RESEARCH ARTICLE

Neuromechanistic Model of Auditory Bistability

James Rankin[1] *, Elyse Sussman[2,3], John Rinzel[1,4]

1 Center for Neural Science, New York University, New York, New York, United States of America,
2 Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York, United States of America, **3** Department of Otorhinolaryngology-HNS, Albert Einstein College of Medicine, Bronx, New York, United States of America, **4** Courant Institute of Mathematical Sciences, New York University, New York, New York, United States of America

* james.rankin@nyu.edu

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Sequences of higher frequency A and lower frequency B tones repeating in an ABA- triplet pattern are widely used to study auditory streaming. One may experience either an integrated percept, a single ABA-ABA- stream, or a segregated percept, separate but simultaneous streams A-A-A-A- and -B—B–. During minutes-long presentations, subjects may report irregular alternations between these interpretations. We combine neuromechanistic modeling and psychoacoustic experiments to study these persistent alternations and to characterize the effects of manipulating stimulus parameters. Unlike many phenomenological models with abstract, percept-specific competition and fixed inputs, our network model comprises neuronal units with sensory feature dependent inputs that mimic the pulsatile-like A1 responses to tones in the ABA- triplets. It embodies a neuronal computation for percept competition thought to occur beyond primary auditory cortex (A1). Mutual inhibition, adaptation and noise are implemented. We include slow NDMA recurrent excitation for local temporal memory that enables linkage across sound gaps from one triplet to the next. Percepts in our model are identified in the firing patterns of the neuronal units. We predict with the model that manipulations of the frequency difference between tones A and B should affect the dominance durations of the stronger percept, the one dominant a larger fraction of time, more than those of the weaker percept—a property that has been previously established and generalized across several visual bistable paradigms. We confirm the qualitative prediction with our psychoacoustic experiments and use the behavioral data to further constrain and improve the model, achieving quantitative agreement between experimental and modeling results. Our work and model provide a platform that can be extended to consider other stimulus conditions, including the effects of context and volition.

Author Summary

Humans have an astonishing ability to separate out different sound sources in a busy room: think of how we can hear individual voices in a bustling coffee shop. Rather than

voices, we use sound stimuli in the lab: repeating patterns of high and low tones. The tone sequences are ambiguous and can be interpreted in different ways—either grouped into a single stream, or separated out into different streams. When listening for a long time, one's perception switches every few seconds, a phenomenon called auditory bistability. Based on knowledge of the organization of brain areas involved in separating out different sound sources and how neurons in these areas respond to the ambiguous sequences, we developed a computational model of auditory bistabilty. Our model is less abstract than existing models and shows how groups of neurons may compete in order to dictate what you perceive. We predict how the difference between the two tone sequences affects what you hear over time and we performed an experiment with human listeners to confirm our prediction. The model provides groundwork to further explore the way the brain deals with the busy and often ambiguous world of sound.

Introduction

Auditory scene analysis involves segregating a complex scene into individual objects or streams [1]. A common stimulus used to study streaming in psychoacoustic experiments involves alternating tone sequences organized in repeating ABA- triplets [2]. The sequence can be perceived integrated in to one, or segregated into two streams, see Fig 1A and 1B. A key point of comparison in studies of streaming has been the van Noorden diagram, which describes the predominance of different interpretations across ranges of two stimulus parameters: the difference (Δf) in frequency between the A and the B tones and the tone presentation rate (PR). The diagram describes parameter regions where integrated is predominant (smaller Δf), where segregated is predominant (larger Δf), or where perception is ambiguous between the two (an intermediate range, also dependent on PR). Over a large range of stimulus parameter values the initial percept is integrated, but as time proceeds a perceptual switch to the segregated interpretation becomes more likely, a phenomenon called the build up of stream segregation [1, 3, 4]. At intermediate values of Δf the build up to segregation takes a few to ten seconds, but can take extend to tens of seconds at small Δf and occur almost instantaneously at large Δf. Behavioral studies have also looked at the effects of attention [5–7] and context [8–10] for build up.

Several computational models have focused on the early phase of auditory streaming, reproducing some characteristic features of build up [11–13], of the van Noorden diagram [14–16] or of both [17]. Various modelling approaches have been employed to reach these goals, including a coupled oscillator networks with frequency and time dimensions [15], a neural field description with a continuous representation of tonotopy [16] and a statistical model treating build up as an alternating renewal process [13]. The models are typically quite abstract, although some include a hybrid of auditory pathway qualitative features and signal processing [11, 14, 17]. However, these models were not neuromechanistic in nature; they did not account for neuronal dynamics based on explicit neurophysiological mechanisms nor did they study dynamics beyond a first perceptual switch and the build up phase.

During long stimulus presentations, on the order of several minutes, the initial switch to segregated is typically followed by further switches back and forth between segregation and integration [18–21], as illustrated by a subject's reporting of current percept in Fig 1C. Imaging studies have shown activation of a thalamocortical network [22] and the cerebellum [23] around the time of perceptual switches and an MEG study [18] localised to auditory cortex implicates non-primary auditory areas in maintaining perceptual streams. Psychoacoustic studies have shown that the initial perceptual phase is typically longer than the subsequent
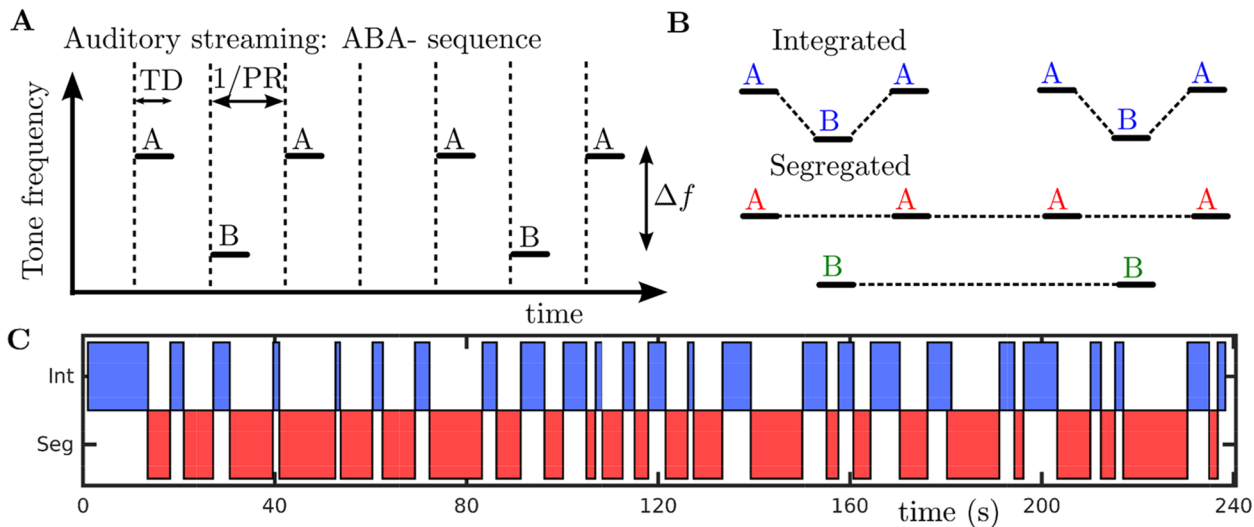
Fig 1. Stimulus paradigm and two possible percepts. A: Repeating ABA- triplet sequences (two triplets shown) consist of higher frequency pure tones A interleaved with lower frequency pure tones B of duration TD separated by a frequency difference Δf. The time between tone onsets (dashed vertical lines) is inverse of the presentation rate $1/PR$ (the "-" in "ABA-" represents a silence of duration $1/PR$). Throughout this paper tone duration will be set to $TD = 1/PR$ such that offset of an A tone abuts the onset of the next B tone. **B**: The stimulus is perceived as either integrated into a single stream ABA-ABA- or as two separate streams A-A-A-A- and -B—B–. **C**: Subject reports of integrated and segregated from a single 4-minute trial (480 triplets) at $\Delta f = 5$ st and $PR = 8$ Hz.

doi:10.1371/journal.pcbi.1004555.g001

percept durations [21] and that mean percept durations remain stable from the second perceptual phase onward [19]. The phenomenological CHAINS model presented in [24] is an implementation of the two stage proposal outlined in [25], addressing algorithmically the formation of the different perceptual patterns that enter into competition and capturing the later competition between these organizations. Aside from the CHAINS model [24], the available auditory modelling literature has not addressed the persistent perceptual alternations following the build up phase for the auditory streaming paradigm. The persistent, irregular alternations after the first perceptual switch will be the focus of this study.

Numerous neurophysiology studies using short stimulus presentations (say, less than 20 s) are of particular relevance to our modeling approach. Trial averaged recordings show a dependence on Δf and PR in the primary auditory cortex (A1) of awake monkeys [26–28], in the forebrain of songbirds [29] and in the cochlear nucleus of anaesthetized guinea pigs [30]. A1 responses were interpreted in a statistical model using neurometric functions that mimic the time course of build up in [28, 30]. In other studies data have been interpreted according to a conceptual model [26, 29] based on the tonotopic organization in auditory cortex. When the A and B tones are far enough apart in frequency, cortical activity is sufficiently separated in tonotopic space to drive segregated representations of the high and low tone sequences. However, when the A and B tones are closer, there is enough receptive field overlap for both tones to drive a common, intermediate population that could encode the integrated percept. There is no evidence (nor do the authors conclude) that the neural substrate for the first perceptual switch is to be found in A1 and further experiments in the behaving songbird suggest stimulus features, but not perceptual choice, are encoded in A1 [31].

In this study we develop a neuromechanistic model of auditory streaming that focuses on accurately reproducing the dynamics of the alternations after the build up phase (i.e. after the first perceptual switch). Our formulation is directly motivated from physiological studies of auditory streaming [26, 28] and general models of perceptual bistability [32–34]. A reduction from a continuous feature representation to percept based inputs and competition was

proposed in [32], and this approach has become relatively standard in general models of perceptual rivalry. By contrast, input in our model directly represents the dynamics of sensory features and incorporates neuronal responses of pre-competition stages. It features a tonotopic organization with three units assumed to be downstream and receive input from A1. Two peripheral units receive input from regions of A1 centred at locations with best frequencies A or B, and a third unit receives input from tonotopically intermediate A1 locations, say centred at $(A + B)/2$. The perceptual interpretations are classified through criteria on the firing patterns, for example, dominance of the central unit, receiving input via A1 from both the A and the B tones, corresponds to integrated [26]. A dynamical systems framework is used with firing rate based neuronal competition mediated by mutual inhibition, adaptation and noise [32, 34]; this combination of neural mechanisms has proved successful in accounting for many of the characteristic behaviors of perceptual bistability. Our model incorporates in its inputs the onset and transient dynamics of A1 responses with gaps between triplets. The inclusion of recurrent NMDA-like excitation allows for some neuronal memory that links each tone and each triplet to the next.

A starting hypothesis is that the perceptual organization known from [2] for initial dominance phases extends to long presentations; this allows for the tonotopic distribution of input and lateral connectivity to be constrained. As explained in more detail in the *Discussion*, comparisons with van Noorden's experimental findings are frequently made in experiments and models that do not take into account time-varying stimuli, attention and the termination of trials at the first perceptual switch. Accordingly, the starting assumption with regards to the van Noorden organization is re-evaluated based our own experimental findings. A further point of comparison from the empirical literature will be the log-normal- or gamma-like distribution of dominance durations characteristic of perceptual bistability [35]. Two existing auditory studies have reported durations resembling the characteristic distributions [19, 24]. Here, we will use the dominance duration distributions from our own experiments outlined below to constrain model parameters.

In parallel and interacting with the model development we have established in psychoacoustic experiments how parameter manipulations (here Δf) affect the dominance of competing percepts. Importantly, we have confirmed that a generalization about parameter manipulations from visual bistability extends to the auditory case. Levelt's Proposition II (LPII) described the specific way that asymmetric contrast manipulations in binocular rivalry can shift the proportion of dominance from one perceptual interpretation to another [36, 37]. In [38], LPII was generalized for three bistable visual stimuli varying parameters either side of equidominance. Equidominance coincides with a choice of stimulus parameters where each of the two interpretations are dominant for an equal proportion of time. It was shown that as a parameter is varied, the durations of the stronger percept (dominant for a larger proportion of time) are affected more than the durations of the weaker percept. Although some existing studies recorded perceptual durations for a similar stimulus paradigm (with minutes-long trials), data was not been presented so as to allow for a comparison with gLPII. For example, [21] considered measures that did not separate the two percept types and [24] did not separate the first phase from subsequent phases in their analysis. Our experiment also provides a means to test model predictions, to constrain its parameters and to evaluate its underlying assumptions. Indeed, the model was developed concurrently with the design and execution of the experiments, a process that has allowed us to develop a deeper understanding as one discipline feeds into and informs the other. The results in this paper are organized to reflect this process. The model is described in detail in *Neuromechanistic model of auditory bistability*, initial predictions are described in *Parametric dependence of perceptual dominance: model*, experimental results are in *Parametric dependence of perceptual dominance: experiment*, and a comparison with the model is presented in and *Refined model better captures experimental results*.

Results

Neuromechanistic model of auditory bistability

The neuronal circuits for competition and perceptual encoding are assumed to be downstream and taking inputs from A1 (see the Discussion, specifically *Model of neuronal competition beyond A1*, for comments on the possible cortical location of our model). The periodic inputs mimic the A1-responses to ABA- sequences reported in [28]. A firing rate description is used where competitive interactions emerge through a combination of excitatory and inhibitory connections, a slow fatigue process and intrinsic noise. We provide a brief outline of the model's architecture, mechanisms and inputs here; the full model equations and further details are given in *Model equations and details*.

The schematic in Fig 2A shows units r_A, r_B and r_{AB} that respectively pool inputs from regions of A1 centred at locations with best frequencies A, B and somewhere between, say (A + B)/2. The frequency difference between A and B is Δf. The associated variables r_k represent the mean firing rate of a population of neurons centred at the corresponding tonotopic location k ($k = \{A, AB, B\}$). The inputs I_A and I_B, described in more detail below, are distributed between the units (arrows) with less of the input I_B (I_A) reaching r_{AB} and r_A (r_B) with increasing Δf as controlled by $w(\Delta f)$ (plotted in panel C). The amount of mutual inhibition (circular-terminated vertices) between the units also depends on Δf as controlled by $C_i(\Delta f)$ (plotted in panel C for local i_{lcl} and global i_{gbl} inhibition cases, see below). For each unit r_k the intrinsic dynamics are illustrated in panel B and decribed by a differential equation like the following,

$$
\begin{aligned}
\tau_r \dot{r}_{AB} = \quad & -r_{AB} + F(\beta_e d_{AB} e_{AB} \\
& -C_i(0)r_{AB} - C_i(\Delta f/2)(r_A + r_B) - ga_{AB} + w(\Delta f/2)(I_A + I_B) + \chi_{AB}).
\end{aligned}
\tag{1}
$$

By way of an example, we describe this equation for r_{AB} in detail; the equations for r_A and r_B take the same general form. The dynamics evolve on a cortical timescale τ_r with a decay term and a sigmoidal function F that transforms all local and lateral inputs to a firing rate, see Fig 2D. Recurrent excitation has strength β_e and the excitation variable e_{AB} evolves on a timescale τ_e, slower than the cortical dynamics. The inclusion of slow NMDA-like excitation was necessary in order to maintain activity during the silent phases between successive inputs. The recurrent excitation term also features slow synaptic depression d_{AB} operating on a timescale τ_d. The function $C_i(\Delta f)$ controls the strength of local inhibition at $\Delta f = 0$ and lateral inhibition from r_A and r_B, a distance $\frac{\Delta f}{2}$ away. Net local excitation is ensured by setting $\beta_e > C_i(0) = \beta_i$. A simplification, that inhibition is instantaneous $\tau_i \approx 0$, removes the need for an explicit inhibition variable i_k and inhibition is proportional to the cortical variables r_k. Simulations were run (not shown) to verify that this simplification did not have a significant effect on the model's dynamics. Intrinsic spike frequency adaptation has strength g and the adaptation variable a_{AB} evolves on a slow timescale τ_a. Intrinsic additive noise χ_{AB} is an independent Ornstien-Uhlenbeck process for each r_k.

Inputs to the model are illustrated in Fig 3. The onset-plateau shape of A1 responses to a single tone mimic [28] and are shown in panel A for an isolated 125 ms A tone at locations A, AB and B with $\Delta f = 4$ st. The inputs are illustrated here for a presentation rate (inverse of the tone durations) of $PR = 8$ Hz ($= 1/125$); no gaps are assumed between the tones ($TD = 1/PR$) and tone durations are assumed vary with PR accordingly. The input spread across tonotopy is mediated by $w(\Delta f)$ as plotted in Fig 2C. The overall input for an ABA- triplet at the three locations is illustrated in panel B. The same curves are plotted for two triplets distributed across the tonotopic axis in panel C. Note that there is a full response to the A tone at the location A, an
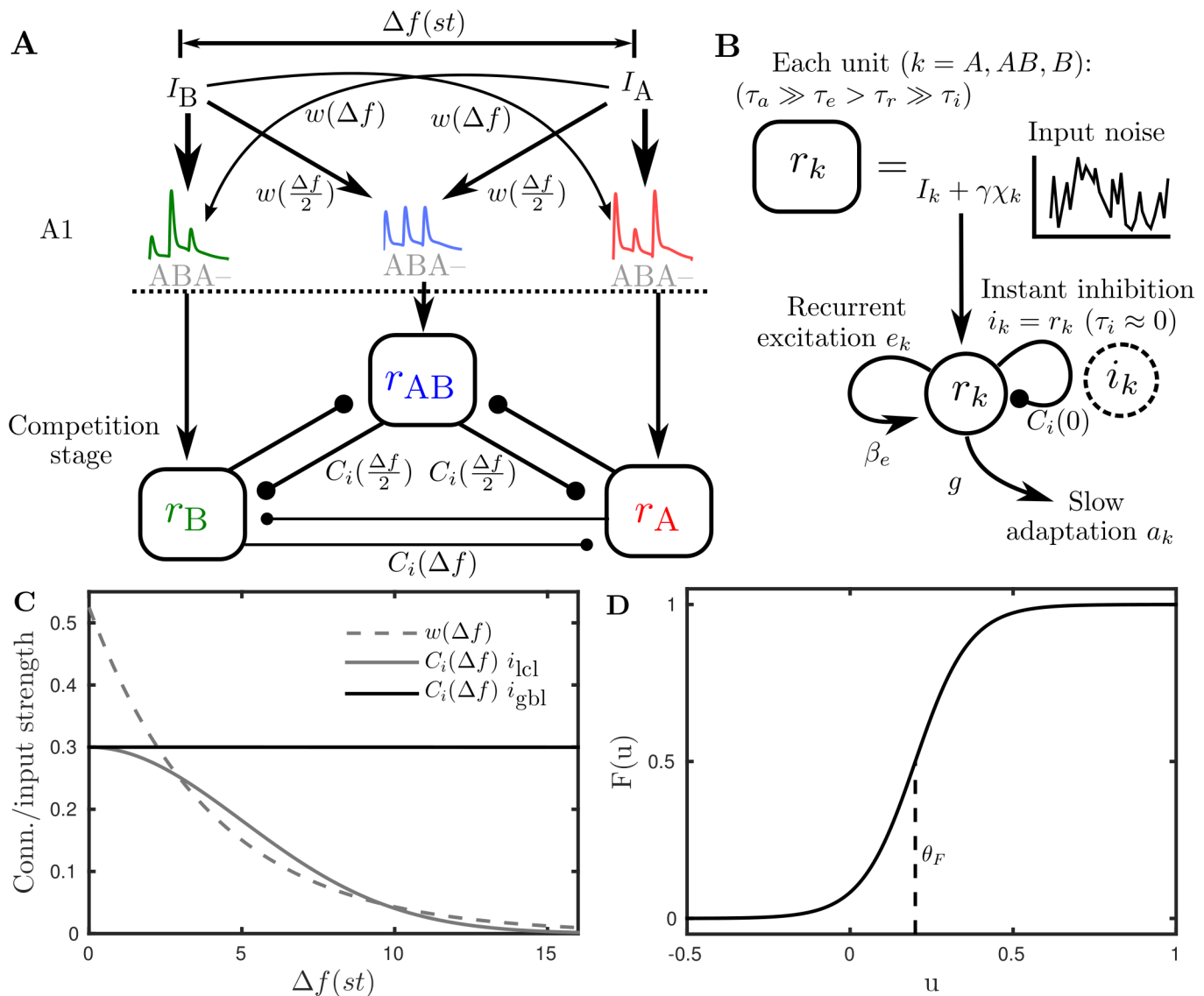
Fig 2. Model architecture. A: The tone frequency difference between A and B is Δf. The spread of inputs I_A and I_B across A1 is governed by the decaying input gain function $w(\Delta f)$. Example A1 response patterns to the ABA- stimulus are shown (see Fig 3 and associated text for more details); these form the inputs to three neuronal populations r_A, r_B and r_{AB} at the competition stage. Lateral inhibition strength can depend on Δf (i_{lcl} case) or be independent of Δf (i_{gbl} case) as governed by $C_i(\Delta f)$. **B**: Each population has a slow adaptation a_k on a timescale τ_a with strength g, recurrent excitation e_k on an intermediate timescale τ_e with strength β_e and an independent noise source χ_k with strength γ. Slow synaptic depression d_k on the recurrent excitation e_k is not shown. Recurrent inhibition i_k with strength $C_i(0)$ is instantaneous allowing for the simplification $i_k = r_k$. See Eq (1) for an illustrative single-unit equation and *Model equations and details* for the full model (Eq (4)). **C**: The Δf-dependent profiles for the input spread $w(\Delta f)$ (exponential decay (Eq (7))) and lateral inhibition $C_i(\Delta f)$ (Gaussian decay (Eq (6)) for i_{lcl} or constant β_i for i_{gbl}). **D**: Sigmoidal firing rate function $F(u)$ (Eq (5)) with maximal slope $k_F/4$ at the threshold θ_F.

doi:10.1371/journal.pcbi.1004555.g002

intermediate responses to both tones at the location AB and a full response to the B tones at the location B.

In the first part of the modeling results the slow synaptic depression variable d_{AB} is frozen at 1 and inhibition is localized in Δf (see corresponding $C_i(\Delta f)$ curve i_{lcl} in Fig 2C). The symbols e_{fix} and i_{lcl} are used to make these mechanistic distinctions. In the second part of the modeling results there is a dynamic slow decay d_{AB} on the excitation and inhibition is global in Δf (see corresponding $C_i(\Delta f)$ curve i_{gbl} in Fig 2C). The symbols e_{dyn} and i_{gbl} are used to make these
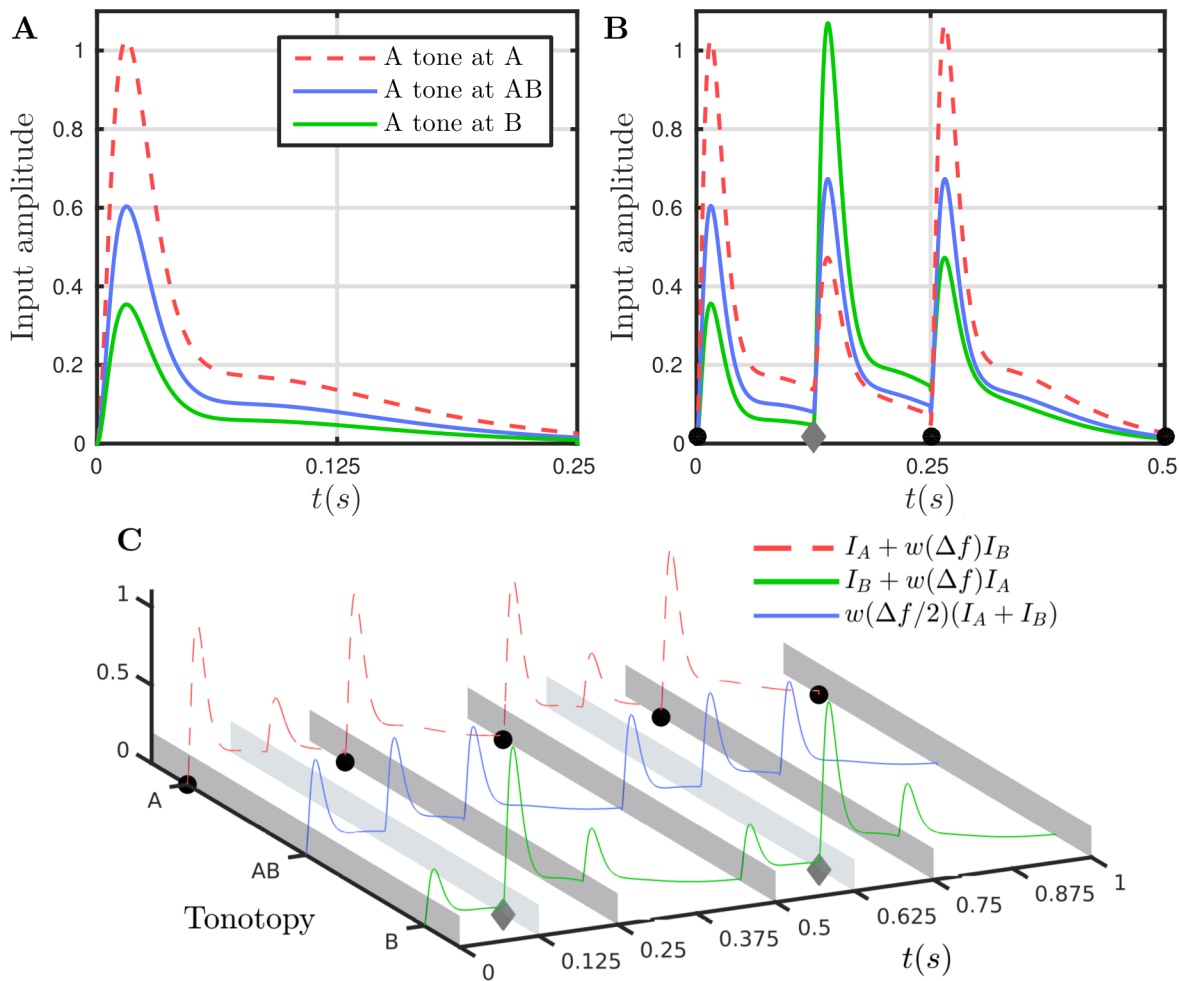
Fig 3. Model inputs. A: Input time courses are represented by double alpha functions (see *Model equations and details* and Eq (8)) that capture the onset and plateau characteristics of A1-responses from [28]. For a single 125 ms tone of frequency A less input will arrive at locations AB and B than at A as described by Eq (7) and plotted here for $\Delta f = 4$ [26]. **B**: Inputs (see legend in C) to the respective populations r_A, r_B and r_{AB} for an ABA- triplet of 0.5 s (tone duration and post-triplet silence "-" of 125 ms, i.e. $PR = 8$ Hz). Tone onsets: black circle for A-tone, gray diamond for B-tone. **C**: As B with curves distributed across the model's tonotopy. The A-tone input is full amplitude at the A location, less at the AB location and further less at the B location, correspondingly for the B-tone input.

mechanistic distinctions. A comparison between the cases with (e_{fix}, i_{lcl}) and (e_{dyn}, i_{gbl}) is made in *Refined model better captures experimental results*.

The full set of fifteen model equations (three principal equations like Eq (1) and twelve simple equations for the e_k-, a_k-, d_k- and χ_k-variables) are given in *Model equations and details* along with the full expressions for F, $w(\Delta f)$, $C_i(\Delta f)$, I_A and I_B. All model parameter values are given in Table 1.

Encoding of percepts and dynamics of alternations: model

We first discuss the output from individual simulations of the model and illustrate how the model's firing rate variables can encode the competing percepts for comparison with experimental data. Fig 4 shows time histories for a model simulation, panels A–D represent the first 20 s of a 240 s simulation at $\Delta f = 5$ st and $PR = 8$. These stimulus parameters correspond to the ambiguous region where perception is bistable and regular alternations take place [19]. Panel

Table 1. Model parameters as defined in *Neuromechanistic model of auditory bistability* and forming part of the general model equations given in Eq (4).

Parameter	Description	Model (e_{fix}, i_{lcl})	Model (e_{dyn}, i_{gbl})
Firing rate:			
θ_F	Sigmoid threshold	0.2	unchanged
k_F	Sigmoid slope	12	unchanged
Input:			
Λ_2	Input plateau fraction	1/6	unchanged
α_1	Input impulse rise time	15 ms	unchanged
α_2	Input plateau stay time	82.5 ms	unchanged
I_P	Input amplitude	0.525	0.47
σ_P	Lateral input decay	8 st	8.5 st
Intrinsic:			
g	Adaptation strength	0.065	unchanged
γ	Noise strength	0.075	unchanged
β_i	Inhibition strength	0.3	unchanged
σ_i	Inhibition decay constant	10 st	$\sigma_i \to \infty$ (global)
β_e	Recurrent excitation strength	0.7	0.85
κ	Excitation decay strength	0	0.25
Timescales:			
τ_r	Cortical timescale	10 ms	unchanged
τ_a	Adaptation timescale	1.4 s	unchanged
τ_e	NMDA excitation timescale	70 ms	unchanged
τ_X	Noise timescale	100 ms	unchanged
τ_d	Excitation decay timescale	N/A	3 s

doi:10.1371/journal.pcbi.1004555.t001

A shows the primary population's variables evolving on the cortical timescale τ_r. When central unit r_{AB} is active (e.g. during the first 3 s), the peripheral units r_A and r_B are suppressed. During this time the adaptation a_{AB} builds up until the peripheral units become active, see panel C. Conversely, when either or both of the peripheral units r_A and r_B are active, the central unit r_{AB} is suppressed. When r_{AB} is suppressed, typically one of r_A or r_B will be more active, reflecting the fact that either the A tones or the B tones are in the foreground during the segregated percept. The possibility of considering perception as tristable or multistable is discussed further in the discussion. The NMDA-like excitation variables, evolving on a slower timescale τ_e than the cortical timescale τ_r, shown in panel B are effectively a temporal smoothing of the cortical variables in panel A. Note that the response transients for individual inputs seen in panel A (look at the sub-threshold activity) are integrated across several inputs in panel B (seen most clearly around 10 s). In our model, activation of the peripheral units encodes the segregated percept, while activation of the central unit encodes the integrated percept. This classification of integrated relies on an implicit assumption of a stimulus featuring a combination of A and B tones where activation of the central unit reflects a linking of the two sequences. An AND-like computation (e.g. classify integrated as when As and Bs are present AND central unit is active) could be implemented to distinguish between an ABA- stimulus and an unambiguous sequence CCC- where C = (A + B)/2. Moreover, the CCC- input should not be confused with integration since the central unit would suppress the peripheral units indefinitley and no alternations would take place.

From a practical point of view, our criteria for classifying the integrated and segregated percepts are based, as follows, on the time courses of the model output plotted in Fig 4. Vertical
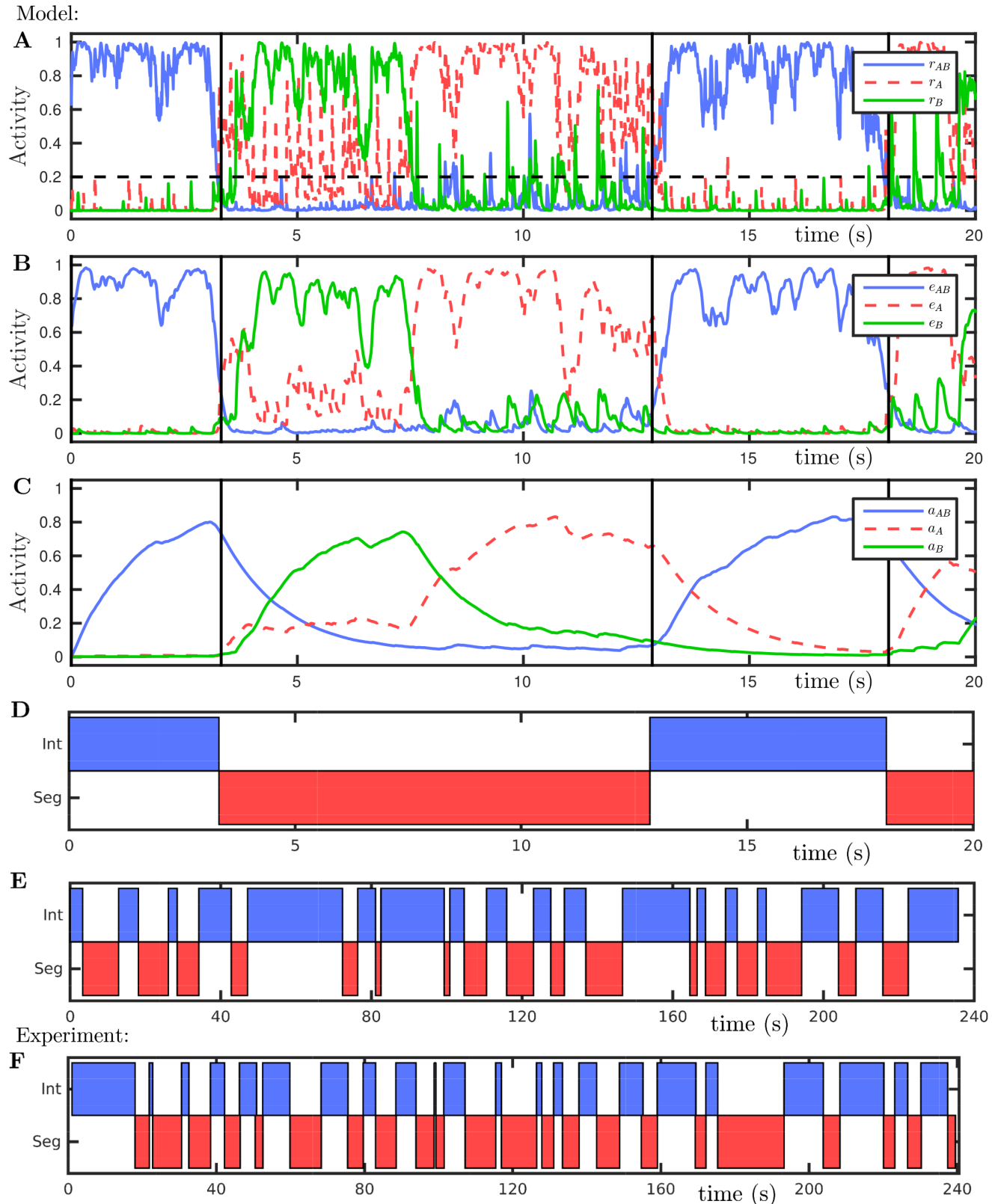
Model:

Fig 4. Time courses of model responses (A–C), predicted percepts (D–E), and an example of perceptual reports from our psychoacoustic experiments (F). A: Population firing rate time courses with $\Delta f = 5$ st and $PR = 8$ Hz. The firing rate function threshold θ_F is a horizontal dashed black line.

Abrupt perceptual switches are seen when AB firing decreases/increases drastically (vertical black lines); see text for the exact criterion for a switch. **B**: As in A, here for the synaptic excitation variables. **C**: As in A for the adaptation variables. **D**: Percept as encoded from A, see text. **E**: Encoded percept for the full 240 s simulation; panels A–D show only first 20 s. **F**: Time course of continuous percept reporting in psychoacoustic experimental for a 240 s trial at $\Delta f = 5$ st.

black lines in panels A–C indicate switches in dominance between the central unit r_{AB} and the peripheral units r_A and r_B. These switches can be computed algorithmically in the following way. First the output variables r_{AB}, r_A and r_B are processed with a temporal moving average filter with 50 ms width to obtain smoothed readouts $\widetilde{r_{AB}}$, $\widetilde{r_A}$ and $\widetilde{r_B}$, these look similar to the excitation variable traces shown in panel B. The output is encoded as integrated when $\widetilde{r_{AB}} > (\widetilde{r_A} + \widetilde{r_B})/2$ and segregated otherwise. The mutual inhibition between the populations gives rise to sharp well defined crossovers with steep gradient at points where $\widetilde{r_{AB}} = (\widetilde{r_A} + \widetilde{r_B})/2$. In this way the model output is encoded as integrated when r_{AB} is dominant and segregated when r_A and r_B are dominant, as plotted in panel D (first 20 s) and panel E (full 240 s). This allows for a direct comparison between the model simulations and perceptual reports from experiment, an example of which is shown in panel F; the experiment will be discussed in *Parametric dependence of perceptual dominance: experiment* and a full comparison with the model made in *Refined model better captures experimental results*.

Statistics of dominance durations: model

Percept durations for perceptually bistable stimuli have been shown to be fit well by gamma or log-normal distributions, for example, see [39–41]. As has become standard in the analysis of percept durations for bistable stimuli, the switching times are normalized by the mean for each percept type (integrated or segregated) [40]. The coefficient of variation (CV), which is the ratio of the standard deviation s with the mean $\bar{x}$ ($CV = \frac{s}{\bar{x}}$), is used as a measure of variability in the percept durations, where smaller values indicate more adaptation-driven alternations, and larger values more noise-driven alternations [34]. During the design of the model some pilot data was used to constrain the model, but informality in its collection preclude its publication. Nevertheless, we do make a comparison with data from the final experiment as reported in detail *Parametric dependence of perceptual dominance: experiment*, which were collected after the development and tuning of model parameters.

In order to look at the distribution of perceptual durations in the model, 1000 durations were randomly sampled from 2225 durations computed from a total of 50 4-minute simulations at $\Delta f = 5$ st and $PR = 8$ Hz. Model parameters controlling the balance between adaptation and noise were constrained to match pilot data collected from 6 subjects at $\Delta f = 5$ and $PR \approx 8$, which showed a mean duration of 4.3 s, a CV of 0.73 and a fit to a log-normal distribution (but not a gamma distribution). After parameter tuning, the model produced a mean duration of 5.1 s and the distribution has a CV of 0.72. These model results were consistent with our pilot data (the mean duration is longer but matching the CV was considered more important, given that durations are normalized for most of the analysis in this paper). Furthermore, model data showed a fit to a log-normal, but not a gamma distribution, see Fig 5A (details of the statistical tests for fits to standard distribution are given in *Comparison to standard statistical distributions*).

We make a comparison with the distributions from subsequent experiments (reported in full in *Parametric dependence of perceptual dominance: experiment*) where, similarly, 1000 durations were randomly sampled from 1575 durations collected across a total of 45 4-minute trials (3 repetitions with $N = 15$ subjects) at $\Delta f = 5$ st and $PR = 8$ Hz. The mean duration in the experiment was 7.7s and the distribution has a CV of 0.83. The data is best fit by a log-normal distribution, but not by a gamma distribution, see Fig 5B. A similar CV and fit to log-normal but not gamma
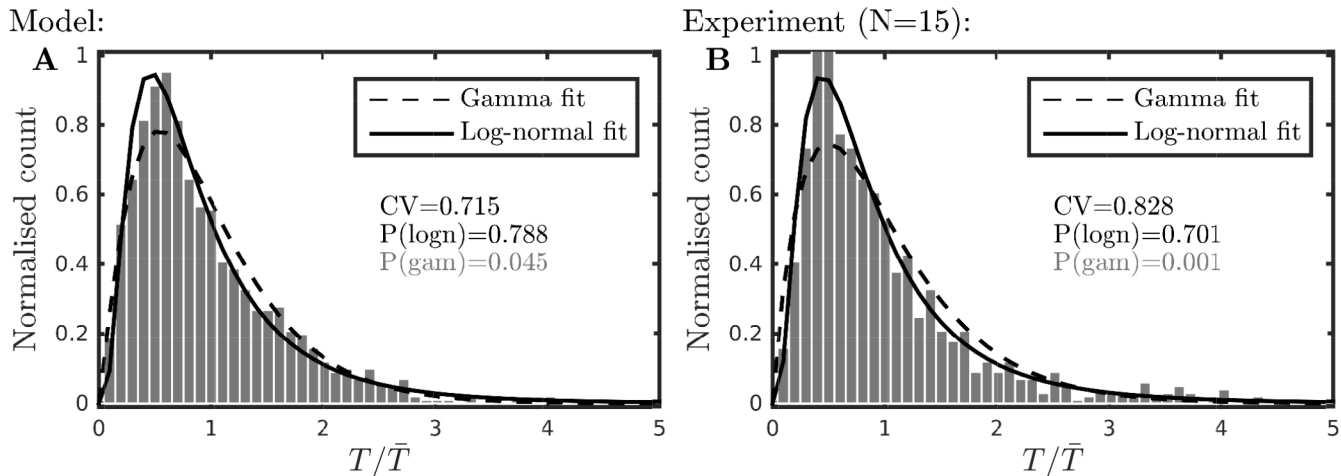
Fig 5. Statistics of dominance durations. A: Histogram of 1000 durations from model simulations at $\Delta f = 5$ combined across perceptual type after normalising by the mean, see text. Curves show best-fit by gamma and log-normal distributions, P-values from one-way KS test are shown (in gray if the distribution can be rejected at the 0.05 significant level). **B**: As in A, here for the experimental condition $\Delta f = 5$; normalized data combined across subjects.

doi:10.1371/journal.pcbi.1004555.g005

was obtained for the other Δf conditions. A two-way KS test comparing the normalized distributions for the model and experiment (plotted in Fig 5A and 5B) showed that the data could be drawn from the same underlying distribution ($P = 0.19$). Although there is a difference of 0.11 in CV values, this statistical test shows that the model has captured the dynamics of alternations in the experiment accurately. In the literature, mean durations across subjects ranging from ≈ 4 s to ≈ 9 s were reported in [19] with equivalent stimulus parameters. To our knowledge, there are no existing studies reporting the coefficient of variation in auditory bistabilty, but in the literature on bistable visual perception values for CV have been reported ranging from 0.48 to 0.67 across several paradigms [42] ($N = 5$–11 subjects) and from $= 0.55$ to 0.70 in binocular rivalry [40] ($N = 3$ subjects). The large CV value (> 0.7) in our auditory bistability experiment suggests that alternations, although driven by a combination of adaptation and noise, are more noise driven than the visual paradigms for which the CV has been reported [40, 42].

Perceptual organization for stimulus parameters: model

The experiments in [2] used wide ranges of presentation rate *PR* and frequency difference Δf to investigate how the predominance of the different interpretations can change. Three qualitative regions were mapped out in van Noorden's experiments: at low Δf integration is dominant, at large Δf segregation is dominant and at intermediate Δf perception is ambiguous. Fig 6A is a cartoon of the organization based upon [2]. Note that there is a further dependence of the boundaries between the regions on *PR*. The experiments in [2] focused only on the first transition and involved either dynamically varying stimuli or the deployment of the subject's attention. The model we present specifically captures the dynamics of regular alternations after the initial transition, has a fixed stimulus and does not account for attention. Nevertheless, as a starting assumption, the same qualitative organization in terms of Δf and *PR* is assumed for the model. Model parameters controlling the spread of input and lateral inhibitory connections, defined by the functions plotted in Fig 2C, were tuned to obtain the appropriate organization. In particular, it was necessary to include lateral inhibition $C_i(\Delta f)$ that decays with Δf as indicated by i_{lcl}.

In order to define regions that are predominantly integrated, segregated or ambiguous we look at the total proportion of time segregated averaged over twelve 4-minute simulations at
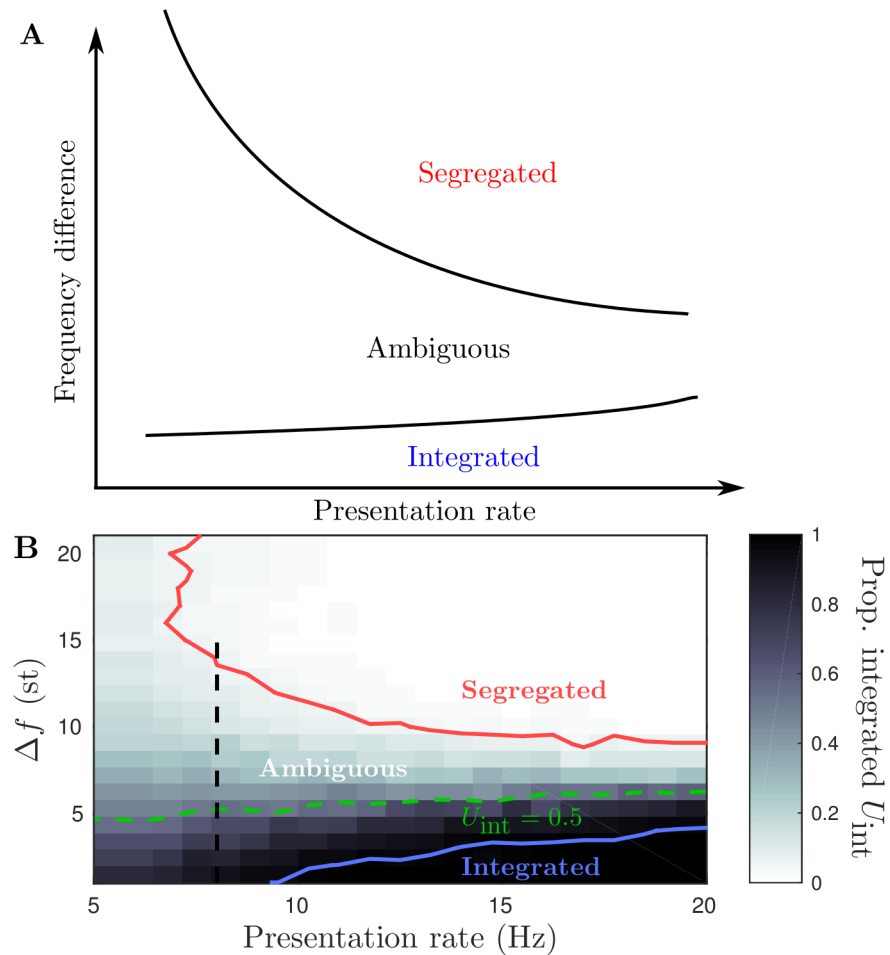
Fig 6. Perceptual organization for stimulus parameters. A: Schematic diagram of the perceptual regions in terms of presentation rate and frequency difference, redrawn after [2]. **B**: Grayscale map of proportion (of time) integrated U_{int} (see Eq (2)), segregated region is above red contour at $U_{int} = 0.05$, integrated region is below blue contour at $U_{int} = 0.95$, ambiguous region lies in between with equidominance at $U_{int} = 0.5$ along the dashed green contour. Vertical dashed line at $PR = 8$ corresponds to the frequency difference sweep used later in Figs 8 and 9.

points on a 21×21 grid with $PR \in [5, 20]$ and $\Delta f \in [1, 22]$. The proportion of time integrated U_{int} is defined by

$$U_{int} = \frac{V^{int}}{V^{seg} + V^{int}}, \qquad (2)$$

where V^{int} is the total time that integrated is dominant (similarly V^{seg} for segregated) across the twelve simulations at given $(PR, \Delta f)$-values.

Fig 6B shows a grayscale map of U_{int} across the $(PR, \Delta f)$-plane. We specify the integrated region by $U_{int} > 0.95$, the ambiguous region by $0.05 < U_{int} < 0.95$ and the segregated region by $U_{int} < 0.05$. The 0.05 and 0.95 contours are plotted in Fig 6B. The model produces the correct organization with a predominance of integrated for small Δf, segregated for large Δf and ambiguous at intermediate ranges. Furthermore, the ambiguous region's Δf-range contracts at larger presentation rates, consistent with [2].

Parametric dependence of perceptual dominance: model

Before presenting the model's predicted behavior for the dependence of percept durations varying a stimulus parameter Δf, we consider two different scenarios for a general experiment on bistable perception. The mean durations of two competing percepts are T_1 and T_2. Suppose that S is a parameter affecting the proportion of time P_1 that percept 1 is dominant, and that P_1 is close to 1 when S is small and decreases monotonically through 0.5 as S is increased. In Fig 7 panels A and D show this general relationship between P_1 and S. When $S = S_{eq}$ the percepts are dominant for an equal proportion of time ($P_1 = 0.5$). We call S_{eq} the *equidominance point*, where $T_1 = T_2$. Panels B and E illustrate two different ways that T_1 and T_2 can vary to obtain this general relationship for P_1. In the upper panels, the stimulus affects the weaker percept more, that is, either side of S_{eq} the percept with shorter duration (the *weaker percept*) changes more with S. For decreasing $S < S_{eq}$, T_2 decreases more than T_1 increases and for increasing $S > S_{eq}$, T_1 decreases more than T_2 increases. In the lower panels, the stimulus affects the stronger percept more, that is, for decreasing $S < S_{eq}$, T_1 increases more than T_2 decreases and for increasing $S > S_{eq}$, T_2 increases more than T_1 decreases.

We further distinguish between these three cases by defining η as the normalized total duration above equidominance

$$\eta(S) = \frac{T_1(S) + T_2(S) - 2T_{eq}}{T_{eq}}. \tag{3}$$

This measure is negative when the weaker percept duration is affected more and positive when the stronger percept duration is affected more; see Fig 7C and 7F. Equivalently, one could show that the overall rate of alternation $1/\eta$ is minimal at equidominance when the weaker is affected more or maximal at equidominance when the stronger is affected more. The latter case was shown to be consistent with three visual bistable paradigms in [38].

We now make a comparison between the model and the scenarios proposed above. In particular, we are interested to see if the model predicts behavior that is consistent with gLPII, where the stronger percept is affected more than the weaker either side of equidominance. We recall that the model was set up and constrained to match pilot data for the distribution of percept durations and the van Noorden organization, as discussed in the preceding sections. The model's ability to account for the available data provides confidence in its predictive power. Here we vary Δf, a parameter that affects the proportion of time integrated in a way consistent with the examples from Fig 7, see Fig 8A. Fig 8B shows the normalized mean durations as a function of Δf, where the durations for integrated are longer at small Δf and the durations for segregated are longer for large Δf, with equidominance reached at $\Delta f = 5$. The normalization of durations, analogous to that used for the experimental data presented later, is explained in *Computation of normalized durations*. On each side of equidominance the upper branches increase more than the lower branches decrease, which is consistent with Fig 7E and Δf affecting the stronger percept more than the weaker. This is further illustrated in Fig 7F, where η is positive either side of equidominance. We also note an asymmetry, with a sharper increase of the integrated durations for $\Delta f < 5$ than in the segregated durations for $\Delta f > 5$. The predicted behavior from the model is consistent with gLPII [38].

Parametric dependence of perceptual dominance: experiment

We carried out behavioral experiments to characterize the effect of varying Δf on the balance between the durations of the integrated and segregated percepts, allowing for a comparison with the model prediction and gLPII. As explained below we consider the durations of percepts
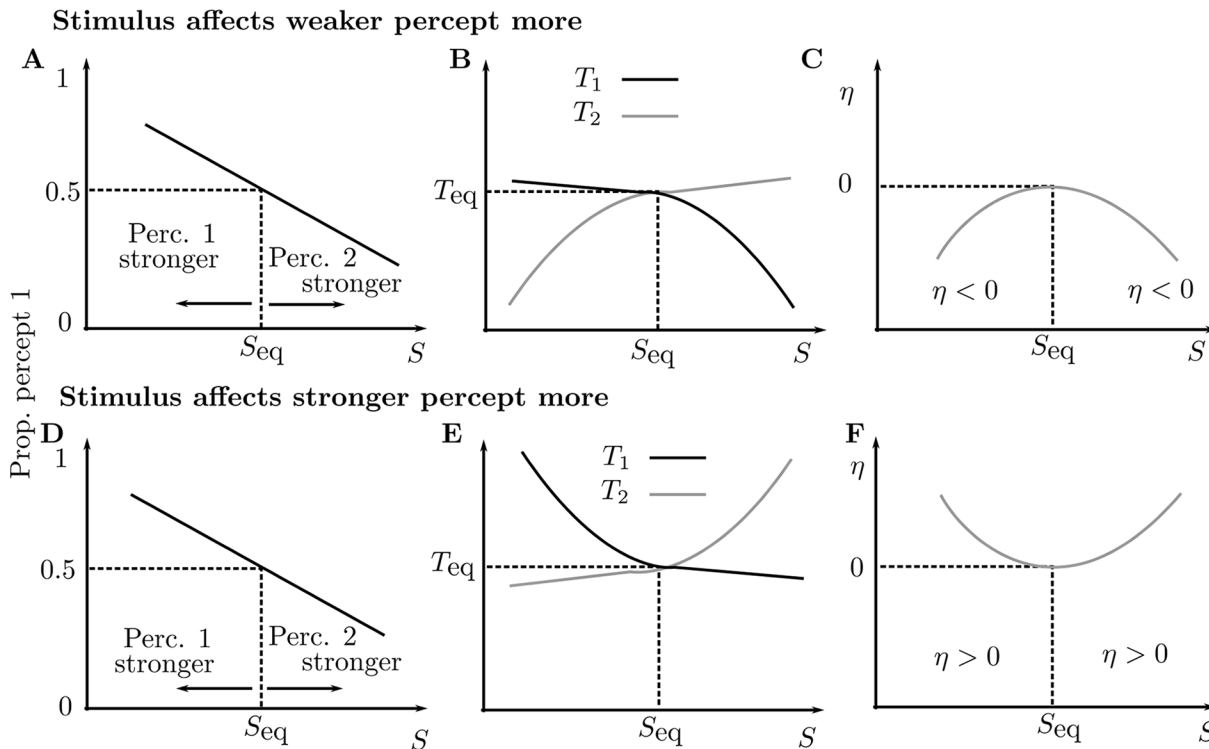
Fig 7. Scenarios for parametric dependence of perceptual dominance. Schematic diagrams illustrate how the mean percept durations may change as a stimulus parameter S is varied and dominance shifts gradually from percept 1 to percept 2. The upper row illustrates the weaker percept being affected more and the lower row the stronger percept being affected more. **A,D**: Proportion of time when percept 1 is dominant decreases monotonically through equidominance (0.5) (dashed lines) in both scenarios. When $S < S_{eq}$ percept 1 is stronger, when $S > S_{eq}$ percept 2 is stronger. **B,E**: Percept durations are equal ($T_1 = T_2 = T_{eq}$) at equidominance ($S = S_{eq}$) in both scenarios. When the weaker percept is affected more the lower branches decrease more on either side of equidominance (**B**). When the stronger percept is affected more the upper branches increase more on either side of equidominance (**E**). **C,F**: The measure η (defined by Eq (3)) is zero at equidominance ($S = S_{eq}$) for both scenarios. It decreases on either side of equidominance when the weaker percept is affected more (**C**) and increases on either side of equidominance when the stronger percept is affected more (**F**).

doi:10.1371/journal.pcbi.1004555.g007

following the first switch and the durations are normalized for each subject. In Fig 9A–9C solid curves with error bars show the three measures described and illustrated in Fig 7 for the auditory streaming experiment over the eight Δf-values indicated on the x-axis. Fig 9A shows that for increasing Δf the proportion of time integrated decreases monotonically; a one-way repeated measure ANOVA shows a significant effect ($F(7, 77) = 24.656$, $P < 0.001$, Greenhouse-Geisser corrected P-values are reported where appropriate, see *Repeated measures ANOVAs* for further details). Our data show that Δf shifts the balance from integrated at low Δf to segregated at large Δf. Therefore, this experiment varying Δf is suitable for a comparison with our qualitative modeling predictions and gLPII. Fig 9B shows that with increasing Δf mean duration integrated $\bar{T}^{int}$ decreases and mean duration segregated $\bar{T}^{seg}$ increases with a crossover occurring between $\Delta f = 3$ and $\Delta f = 5$; a two-way repeated measure ANOVA shows a significant interaction between percept type and Δf ($F(7, 77) = 16.225$, $P < 0.001$). In the analysis that follows we take the case $\Delta f = 5$ as the equidominance point for the data, where integrated and segregated have approximately equal durations. Visual inspection indicates that either side of $\Delta f = 5$ the upper branches of this X-shaped diagram increase more than the lower branches decrease. The measure η defined in Eq (3) tests this qualitative description. Fig 9C shows that η is positive either side of $\Delta f = 5$ and shows that Δf affects the stronger percept (percept dominant for larger proportion of time) more than the weaker either side of equidominance, compare

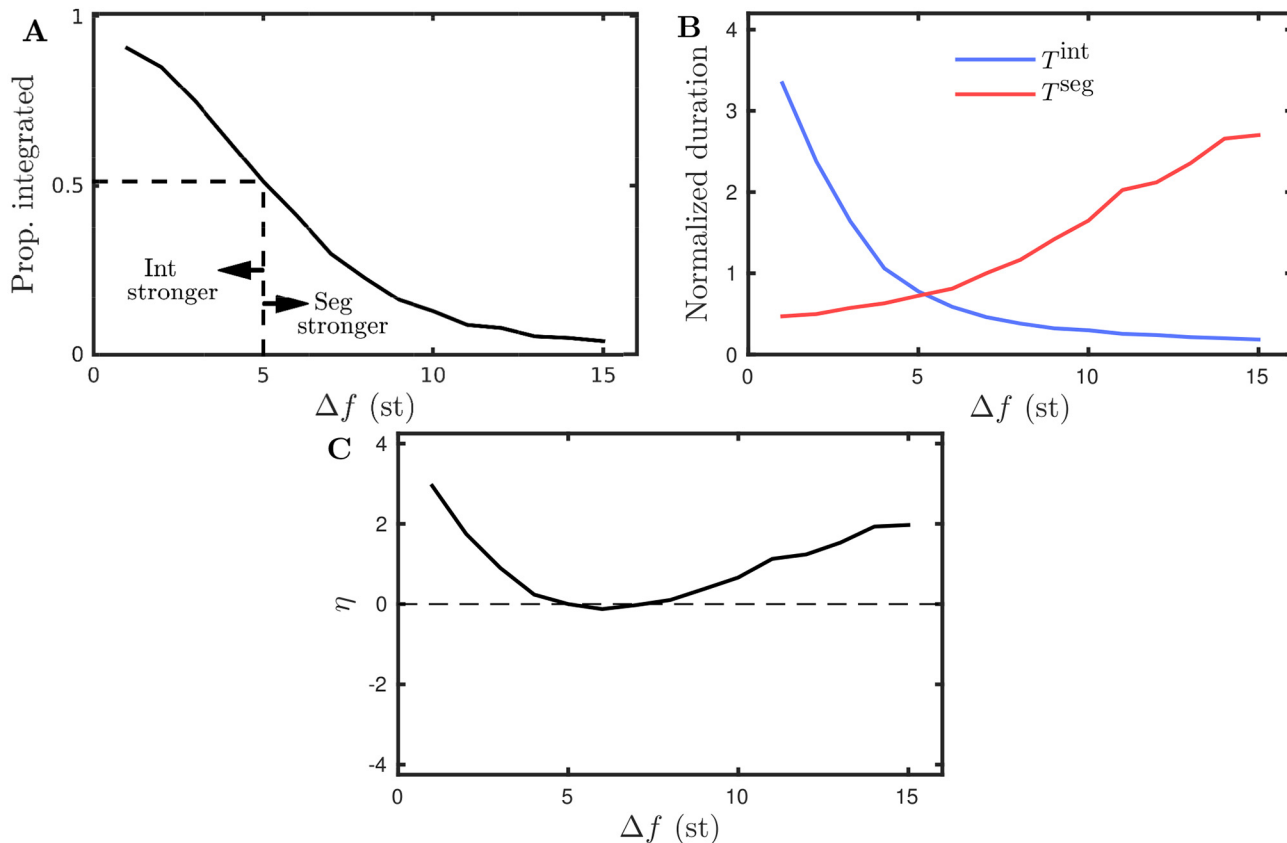**Fig 8. Parametric dependence of perceptual dominance: model prediction with (e_{fix}, i_{lcl}).** **A**: Proportion integrated computed across 50 simulations of 240 s at each of 15 values of $\Delta f \in [1, 15]$ with fixed $PR = 8$ (this parameter range corresponds to the dashed vertical black line in Fig 6A). There is a shift from integrated being stronger to segregated being stronger with equidominance (indicated by dashed lines) at $\Delta f \approx 5$. **B**: Normalized durations integrated and durations segregated with a crossover at $\Delta f \approx 5$ where $T^{int} = T^{seg}$. **C**: The measure η given by Eq (3), which equals 0 at equidominance. The results are consistent with Δf affecting the stronger percept more (Fig 7D–7F).

doi:10.1371/journal.pcbi.1004555.g008

Fig 9A–9C with Fig 7D–7F. A one-way repeated measures ANOVA on η with Δf as factor does not reach significance ($F(7, 77) = 0.878$, $P = 0.463$). Further details and ANOVA tables reported in *Repeated measures ANOVAs*.

We found a substantial variation across individuals in their mean durations. For each subject a global mean duration T^{glob} was computed across all trials and conditions (for summary statistics of T^{glob}, see *First durations and normalization*). One clear outlier, with $T^{glob} = 51.4$ s (consistently longer than the group across all conditions) is excluded from further analysis. The remaining 15 subject's T^{glob}-values span the range 2.6–14.8 s, which shows that some subjects switch much more rapidly than others. In the analysis that follows each subject's durations are normalized by their T^{glob}. Furthermore, to avoid biasing the results towards the faster switchers who record more durations, each subject contributes a single mean integrated $\bar{T}^{int}(\Delta f)$ and mean segregated $\bar{T}^{seg}(\Delta f)$ score at each value of Δf. The mean for each subject at each Δf is taken across all durations from three repetitions. Refer to *Computation of normalized durations* for further details. We note that, before normalization the mean durations integrated and segregated at $\Delta f = 5$ are respectively 7.86 s and 7.65 s, which is within the range reported in [19] where similar stimuli and procedure were used.

Existing studies of bistable (or multistable) auditory perception with long presentations (on the order of minutes) have established that first durations are typically longer than subsequent
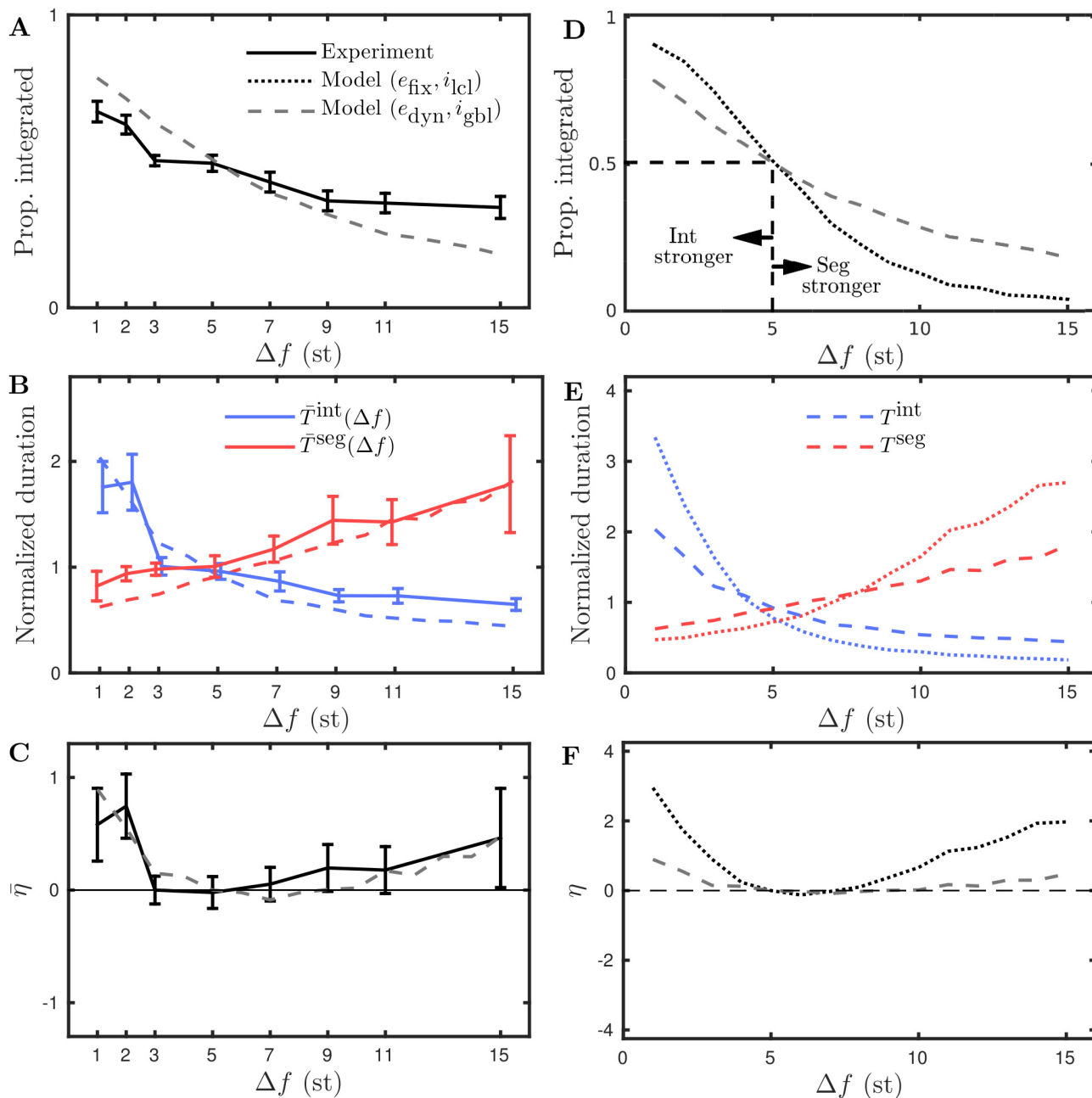
Fig 9. Parametric dependence of perceptual dominance: experiment. Comparison of experimental and computational results. **A–C**: Proportion integrated, durations integrated and segregated, and the measure η plotted against Δf. Experimental data are solid curves with data points at the Δf-values indicated on the x-axis, error bars show standard error of the mean with $N = 15$ subjects except at $\Delta f = 1$ ($N = 13$) and $\Delta f = 15$ ($N = 14$). Model data with dynamic recurrent excitation and global inhibition (e_{dyn}, i_{gbl}) plotted for comparison. **D–F**: Model data with (e_{dyn}, i_{gbl}) plotted with model data for fixed recurrent excitation and spatially localized inhibition (e_{fix}, i_{lcl}) for comparison.

durations and should be treated separately [19, 21, 24]. In line with existing studies, across the range of Δf values tested, we found the first durations to be consistently longer than subsequent durations, with the difference being largest for small Δf and decreasing with Δf, for further details see *First durations and normalization*. In the analysis presented, we study only the subsequent percept durations.

Refined model better captures experimental results

The experimental data plotted in Fig 9A–9C agrees with the following qualitative predictions from the model that were illustrated in Fig 8A–8C (replotted in Fig 9D–9F as dotted curves, (e_{fix}, i_{lcl}) in the figure legend). Proportion integrated decreases monotonically with Δf. Normalized durations integrated $\bar{T}^{int}$ and segregated $\bar{T}^{seg}$ form an X diagram with crossover at equidominance $\Delta f = 5$ with the stronger percept (upper branches) increasing more than the weaker percept (lower branches) decrease either side of $\Delta f = 5$. Accordingly the measure η is zero at $\Delta f = 5$ and increases either side of $\Delta f = 5$.

There are quantitative differences between the experiment and the model predictions from Fig 8A–8C. The model predicted a larger range of variation of proportion integrated than the experiment. The model predicted larger increases in duration for the stronger percept either side of equidominance and also larger decreases in the duration of the weaker percept either side of equidominance. Overall this results in the model predicting a stronger effect on η than was observed in the experiments.

Two mechanistic changes were made to the model that dramatically improve the quantitative fit between model and experiments, but do not alter the basic qualitative predictions. The first mechanistic change is a shift from localized inhibition that decays with Δf to global inhibition independent of Δf (compare i_{lcl} and i_{gbl} curves in Fig 2C). This change along with a reduction in input amplitude I_p from 0.525 to 0.425 improved the agreement between the upper branches in Fig 9B. The second mechanistic change introduces a minimum timescale for perceptual dominance immediately following a switch, which improves the agreement between the lower branches in Fig 9B. This is achieved by an increase in the strength of recurrent excitation whilst also introducing a slow synaptic depression. For further details of this extension to the model, see *Model equations and details*. We refer to this updated version of the model with dynamic recurrent excitation and global inhibition as (e_{dyn}, i_{gbl}). A direct comparison is made with the experimental data in Fig 9A–9C and with the earlier version of the model (e_{fix}, i_{lcl}) in Fig 9D–9F.

We note that the attained level of agreement between the model and experiment is not contingent on the normalization of percept durations. The normalization served to remove variability within subjects. To better see the effect of the normalization compare Fig 9B with non-normalized durations as given in *Computation of normalized durations*.

Further predictions for constrained model

After introducing modifications that allow the model to account for our experimental data, we can reevaluate the perceptual dominance assumption based on the van Noorden diagram. For the model with (e_{dyn}, i_{gbl}), Fig 10A shows a grayscale map of proportion integrated over a range of PR and Δf-values. Solid curves show 95% boundary (bottom right corner), outside which integrated is considered as dominant and the 5% boundary (top right corner), outside which integrated is considered as dominant. In the model with (e_{dyn}, i_{gbl}), the qualitative organization is the same as for the model with (e_{fix}, i_{lcl}), but the regions of perceptual dominance are much smaller; compare Fig 10A with Fig 6B. The model with (e_{dyn}, i_{gbl}) (i.e. constrained by our experimental data) produces alternations over a much broader range of Δf and PR than could be expected with the van Noorden perceptual dominance assumption.

The model predicts a stronger effect of gLPII at higher presentation rates. Fig 10B and 10C show durations integrated and segregated plotted over a range of Δf- and PR-values. This model data has not been normalized, which will allow for future quantitative comparison with the predicted behavior. Dashed curves correspond to $PR = 8$ Hz, as plotted previously in Fig 9, and solid curves at larger $PR = 15$ Hz. The results show a modest increase in durations at
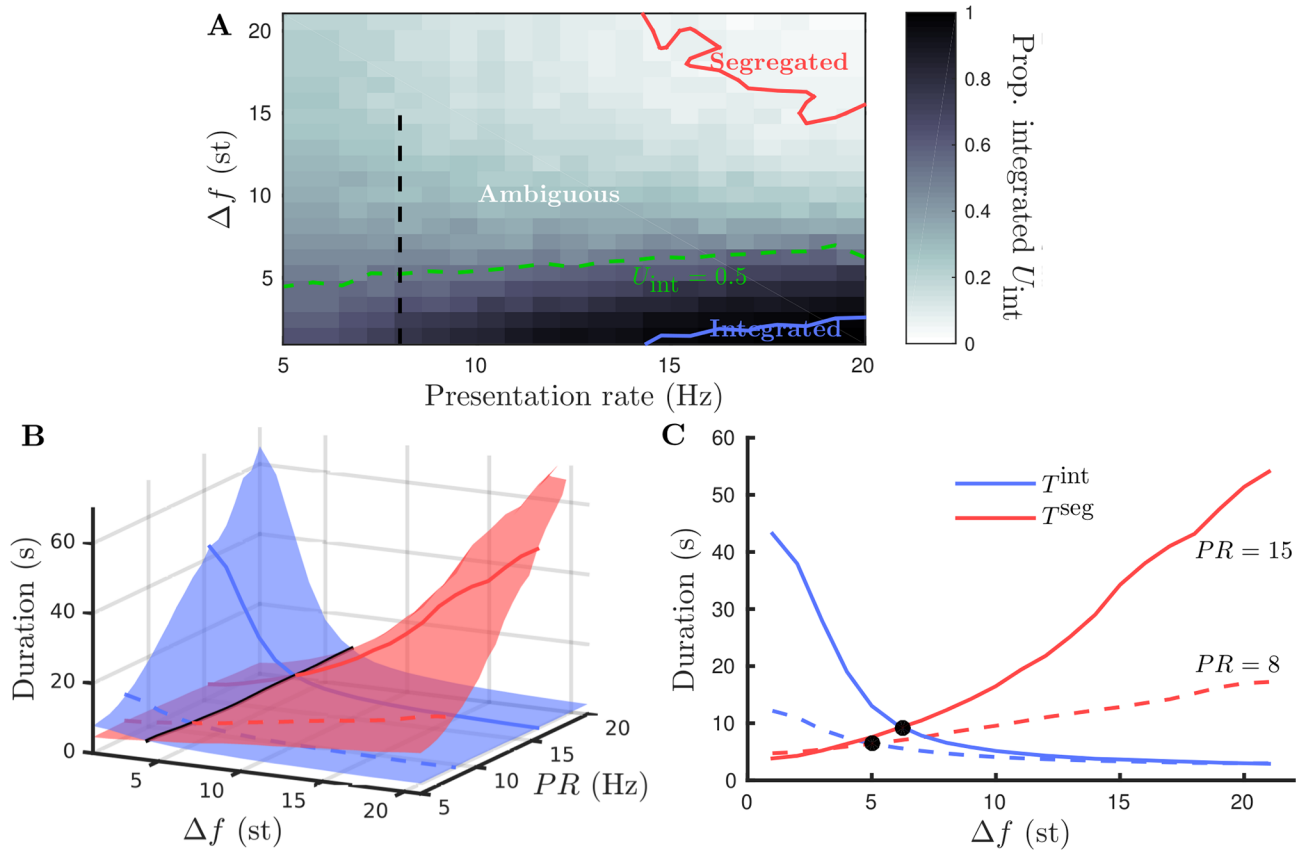
Fig 10. Perceptual organization for stimulus parameters in model (e_{dyn}, i_{gbl}). A: Proportion integrated U_{int} varying Δf and PR plotted as a grayscale map. Solid are contours at $U_{int} = 0.05$ (red) and $U_{int} = 0.95$ (blue) demarcating regions where segregated and integrated are considered dominant, respectively. A dashed green contour in the ambiguous region indicates equidominance $U_{int} = 0.5$. Vertical dashed line at $PR = 8$ corresponds to the frequency difference sweep used in Figs 8 and 9. **B**: Surface plots of mean duration integrated and mean duration segregated (not normalized) across the same range as A. Black curve is the intersection of the two surfaces (equidominance). Dashed and solid curves indicate the cross sections plotted in C at $PR = 5$ and $PR = 15$, respectively. **C**: Durations integrated and segregated varying Δf for fixed PR as indicated; in each case the equidominance point is a black dot.

equidominance and a slight upward shift in Δf of the equidominance point with increasing PR. The weaker percept durations (see lower branches in panel C) are relatively unaffected by an increase in PR. The bias towards affecting the stronger percept durations (see upper branches in panel C) increases further at larger PR-values.

The most relevant comparison with published experiments could be with [24]. Results separating first and subsequent post build up durations showed a dramatic increase in mean duration (combined across percept types) for large Δf and PR, and our model predicts similar behavior [24]. Values of Δf below 4 st were not tested in [24] to see if a similar trend predict persists for large PR and small Δf. Where integrated and segregated durations are reported separately in [24], no distinction is made between first and subsequent durations, precluding a direct comparison. Another study [21] with minutes-long presentations and a wide ranges of Δf and PR tested did not distinguish between percept types, thus also precluding a direct comparison.

Discussion

We developed and presented a neuromechanistic model of bistable auditory perception for the streaming paradigm of ABA- triplet stimuli. Our firing rate model accounts for the dynamics of alternations after the build up of streaming, matching the statistics from our psyshoacoustic

experiments. The model's architecture incorporates the tonotopic organization of auditory cortex. We suppose that the neuronal computation and our model's units reside downstream from primary auditory cortex (A1); inputs mimic the tonotopic spread of activity in A1 and incorporate the transient (onset prominent) firing patterns known from physiological experiments [26, 28]. Alternations arise through the interplay of mutual inhibition, noise, and slow adaptation. This combination captures the characteristics of exclusivity, randomness, and inevitability as described by [35] and demonstrated for auditory bistability in [19]. Our model has some novel features: It incorporates NMDA-like recurrent excitation that supports some neuronal memory, enabling carry-over tone-to-tone within a triplet and across the tone gaps between triplets; percepts and dominance emerge naturally from the tonotopic architecture rather than having percepts pre-assigned to units that compete for dominance as in many previous models for perceptual bistability [33, 34]. With modeling and experiments we have confirmed that the auditory streaming paradigm conforms to a generalization of Levelt's proposition II (gLPII): with tone frequency difference, Δf, as a control parameter we find that on either side of equidominance the stronger percept durations are affected more than the weaker percept durations, in agreement with related experiments in bistable visual perception [38].

Model of neuronal competition beyond A1

The neuronal circuits for competition and perceptual encoding are assumed to be located in downstream and to receive inputs from A1. Feature dependent activity and temporal forward masking can account for several aspects of auditory streaming build up as demonstrated in recordings from pre-attentive areas across a range of species, for example, studies of primate A1 [26–28], European starling forebrain [29] and guinea pig cochlear nucleus [30]. Spiking patterns in the European starling forebrain (assumed homologous to A1) have been shown to reflect stimulus characteristics but not perceptual decision [31]. Another study in ferrets has also demonstrated a disconnect between perceptual behavior in humans and the encoding of stimulus features in ferret A1 [43]. Recent fMRI studies (human subjects) provide evidence that links activation in both A1 and sub-cortical areas to perceptual reversals for streaming experiments [22, 44], however, the associated low temporal resolution cannot preclude this activation being a result of top-down modulation from higher auditory centers or other mixed/non-auditory areas. Behavioral and MEG studies (human subjects) with long stimulus presentations of streaming stimuli [18] and an informational masking paradigm [45] provide evidence that the auditory core encodes stimuli features independent of perceptual awareness, whilst perception is represented in longer latency responses associated with later stages of processing in non-primary auditory cortex [45]. Studies with human fMRI have shown that tonotopic organization extends beyond core regions (including A1) to some secondary auditory areas, see [46, 47] and [48] for a review. In our study we explored two possibilities, that the lateral inhibition in our competition stage depends on Δf (i.e. tonotopic distance between the units), or that the inhibition is global and independent of Δf (consistent with there being no tonotopy in the competition stage's recurrent connections). Based on the available literature we are not able to pinpoint a more specific location for our model. We have proposed a minimal description of the competition involved in auditory bistability, which likely involves multiple brain areas beyond A1.

Model formulation and fitting of parameters

The reduction of a continuous representation of a feature space to individual units encoding different percepts was first proposed in [32] and simplified models where the competing percepts are assumed have become relatively standard in modeling studies of perceptual bistability

(see, for example, [33, 34, 42, 49]). In these models inputs are percept specific rather than incorporating a realistic feature dependence. In our model, the competing percepts arise from the tonotopic stimulus feature dependence without assuming that abstracted populations encode each percept. This, along with the periodic nature of the inputs distinguishes our work from typical competition models. Most existing computational models of auditory streaming focused on the build up phase and, although some include a hybrid of auditory pathway physiological detail and electrical signal processing [14, 17], they are typically abstract in construction [11, 12, 15, 16]. Both build up and subsequent alternations were modelled in [24] with competition taking place between abstracted units representing patterns formed in the build up phase. The focus on alternations after the build up phase, the neuromechanistic formulation that we propose and the emergence of competing percepts within the tonotopic organization distinguish our work from existing auditory models.

Model parameters controlling the balance between adaptation and noise were tuned to match the statistics of perceptual durations in terms of mean durations, fit to a log-normal distribution and coefficient of variation. The level of agreement obtained demonstrates that the dynamics of auditory bistability were accurately captured by the model. Model parameters controlling the spread of input and range of inhibitory connections in tonotopic space were constrained based on a starting hypothesis that van Noorden's proposed regions of perceptual dominance [2] are also applicable to the perceptual bistability we study here. In order to obtain the van Noorden organization a tonotopic dependence of lateral inhibition within our model's competition stage was included. The assumption of a tonotopic dependence, beyond that inherited from A1 through its inputs, in the lateral recurrent connections in the competition stage is linked to the hypothesis for the van Noorden organization. Below we discuss this assumption in light of our own data whilst highlighting a number of differences between van Noorden's experimental paradigm and the situation considered here. Nevertheless, the reproduction of the correct perceptual organization suggests that stimulus features have been accurately captured along with the dynamics of alternations. On this basis the model was used to predict the relationship between frequency difference Δf and the relative durations for the integrated and segregated percepts. Data from experiments designed to test the model prediction were later used as a more concrete means to constrain the model.

Experiment to test Levelt's II in auditory bistability

We asked whether gLPII extends to auditory bistability. We chose Δf as a control parameter since it affects the proportion of dominance between the percepts [21, 24]. Indeed, our model satisfies gLPII (compare Fig 8 with schematics in Fig 7D–7F), which cannot be generically expected given differences in mechanisms and architecture with the models that show gLPII-like behavior for visual bistability [32, 38, 50]. Satisfaction of gLPII for auditory bistability was a joint prediction from our model and the literature on bistable perception.

Results from our psychophysical experiments also satisfied gLPII (Fig 9A–9C), supporting the predictive power of our basic model. Repeated measures ANOVAs demonstrated a significant effect of varying Δf on proportion of time integrated, a significant interaction between percept type and Δf, but no significant effect of varying Δf on η. Given that η was positive away from equidominance, we can conclude that auditory bistability conforms to gLPII, however, the bias towards affecting the stronger percept is modest. A recent study [24] reported durations of integrated and segregated percepts over a range of Δf-values, but not a sufficient number of conditions were tested at a fixed presentation rate to allow for a direct comparison with gLPII.

In order to improve the quantitative agreement between the model and experiment, we made two mechanistic changes and minor parameter adjustments in the model. Firstly, by

relaxing the hypothesis of conforming to the qualitative van Noorden organization and shifting to global inhibition the overestimation was reduced for durations of the stronger percept, see Fig 9E. The implication is that tonotopic dependence of inhibition is not required at the level where percepts are encoded although localization of inhibition and excitation are present in A1 for conveying stimulus features and dependence on Δf. We searched many different parameter combinations testing the relative balance between the Δf-footprints of inhibition and input but did not identify another means, within the scope of the model's features, to account for the experimental data. In the conception of our model, we assumed that the competition in our model takes place at a stage with tonotopy-dependent inhibition between units (plausible for non-primary auditory areas). However, given that our model better accounts for our experimental findings with (tonotopy independent) global inhibition, it is also plausible that the competition take place in a non-tonotopically organised area of auditory cortex or a non-auditory area downstream of A1 (for example, intraparietal sulcus (IPS) as implicated in an fMRI study of auditory streaming [51]). Although tonotopy is still inherited through the inputs to our competition stage, it does not necessarily play a role in the recurrent synaptic inhibition. Secondly, we increased the strength of recurrent excitation whilst introducing a slow synaptic depression that could prolong a dominance duration beyond a minimum value. The modification of the recurrent excitation mechanism effectively prevents immediate switches away for a unit that has recently become active. This change reduced the underestimation of durations for the weaker percept, see Fig 9E. Self excitation also played a role in reducing the probability of short durations in [24].

van Noorden organization

Regions of temporal coherence (dominance of integrated), fission (dominance of segregated) and ambiguity over ranges of Δf and PR were established in [2] and the resulting van Noorden diagram has become a key point of comparison in much of the research that has followed. However, comparisons to the van Noorden diagram are often made in experiments and models where several features particular to van Noorden's experiments are not take into account. The procedure used by van Noorden included a slowly increasing or decreasing frequency difference, and subjects were instructed to attend to ("hold") a specific interpretation during presentation. Trials were terminated after the first perceptual switch. Although the van Noorden diagram has provided a rough, qualitative overview of perceptual organization, a full comparison with these experiments will await the development and application of a model that can treat early and later phases of streaming, time varying stimulus parameters and that includes some mechanism for attention.

We used the organization of perceptual dominance and ambiguity as described/established by van Noorden [2] as a qualitative guide in the initial determination of our model's parameter values, particularly those that control tonotopic spread of input (from A1) and the tonotopic footprint for the strength of inhibition. The data from our subsequent psychoacoustics experiments (to investigate gLPII) provided a more concrete means to constrain the model. Once constrained by our experimental data, the model predicted that alternations occur over a much broader range of Δf and PR than expected from the original van Noorden diagram. The hypothesis that the post build-up phase of streaming conforms to the van Noorden organization was not satisfied in a qualitative way. Our results support recent experiments that found alternations taking place over the entire range of Δf and PR typically tested in streaming experiments [21]. A further testable model prediction, which should be the subject of future experiments, is that the bias towards affecting the stronger percept durations either side of equidominance increases with presentation rate.

Attention

Could attention be thought of as a top-down modulation of the effective input to neuronal groups whose dominance corresponds to the attended percept? Attention can significantly affect the proportion of dominance during alternations between the competing percepts [19]. Interestingly, although the proportion increased for the attended percept the effect was achieved by decreased mean durations of the weaker (unattended) percept; results were reported only for an equidominance condition (Δf = 5 st). If there is a parallel it would be with the hypothetical case of Fig 7A–7C, where the weaker percept is affected more than the stronger percept. In contrast, our study has shown that stimulus manipulations affected the stronger percept more than the weaker percept on either side of equidominance (Fig 9)—consistent with gLPII (Fig 7D–7F). Our observations suggest an opposite effect in auditory bistability for attention and Δf-manipulations. We conclude that in order to model effects of attention we should consider more than static manipulations of stimulus parameters. Simply increasing the stimulus strength to favor the attended percept will be inadequate. Rather, we suggest that inputs should be dynamic and depend on perceptual state. An exploration of the neural mechanisms underlying attention effects in auditory streaming will be the focus of future work.

Toward a more general modeling framework

In most studies with the ABA- paradigm, subjects report integrated and segregated percepts, but some psychoacoustic studies have documented other interpretations [21, 52–54]. Our model's dynamic responses were categorized with a simple criterion to be either integrated or segregated. The categorization could be extended to include the possibility of the A or B tones being in the foreground within the segregated percept as was considered in recent psycho-acoustic experiments [53]. The extension of phenomenological, percept-based, competition models to tristable perception was considered in [49] for visual motion plaids, and a similar extension for auditory stimuli would be of interest. We did not formulate a neuromechanistic description for distinguishing a wider variety of patterns (as reported in experimentally [54]), which remains an outstanding challenge for modeling auditory perception (see [24] for a proposed algorithmic approach).

The temporal coherence in the relative onsets of the A and B tones in the streaming paradigm can have a significant effect on perception [55]. As has been demonstrated for short stimulus presentations, there is a bias towards grouping of stimulus elements with a common onset even when they separated by large frequency differences [43] (see [56] for an example with harmonicity-based ambiguity where this is not the case). We think our framework could be extended to consider coherence by including another feature dimension for, say, assessing temporal coincidence.

Recent imaging studies have shown changes in activation around the time of auditory perceptual switches not only in auditory areas, but in non-auditory areas such as IPS [51] and cerebellum [23], and in subcortical regions such as the medial geniculate body [23] and inferior colliculus [44]. Although it is likely that subcortical areas play an important role in early stimulus adaptation and build up, their exact role in auditory bistability is still far from clear. The extension of our model to incorporate a hierarchical structure with descending feedback and further subcortical detail would allow for an exploration of whether activity in subcortical areas is a driving force in perceptual alternations or the result of modulation through descending connections.

The present model has been specifically designed around a stimulus involving two sets of tones at a fixed frequency separation. It is formulated with discrete neuronal units that are tuned to tone frequencies A, B or the intermediate (A + B)/2. The amount of input going to

each unit and the inhibition between the units was defined through continuous functions. This description could be directly extended to a continuous representation of tonotopic space that would allow for an arbitrary number of inputs and for inputs with Δf varying over time. Such a representation was used in the modelling study [16] to look at build up and at rising/falling tone sequences perceived as *crossing* or *bouncing* as reported in experiments [2]; elsewhere, the choice of a continuous feature has proved effective in capturing the dynamics of multistable motion perception [57]. A model with a continuous tonotopic representation would provide a basis to look more closely at the experiments of van Noorden involving Δf-time-varying stimuli and bridge the gap with the fixed stimuli considered here.

Conclusion

Our study's foremost contribution is a physiologically-based model of auditory bistability that captures key features of the perceptual alternation dynamics for the ABA- streaming paradigm. Our formulation explicitly incorporates sound-evoked A1 responses as input to a competition network. Perceptual states are identified in the neuronal dynamics rather than pre-assigned to phenomenological units. The inclusion of NMDA-like synaptic dynamics enables temporal binding from one triplet to the next in order to deal with gaps in the periodic sound input. We applied the model to predict the dependence of mean dominance durations on the tone frequency difference Δf and carried out psychoacoustic experiments for comparison. Based on our findings we refined the model to include slow depression of the NMDA excitation and a broadened footprint for inhibition. The quantitative agreement between the model and experiments across several important signatures, including the steady state statistics of percept durations, demonstrates the model's success in capturing the dynamics of auditory bistability.

In parallel with the model development we established in our experiments that generalized Levelt's Proposition II holds for the alternations in auditory bistability that follow the first perceptual switch. We showed that varying Δf away from equidominance increases the duration of the stronger (dominant a larger fraction of time) percept rather than that of the weaker. These findings further establish a further commonality between auditory and visual bistable dynamics [38]. Although the measured effect in favor of increasing the stronger percept duration was small, it was opposite to that of attention as shown in [19] (strongly in favor of decreasing the weaker percept's duration). Our model provides a platform to further investigate these contrasting effects and to explore the mechanisms of attention in auditory streaming.

Materials and methods

Ethics statement

The psychoacoustics experiments were approved by the University Committee on Activities Involving Human Subjects at New York University (IRB #12-8810).

Model equations and details

We consider a three population network—a discrete idealization of a tonotopically organized array. Two populations pool inputs from regions of A1 centred at locations with best frequencies A and B, whilst a third population pools input centred at a tonotopically intermediate location, say (A + B)/2; see Fig 2.

The firing rate variables r_k are indexed by $k = \{A, AB, B\}$ for each population shown in Fig 2A with the associated adaptation a_k and recurrent excitation e_k variables (note that, throughout the paper, the symbol "e" is used exclusively for excitation variables and associated constants whilst the symbol "exp()" is used for the exponential function). The system of first order

differential equations is as follows:

$$
\begin{aligned}
\tau_r \dot{r}_{AB} &= -r_{AB} + F(\beta_e d_{AB} e_{AB} - C_i(0) r_{AB} - C_i(\Delta f/2)(r_A + r_B) \\
&\quad -g a_{AB} + w(\Delta f/2)(I_A + I_B) + \chi_{AB}), \\
\tau_r \dot{r}_A &= -r_A + F(\beta_e d_A e_A - C_i(0) r_A - C_i(\Delta f/2) r_{AB} - C_i(\Delta f) r_B \\
&\quad -g a_A + I_A + w(\Delta f) I_B + \chi_A), \\
\tau_r \dot{r}_B &= -r_B + F(\beta_e d_B e_B - C_i(0) r_B - C_i(\Delta f/2) r_{AB} - C_i(\Delta f) r_A \\
&\quad -g a_B + I_B + w(\Delta f) I_A + \chi_B), \\
\tau_a \dot{a}_{AB} &= -a_{AB} + r_{AB}, \\
\tau_a \dot{a}_A &= -a_A + r_A, \\
\tau_a \dot{a}_B &= -a_B + r_B, \\
\tau_e \dot{e}_{AB} &= -e_{AB} + r_{AB}, \\
\tau_e \dot{e}_A &= -e_A + r_A, \\
\tau_e \dot{e}_B &= -e_B + r_B, \\
\tau_d \dot{d}_{AB} &= -d_{AB} + (1 - \kappa \, r_{AB}), \\
\tau_d \dot{d}_A &= -d_A + (1 - \kappa \, r_A), \\
\tau_d \dot{d}_B &= -d_B + (1 - \kappa \, r_B),
\end{aligned}
\tag{4}
$$

with time constants τ_r (cortical), τ_a (spike frequency adaptation), τ_e (NMDA-excitation) and τ_d (slow depression of excitation). Each population shares a common firing rate function F is given by

$$
F(u) = \frac{1}{1 + \exp\left(k_F(-u + \theta_F)\right)},
\tag{5}
$$

where θ_F is a threshold parameter and k_F is a slope parameter, see Fig 2D. The strength of adaptation is g.

Inhibition is assumed to be instantaneous ($\tau_i = 0$ in $\tau_i \dot{i}_k = -i_k + r_k$ for each k) such that the inhibitory terms in Eq (4) appear as proportional to the r-variables ($i_k = r_k$). Simulations were run (not shown) to verify that this simplification did not have a significant effects on the model's dynamics. In the first phase of model simulations the strength of lateral inhibition is assumed to have Gaussian decay with Δf

$$
C_i(\Delta f) = \beta_i \exp\left(-\frac{\Delta f^2}{2\sigma_i^2}\right),
\tag{6}
$$

where the inhibition strength is β_i with decay constant σ_i; the symbol i_{lcl} is used to refer to this case. In the second phase of the model simulations inhibition is assumed to be global such that $C_i(\Delta f) = \beta_i$ (equivalently $\sigma_i \to \infty$ in Eq (6)); the symbol i_{gbl} is used to refer to this case.

Recurrent excitation is assumed to occur on an intermediate timescale τ_e (representative of NMDA synapses) between τ_r and τ_a. Excitation is assumed to be local only with strength $\beta_e > \beta_i$ such that there is net local excitation for each unit. In the first phase of the modeling results the variables d_k are frozen at $d_k = 1$ ($\kappa = 0$) in (4); the symbol e_{fix} is used to refer to this case. In the second phase of the modeling results the strength of the recurrent excitation is increased from $\beta_e = 0.7$ to $\beta_e = 0.85$ and a slow ($\tau_d = 3$ s) synaptic depression with strength $\kappa = 0.25$ is turned on; the symbol e_{dyn} is used to refer to this case. The effect of increased excitation

strength coupled with a slow synaptic depression is to introduce a minimum active phase (i.e. a minimum percept duration) once a given r_k becomes dominant.

The spread of input is defined via the weighting function

$$w(\Delta f) = I_p \exp\left(\frac{-\Delta f}{\sigma_p}\right) \tag{7}$$

where σ_p is a spatial decay parameter and I_p is the pulse amplitude. The input terms I_A and I_B appearing in Eq (4) are defined by Eq (8). The particular form of the periodic inputs are based on recorded responses from A1 with such ABA- stimuli [28]; see Fig 3. We capture the basic form of these responses with a pair of onset response functions, one with larger amplitude and early rise that captures the initial onset and a second with smaller amplitude and late rise that captures the plateau:

$$I(t) = H(t)\left[\frac{\exp(2)}{\alpha_1^2}t^2\exp\left(\frac{2t}{\alpha_1}\right) + \Lambda_2\frac{\exp(2)}{\alpha_2^2}t^2\exp\left(\frac{2t}{\alpha_2}\right)\right], \tag{8}$$

with plateau amplitude fraction Λ_2 and rise times $\alpha_1 < \alpha_2$; see Fig 3 (thick black curve). The constant term $\frac{\exp(2)}{\alpha_{\{1,2\}}}$ normalises the amplitude at $t = \alpha_{\{1,2\}}$ of the individual onset functions to 1. A standard Heaviside function H ensures no response before an input tone at $t = 0$. Rise times of $\alpha_1 = 15$ ms and $\alpha_2 = 82.5$ ms and an amplitude $\Lambda_2 = 1/6$ were chosen to approximately match the rise time and relative onset-to-plateau ratio observed in [28].

Additive noise is introduced with independent stochastic processes χ_A, χ_B and χ_{AB} added to the inputs of each population as in [34] and [58]. Input noise is modeled as an Ornstien-Uhlenbeck process:

$$\dot{\chi}_k = -\frac{\chi_k}{\tau_d} + \gamma\sqrt{\frac{2}{\tau_X}}\xi_k(t), \tag{9}$$

where $\tau_X = 100$ ms (a standard choice [34, 58]) is the timescale, γ the strength and $\xi(t)$ a white noise process with zero mean. Note these terms appear inside the firing rate function F such that firing rates r_k remain positive and do not exceed 1.

All model parameters defined throughout this section are given in Table 1. The values of θ_F and k_F are standard for neuronal competition models [34, 58]. The input shape parameters Λ_2, α_1 and α_2 are estimated from [28]. In the first modeling phase (e_{fix}, i_{lcl}), the Δf-dependent functions controlling the spread of input (Eq (7)) and decay of lateral inhibition (Eq (6)) were empirically chosen to produce a van-Noorden like organization for the stimulus parameters Δf and PR as shown in Fig 6. The parameters I_p, σ_p and σ_i were chosen accordingly. In the second phase of the modeling (e_{dyn}, i_{gbl}), these were readjusted and a slow synaptic depression introduced (with parameters κ and τ_d) to better account for our experimental data. The strength parameters for adaptation g, noise γ, inhibition β_i and excitation β_e were set to provide a balance between adaptation and noise [34], to produce the correct mean durations and switching statistics to match our pilot data, see Fig 5. The cortical timescale ($\tau_r = 10$ ms) and the spike frequency adaptation timescale ($\tau_a = 1,400$ ms) fall within typical ranges [59]; similar values have been used in competition models [34, 58]. The NMDA-like timescale ($\tau_e = 70$ ms) for excitation is motivated from [60, 61] and similar values have been used in working memory and decision making models [62, 63]. In the second phase of the modeling results, excitation is assumed to decay through slow synaptic depression on a timescale $\tau_d = 3,000$ ms, see below. Slow depression on the order several

seconds has been reported in primary visual cortex [64] and several slow timescales for adaptation, also on the order of several seconds, have been reported in auditory cortex responses [65].

Stimuli and procedure for psychoacoustic experiments

The stimuli consist of repeating 125 ms pure tone ABA- triplets where the '-' indicates a silence also lasting 125 ms; each ABA- sequence is 0.5 s in duration (similarly, 125 ms tones were used in [28] and 120 ms tones were used in [19]). The higher frequency A tones are a variable Δf semitones above the lower frequency B tones. Cosine squared ramps are used with 5 ms rise and fall times. During 4 minute trials the tone sequence is played binaurally through etymotic headphones at 65 dB SPL. Eight Δf conditions are used $\Delta f \in \{1, 2, 3, 5, 7, 9, 11, 15\}$, with each 8-trial block consisting of a single presentation of each condition. The frequencies used for the A and B tones are taken to be semitone intervals between 392 Hz and 932 Hz, a range spanning 15 semitones. Within a trial block a total of 16 frequencies are used, two for each Δf condition, such that no frequency appears twice within the block; see Table 2. In this way, irrespective of the order of presentation of the 8 conditions, no specific frequency is repeated from one presentation to the next so as to avoid any residual effects of adaptation between trials. Furthermore, a minimum 30 s interval between trials (and 180 s after the fourth trial) was used after which subjects could the next trial when ready. The increments of Δf have been chosen to give a good coverage of the range of responses based on trial data.

Sixteen subjects (9 female) with a mean age 24 years took part in the experiment for modest reimbursement. Procedures were in compliance with guidelines for research with human subjects and approved by the University Committee on Activities Involving Human Subjects at New York University. All subjects provided written informed consent. Each subject completed three eight-trial blocks giving a total of 24 trials per subject. An 8 × 8 latin square design was used to determine the order of conditions for each subject and block repetition. Six randomized and unique grids were generated (two blocks of 8 subjects and three repetitions). The blocks were run across two sessions on different days, one block in the first session, two blocks in the second session.

Subjects sat in an acoustically shielded chamber and indicated their perceptual responses with button presses on a keyboard. In a 2AFC task subjects were instructed to report the integrated percept when they heard the A and the B tones together in an alternating or galloping rhythm and the segregated percept when they heard two separate streams, one with only A tones and one only B tones. The percepts were explained to the subjects with auditory and visual illustrations to ensure that the subjects understood the two interpretations and could clearly distinguish between them. Subjects were instructed to passively report their percepts without attempting to hear one perceptual organization over another. Subjects reported their

Table 2. Frequency conditions used.

Δf (st)	A (Hz)	B (Hz)
1	622.3	587.3
2	640.5	570.6
3	659.3	554.4
5	698.5	523.3
7	740.0	493.9
9	784.0	466.2
11	830.6	440.0
15	932.3	392.0

percepts by holding specific keys associated with of the percepts. The state of the two response buttons was recorded with a sampling rate of 100 Hz.

In this paper we considered bistability between integrated and segregated percepts for ABA- triplets, using a 2AFC task in our experiments. In studies where response keys are provided for integrated and segregated and subjects are instructed to press neither key when their responses are "indeterminate" such responses are recorded for a very small fraction of presentation time [19, 24]. Here the task was 2AFC and no instruction for "indeterminate" responses was provided. We chose to use a simplified task given the wide range of Δf-values tested. Furthermore, in order to make a direct comparison with gLPII, it was necessary to impose categorization between only two competing interpretations as done in the three visual paradigms considered in [38].

Durations shorter than 0.5 s (one triplet) were excluded from the analysis. Across all conditions neither key was depressed for 1.4% of total presentation time (including the time immediately after stimulus onset before any key is pressed) and both keys were depressed 0.9% of total presentation time. Times with neither or both keys depressed were at the transitions between the two percepts (any trials where neither or both keys were depressed for more than 10 s were inspected visually). Given the 2AFC task, each percept duration was computed from the button press onset associated with one percept type up to the button press onset of the opposite percept type. The final (incomplete) duration was discarded for each trial.

First durations and normalization

In an initial phase of our analysis, we establish that first durations are longer than subsequent durations across the range of Δf-values tested. At each condition we look at, separated our into first and subsequent, the mean duration taken across both percept types (integrated of segregated), $N = 16$ subjects and 3 repetitions. For the first percept this gives 48 first durations at each condition and for subsequent percepts there is a variable number of durations at each condition (mean 224 individual durations). The results are plotted in Fig 11A where, consistent with the literature, first duration curve is well separated from the subsequent durations curve across all conditions, the different being largest for small Δf and decreasing with Δf. In all the analysis that follows the first duration is excluded.

We next look at how subjects responses are distributed in terms of their global mean percept duration T^{glob}, grouping data across all conditions, see (10). This information can be used to identify any outliers and will be used to normalize responses across subjects. For each of 16 subjects we take the mean percept duration taken across both percept types, all repetitions and all Δf conditions. We plot summary statistics of the subject distribution of T^{glob}-values in Fig 11B. One clear outlier, with $T^{\mathrm{glob}} = 51.4$ s (consistently longer than the group across all conditions) is excluded from further analysis. The remaining 15 subject's T^{glob}-values spans the range 2.6–14.8 s. The result is that many more durations are recorded from subjects with small T^{glob} (fast switchers) than large T^{glob} (slow switchers). In further analysis, if we were to combine data across subjects including all durations, any results would be skewed towards fast switchers who contribute more durations. To account for this, in the analysis that follows, each subject contributes a mean integrated and a mean segregated duration averaged across all durations for the three repetitions at each condition. Other measures such as proportion of time integrated are computed in a similar fashion. Furthermore, due to the large range of T^{glob}-values, representing a near-order-of-magnitude difference from smallest to largest, we also scale each subject's durations by T^{glob} before combining across subjects. Further details of the normalization are given in *Computation of normalized durations*.
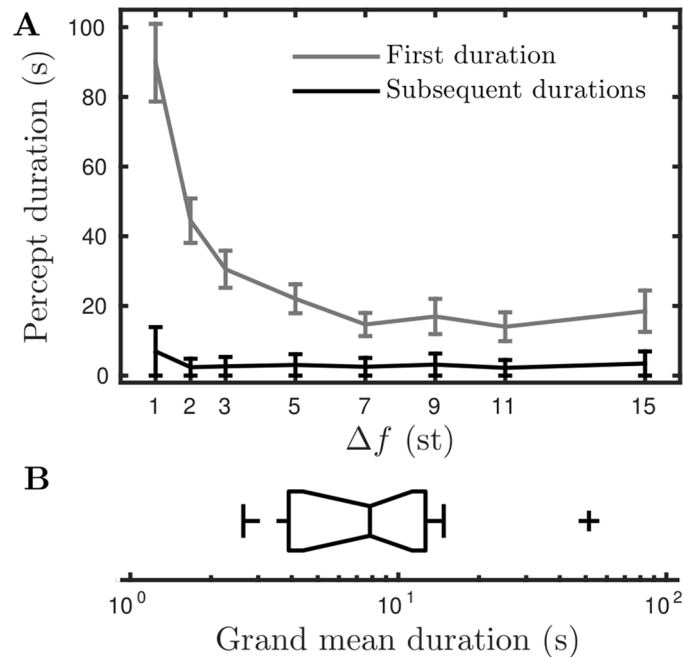
Fig 11. First and subsequent durations, summary statistics of the subject's grand mean durations. A: Mean first percept duration and mean subsequent percept duration for all durations combined across both percept types, $N = 16$ subjects and $R = 3$ repetitions. Error bars show standard error of the mean. **B**: Standard Tukey box plot of T^{glob} for $N = 16$ subjects (box shows quartiles, whiskers are most extreme data points within $1.5 \times$ iqr of the upper and lower quartiles, "+" are individual outliers).

Computation of normalized durations

The normalization of percept durations, as described below, was used to remove some of the subject variability in experiments. For reference, the percept durations for the experiment and model with (e_{dyn}, i_{gbl}) are plotted without normalization in Fig 12. To see the effect of the normalization compare Fig 9B and 9C with Fig 12A and 12B.
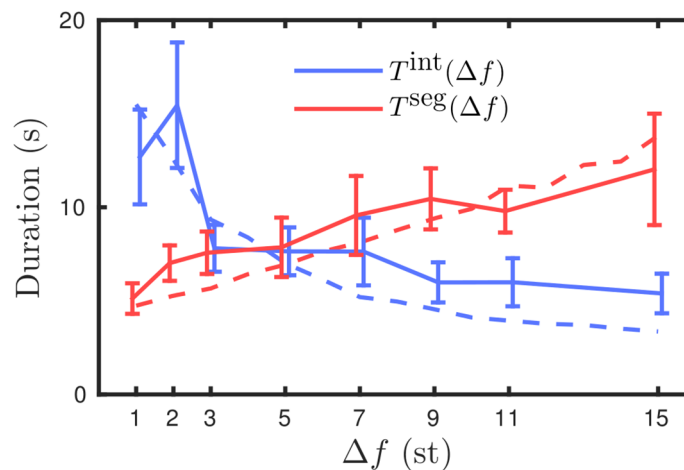
Fig 12. Non-normalized percept durations. Durations integrated and segregated without normalization plotted against Δf for experiment and model with (e_{dyn}, i_{gbl}), as Fig 9B.

For the experiment, we index N subjects by $n = 1, \ldots, N$, $R = 3$ repetitions by $r = 1, \ldots, R$ and $C = 8$ conditions by $c = 1, \ldots, C$. For each of the $N \times R \times C$ trials the we index all D durations (integrated and segregated) recorded in that trial by $d = 1 \ldots D$ or separately the D_{int} integrated durations by $d_{\text{int}} = 1, \ldots, D_{\text{int}}$ and the D_{seg} segregated durations by $d_{\text{seg}} = 1, \ldots, D_{\text{seg}}$ where $D = D_{\text{int}} + D_{\text{seg}}$. The first duration and (incomplete) last duration are excluded. On a given trial all durations $T_{n, r, c, d}$ can be indexed by d or indexed separated into durations integrated $T^{\text{int}}_{n,r,c,d_{\text{int}}}$ and durations segregated $T^{\text{seg}}_{n,r,c,d_{\text{seg}}}$.

For a given subject k the global mean percept duration is taken across R repetitions, C conditions and D durations

$$T^{\text{glob}}_k = \frac{1}{RCD} \sum_{r=1}^{R} \sum_{c=1}^{C} \sum_{d=1}^{D} \left(T_{k,r,c,d} \right). \tag{10}$$

Summary statistics of T^{glob} are plotted in Fig 11B for $N = 16$ subjects.

For a given subject k and condition j the normalized mean duration integrated across R repetitions is

$$\bar{T}^{\text{int}}_k(j) = \frac{1}{RD_{\text{int}}} \sum_{r=1}^{R} \sum_{d_{\text{int}}=1}^{D_{\text{int}}} \left(\frac{T^{\text{int}}_{k,r,j,d_{\text{int}}}}{T^{\text{glob}}_k} \right), \tag{11}$$

and, similarly, the noramlized mean duration segregated is

$$\bar{T}^{\text{seg}}_k(j) = \frac{1}{RD_{\text{seg}}} \sum_{r=1}^{R} \sum_{d_{\text{seg}}=1}^{D_{\text{seg}}} \left(\frac{T^{\text{seg}}_{k,r,j,d_{\text{seg}}}}{T^{\text{glob}}_k} \right). \tag{12}$$

For a given condition j the the normalized mean duration integrated across subjects is

$$\bar{T}^{\text{int}}(j) = \frac{1}{N} \sum_{k=1}^{N} \bar{T}^{\text{int}}_k(j), \tag{13}$$

and the standard error of the mean is the standard deviation of $\bar{T}^{\text{int}}_k(j)$ divided by $\sqrt{N}$. Similarly,

$$\bar{T}^{\text{seg}}(j) = \frac{1}{N} \sum_{k=1}^{N} \bar{T}^{\text{seg}}_k(j). \tag{14}$$

An analogous normalization was used for the model data plotted across a range of Δf-values in Figs 8B, 9D and 9E. At each Δf-value, durations integrated and segregated were extracted from 36 four minute simulations excluding the first and (incomplete) last duration. A global mean percept duration T^{glob} was computed combining durations integrated and segregated across all simulations at each of the 8 Δf-values used in the experiments. Durations were normalized by T^{glob} before taking a mean to produce the normalized mean durations integrated T^{int} and segregated T^{seg} plotted in the figures.

Comparison to standard statistical distributions

Note that the notations used to define these standard distributions are local to this section and not used elsewhere in the paper. The probability density function (pdf) for the gamma distribution is given by

$$f(x|k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\left(-x/\theta \right), \tag{15}$$

Table 3. One-way repeated measures ANOVA of proportion integrated for the factor Δf (eight conditions, see Fig 9A). Analysis shows a significant effect of Δf on proportion integrated. Data for N = 12 subjects, see text. Mauchly test for sphericity reaches significance, Greenhouse-Geisser correct P-value reported in the text.

Source	Type III SS	df	MS	F	Sig.	G-G
DF	1.129	7	0.161	24.656	< 0.001	< 0.001
Error(DF)	0.504	77	0.007			

doi:10.1371/journal.pcbi.1004555.t003

where k is the shape and θ the scale parameter. The mean of the distribution is equal to k/θ and the variance is equal to k/θ^2. The pdf for the log-normal distribution is given by

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \tag{16}$$

where μ is the location parameter and σ the scale parameter. The mean of the distribution is equal to $\exp(\mu + \sigma^2/2)$ and the variance is equal to $(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$.

In *Statistics of dominance durations: model*, the distributions from the model and from the experiment shown in Fig 5A and 5B were compared with the standard distributions given by Eqs (15) and (16) using a one-way Kolmogorov-Smirnov KS test. The null hypothesis is that the test data are drawn from the standard distribution and a significant result ($P < 0.05$) indicates that the test data are not drawn from the comparison distribution. In a further analysis, the model and experimental data were also compared in a two-way KS test to see if they come from the same distribution (without specifying what that distribution might be). The null hypothesis is that they have the same underlying distribution and a significant result ($P < 0.05$) indicates that they are drawn from different distributions.

Repeated measures ANOVAs

Three subjects that did not contribute scores at all eight conditions (two at Δf = 1 st, one at Δf = 15 st) were excluded from the following analysis. A significance level of 0.05 is used throughout. In the text throughout the manuscript, the Greenhouse-Geisser (GG) corrected P-values are reported as appropriate where the data reached significant in a Mauchly sphericity test.

A one-way repeated measures ANOVA on proportion of time integrated (see Fig 9A) was performed with Δf = {1, 2, 3, 5, 7, 9, 11, 15} as the within subjects factor. The analysis reported in Table 3 shows a significant effect of Δf on proportion integrated $F(7, 77) = 24.656$, $P < 0.001$. Pairwise comparisons with Bonferroni-corrected significance levels showed that

Table 4. Two-way repeated measures ANOVA of log noramlized durations for Percept type (integrated or segregated), Δf (eight conditions) and their interaction, see Fig 9B. Analysis shows a significant interaction for Percept * Δf. Data for N = 12 subjects, see text. Mauchly test for sphericity reaches significance for the Percept * Δf interaction, Greenhouse-Geisser corrected P-value reported in the text.

Source	Type III SS	df	MS	F	Sig.	G-G
Percept	1.451	1	1.451	3.158	0.103	0.103
Error(Percept)	5.053	11	0.459			
Δf	0.795	7	0.114	1.053	0.402	0.390
Error(Δf)	8.304	77	0.108			
Percept * Δf	7.603	7	1.086	16.225	< 0.001	< 0.001
Error(Percept * Δf)	5.155	77	0.067			

doi:10.1371/journal.pcbi.1004555.t004

Table 5. One-way repeated measures ANOVA for the measure η with Δf as a factor (eight conditions) see Fig 9C. Analysis shows that effect of Δf on η does not reach significance. Mauchly test for sphericity reaches significance, Greenhouse-Geisser corrected P-value reported in the text. Data for $N = 12$ subjects, see text.

Source	Type III SS	df	MS	F	Sig.	G-G
DF	5.715	7	0.816	0.878	0.528	0.463
Error(DF)	71.597	77	0.930			

doi:10.1371/journal.pcbi.1004555.t005

each individual condition has significant differences with at least two (and up to 5) other conditions.

A two-way repeated measures ANOVA on log normalized durations $\log\left(\bar{T}_k^{\mathrm{int}}(j)\right)$ and $\log\left(\bar{T}_k^{\mathrm{seg}}(j)\right)$ (see Eqs (11) and (12)) was performed with Percept type (integrated or segregated) and frequency difference Δf as within subjects factors. The use of a log transformation of the durations is standard for this type of experiment [19, 21]. The analysis reported in Table 4 shows a highly significant interaction for Percept * Δf, $F(7, 77) = 16.225$, $P < 0.001$. As for the individual factors, Percept does not reach significance $F(1, 11) = 3.158$, $P = 0.103$, neither does Δf $F(7, 77) = 1.053$, $P = 0.402$. We note that, given the interaction between Percept and Δf it is hard to interpret effects of the individual factors. Pairwise comparisons with Bonferroni-corrected significance levels reveal no significant differences between individual conditions in this two-way analysis.

A one-way repeated measures ANOVA on η with Δf as a within subjects factor does not show a significant effect $F(7, 77) = 0.878$, $P = 0.463$, see Table 5.

Supporting Information

S1 Data. Percept durations from psychoacoustic experiments. Durations from each trial are provided in Matlab data format in the supporting information file `S1_data.zip`. (ZIP)

Acknowledgments

Author Contributions

Conceived and designed the experiments: JRa ES JRi. Performed the experiments: JRa. Analyzed the data: JRa. Wrote the paper: JRa JRi. Conceived the model: JRa JRi. Implemented the model: JRa.

References

1. Bregman AS. Auditory streaming: Competition among alternative organizations. Perception & Psychophysics. 1978; 23(5):391–398. doi: 10.3758/BF03204141

2. van Noorden LPAS. Temporal coherence in the perception of tone sequences. Ph.D. thesis, Eindhoven University of Technology; 1975.

3. Anstis SM, Saida S. Adaptation to auditory streaming of frequency-modulated tones. Journal of Experimental Psychology: Human Perception and Performance. 1985; 11(3):257–271.

4. Bregman AS. Auditory scene analysis: The perceptual organization of sound. MIT press; 1994.

5.   Carlyon RP, Cusack R, Foxton JM, Robertson IH. Effects of attention and unilateral neglect on auditory stream segregation. Journal of Experimental Psychology: Human Perception and Performance. 2001; 27(1):115–127. PMID: 11248927

6.   Carlyon RP, Plack CJ, Fantini DA, Cusack R. Cross-modal and non-sensory influences on auditory streaming. Perception. 2003; 32(11):1393–1402. doi: 10.1068/p5035 PMID: 14959799

7.   Macken WJ, Tremblay S, Houghton RJ, Nicholls AP, Jones DM. Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. Journal of Experimental Psychology: Human Perception and Performance. 2003; 29(1):43–51. PMID: 12669746

8.   Beauvois MW, Meddis R. Time decay of auditory stream biasing. Perception & Psychophysics. 1997; 59(1):81–86. doi: 10.3758/BF03206850

9.   Rogers WL, Bregman AS. Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. Perception & Psychophysics. 1998; 60(7):1216–1227. doi: 10.3758/BF03206171

10.   Rahne T, Sussman E. Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. European Journal of Neuroscience. 2009; 29(1):205–211. doi: 10.1111/j.1460-9568.2008.06561.x PMID: 19087164

11.   McCabe SL, Denham MJ. A model of auditory streaming. Advances in Neural Information Processing Systems. 1996; 101(3):1611–1621.

12.   Beauvois MW, Meddis R. A computer model of auditory stream segregation. The Quarterly Journal of Experimental Psychology. 1991; 43(3):517–541. doi: 10.1080/14640749108400985 PMID: 1775655

13.   Steele SA, Tranchina D, Rinzel J. An alternating renewal process describes the buildup of perceptual segregation. Frontiers in Computational Neuroscience. 2014; 8(166):1–13.

14.   Elhilali M, Shamma SA. A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. The Journal of the Acoustical Society of America. 2008; 124(6):3751–3771. doi: 10.1121/1.3001672 PMID: 19206802

15.   Wang D, Chang P. An oscillatory correlation model of auditory streaming. Cognitive Neurodynamics. 2008; 2(1):7–19. doi: 10.1007/s11571-007-9035-8 PMID: 19003469

16.   Almonte F, Jirsa VK, Large EW, Tuller B. Integration and segregation in auditory streaming. Physica D: Nonlinear Phenomena. 2005; 212(1):137–159. doi: 10.1016/j.physd.2005.09.014

17.   Beauvois MW, Meddis R. Computer simulation of auditory stream segregation in alternating-tone sequences. The Journal of the Acoustical Society of America. 1996; 99(4):2270–2280. doi: 10.1121/1.415414 PMID: 8730073

18.   Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. Neuromagnetic correlates of streaming in human auditory cortex. The Journal of Neuroscience. 2005; 25(22):5382–5388. doi: 10.1523/JNEUROSCI.0347-05.2005 PMID: 15930387

19.   Pressnitzer D, Hupé JM. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Current Biology. 2006; 16(13):1351–1357. doi: 10.1016/j.cub.2006.05.054 PMID: 16824924

20.   Denham SL, Winkler I. The role of predictive models in the formation of auditory streams. Journal of Physiology-Paris. 2006; 100(1):154–170. doi: 10.1016/j.jphysparis.2006.09.012

21.   Denham SL, Gyimesi K, Stefanics G, Winkler I. Perceptual bistability in auditory streaming: How much do stimulus features matter? Learning & Perception. 2013; 5(2):73–100. doi: 10.1556/LP.5.2013.Suppl2.6

22.   Kondo HM, Kashino M. Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. The Journal of Neuroscience. 2009; 29(40):12695–12701. doi: 10.1523/JNEUROSCI.1549-09.2009 PMID: 19812344

23.   Kashino M, Kondo HM. Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012; 367(1591):977–987. doi: 10.1098/rstb.2011.0370

24.   Mill RW, Bőhm TM, Bendixen A, Winkler I, Denham SL. Modelling the emergence and dynamics of perceptual organisation in auditory streaming. PLoS Computational Biology. 2013; 9(3):e1002925. doi: 10.1371/journal.pcbi.1002925 PMID: 23516340

25.   Winkler I, Denham SL, Mill R, Bőhm TM, Bendixen A. Multistability in auditory stream segregation: a predictive coding view. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012; 367:1001–1012. doi: 10.1098/rstb.2011.0359

26.   Fishman YI, Reser DH, Arezzo JC, Steinschneider M. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hearing Research. 2001; 151(1):167–187. doi: 10.1016/S0378-5955(00)00224-0 PMID: 11124464

27. Fishman YI, Arezzo JC, Steinschneider M. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. The Journal of the Acoustical Society of America. 2004; 116(3):1656–1670. doi: 10.1121/1.1778903 PMID: 15478432

28. Micheyl C, Tian B, Carlyon RP, Rauschecker JP. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron. 2005; 48(1):139–148. doi: 10.1016/j.neuron.2005.08.039 PMID: 16202714

29. Bee MA, Klump GM. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. Journal of Neurophysiology. 2004; 92(2):1088–1104. doi: 10.1152/jn.00884.2003 PMID: 15044521

30. Pressnitzer D, Sayles M, Micheyl C, Winter IM. Perceptual organization of sound begins in the auditory periphery. Current Biology. 2008; 18(15):1124–1128. doi: 10.1016/j.cub.2008.06.053 PMID: 18656355

31. Itatani N, Klump GM. Neural correlates of auditory streaming in an objective behavioral task. Proceedings of the National Academy of Sciences. 2014; 111(29):10738–10743. doi: 10.1073/pnas.1321487111

32. Laing CR, Chow CC. A spiking neuron model for binocular rivalry. Journal of Computational Neuroscience. 2002; 12(1):39–53. doi: 10.1023/A:1014942129705 PMID: 11932559

33. Wilson HR. Computational evidence for a rivalry hierarchy in vision. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(24):14499–14503. doi: 10.1073/pnas.2333622100 PMID: 14612564

34. Shpiro A, Moreno-Bote R, Rubin N, Rinzel J. Balance between noise and adaptation in competition models of perceptual bistability. Journal of Computational Neuroscience. 2009; 27:37–54. doi: 10.1007/s10827-008-0125-3 PMID: 19125318

35. Leopold DA, Logothetis NK. Multistable phenomena: changing views in perception. Trends in Cognitive Sciences. 1999; 3(7):254–264. doi: 10.1016/S1364-6613(99)01332-7 PMID: 10377540

36. Levelt WJ. On binocular rivalry. vol. 2. Mouton The Hague; 1968.

37. Brascamp J, Klink P, Levelt WJ. The'laws' of binocular rivalry: 50 years of Levelt's propositions. Vision research. 2015; 109:20–37. doi: 10.1016/j.visres.2015.02.019 PMID: 25749677

38. Moreno-Bote R, Shpiro A, Rinzel J, Rubin N. Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. Journal of Vision. 2010; 10(11):1–18. doi: 10.1167/10.11.1 PMID: 20884496

39. Borsellino A, De Marco A, Allazetta A, Rinesi S, Bartolini B. Reversal time distribution in the perception of visual ambiguous stimuli. Kybernetik. 1972; 10(3):139–144. doi: 10.1007/BF00290512 PMID: 5021011

40. Logothetis NK, Leopold DA, Sheinberg DL. What is rivalling during binocular rivalry? Nature. 1996; 380:621–624. PMID: 8602261

41. Zhou Y, Gao J, White K, Merk I, Yao K. Perceptual dominance time distributions in multistable visual perception. Biological Cybernetics. 2004; 90(4):256–263. doi: 10.1007/s00422-004-0472-8 PMID: 15085344

42. Pastukhov A, García-Rodríguez PE, Haenicke J, Guillamon A, Deco G, Braun J. Multi-stable perception balances stability and sensitivity. Frontiers in Computational Neuroscience. 2013; 7(17):1–18.

43. Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron. 2009; 61(2):317–329. doi: 10.1016/j.neuron.2008.12.005 PMID: 19186172

44. Schadwinkel S, Gutschalk A. Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. Journal of neurophysiology. 2011; 105(5):1977–1983. doi: 10.1152/jn.00461.2010 PMID: 21325685

45. Gutschalk A, Micheyl C, Oxenham AJ. Neural correlates of auditory perceptual awareness under informational masking. PLoS Biology. 2008; 6(6):e138. doi: 10.1371/journal.pbio.0060138 PMID: 18547141

46. Humphries C, Liebenthal E, Binder JR. Tonotopic organization of human auditory cortex. Neuroimage. 2010; 50(3):1202–1211. doi: 10.1016/j.neuroimage.2010.01.046 PMID: 20096790

47. Striem-Amit E, Hertz U, Amedi A. Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding FMRI. PLoS One. 2011; 6(3):e17832–e17832. doi: 10.1371/journal.pone.0017832 PMID: 21448274

48. Moerel M, De Martino F, Formisano E. An anatomical and functional topography of human auditory cortical areas. Frontiers in neuroscience. 2014; 8. doi: 10.3389/fnins.2014.00225 PMID: 25120426

49. Huguet G, Rinzel J, Hupé JM. Noise and adaptation in multistable perception: Noise drives when to switch, adaptation determines percept choice. Journal of Vision. 2014; 14(3):19. doi: 10.1167/14.3.19 PMID: 24627459

50. Moreno-Bote R, Rinzel J, Rubin N. Noise-Induced Alternations in an Attractor Network Model of Perceptual Bistability. Journal of Neurophysiology. 2007; 98(3):1125–1139. doi: 10.1152/jn.00116.2007 PMID: 17615138

51. Cusack R. The intraparietal sulcus and perceptual organization. Journal of cognitive neuroscience. 2005; 17(4):641–651. doi: 10.1162/0898929053467541 PMID: 15829084

52. Bendixen A, Denham SL, Gyimesi K, Winkler I. Regular patterns stabilize auditory streams. The Journal of the Acoustical Society of America. 2010; 128(6):3658–3666. doi: 10.1121/1.3500695 PMID: 21218898

53. Hupé JM, Pressnitzer D. The initial phase of auditory and visual scene analysis. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012; 367(1591):942–953. doi: 10.1098/rstb.2011.0368

54. Denham SL, Bohm T, Bendixen A, Szalárdy O, Kocsis Z, Mill R, et al. Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. Frontiers in Neuroscience. 2014; 8(25):1–15.

55. Shamma SA, Elhilali M, Micheyl C. Temporal coherence and attention in auditory scene analysis. Trends in neurosciences. 2011; 34(3):114–123. doi: 10.1016/j.tins.2010.11.002 PMID: 21196054

56. Micheyl C, Kreft H, Shamma S, Oxenham AJ. Temporal coherence versus harmonicity in auditory stream formation. The Journal of the Acoustical Society of America. 2013;1 33(3):EL188–EL194.

57. Rankin J, Meso A, Masson GS, Faugeras O, Kornprobst P. Bifurcation study of a neural field competition model with an application to perceptual switching in motion integration. Journal of Computational Neuroscience. 2014; 36(2):193–213. doi: 10.1007/s10827-013-0465-5 PMID: 24014258

58. Seely J, Chow CC. Role of mutual inhibition in binocular rivalry. Journal of Neurophysiology. 2011; 106 (5):2136–2150. doi: 10.1152/jn.00228.2011 PMID: 21775721

59. Wang XJ, Liu Y, Sanchez-Vives MV, McCormick DA. Adaptation and temporal decorrelation by single neurons in the primary visual cortex. Journal of Neurophysiology. 2003; 89(6):3279–3293. doi: 10.1152/jn.00242.2003 PMID: 12649312

60. Hestrin S, Sah P, Nicoll RA. Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices. Neuron. 1990; 5(3):247–253. doi: 10.1016/0896-6273 (90)90162-9 PMID: 1976014

61. Spruston N, Jonas P, Sakmann B. Dendritic glutamate receptor channels in rat hippocampal CA3 and CA1 pyramidal neurons. The Journal of Physiology. 1995; 482(2):325–352. doi: 10.1113/jphysiol.1995.sp020521 PMID: 7536248

62. Brunel N, Wang XJ. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. Journal of Computational Neuroscience. 2001; 11(1):63–85. doi: 10.1023/A:1011204814320 PMID: 11524578

63. Wang XJ. Probabilistic decision making by slow reverberation in cortical circuits. Neuron. 2002; 36 (5):955–968. doi: 10.1016/S0896-6273(02)01092-9 PMID: 12467598

64. Varela JA, Sen K, Gibson J, Fost J, Abbott L, Nelson SB. A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. The Journal of Neuroscience. 1997; 17(20):7926–7940. PMID: 9315911

65. Ulanovsky N, Las L, Farkas D, Nelken I. Multiple time scales of adaptation in auditory cortex neurons. The Journal of Neuroscience. 2004; 24(46):10440–10453. doi: 10.1523/JNEUROSCI.1905-04.2004 PMID: 15548659