen 2007, Bogacz et al 2006). The DDM is based on signal detection theory (SDT), which was initially developed to account for the variability in perceptual detection performance due to the noisy representation of sensory information as well as to quantify the effect of varying stimulus parameters on perception (Green & Swets 1966). The DDM extends signal detection theory to relate discrimination performance, both accuracy and the time it takes to make a decision, to the perceptually-relevant stimulus parameters (Ratcliff & Tuerlinckx 2002, Ratcliff & Smith 2004, Ratcliff & McKoon 2008). By quantifying how subjects make their decisions under a variety of listening duration conditions, the DDM allows us to infer how the sensory information and the decision-making process varies under different stimulus timing and auditory perceptual grouping conditions.

While SDT allows us to distinguish stimulus or task-related factors that affect the sensory representation or the decision criterion, it does not account for how

long subjects take to make a decision. The general class of sequential sampling models extends the SDT to relate the stimulus parameters to both the accuracy as well as the response times of each decision (Ratcliff & Tuerlinckx 2002, Ratcliff & Smith 2004, Ratcliff & McKoon 2008). The sequential sampling models, which include the DDM, make the key assumption that decisions are based on the accumulation of multiple samples of sensory evidence. The accumulation of sensory evidence over time is a mechanism to handle the inherent variability in the sensory representation. This variability could be the result of a noisy, stochastic stimulus or from a noisy neural representation (Gold & Shadlen 2007, Shadlen & Kiani 2013, Roitman & Shadlen 2002, Mazurek et al 2003).

In the DDM, beginning at the start of the trial, momentary sensory evidence is accumulated through time to form a decision variable. Variants of the DDM make different assumptions of the nature of the sensory evidence. In the standard DDM, the sensory evidence accumulated through time is set at a fixed, constant mean level for the duration of the trial (Luce 1986, Smith 1995, Ratcliff & Tuerlinckx 2002, Ratcliff & Smith 2004). In other DDM variants, such as the leaky integration model, the accumulated sensory evidence is “leaked” away at a constant rate for the duration of the trial, allowing for recency effects, where sensory information that was processed closer in time to when the decision was made would have a greater contribution to the decision variable than sensory information that came much earlier in the trial (Usher & McClelland 2001). These models where the key parameters do not vary during the duration of a trial, or are “stationary”, have less

complex expressions that relate accuracy to the time it takes to make a decision, thus are easier to fit than more general DDM models (Ratcliff & Tuerlinckx 2002, Shadlen et al 2006). The more general DDM models relax the stationarity constraint and have time-varying parameters that can vary during a trial, which increase the complexity of the model and therefore make them more difficult to fit to the behavioral data. In these models, the sensory evidence can vary as a function of time according to a power-law (Eckhoff et al 2008) or has a mean level of sensory evidence that varies as the trial progresses, consistent with an urgency or confidence signal that pushes subjects to commit to a decision rather than perseverating on the process of making a decision (Cisek et al 2009, Churchland et al 2008, Ditterich 2006, Hanks et al 2011). The goal of using the different variants of the DDM is to better quantify how auditory grouping based on the temporal structure of the stimulus affects the nature of the sensory representation. By fitting the subjects' performance data collected under different stimulus temporal structure conditions to the different DDM variants, the model that consistently fits best would suggest that the neural representation might have features that are consistent with the assumptions of the best fitting model.

In the DDM, sensory evidence is accumulated as the decision variable through time. A decision is made when the accumulation process ends. The termination of the accumulation process can either be due to the experimental control of the stimulus duration, as in "fixed duration" or "variable duration" discrimination tasks, or be due to the subject deciding that he or she is ready to

commit to a decision even while the stimulus is still present. In the case of the experimentally controlled stimulus duration, at the end of the stimulus presentation, the level of the accumulated evidence is compared to a decision threshold to determine which one of two possible choices is made. In the subject-controlled viewing or listening duration case, the sensory evidence accumulation is terminated when the accumulated evidence reaches one of two decision boundaries, corresponding to each of the two choices in a two-choice discrimination task. These boundaries are the criteria by which the accumulated evidence is converted to a decision. Once the accumulated evidence reaches one of these boundaries, a decision is rendered according to the boundary that was reached and typically converted to a motor command, and eventually a behavioral report. The accumulated evidence boundaries reflect strategic considerations based on maximizing benefit and minimizing cost for the decision outcomes as well as other considerations such as the urgency of making a decision or confidence of the level of accumulated evidence (Ditterich 2006, Churchland et al 2008, Hanks et al 2011, Drugowitsch 2012). While the decision outcome reflects which decision boundary was reached, the time from the start of the accumulation process, typically the start of the trial, to when the decision variable reached a decision boundary, represents the decision time. The overall reaction time is the sum of the decision time and the non-decision time, which represents the time for the decision to be converted into a behavioral report.

Because the temporal structure of an auditory stimulus can affect how the auditory system perceptually represents the stimulus, the goal of this dissertation is to test whether the temporal structure affects the accuracy and the time it takes to make a perceptual decision as well as to explore how the temporal structure might affect decision performance. By using the DDM model, I can quantify the effect of auditory perceptual grouping as well as the processing of the stimulus temporal structure on both the process of extracting sensory information from the stimulus to guide a perceptual judgment and the decision-process that converts the extracted sensory information into the judgment. In this dissertation, I examine whether differences in presentation rate of the tone sequence stimulus can account for the effects on decision performance, consistent with the stimulus temporal structure affecting the process that extracts decision-related sensory information from the stimulus. I also test whether the temporal structure of the stimulus affects the process that converts the sensory information into a decision. Finally, I test whether auditory perceptual grouping specifically has an effect on the process of extracting sensory information from the stimulus that is eventually used to make the perceptual judgment.

### **Summary**

How do perceptual decision-making and auditory perceptual grouping interact on a perceptual, computational, and neural level? The work in this dissertation lays the groundwork to investigate the neural basis for auditory perceptual decisions by examining the perceptual and computational effects of the



# Chapter 2

## Integration of Auditory Contour Information is Invariant to Time Gaps and Perceptual Grouping of Stimulus

### Introduction

Time plays a key role in auditory perception. The timing of acoustic events in a stimulus, such as notes in a melody, can affect the perceptual organization of the stimulus (Moore & Gockel 2013, Shamma et al 2011). A rapid sequence of tones, for example, are more likely to be perceptually grouped into a single auditory “stream” than a slow sequence of tones (Van Noorden 1975, Bregman 1994). The auditory system transforms low-level sensory input into a high level perceptual representation by grouping acoustic events in a complex auditory scene into auditory objects (Griffiths & Warren 2004). The processing of low-level temporal cues to organize perceptually the incoming acoustic signal is fundamental (Bizley & Cohen 2013) to allow for us to perceive sounds in the auditory environment.

Many studies have characterized the factors that affect auditory perceptual organization (Moore & Gockel 2013) and grouping can affect perceptual judgment performance (Micheyl & Oxenham 2010, Thompson et al 2011, Borchert et al 2011,

Roberts et al 2002, Bey & McAdams 2002). However, because of the task- and stimulus-dependence of these grouping-related effects, the mechanisms that contribute to these effects are unclear. For example, perceptual judgments of timing differences are more accurate if the comparison is made within a stream than comparisons between streams (Roberts et al 2002). In contrast, the grouping of stimuli into different streams can facilitate the identification of a tone sequence if the sequence is contained within one stream, effectively, segregating the target from the distractor (Bey & McAdams 2002). These studies raise the question: by what mechanism does perceptual grouping affect the processing of sensory evidence for auditory perceptual judgments?

To address this, we examined the role of grouping on auditory perceptual judgments by asking whether time gaps in a stimulus sequence, which can affect the perceptual grouping of the stimulus, affect how sensory evidence converted into a decision. We devised a task that allows us to examine, under different grouping conditions, the representation of sensory evidence used to make the perceptual judgment and the process by which the sensory evidence is converted into the decision. Subjects made a frequency-change judgment based on tone sequences increasing or decreasing in frequency while we manipulated the interburst interval, or the time between tones of the stimulus. We examined how subjects processed the sensory evidence to form their decisions as well as modeled the effect of IBI on their decision-making process. We found that subjects accumulated sensory evidence over time to form their judgment and while IBI and perceived grouping did not

affect the accumulation rate, subjects accumulated less total sensory evidence for long IBIs consistent with a collapsing decision boundary.

## **Methods**

### **Experimental Setup**

Subjects rested their chin on a chin rest approximately 2' from the speaker while seated in a soundproof booth (IAC) during the experiment. The task was developed using the MATLAB-based Snow-Dots (<http://code.google.com/p/snow-dots>) psychophysics software. On each trial, task parameters from the Snow-Dots environment were used to generate the stimulus on a digital signal processor (Tucker Davis Technologies RX6). The stimulus was presented free-field using a calibrated speaker (Yamaha MSP-7). Subjects responded by a button press on a gamepad (Microsoft Sidewinder) and task instructions and feedback was displayed on a LCD flat panel monitor (DELL E171FP).

### **Stimulus**

The stimulus was a tone-burst sequence (Figure 1A-D). Each tone burst was presented at 65 dB SPL and had a duration of 30 ms (onset/offsets were smoothed with a 5-ms  $\cos^2$  gate). The time between the offset of one tone to the onset of the next tone (i.e., the “inter-burst interval”; IBI) ranged between 10 and 150 ms (Figure 1D). The range of IBIs chosen were based on pilot studies in which subjects tended to report that the stimulus sounded continuous when the IBI was relatively small (<50 ms) but like discrete tones for larger IBIs (Figure 2). The properties of a tone-

burst sequence was a function of (1) sequence direction (increasing or decreasing frequency), (2) sequence duration, (3) IBI, and (4) coherence (Figure 1A-C); the stimulus direction, duration, IBI, and coherence varied on a trial-by-trial basis depending on the specific task constraints (see task descriptions below).

At the beginning of each trial, the frequency for the first tone burst in the sequence was randomly sampled from a uniform distribution corresponding to the direction of the stimulus. A monotonically increasing or decreasing frequency sequence was first generated by adding or subtracting a fixed frequency increment ( $\Delta f = 7.5$  hz) to the previous tone-burst frequency.

Stimulus coherence was the proportion of tones in a tone-burst sequence whose frequencies were a fixed frequency increment from the previous tone. A tone-burst sequence in which all of the tone bursts were monotonically increasing or decreasing had a coherence of 100%. For sequences with coherence <100%, we first generated the 100% coherent stimulus and then randomly shuffled the temporal order of a subset of tone bursts. A 50% coherent stimulus, for instance, was generated by randomly shuffling 50% of the tone bursts. A 0% coherent stimulus had 100% of the tone bursts shuffled. By generating the stimulus in this manner, we could ensure that each sequence traversed a fixed frequency range for both increasing and decreasing stimuli. This is to reduce the likelihood of subjects basing their decisions on the frequency content of the stimulus instead of the pattern of frequency changes.

## **Tone-Sequence Frequency-Direction Discrimination Task**

In the tone-sequence frequency-direction discrimination task (Figure 3), subjects reported whether the frequency of the tone-burst sequence was increasing or decreasing. Subjects were tested in three variants of the task: “response-time”, “interrogation”, and “hybrid”. Prior to participating in the experiments, subjects provided informed consent. The University of Pennsylvania Institutional Review Board approved the experimental protocol and informed-consent procedures.

### *Response-time task*

We tested 6 subjects (5 male, 1 female) in 4 weekly 1.5 hour sessions. Subjects were told to respond as quickly as possible, without sacrificing accuracy, after they reached their decision. Each session contained 4 blocks of trials. Subjects took a short break between each block. During each block, we varied, on a trial-by-trial basis, IBI (10, 60, or 150 ms), coherence (0, 10, 25, 50, or 100%), and stimulus direction (increasing or decreasing). Each combination of these sequence properties was presented 5 times within a block, for a total of 150 trials per block. Subjects were given feedback at the end of each trial.

### *Interrogation Task*

We tested 5 subjects (4 male, 1 female) in 6 weekly 1.5 hour sessions. On each trial, the stimulus duration was sampled from a truncated exponential distribution ( $\lambda = 2000$  ms for all IBIs; upper and lower stimulus duration limits by

IBI: 10 ms, [160, 1400] ms; 60 ms, [360, 3150] ms; 150 ms, [720, 6300] ms) so that subjects could not anticipate the end of the stimulus. The truncation limits are standardized across IBI by the minimum (4) and maximum (35) number of tones presented. After the end of the stimulus, a response cue flashed on the screen, and subjects had 800 ms to respond. Each session contained four blocks of trials. Subjects took a short break between each block. During each block, we varied, on a trial-by-trial basis, IBI (10, 60, or 150 ms), coherence (0, 10, 25, 50, or 100%), and stimulus direction (increasing or decreasing). Each combination of these sequence properties was presented 5 times within a block, for a total of 150 trials per block. Subjects were given feedback at the end of each trial.

### *Hybrid Task*

We also tested 5 subjects (3 male, 2 female) in 4 weekly 1.25 hour sessions. Like on the interrogation task, on each trial, stimulus duration was sampled from a truncated exponential distribution ( $\lambda = 2000$  ms) so that subjects could not anticipate the end of the stimulus. The truncation limits are standardized across IBIs by the minimum (4) and maximum (35) number of tones presented. In addition to making a judgment regarding whether stimulus direction, subjects also reported whether they perceived the stimulus as “one, continuous sound” or “as a series of discrete sounds”. Subjects reported their responses during two 800-ms response periods, and a response cue indicated the response order. The order of the stimulus-

direction judgment and stimulus-continuity judgments was alternated on a block-by-block basis.

Prior to participating in the hybrid task, each subject's "50%-IBI threshold" and "coherence threshold" was measured and incorporated into the parameters used to generate the tone-burst sequence. Each subject's IBI threshold, defined to be the IBI value for which 50% of the responses rated the stimulus as continuous, was measured using a 1-up/1-down adaptive procedure. The IBI threshold was measured daily, before the start of each session. We found, in early pilot experiments, that each subject's IBI threshold varied across days. We measured each subject's coherence threshold using a 2-up/1-down adaptive procedure. Threshold was defined to be 70.7% correct performance, which corresponds to a  $d'$  of 0.77. Each subject's coherence threshold was calculated for tone-burst sequences using the initial session's IBI threshold. Coherence threshold was calculated prior to a subject's participation in the hybrid task and kept constant for all subsequent hybrid-task sessions.

Each session contained four blocks of trials. Subjects took a short break between each block. During each block, we varied, on a trial-by-trial basis, IBI (50%-threshold - 15 ms, 50%-IBI threshold, and 50%-IBI threshold + 15 ms) and stimulus direction (increasing or decreasing). The sequence coherence was set to each subject's coherence threshold. For each stimulus direction condition, the threshold IBI condition was repeated 40 times per block, and the two off-threshold IBI conditions were repeated 10 times per block. The trials were randomly presented

within each block. Subjects were given feedback regarding whether they made the correct stimulus direction judgment at the end of each trial.

## **Data Analysis**

Each subject's behavioral data was fit to a drift diffusion model (DDM) to quantify the effects that sequence coherence and IBI had on the decision-making process. We computed psychophysical kernels for each IBI using data pooled across all of the subjects to test whether subjects were accumulating sensory evidence across time and whether the time course of integration was affected by IBI. We also tested whether subjects could use the frequency of the stimulus as an alternative strategy to decide whether the stimulus was increasing or decreasing in frequency.

### *Drift Diffusion Model*

The DDM relates a subject's accuracy on a two-alternative forced-choice (2AFC) task to the amount of time that the subject used to reach their decision. Central to the drift diffusion model is that sensory evidence is accumulated through time as the decision variable (Bogacaz 2006, Gold & Shadlen 2007). For the response-time task, a categorical decision is generated, according to the DDM, when the accumulated sensory evidence reaches a response threshold, or the decision boundary. Similarly, according to the DDM, for the interrogation task, the value of the accumulated sensory evidence is compared to its value at the start of the trial to determine the choice that is reported. Because of the different assumptions of how

the decision is generated based on the accumulated sensory evidence for the two types of tasks, we used one mathematical formulations of the DDM to model the reaction-time and a different formulation of the DDM for interrogation and hybrid data.

For the response-time data, the DDM (Palmer et al 2005) is parameterized by the drift rate coefficient, two decision boundaries, and a non-decision time. The drift rate coefficient scales the coherence of the stimulus proportionately to obtain the drift rate, which represents the amount of momentary evidence available at a given moment. The drift rate is normally distributed with the mean given by  $\mu = kC$ , where  $k$  is the fitted drift rate coefficient and  $C$  coherence, and unit variance. As the accumulated sensory evidence reaches one of the two decision boundaries, the corresponding choice is made. The non-decision time is the component of response time corresponding to motor preparation time and low-level sensory processing latencies. The probability of the accumulated evidence reaching boundary A is  $P_a = \frac{e^{2\mu B} - 1}{e^{2\mu B} - e^{2\mu A}}$ . The mean decision time for choice A, or the time from stimulus onset to the accumulated evidence reaching boundary A, is  $T_A = \frac{A+B}{\mu} \coth[(A+B)\mu] - \frac{B}{\mu} \coth[B\mu]$ . Similarly, for choice B, the mean decision time is  $T_B = \frac{A+B}{\mu} \coth[(A+B)\mu] - \frac{A}{\mu} \coth[A\mu]$ . The response time for a given choice is the sum of the decision time and the non-decision time.

For the interrogation and hybrid tasks, the DDM is parameterized by the drift rate and lapse rate (Eckhoff et al 2005). Like the reaction time model, the drift rate

is given by  $\mu = kC$ , where  $k$  is the fitted drift rate coefficient and  $C$  coherence. For a given listening duration, the probability of choosing choice A is given by  $P_A(t) = \frac{1}{2} + \left(\frac{1}{2} - l\right) \text{Erf}\left(\frac{\mu}{\sqrt{2V}}\right)$

All model fits were based on maximizing the likelihood of the parameters given the data. Subjects' choices were modeled as binomial errors and mean response times were modeled as Gaussian errors. The models were implemented and fit using Matlab. To avoid local maxima, the fitting procedure was initialized with multiple random starting points in the parameter space. Parameter confidence intervals were estimated using a bootstrap method.

### *Simulated DDM*

We examined how decision boundary dynamics of the DDM affects choice, response-times, and the psychophysical kernels by simulating the DDM with different types of decision boundary dynamics. We simulated the temporal evolution of the accumulated sensory evidence on each trial by numerically integrating the stochastic differential equation:  $dX = kC*dt + dW$ . Here,  $k$  corresponds to the drift rate coefficient,  $C$  is the coherence of the trial being simulated,  $dW$  is a standard Gaussian process representing the noise in the sensory evidence, when combined, represents  $dX$ , the momentary sensory evidence. For each coherence and IBI condition, we simulated 5000 trials for 5000 ms using a variable simulation time step that varied according to the IBI in order to run the simulations in signal time.

We computed the simulated choices and response times by comparing the accumulated sensory evidence,  $X$ , on each trial to a decision boundary. We tested two types of boundaries, one that did not vary in time (“fixed bound”) and one that decreased linearly to zero as the trial progressed (“collapsing bound”). The fixed boundary had one parameter, the bound height. The collapsing bound had two parameters, the bound height at the start of the trial and the time (in elapsed time, not signal time) when the bound reached zero. In addition to the type of bound, we also tested two different relationships between the boundary and IBI. One way only had one set of boundary parameters *for all* IBI conditions, whereas the other way had one set of boundary parameters *for each* IBI condition. In summary we tested whether a fixed bound or a collapsing bound with one set of boundary parameters for all IBIs or with a set of boundary parameters for each IBI.

To find the best fitting parameters, we maximized the likelihood of the model given the parameters using, as before, a binomial likelihood for choices and a Gaussian likelihood for the mean response times. We used a derivative-free optimization method (`fminsearch` in Matlab) to compute the maximum likelihood fits. To compare the different models we computed the Bayesian information criterion ( $BIC = -2 \cdot \log(F(x|\hat{\theta})) - p \cdot \log(n)$ ;  $n$ : number of observations;  $p$ : number of parameters;  $\hat{\theta}$ : model parameters) for each of the tested models. to account for the differences in the number of parameters for the models we tested.

To generate the kernel predictions for the best fitting model, the simulation paths for the 0% coherence condition were averaged across trials separated by choice and then averaged across time.

### *Psychophysical kernel*

The psychophysical kernels were estimated using a logistic regression that related the stimulus to the subject's choice on a particular trial: (Equation:  $\text{logit}(y) \sim b_0 + b_1*x_1 + \dots + b_n*x_n + \text{epsilon}$ ,  $\text{epsilon} \sim N(0,1)$ ). The covariates  $(x_1, \dots, x_n)$  represented the difference in frequency between consecutive tones for a tone-burst sequence presented on each trial. The estimated regression coefficients  $(b_1, \dots, b_n)$  represented the degree to which a subject weighed, at a given moment in time, the difference in frequency of a tone burst pair to form a judgment. We pooled choice data and stimuli across all 6 subjects in the response-time task to compute the psychophysical kernels for each of the IBI conditions. The kernels were computed using only 0% coherence trials and not on any coherent patterns of frequency changes in the stimulus.

A permutation test was used to test whether the observed kernels were significantly different from a null kernel, assuming that there was no systematic relationship between the stimulus and subject's response. We generated  $N$  synthetic, test datasets by randomly permuting the reported choice on each trial and the associated stimuli. Each synthetic dataset was used to calculate a kernel. The collection  $N$  kernels for each IBI approximated the distribution of kernel weights

assuming that the null hypothesis was true. The kernel weights calculated based on the actual, observed data was compared to the null distribution of kernel weights to compute a p-value.

The kernel widths were calculated for each kernel at its half-maximum value. Bootstrap confidence intervals were calculated by resampling trials within each subject's dataset with replacement and then pooling across subjects to generate a pooled bootstrap dataset. Kernels were estimated using these pooled bootstrap dataset and this procedure was repeated 2000 times. Kernel widths were estimated for each of the bootstrap kernels and confidence intervals were calculated based on the 2.5% and 97.5% percentiles of the bootstrap kernel width distribution.

#### *Absolute Frequency Decision Model*

We tested whether subjects used an alternative strategy to solve the task by comparing the absolute frequency of the stimulus to an optimal criterion based on an ideal observer. The tone frequencies immediately preceding when the decision was reported was sorted by the reported decision outcome. The frequency distribution corresponding to "increasing frequency" choices was compared to the frequency distribution corresponding to "decreasing frequency" choices by an ROC analysis for all trials of a given IBI. The frequency threshold corresponding to the point on the operating curve that minimizes false positives and maximizes true positives is used to calculate is used as the decision rule to make a judgment whether the stimulus presented on a given trial was increasing or decreasing in

frequency. In addition to the tone immediately preceding the moment that the decision was reported, we tested up to 9-tones back from when the decision was made. We also tested this strategy using the average frequency of up to 9-tones back from when the decision was reported.

## **Results**

We tested how the time-course of sensory evidence integration in an auditory task was affected by a temporal manipulation that affected the perceived grouping of the stimulus. We used a two-alternative forced-choice task that required subjects to report whether frequency direction of a tone burst sequence was increasing or decreasing (Figure 1A). To control task difficulty, we manipulated stimulus “coherence,” corresponding to the fraction of tones that increased or decreased systematically in frequency (Figure 1A-C). To manipulate the sequential-grouping cues of the stimulus, we altered the stimulus presentation rate (i.e., the inter-tone-burst-interval [IBI], Figure 1D). As detailed below, we used three versions (Figure 3) of the decision task to test how perceptual grouping affects the time-course of the decision: 1) a “response-time” task where subjects could respond at any time after the beginning of the stimulus presentation, 2) an “interrogation” task where the stimulus duration was experimentally varied on each trial, and 3) a “hybrid” task where subjects reported both the continuity of the stimulus and the stimulus direction.

## Response-time Task

The response-time task tested whether subjects' accuracy and reaction time varied with task difficulty (coherence) and stimulus presentation rate (IBI). Figure 4A-B shows the behavioral data collected from the response-time task for one subject. As can be seen, this subject was more accurate and responded faster when the decisions were easy (i.e.,  $\pm 100\%$  coherence trials) than for difficult trials (i.e., 0% coherence trials).

We quantified the effect of coherence and IBI on the decision process by fitting the choice and response-time data for each subject to a drift diffusion model (DDM). (The solid curves in Figure 4A, B represent best-fitting DDM psychometric and chronometric functions.) The best-fitting DDM parameters for the 6 subjects are shown in Figure 5A-C. IBI significantly modulated the fit boundary heights (or response criteria) ( $F(2,10) = 53.73, p < 0.05$ ) and the drift rates (rate of sensory evidence accumulation) ( $F(2,10) = 239.6, p < 0.05$ ). The nondecision times were also significantly affected by IBI ( $F(2,10) = 6.336, p < 0.05$ ).

The differences in the fitted drift rates and boundary heights across IBIs could be due to IBI affecting the accumulation rate, the response criteria for making a response, or both. The stimulus presentation rate has a clear effect on the accumulation rate of sensory evidence. The faster the stimulus is played (equivalently the shorter the IBI), then the faster the information will accumulate. Similarly, if subjects made their decisions by integrating sensory evidence over a fixed time window, the faster the presentation rate, the more information will

accumulate over this period and thus have higher corresponding bound heights. In order to disambiguate the effects of IBI on the decision-making computation that cannot be explained by the differences in the presentation rate across IBIs, we fitted the DDM to transformed chronometric data to control for the presentation rate differences. We transformed RTs for each trial to “signal response time,” (signal RT) by number of elapsed tones from the start of the stimulus presentation to the subject’s response multiplied by the tone duration (30 ms).

If there was no difference in signal RTs across IBI conditions, this would suggest that subjects based their decisions solely on accumulating a fixed number of tones independent of IBI and therefore the accumulation rate would vary with the stimulus presentation rate and the bound height would decrease at long IBIs, corresponding to a fixed signal-time integration window. However, we found that signal RTs (Figure 4C) for short IBI trials in a representative subject was increased across all coherences compared to long IBI trials. This was observed for all 6 subjects. This suggests that subjects are not merely integrating each tone to a common bound so to quantify this, we fitted the signal RT data to the DDM.

Indeed, when the DDM was fit to the signal RT data (See Figure 4C for DDM fit to example subject, Figure 5D-F for fitted parameters by subject), we found that the accumulation rate was not modulated ( $F(2,10) = 3.122, p > 0.05$ ) by IBI, whereas response thresholds (decision criteria for the accumulated information) decreased reliably ( $F(2,10) = 306.2, p < 0.05$ ) with increasing IBI. The nondesision times increased as IBI increased ( $F(2,10) = 21.15, p < 0.05$ ). This suggests that after

controlling for the presentation rate of the stimulus, the quality or the signal-to-noise of the sensory evidence did not change as the IBI varied and that subjects were actually varying the amount of information they accumulated before responding.

We wanted to then examine how subjects actually converted the information in the stimulus into the decision and whether we can validate our previous finding in a model-independent manner. First we tested whether subjects could base their judgments on the absolute frequency of the stimulus. We found that this absolute frequency-based strategy is incompatible with the observed performance (Figure 6).

To test the effect of IBI on the time course of the sensory evidence contributing to the decision, we computed the reverse correlation-based psychophysical kernels based on data pooled across the 6 subjects who participated in the response-time task. The psychophysical kernel is computed based on the principles of signal-detection theory (Murray 2011; Knoblauch & Maloney 2008) and, here, it relates the temporal dynamics of the tone frequencies presented on a given trial to the subjects' choice. This analysis is model free in that it does not make any particular assumptions on the strategy that subjects use to convert the stimulus to the decision. These kernels show the relative weighting of each sample of sensory evidence over time that contribute to the decision. In other words, regions of the kernel closest in value to 0 represent the times when the sensory evidence contributed very little towards making the decision, and, similarly, the regions of the

kernel with relatively large values represent the times when the sensory evidence had a large contribution towards making the decision.

Consistent with our alternative absolute frequency strategy analysis, we found that kernels calculated using the tone frequencies alone were not significantly different from zero (not shown). By calculating the kernels using the difference in frequency between adjacent tones, we found that subjects were using frequency differences as sensory evidence in this task. We found that the kernels calculated for each IBI using the frequency differences as the sensory evidence showed significant non-zero weights across multiple, consecutive samples of sensory evidence, consistent with the accumulation of sensory evidence. We calculated the kernels both by aligning the data to the start of each trial (Figure 7A,C), as well as by aligning the data to when the judgments were reported, as seen in the choice-aligned kernels (Figure 7B,D). By calculating the kernels in this way, we can examine the kernel dynamics at the start of each trial, however, as the trial progresses, the kernels become less informative because of fewer number of trials with long response times. We use the choice-aligned kernels to examine the kernel dynamics towards the end of the trial, when presumably subjects are actively deliberating and processing the sensory evidence to make their decision. Overall, our results suggest that the non-zero weights seen in both the start-aligned and choice-aligned kernels show that subjects accumulate sensory evidence contained in the stimulus sequentially, through time, to inform their decisions.

While the kernels plotted in terms of elapsed time and the kernels plotted in terms of signal-time represent the same information, the different time scales allow us to compare the kernel dynamics in terms of the number of samples of sensory evidence with the dynamics in terms of elapsed time. By comparing the dynamics of the kernel across the different time units, we can examine how subjects are deliberating over the incoming sensory evidence across IBI conditions. For example, the time to reach the peak of the kernel, when expressed as elapsed time seem to be similar overall (Figure 7C-D), whereas when expressed as the number of elapsed samples of sensory evidence, there is a dramatic reduction in the time-to-peak for long IBI compared to short IBI (Figure 7A-B).

We quantified whether IBI had an effect on the shape of the kernels by calculating the width of the choice-aligned kernels at half of its peak value for each IBI. We found that the kernel widths estimated in elapsed time units were 840 ms for the 10 ms IBI (Bootstrap 95% CI: [700 ms, 880 ms]), 1080 ms for the 60 ms IBI ([900 ms, 1350 ms]), and 1440 ms for the 150 ms IBI ([1260 ms, 1620 ms]). The kernel widths across IBI suggest that subjects integrate the incoming sensory evidence over a temporal integration window that varies as a function of stimulus presentation rate/IBI.

### **Interrogation Task**

Since the data collected for the response time task depended on each subject's particular speed-accuracy tradeoff, which we can not directly measure,

there is still the possibility that the output of the DDM fitting procedure could result in the trade-off of drift rates with the bound heights across IBI (Ratcliff 2002). This could give us results that do not accurately reflect the differences in the drift rate due to legitimate IBI-related effects. To circumvent this possibility, we tested subjects using the interrogation task, which allowed us to test directly the relationship between IBI and drift rate by experimentally controlling the listening duration of the stimulus and monitoring subjects' performance. In the interrogation task, subjects based their decisions on the sensory evidence that was provided. We used a two-alternative forced choice "interrogation" task (Figure 3B) where we experimentally manipulated the listening duration across trials. The stimulus durations in this task were set randomly at the start of each trial and subjects could respond only after the end of the stimulus presentation.

We found that, for the interrogation task, subjects improved their performance as listening durations increased (Figure 8A-C). We fit the DDM to the time-dependent psychometric data, converted into signal time units, and found (Figure 8D) that subjects integrated the available sensory evidence to form their judgments and that IBI had no effect on the rate at which sensory evidence is integrated. The 95% confidence intervals for the drift rates were overlapping for the IBIs. These results provide further evidence that IBI had no effect on the accumulation rate of sensory evidence, despite that it can have a crucial effect on whether subjects perceive the stimulus as a continuous sound or as a discrete series of tone bursts.

## **Hybrid Task**

In a third experiment (the “hybrid” task, Figure 3C), we directly addressed whether perceptual grouping had an effect on the listener’s decisions. At the start of each session of the hybrid-task we determined the IBI for each subject that elicited equal numbers of grouped and discrete judgments. We adapted the interrogation task by adding a second response interval for subjects to report whether they heard the tone-burst sequence as a single, grouped sound or as a series of discrete sounds. Figure 9A shows the choice data pooled across subjects separated by the reported percept (“grouped” versus “discrete”). We fit the DDM to the data and we found that the 95% confidence intervals of the drift rates for continuous versus the discrete trials were overlapping (Figure 9B). This suggests that the same physical stimulus can elicit ambiguous (on a trial-by-trial) basis perceptual report, and that the rate of accumulation is indistinguishable, regardless of whether the stimulus is perceived as a continuous sound or as a series of discrete sounds.

## **Discussion**

We examined the role of grouping on auditory perceptual judgments by asking whether time gaps (i.e., the IBI) in a stimulus sequence, which can affect the perceptual grouping of the stimulus, can also affect how sensory evidence is converted into a perceptual decision. We used a novel auditory-discrimination task that required human subjects to report whether a stochastic auditory stimulus was increasing or decreasing in frequency. We found that their performance was

consistent with an accumulate-to-bound model, in which incoming auditory evidence was accumulated over time until reaching a fixed criterion that determined choice and RT. We manipulated the IBI of the stimulus and found that as IBI increased, the decision criterion decreased, thereby affecting the speed-accuracy trade-off. However, the accumulation rate of sensory evidence, or the drift rate, was independent of both IBI and how subjects perceived the grouping of the stimulus.

Perceptual grouping can affect perceptual judgments under certain conditions (Micheyl & Oxenham 2010, Thompson et al 2011, Borchert et al 2011, Roberts et al 2002, Bey & McAdams 2002). For example, judgments about timing differences are more accurate if the comparison is made within a stream than comparisons between streams (Roberts et al 2002). In contrast, in other situations, the grouping of stimuli into different streams can facilitate the identification of a tone sequence (Bey & McAdams 2002). However, because of the task- and stimulus dependence of these effects, the mechanisms that may have contributed to these effects are not clear. Our goal was to devise a stimulus and a task that allowed us to vary the temporal structure of the stimulus to affect how it was perceptually grouped by the listener as well as requiring the temporal accumulation of sensory evidence to solve the task. As sequential grouping is affected by the timing of the tones of the stimulus, we expected that decreasing the IBI, which would correspondingly increase the presentation rate of the tones of our stimulus, would also affect increase the rate of sensory evidence accumulation. We also expected that perceptual grouping effects resulting from the IBI manipulation could further

facilitate the processing of the sensory evidence and therefore increase the rate of sensory evidence accumulation. Whereas we found that varying the IBI of the stimulus changed how the stimulus was perceived (see Figure 2), the mechanisms that account for the differences in performance on the frequency-change discrimination task was independent of both the perceptual organization of the stimulus and the timing of the stimulus. In the following sections we discuss first, the interpretation of our data; second, how our results suggest that subjects vary their speed-accuracy-tradeoff depending on the IBI; and third, why was the accumulation of sensory evidence invariant to IBI or the perceptual grouping of the stimulus.

### **Discussion of IBI and Grouping Effects on the Accumulation of Sensory Evidence**

Our finding that both IBI and grouping do not affect the accumulation of sensory evidence is based on several lines of evidence. First, our analysis of the response time data addresses the facile explanation that by decreasing the IBI, the sensory evidence accumulation rate would naturally increase and could explain the decrease in RT for shorter IBI. We controlled for any presentation-rate related effects by using signal RT or signal time, which measures the listening duration in terms of the total duration of the tones presented but does not include the the silent time between the tone bursts. By controlling for the presentation rate, we found that there was no significant difference in the drift rates across IBI for the response-time task. Second, subjects effectively controlled the listening duration in the response-

time task by arbitrarily adjusting their speed-accuracy trade-off strategy, which cannot be directly measured. Because of the lack of explicit control over each subject's speed-accuracy tradeoffs, the output of the DDM fitting procedure could result in the trade-off of drift rates with the bound heights across IBI (Ratcliff 2002). This trade-off between model drift rates and bound heights could give us results that do not accurately reflect the differences in the drift rate due to legitimate IBI-related effects. To circumvent this possibility, we tested subjects using the interrogation task, which allowed us to test directly the relationship between IBI and drift rate by experimentally controlling the listening duration of the stimulus and monitoring subjects' performance. In the interrogation task, subjects based their decisions on the sensory evidence that was provided, and we found that across IBI, there was also no difference in the drift rate, after controlling for the presentation rate, consistent with our initial findings using the RT task. Third, we wanted to test whether grouping *per se* had an effect on the perceptual judgment that could be explained by differences in IBI. This was tested directly with the hybrid task. In this task, we compared the drift rate for decisions that, for the same IBI stimuli, differed by how subjects' perceived the stimulus. At the start of each session of the hybrid-task, to aid our grouping-based comparison, we determined the IBI for each subject that elicited equal numbers of grouped and discrete judgments. Again, subjects did not differ in the rate of sensory evidence accumulation, regardless of how they perceived the stimulus.

We note that there are a couple of limitations to our overall approach. First, the measure we used to quantify the relationship between IBI and the perceptual organization of the stimulus has been criticized for its dependency on the subjects' subjective report (Thompson et al 2010). By asking whether subjects perceived a grouped stimulus or not lacks an objective standard by which stimuli can be compared. An alternative approach is to simultaneously measure the mismatch negativity (MMN) of event related potentials in humans, whose amplitude has been used as an objective index of the perceptual organization of the stimulus (Sussman 2004, Steinschneider & Sussman 2006, Fujioka et al 2004). Second, another possible limitation is that we pooled data across subjects for many of our analyses, including the psychophysical kernel analyses as well as the fixed duration and hybrid task DDM fits. While it would have been ideal to collect a large number of trials for each condition for each subject, it would not have been practical to do so. This limits our findings because there may have been subtle per-subject effects that could have been the result of IBI or grouping-related effects on the sensory evidence accumulation process but could not be detected due to the lack of statistical power for per-subject inference and the wash-out effects of pooling across multiple subjects.

### **Accumulation of Sensory Evidence & DDM findings**

The accumulation of sensory evidence is a well-established computational model that describes how multiple samples of sensory evidence presented over time are converted into a decision (Gold & Shadlen 2007, Kiani & Shadlen 2013) and is consistent with our findings. We found that subjects accumulated sensory evidence

to different decision criteria, which varied by IBI. One possibility for subjects to base their decisions on varying amounts of accumulated sensory evidence for different IBIs, could be that subjects implicitly vary their speed-accuracy trade-off criteria based on the elapsed time. That is, after a fixed period of time, the costs of accumulating more sensory evidence outweighs any benefits associated with improved accuracy. A number of studies have found that human and non-human-primate subjects make perceptual judgments based on an accumulation-to-bound process with a time-varying bound (Ditterich 2006, Churchland et al 2008, Cisek et al 2009, Hanks et al 2011, Drugowisch et al 2012). These models incorporate the idea that subjects weigh the costs and benefits of accumulating additional sensory evidence as a trial progresses and adjusts the decision boundary accordingly to determine when to stop accumulating evidence and commit to a decision. We found that a model in which sensory evidence accumulation is subject to a time-varying, collapsing bound (Churchland et al 2008, Hanks et al 2011, Drugowisch et al 2012) can better explain our response-time data and generate kernel predictions that are more consistent with our observed kernels than a fixed bound model. This suggests that, for our task, subjects' performance agrees with a model in which they implicitly vary their speed-accuracy tradeoff by comparing the costs and benefits of the accumulating additional sensory evidence on the basis of elapsed time and not the number of samples of sensory evidence.

### **Invariance of Drift Rate to IBI & Grouping**

Here, we discuss two possible but non-exclusive explanations for the accumulation of sensory evidence to IBI and to the perceptual grouping of the stimulus. The first possibility is that the neural mechanisms that accumulate sensory evidence are invariant to IBI. Recent work suggests that the accumulation of sensory evidence indeed is insensitive to IBI, or, more generally, periods without stimulus present (Kiani et al 2013, Brunton & Brody 2013). Trial-by-trial variability in the choice and response-times of an auditory spatial discrimination task was the result of noise in the representation of sensory evidence and not noise in the sensory evidence accumulation process (Brunton & Brody 2013). If the sensory-evidence accumulation process for our task were noisy, then the variability in the accumulated evidence would scale with the listening duration of each trial. As more tones are presented for a given period of time for short IBI than for long IBI, the overall signal-to-noise ratio of the sensory evidence, and therefore the rate of sensory evidence accumulation, would decrease for long IBI. We did not observe an IBI dependent decrease in the drift rate across the two tasks (Fig 5E, for response-time task and Fig 7D for interrogation task) suggesting that the accumulation process for our task may be an instance of a noiseless integrator.

Such a noiseless integration process may be implemented by a multi-stable attractor network (Kiani et al 2013, Pouget & Latham 2002, Koulakov et al 2002) in which the level of the accumulated sensory evidence does not decay or become corrupted by noise during periods without any stimulus present. Whether such a

network is instantiated by actual neural circuits for our task to noiselessly accumulate sensory evidence remains to be explored.

In addition to the noiseless accumulation mechanism, it is also possible that the neural circuits involved in auditory-perceptual grouping are separated from the neural circuits that process and accumulate the sensory evidence for the frequency-change judgment. The computations for the frequency direction judgment and grouping judgment may be segregated along the dorsal and ventral auditory processing pathways. The ventral pathway is thought to be involved in extracting stimulus features to identify auditory objects while the dorsal pathway is thought to extract spatial information and integrate sensory and motor representations for speech perception and production (Rauschecker & Scott 2009, Bizley & Cohen 2013, Rauschecker 2011, Hickok & Poppel 2007). Belt regions of auditory cortex, which receive input from core auditory cortex, show the first signs of a functional difference in their neuronal tuning consistent with the dual pathway hypothesis. The ventral pathway extends from the anterolateral belt region of the auditory cortex to ventrolateral PFC, whereas the dorsal pathway originates in the caudal belt region of the auditory cortex and terminates in the dorsolateral PFC, via the parietal lobe.

On one hand, the sensory evidence for the frequency-change judgment may be processed along the ventral processing pathway. The anterolateral belt regions of auditory cortex, which are part of the ventral stream, is sensitive to complex combinations of stimulus features such as FM (Tian & Rauschecker 2000) and have

long temporal integration windows compared to primary auditory cortex (Scott et al 2011). The conversion of the sensory evidence to a categorical judgment by the accumulating sensory evidence is likely to take place downstream of anteriolateral belt auditory cortex in ventrolateral PFC, consistent with previous reports of auditory choice-related activity in vlPFC (Russ et al 2008).

On the other hand, dorsal stream areas, such as intraparietal sulcus, may be involved in the perceptual organization of incoming auditory stimuli into different streams (Teki et al 2011, Cusak 2005). Whether these areas are processing temporal information, consistent with a “when” processing pathway in the parietal lobe (Davis et al 2009), to form perceptually organized high-level representations or are combining sensory and motor information for guiding behavior remains to be explored in the future.

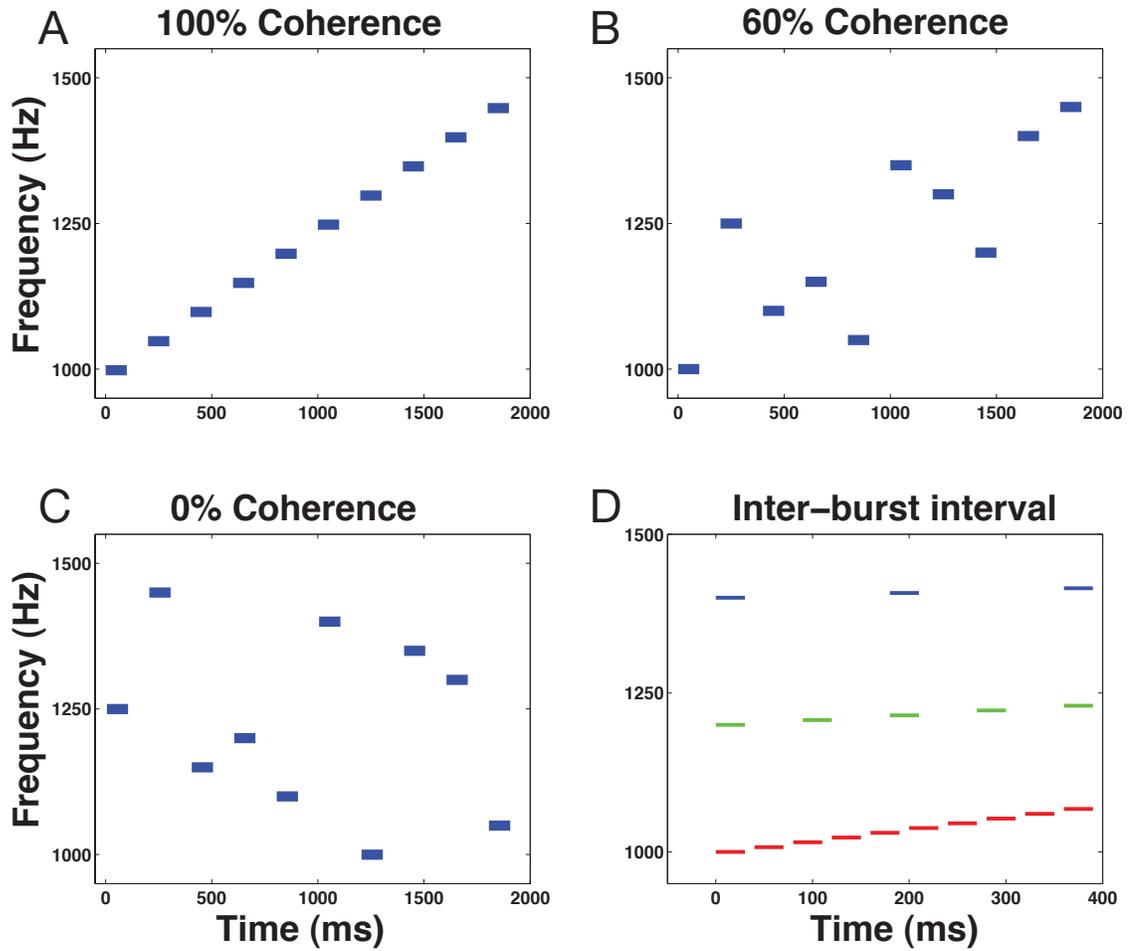
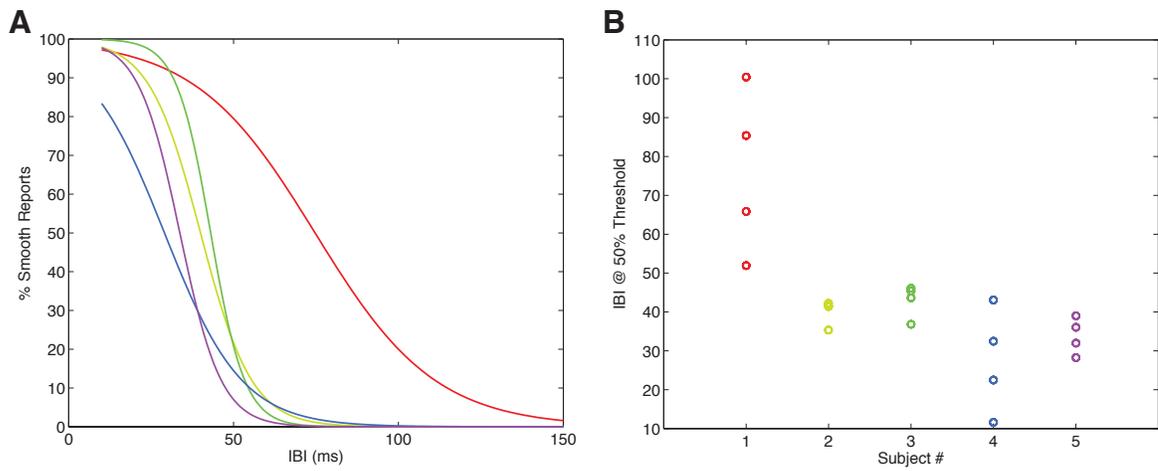


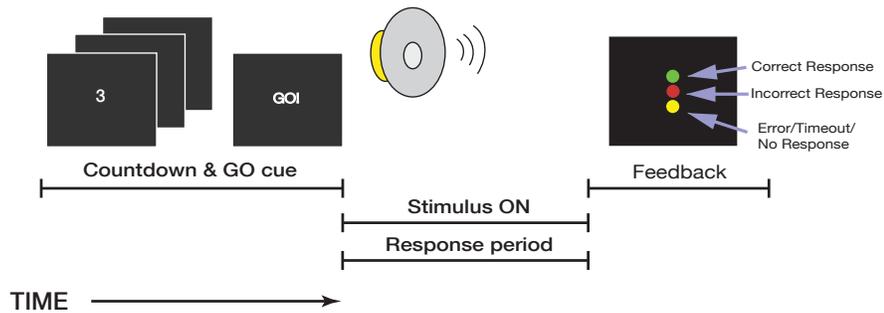
Figure 1  
 Stimulus parameters. A-C: Examples of stimuli varying from 100% - 0% coherence, increasing in frequency. D: The inter-tone burst interval (IBI) is the silent period between successive tone bursts.



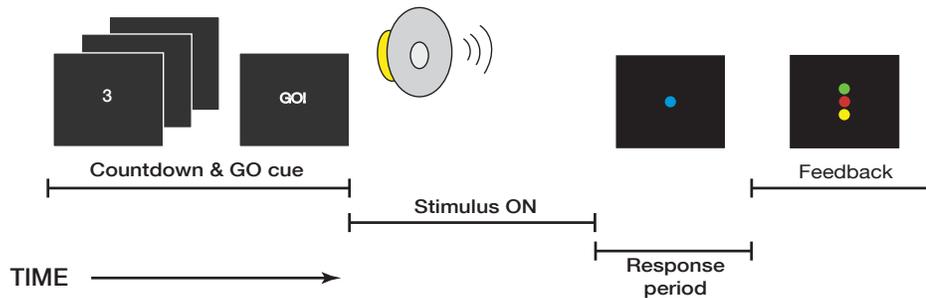
**Figure 2**

IBI affects subjects' reports of perceived grouping of stimuli. Subjects reported whether they perceived the stimulus as a continuous, grouped sound or as a series of discrete tones. Each subject was tested with a method of constant stimulus psychometric procedure across four separate sessions. A: Psychometric function for each subject (indicated by the color), fit to response data and IBI, pooled across sessions. B: Psychometric threshold IBI for each subject, corresponding to *continuous* responses on 50% of trials, plotted for each session.

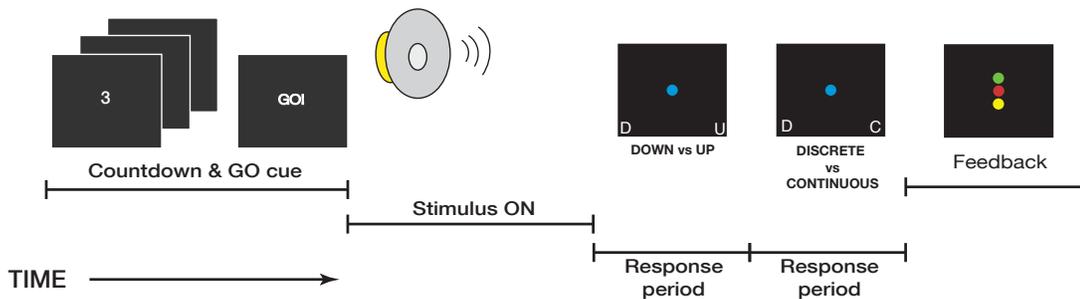
**A. Response Time UP vs DOWN task**



**B. Variable Duration UP vs DOWN task**



**C. Variable Duration UP vs DOWN & stimulus DISCRETE vs CONTINUOUS task**



**Figure 3**

Task variants. Subjects, in a 2AFC paradigm, discriminated between tone sequences increasing or decreasing in frequency. A countdown (“3”, “2”, “1”) and a “GO” cue preceded the stimulus presentation period. After the subjects responded with their decision, feedback regarding the judgment was given (“green” – correct, “red” – incorrect, “yellow” – response timeout). A: In the response-time task, subjects could respond as soon as the the stimulus was presented. The response time is defined as the time interval from the start of the stimulus presentation to when the subject responded with their choice on a game pad. B: In the interrogation task, subjects could only respond after the end of the stimulus and the stimulus duration was experimentally varied on each trial. C: The hybrid task adapts the interrogation task to include an additional response period in which subjects reported that they perceived the stimulus as a single, continuous sound or as a series of discrete tones.

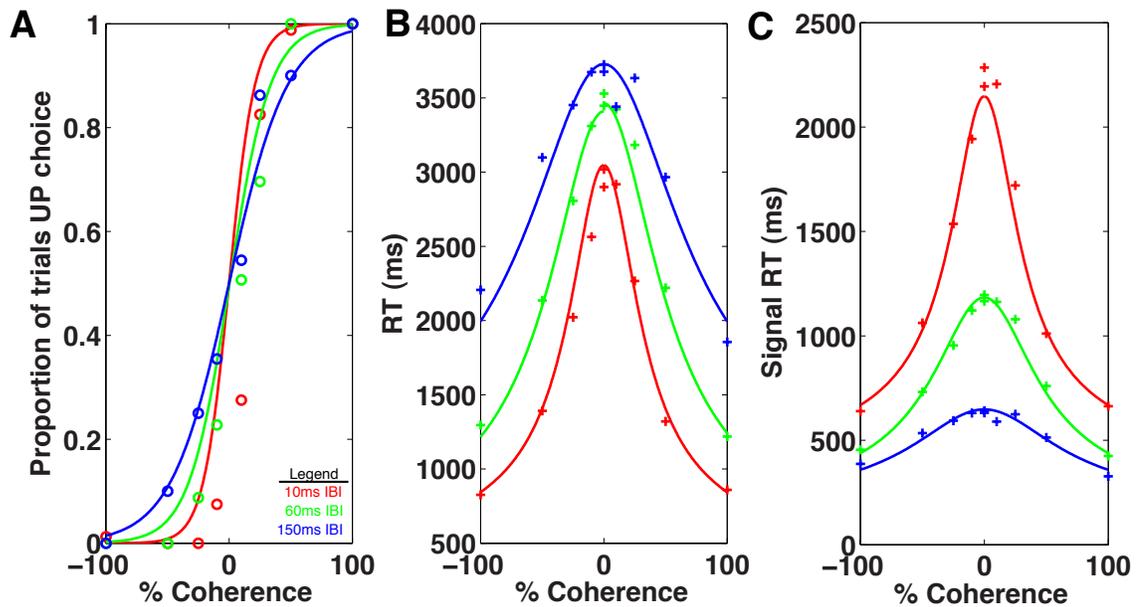


Figure 4

Response-time psychometric and chronometric data for one representative subject averaged across the repeated presentations for the particular inter-tone burst-interval (IBI) & coherence condition. A: Choice accuracy data (circles), and drift diffusion model (DDM) fits (solid line) for three IBIs plotted against signed coherence (negative coherence corresponds to stimuli decreasing in frequency while positive coherence corresponds to stimuli increasing in frequency). B. Response time (RT) data (“+” symbols) and DDM fits (solid line) for three IBIs. C: Signal response time (Signal RT) data (“+” symbols) and DDM fits (solid line) for three IBIs. Signal RT is the tone burst duration (30 ms, same for all IBI conditions) multiplied by the number of elapsed tones.

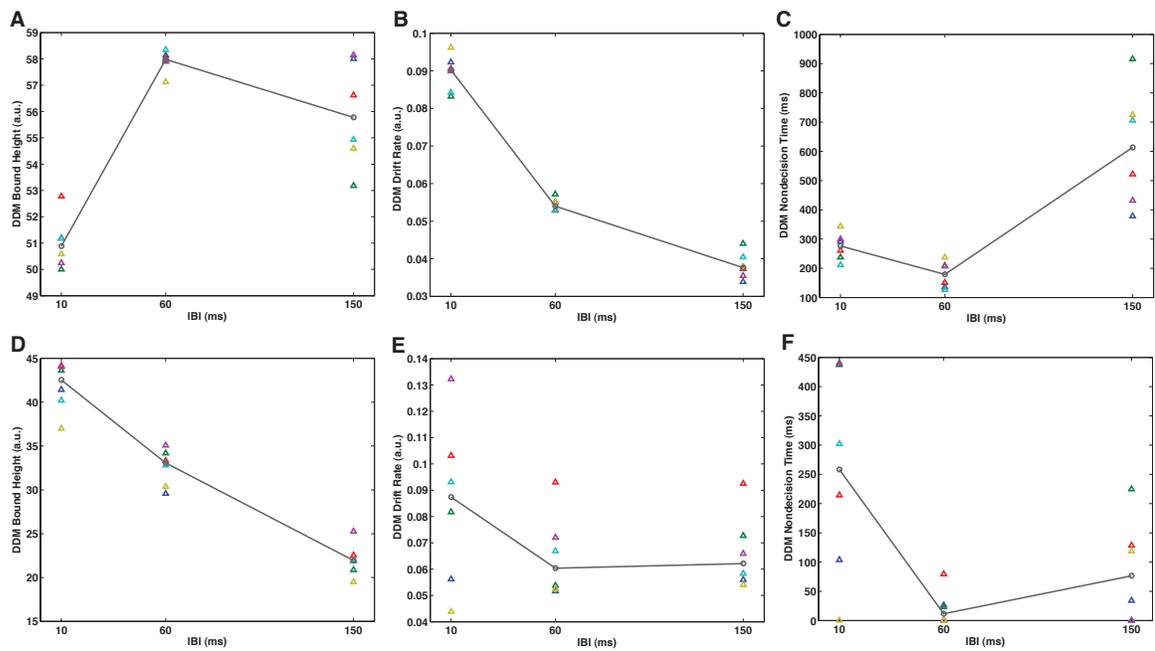


Figure 5

A,D: Decision bound height and B,E: drift rate (accumulation rate) C,F: non-decision time for DDM fitted to (A-C) RT data and (D-F) Signal RT data. Each colored triangle corresponds to one subject. Connected circles are the median values across subjects.

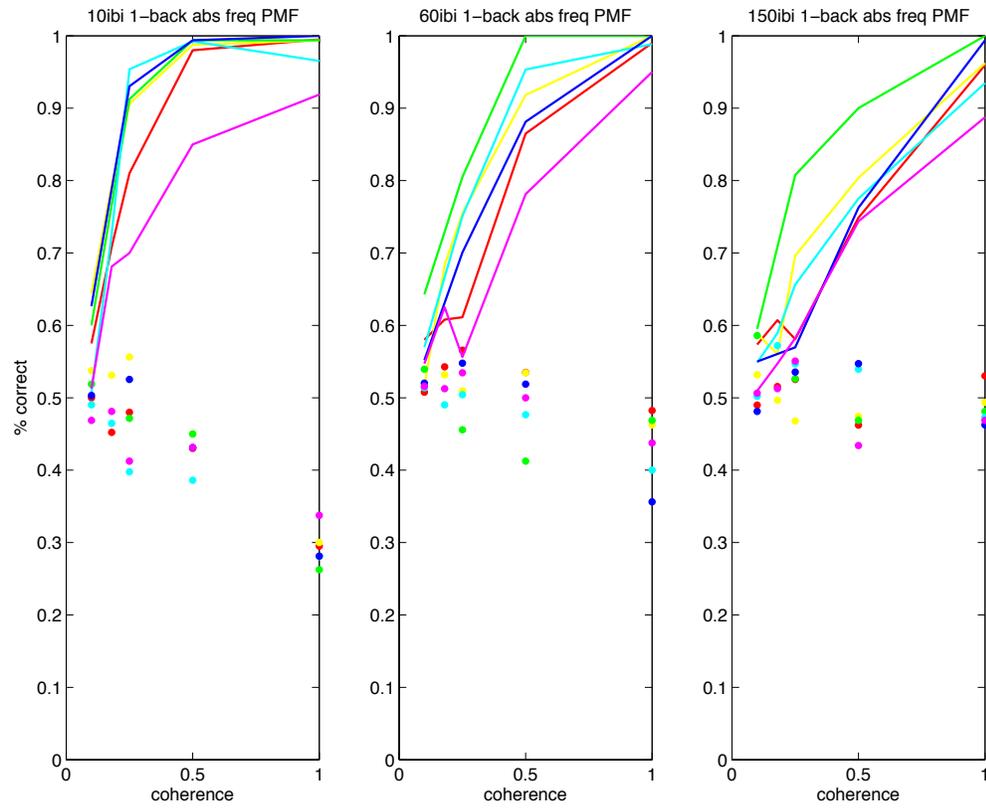
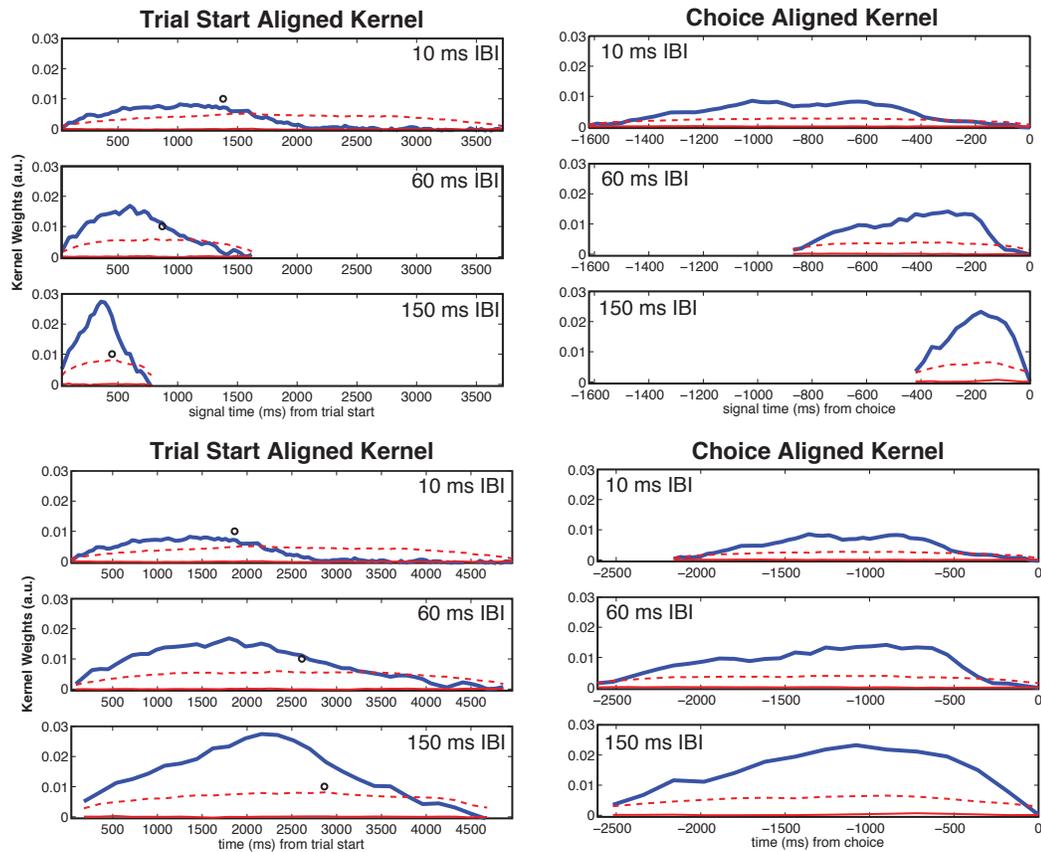
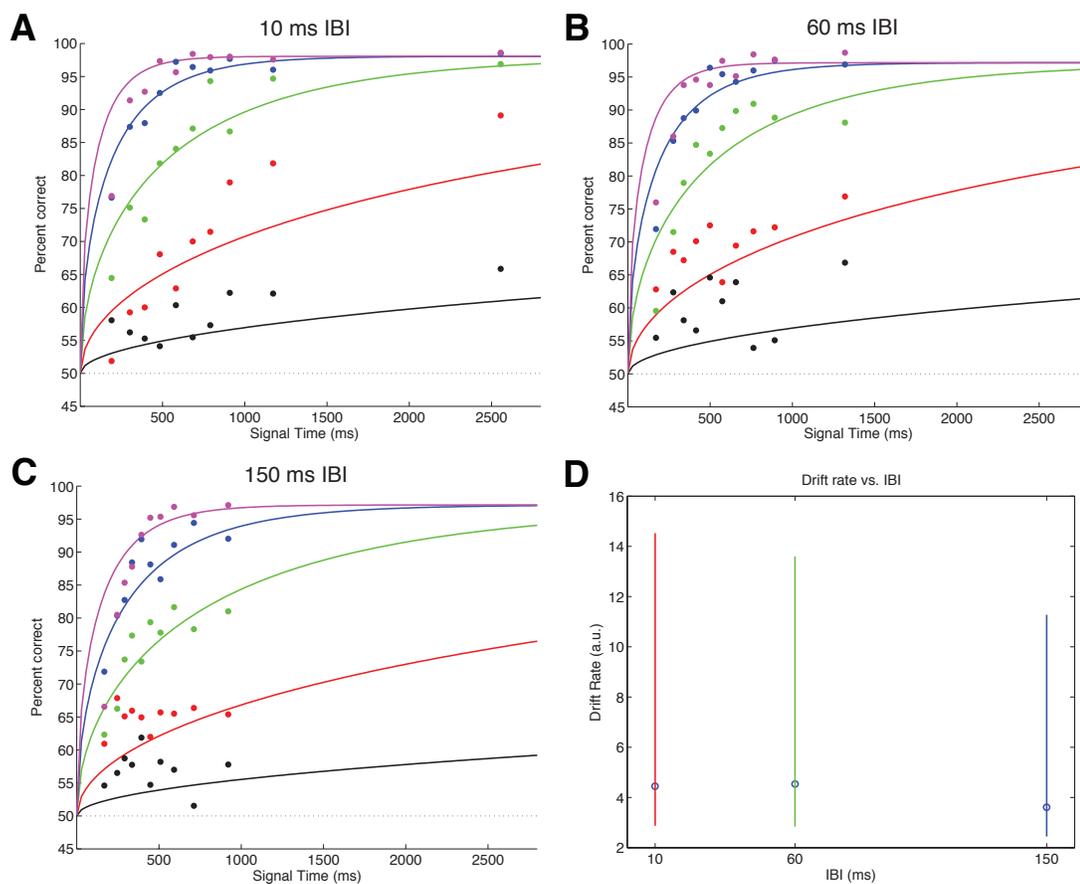


Figure 6  
 Comparison of absolute frequency decision model (dots) with actual data (lines) for all 6 subjects (colors) based on the response-time task data. (Left-Right) 10 ms IBI, 60 ms IBI, 150 ms IBI

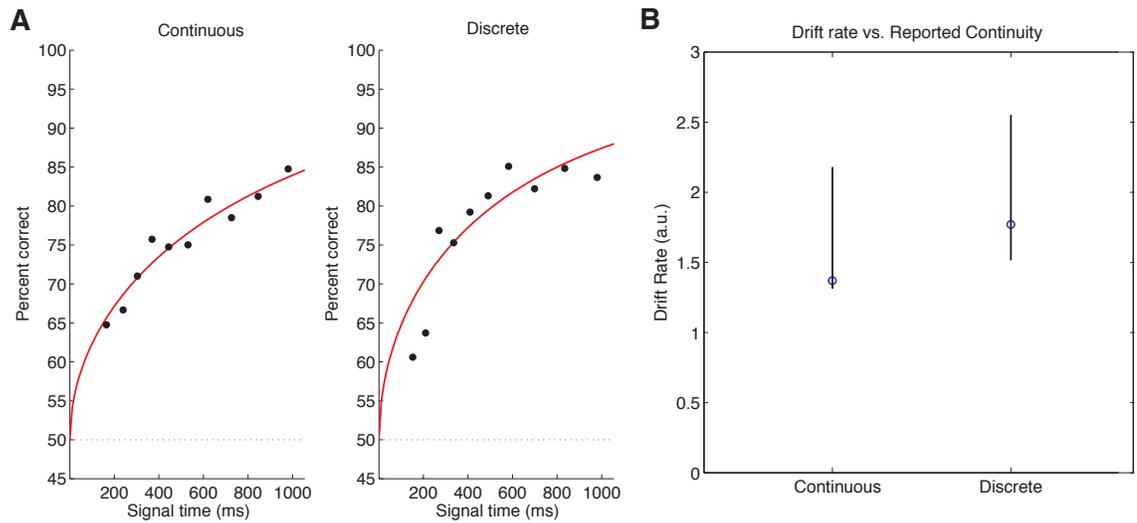


**Figure 7**

Psychophysical kernels based on 0% coherence trials (blue solid curves) aligned by the start of the trial (A,C) and by when the choice was made (B,D). Kernels were computed using logistic regression with data from the response-time task, pooled across 6 subjects. Red solid curve corresponds to mean of bootstrap weights and red dotted curve represents +2 SEM for the bootstrap weights. The kernels in panels A,B are plotted in signal time (tone number multiplied by tone duration, fixed at 30 ms for all IBIs), while the kernels in C, D are plotted in real time.



**Figure 8**  
 Time-dependent psychometric data (dots) and DDM fits (solid curves) for interrogation task on A: 10 ms, B: 60 ms, C: 150 ms IBI trials. The colors in A-C correspond to the coherences tested. D: Plot of fitted DDM drift coefficients for the three IBI conditions (circle) and their associated bootstrapped, 95% confidence intervals.



**Figure 9**

Time-dependent psychometric data (dots) and DDM fits (red solid curves) for the hybrid task in A. Prior to each session, the 50% IBI threshold for perceiving the stimulus as a single, continuous sound versus a discrete series of tones was measured for each subject. Subjects were then run on the hybrid task using this threshold IBI. The data for the threshold IBI was separated by the subjects' report of perceived continuity for analysis. B: Plot of fitted DDM drift rates (circle symbols) for the separately analyzed hybrid task data at the threshold IBI and their associated bootstrapped 95% confidence intervals.

# Chapter 3

## General Discussion & Conclusions

Our study establishes a new auditory decision making paradigm to test how sensory information is converted into a perceptual judgment. We used a novel auditory-discrimination task that required human subjects to report whether a stochastic auditory stimulus was increasing or decreasing in frequency. Using this task, we examined how temporal gaps in a stimulus, which can affect the perceptual grouping of the stimulus, can also affect the computations that convert sensory evidence into a perceptual judgment. By understanding how the brain converts the sensory stimulus into a perceptual decision with our task, we can better understand the computational principles and the neural implementation of how auditory perceptions are represented in the brain. We will discuss first, how timing affects the perceptual grouping of our stimulus, second, how our task provides evidence that the decision making processes are insensitive to IBI and to the perceptual grouping of the stimulus, and finally, future directions to extend our findings.

### IBI & Perceptual Organization

The timing of acoustic events can affect how they are perceived. We characterized how IBI, or the temporal gaps in our tone sequences, affected the perceptual grouping the stimulus. We found that by varying the IBI of our tone

sequences, listeners indeed perceive our stimuli differently. At short IBIs, listeners tend to perceive the stimulus as a continuous, grouped sound, and at long IBIs, listeners perceive the stimulus as a series of discrete tones. By varying the IBI of our stimulus, subjects could perceive the stimulus as grouped or as a discrete series of tones.

Our findings add to the types of stimuli that can be used to study how stimulus features can affect how sounds are perceptually organized. Auditory perceptual organization is typically studied using the “ABA” streaming paradigm (van Noorden 1975, Bregman 1994, Moore & Gockel 2012). The “ABA” stimulus consists of repeating “A” and “B” tones, which different frequencies. Specifically, for the “ABA” stimulus, faster presentation rates or larger frequency separations between the “A” and “B” tones are perceived as separate streams at the two frequencies. For slower presentation rates and smaller frequency separations, listeners perceive the stimulus as a galloping sound, representing one stream. This paradigm set the groundwork for understanding how perceptual grouping of the stimulus is affected by changes in the acoustic features of the stimulus. However, the “ABA” stimulus, as we have established, is not the only stimulus that can be used to study the principles of sequential auditory perceptual grouping.

How might the auditory system process the grouping cues in our stimulus? A number of models (Beauvois & Meddis 1996, Hartman & Johnson 1991) have been developed to explain how frequency-separation can affect the perception of streaming with the “ABA” streaming paradigm. These models are based on the

frequency selectivity and adaptation properties of neurons in primary auditory cortex. Stream segregation occurs when distinct populations of neurons are activated. For example, the population of neurons responding to two tones closely separated in frequency would be overlapping and therefore represent one stream. On the other hand, for two tones separated far apart in frequency, two distinct populations of neurons would be responding to the stimulus, and therefore the stimulus would be perceived as two streams. These population-based models could explain how listeners can perceive a grouped sound based on the frequency content of our stimulus, since coherent tones are separated by 7.5 hz, which is a smaller frequency separation than the bandwidth of primary auditory cortex neurons (Recanzone et al 2000). However, these population-based models do not take into account the relative timing of the different sound elements in a stimulus, which can have a large impact on how the stimulus is perceived.

One model that takes into account the temporal structure of the stimulus is the temporal-coherence model (Elhilali 2009, Shamma et al 2010). This model extends previous frequency-separation based models of streaming to account for the timing between acoustic events in the perceptual organization of sounds. According to the temporal coherence model, sounds that are closely separated in time and have a common rhythm, even if they are separated in frequency, will evoke temporally coherent activity across the neural population. The temporal coherence model can mirror human subjects' performance on a challenging auditory stochastic figure-ground detection task in which subjects detect the presence of a multi-tone

sequence embedded in a simultaneously presented random chord distractor stimulus (Teki et al 2013).

An open question is whether the brain perceptually organizes sounds by using the temporal coherence of neural populations representing the stimulus. One possible mechanism that is consistent with the temporal-coherence model is the phase entrainment of stimulus-evoked oscillations acting as a spectrotemporal filter. This filtering mechanism boosts the excitability of the neural populations corresponding to the temporally coherent stimuli and selectively attenuates activity of neural populations corresponding to the temporally incoherent sounds (Lakatos et al 2013). Whether phase-entrainment can be extended to more complex stimuli and be tested in awake, behaving animals remain to be seen.

In humans, the perceptual organization of a stimulus activates the intraparietal sulcus (Teki et al 2010, Cusack 2005). Using the stochastic figure-ground stimulus in which a tone sequence is embedded in a sequence of random chords, the detection of the target tone-sequence was correlated with activity in IPS, suggesting that this area contains an anatomical substrate for processing the perceptual organization of the stimulus. The IPS receives input from caudal belt regions of auditory cortex (Lewis & van Essen), and in non-human primates lateral intraparietal cortex (LIP) an area in the IPS has been shown to be involved in integrating spatial localization of an auditory stimulus and planning motor behavior (Cohen & Gifford) as well as forming high-level categorical representations independent of PFC activity (Freedman & Assad 2006, Swaminathan et al. 2012),

suggesting that the parietal cortex may play a role in high-level auditory perceptual processes such as the perceptual organization of an acoustic stimulus. Whether IPS plays a role in representing the outcome of the perceptual organization process or in actively processing the stimulus to perceptually organize the sounds sounds remains to be examined in the future.

### Sensory Evidence Accumulation

The accumulation of sensory evidence is a general computational model for explaining how sensory evidence is converted to a perceptual decision. Auditory tasks are no exception to this model. Our goal has been to focus on the role of time in auditory perception and by examining how sensory evidence is processed to form a perceptual judgment, we can examine how auditory processing may process time differently. We found that subjects accumulate sensory evidence to make their frequency-change judgments. In particular, subjects integrated sensory evidence over a fixed period of time independent of IBI, consistent with an accumulation-based perceptual decision making model with a collapsing decision boundary that took into account the costs and benefits of accumulating additional sensory evidence. We also found that the drift rate, which is a measure of the signal-to-noise of the sensory evidence and the rate of sensory evidence accumulation was independent of IBI.

Our finding that subjects integrate sensory evidence over a fixed period, consistent with a collapsing bound decision-making model suggests that listeners

making auditory judgments also adjust their strategy according to the difficulty and speed of the judgments. Subjects in the response-time task appear to compare the costs and benefits of accumulating additional sensory evidence in terms of elapsed time and not by the number of samples of sensory evidence. While we did not impose any explicit costs or rewards, finishing the trials as quickly as possible and as accurately as possible can be seen as an implicit reward for our task. Because trial parameters such as IBI, coherence, and the absolute frequencies of the stimulus sequence varied between trials, in order to efficiently complete the experimental task, subjects' strategies had to take into account the considerable variability in difficulty across trials. While examining the role of the decision bound was not the focus of this work, it would be useful in the future to examine directly each subject's cost function to determine how subjects weighed the costs and benefits to accumulating additional information and to allow for a direct comparison with the cost function for human subjects on a visual discrimination task (Drugowisch et al 2012).

Our main result is that the sensory evidence accumulation process was invariant to IBI and to the grouping of the stimulus. Our findings are consistent with recent work showing that accumulation of sensory evidence is insensitive to gaps in the stimulus (Kiani et al 2013, Brunton & Brody 2013). In one study, subjects in made a decision regarding which one of two stochastic, Poisson pulse trains had the higher rate. The trial-by-trial variability in the choice and response-times in this auditory spatial discrimination task was found to be the result of noise in the

representation of sensory evidence and not noise in the sensory evidence accumulation process (Brunton & Brody 2013). A separate study found that a time gap between two distinct stimulus pulses did not affect subjects' performance on a visual motion direction discrimination task (Kiani et al 2013). If the accumulation process for our task were noisy, the accumulator noise present during the silent IBI periods would be added to the accumulated sensory evidence. As IBI increased the amount of noise accumulated along with the incoming sensory evidence would also increase, leading to an overall increase in the variability in the total, accumulated sensory evidence. This increase in the variability of the accumulated sensory evidence would correspond to a decrease in the drift rate (Ratcliff et al 2002, Voss et al 2004), which we did not observe, suggesting that a noiseless sensory evidence accumulation process may be able to account for the invariance of the drift rate to IBI.

This noiseless integration process may be implemented by a multi-stable attractor network (Kiani et al 2013, Pouget & Latham 2002, Koulakov et al 2002). These network models use biologically plausible assumptions to construct a neural network which replicates some aspect of the target system. For the multi-stable attractor networks, the level of the accumulated sensory evidence does not decay or become corrupted by noise during periods without any stimulus present. Whether actual neural circuits instantiate such a network and how such circuits might function for our task remain to be explored and could provide insights into how the brain converts the auditory stimulus into a perceptual judgment.

In addition to the noiseless accumulation mechanism, it is also possible that the neural circuits involved in auditory-perceptual grouping are separated from the neural circuits that process and accumulate the sensory evidence for the frequency-change judgment. The computations for the frequency direction judgment and grouping judgment may be segregated along the dorsal and ventral auditory processing pathways. The ventral pathway is thought to be involved in extracting stimulus features to identify auditory objects while the dorsal pathway is thought to extract spatial information and integrate sensory and motor representations for speech perception and production (Rauschecker & Scott 2009, Bizley & Cohen 2013, Rauschecker 2011, Hickok & Poppel 2007). Belt regions of auditory cortex, which receive input from core auditory cortex, show the first signs of a functional difference in their neuronal tuning consistent with the dual pathway hypothesis. The ventral pathway extends from the anterolateral belt region of the auditory cortex to ventrolateral PFC, whereas the dorsal pathway originates in the caudal belt region of the auditory cortex and terminates in the dorsolateral PFC, via the parietal lobe.

On one hand, the sensory evidence for the frequency-change judgment may be processed along the ventral processing pathway. The anterolateral belt regions of auditory cortex, which are part of the ventral stream, is sensitive to complex combinations of stimulus features such as FM (Tian & Rauschecker 2000) and have long temporal integration windows compared to primary auditory cortex (Scott et al 2011). The conversion of the sensory evidence to a categorical judgment by the

accumulating sensory evidence is likely to take place downstream of anteriolateral belt auditory cortex in ventrolateral PFC, consistent with previous reports of auditory choice-related activity in vlPFC (Russ et al 2008).

On the other hand, dorsal stream areas, such as intraparietal sulcus, may be involved in the perceptual organization of incoming auditory stimuli into different streams (Teki et al 2011, Cusak 2005). Whether these areas are processing temporal information, consistent with a “when” processing pathway in the parietal lobe (Davis et al 2009), to form perceptually organized high-level representations or are combining sensory and motor information for guiding behavior remains to be explored in the future.

# Bibliography

1. Beauvois, M. W. (1998) The effect of tone duration on auditory stream formation. *Percept Psychophys* 60, 852-61.
2. Beauvois, M. W. & Meddis, R. (1996) Computer simulation of auditory stream segregation in alternating-tone sequences. *J Acoust Soc Am* 99, 2270-2280.
3. Bey, C. & McAdams, S. (2002) Schema-based processing in auditory scene analysis. *Percept Psychophys* 64, 844-54.
4. Bizley, J. K. & Cohen, Y. E. (2013) The what, where and how of auditory-object perception. *Nat Rev Neurosci* 14, 693-707.
5. Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113, 700-65.
6. Borchert, E. M. O., Micheyl, C., & Oxenham, A. J. (2011) Perceptual grouping affects pitch judgments across time and frequency. *J Exp Psychol Hum Percept Perform* 37, 257-69.
7. Bregman, A. S. (1994) *Auditory scene analysis: The perceptual organization of sound*.
8. Bregman, A. S., Ahad, P. A., Crum, P. A., & O'Reilly, J. (2000) Effects of time intervals and tone durations on auditory stream segregation. *Percept Psychophys* 62, 626-36.
9. Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013) Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. *Science* 340, 95-98.
10. Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008) Decision-making with multiple alternatives. *Nat Neurosci* 11, 693-702.
11. Cisek, P., Puskas, G. A., & El-Murr, S. (2009) Decisions in changing conditions: the urgency-gating model. *J Neurosci* 29, 11560-71.
12. Cohen, Y. E., Russ, B. E., & Gifford, 3rd, G. W. (2005) Auditory processing in the posterior parietal cortex. *Behav Cogn Neurosci Rev* 4, 218-31.
13. Cusack, R. (2005) The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience* 17, 641--651.
14. Davis, B., Christie, J., & Rorden, C. (2009) Temporal order judgments activate temporal parietal junction. *J Neurosci* 29, 3182-8.
15. Ditterich, J. (2006) Evidence for time-variant decision making. *Eur J Neurosci* 24, 3628-41.

16. Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012) The cost of accumulating evidence in perceptual decision making. *J Neurosci* 32, 3612-28.
17. Eckhoff, P., Holmes, P., Law, C., Connolly, P. M., & Gold, J. I. (2008) On diffusion processes with variable drift rates as models for decision making during learning. *New J Phys* 10, nihpa49499.
18. Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., & Shamma, S. A. (2009) Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61, 317-29.
19. Elhilali, M. & Shamma, S. A. (2008) A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *J Acoust Soc Am* 124, 3751-71.
20. Freedman, D. J. & Assad, J. A. (2006) Experience-dependent representation of visual categories in parietal cortex. *Nature* 443, 85-8.
21. Gold, J. I. & Shadlen, M. N. (2007) The neural basis of decision making. *Annu Rev Neurosci* 30, 535-74.
22. Green, D. M. & Swets, J. A. (1966) *Signal detection theory and psychophysics*.
23. Griffiths, T. D. & Warren, J. D. (2004) What is an auditory object? *Nat Rev Neurosci* 5, 887-92.
24. Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011) Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J Neurosci* 31, 6339-52.
25. Hartmann, W. M. & Johnson, D. (1991) Stream segregation and peripheral channeling. *Music Perception* 9, 155--183.
26. Hickok, G. & Poeppel, D. (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8, 393-402.
27. Kiani, R., Churchland, A. K., & Shadlen, M. N. (2013) Integration of Direction Cues Is Invariant to the Temporal Gap between Them. *J Neurosci* 33, 16483-9.
28. Kidd, Jr, G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994) Reducing informational masking by sound segregation. *J Acoust Soc Am* 95, 3475-80.
29. Kidd, Jr, G., Mason, C. R., & Richards, V. M. (2003) Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *J Acoust Soc Am* 114, 2835-45.
30. Knoblauch, K. & Maloney, L. T. (2008) Estimating classification images with generalized linear and additive models. *J Vis* 8, 10.1-19.
31. Koulakov, A. A., Raghavachari, S., Kepecs, A., & Lisman, J. E. (2002) Model for a robust neural integrator. *Nat Neurosci* 5, 775-82.

32. Lakatos, P., Musacchia, G., O'Connell, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013) The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750-61.
33. Lewis, J. W. & Van Essen, D. C. (2000) Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J Comp Neurol* 428, 112-37.
34. Luce, R. D. (1986) *Response Times: Their Role in Inferring Elementary Mental Organization*.
35. Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003) A role for neural integrators in perceptual decision making. *Cereb Cortex* 13, 1257-69.
36. Micheyl, C. & Oxenham, A. J. (2010) Objective and subjective psychophysical measures of auditory stream integration and segregation. *J Assoc Res Otolaryngol* 11, 709-24.
37. Moore, B. C. J. & Gockel, H. E. (2012) Properties of auditory stream formation. *Philos Trans R Soc Lond B Biol Sci* 367, 919-31.
38. Murray, R. F. (2011) Classification images: A review. *J Vis* 11, 2.
39. Nelken, I., Rotman, Y., & Bar Yosef, O. (1999) Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* 397, 154-7.
40. Neri, P. & Levi, D. M. (2006) Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision Res* 46, 2465-74.
41. Palmer, J., Huk, A. C., & Shadlen, M. N. (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J Vis* 5, 376-404.
42. Pouget, A. & Latham, P. (2002) Digitized neural networks: long-term stability from forgetful neurons. *Nat Neurosci* 5, 709-10.
43. Ratcliff, R. & McKoon, G. (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput* 20, 873-922.
44. Ratcliff, R. & Smith, P. L. (2004) A comparison of sequential sampling models for two-choice reaction time. *Psychol Rev* 111, 333-67.
45. 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, 438-81.
46. Rauschecker, J. P. (2011) An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear Res* 271, 16-25.
47. Rauschecker, J. P. & Scott, S. K. (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12, 718-24.

48. Rauschecker, J. P. & Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A* 97, 11800-6.
49. Recanzone, G. H., Guard, D. C., & Phan, M. L. (2000) Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J Neurophysiol* 83, 2315-31.
50. Richards, V. M., Carreira, E. M., & Shen, Y. (2012) Toward an objective measure for a "stream segregation" task. *J Acoust Soc Am* 131, EL8-13.
51. Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2002) Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J Acoust Soc Am* 112, 2074-85.
52. Roitman, J. D. & Shadlen, M. N. (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22, 9475-89.
53. Russ, B. E., Orr, L. E., & Cohen, Y. E. (2008) Prefrontal neurons predict choices during an auditory same-different task. *Curr Biol* 18, 1483-8.
54. Scott, B. H., Malone, B. J., & Semple, M. N. (2011) Transformation of temporal processing across auditory cortex of awake macaques. *J Neurophysiol* 105, 712-30.
55. Shadlen, M. N., Hanks, T. D., Churchland, A. K., Kiani, R., & Yang, T. (2006) The speed and accuracy of a simple perceptual decision: a mathematical primer. , 209--37.
56. Shadlen, M. N. & Kiani, R. (2013) Decision making as a window on cognition. *Neuron* 80, 791-806.
57. Shadlen, M. N. & Newsome, W. T. (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol* 86, 1916-36.
58. Shamma, S. A., Elhilali, M., & Micheyl, C. (2011) Temporal coherence and attention in auditory scene analysis. *Trends Neurosci* 34, 114-23.
59. Smith, P. L. (1995) Psychophysically principled models of visual simple reaction time.. *Psychological review* 102, 567.
60. Sussman, E. & Steinschneider, M. (2009) Attention effects on auditory scene analysis in children. *Neuropsychologia* 47, 771-85.
61. Sussman, E. S. (2005) Integration and segregation in auditory scene analysis. *J Acoust Soc Am* 117, 1285-98.
62. Swaminathan, S. K., Masse, N. Y., & Freedman, D. J. (2013) A comparison of lateral and medial intraparietal areas during a visual categorization task. *J Neurosci* 33, 13157-70.

63. Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013) Segregation of complex acoustic scenes based on temporal coherence. *Elife* 2, e00699.
64. Teki, S., Chait, M., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2011) Brain bases for auditory stimulus-driven figure-ground segregation. *J Neurosci* 31, 164-71.
65. Thompson, S. K., Carlyon, R. P., & 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, 1253-62.
66. Usher, M. & McClelland, J. L. (2001) The time course of perceptual choice: the leaky, competing accumulator model. *Psychol Rev* 108, 550-92.
67. Voss, A., Rothermund, K., & Voss, J. (2004) Interpreting the parameters of the diffusion model: an empirical validation. *Mem Cognit* 32, 1206-20.
68. van Noorden, L. P. A. S. (1975) *Temporal coherence in the perception of tone sequences*.