frontiers in Neuroscience

Stimulus pauses and perturbations differentially delay or promote the segregation of auditory objects: psychoacoustics and modeling

James Rankin[1, 2*], Pamela J. Osborn Popp[1], John Rinzel[1, 3]

[1]Center for Neural Science, New York University, USA, [2]Department of Mathematics, University of Exeter, United Kingdom, [3]Courant Institute of Mathematical Sciences, USA

Submitted to Journal:
Frontiers in Neuroscience

Specialty Section:
Auditory Cognitive Neuroscience

ISSN:
1662-453X

Article type:
Original Research Article

Received on:
14 Nov 2016

Accepted on:
23 Mar 2017

Provisional PDF published on:
23 Mar 2017

Frontiers website link:
www.frontiersin.org

Citation:
Rankin J, Osborn_popp PJ and Rinzel J(2017) Stimulus pauses and perturbations differentially delay or promote the segregation of auditory objects: psychoacoustics and modeling. *Front. Neurosci.* 11:198. doi:10.3389/fnins.2017.00198

Stimulus pauses and perturbations differentially delay or promote the segregation of auditory objects: psychoacoustics and modeling

James Rankin[†,‡] (james.rankin@gmail.com),
Pamela J Osborn Popp[†], **John Rinzel**[†,*]

Abstract Segregating distinct sound sources is fundamental for auditory perception, as in the cocktail party problem. In a process called the build-up of stream segregation, distinct sound sources that are perceptually integrated initially can be segregated into separate streams after several seconds. Previous research concluded that abrupt changes in the incoming sounds during build-up — for example, a step change in location, loudness or timing — reset the percept to integrated. Following this reset, the multisecond build-up process begins again. Neurophysiological recordings in auditory cortex (A1) show fast (subsecond) adaptation, but unified mechanistic explanations for the bias toward integration, multisecond build-up and resets remain elusive. Combining psychoacoustics and modeling, we show that initial unadapted A1 responses bias integration, that the slowness of build-up arises naturally from competition downstream, and that recovery of adaptation can explain resets. An early bias toward integrated perceptual interpretations arising from primary cortical stages that encode low-level features and feed into competition downstream could also explain similar phenomena in vision. Further, we report a previously overlooked class of perturbations that promote segregation rather than integration. Our results challenge current understanding for perturbation effects on the emergence of sound source segregation, leading to a new hypothesis for differential processing downstream of A1. Transient perturbations can momentarily redirect A1 responses as input to downstream competition units that favor segregation.

1 Introduction

A valued paradigm for studying auditory streaming involves segregating two interleaved sequences of A tones and B tones, distinguishable by a perceived difference in pure tone frequency and timing. The tones are organized in a repeating ABA_ABA_... pattern[1] ("_" represents silence) (Fig. 1B, top). At first heard as a one stream rhythm (integrated percept), the probability of hearing two streams (segregated percept) gradually builds up over several to tens of seconds (build-up)[2,3,4]. Build-up occurs more rapidly with a large difference in frequency (DF) between A and B and at faster presentation rates. However, abrupt change in the incoming sound (e.g. a step change in location, loudness or timing) can reset perception to integrated[2,5,6], after which multisecond build-up begins again. The first perceptual switch, typically from integrated to segregated is followed by persistent alternations between the two interpretations[7,8]. Build-up progresses not just to the segregation, but to the stable probability of segregation in the subsequent long-term alternations.

Neural responses to triplet stimuli have been studied in primary auditory cortex (A1) of awake monkeys[9,10,11], in forebrain of awake[12,13] or behaving[14] songbirds, and in the auditory periphery of anesthetised guinea pigs[15]. The tonotopic organization of A1 and increased forward masking at higher presentation rates[9,10] can explain the feature dependence of these responses. Studies comparing neural response data with build-up functions from human psychoacoustic experiments have shown that a trial averaged neurometric function can be tuned to match trial averaged behavioral data[11,15,16]. However, no study has claimed that the neural substrate for the perceptual state or switches in perceptual states lies in or before A1. Indeed, the only animal study with neural data recorded from behaving animals[14] concluded that only stimulus features and not perceptual choice is encoded in songbird forebrain (analogous to A1). Responses to tones in A1 show rapid adaptation in the first few hundred milliseconds (1-3 triplets)[11]. In this initial phase, response amplitude adapts and dependence on DF emerges (at first little tonotopic dependence is evident for tones separated by less than an octave). The relationship between this rapid adaptation (~ 500 ms) and the slower build-up process (several seconds) remains unexplained.

In ref. 17 we developed a neuromechanistic model of auditory bistability based on a conceptual model proposed in ref. 9. Far apart A and B tones drive tonotopically segregated representations, but for smaller

‡ Department of Mathematics, University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, UK
† Center for Neural Science, New York University, 4 Washington Place, 10003 New York, NY, USA
* Courant Institute of Mathematical Sciences, 251 Mercer Street, 10012 New York, NY, USA

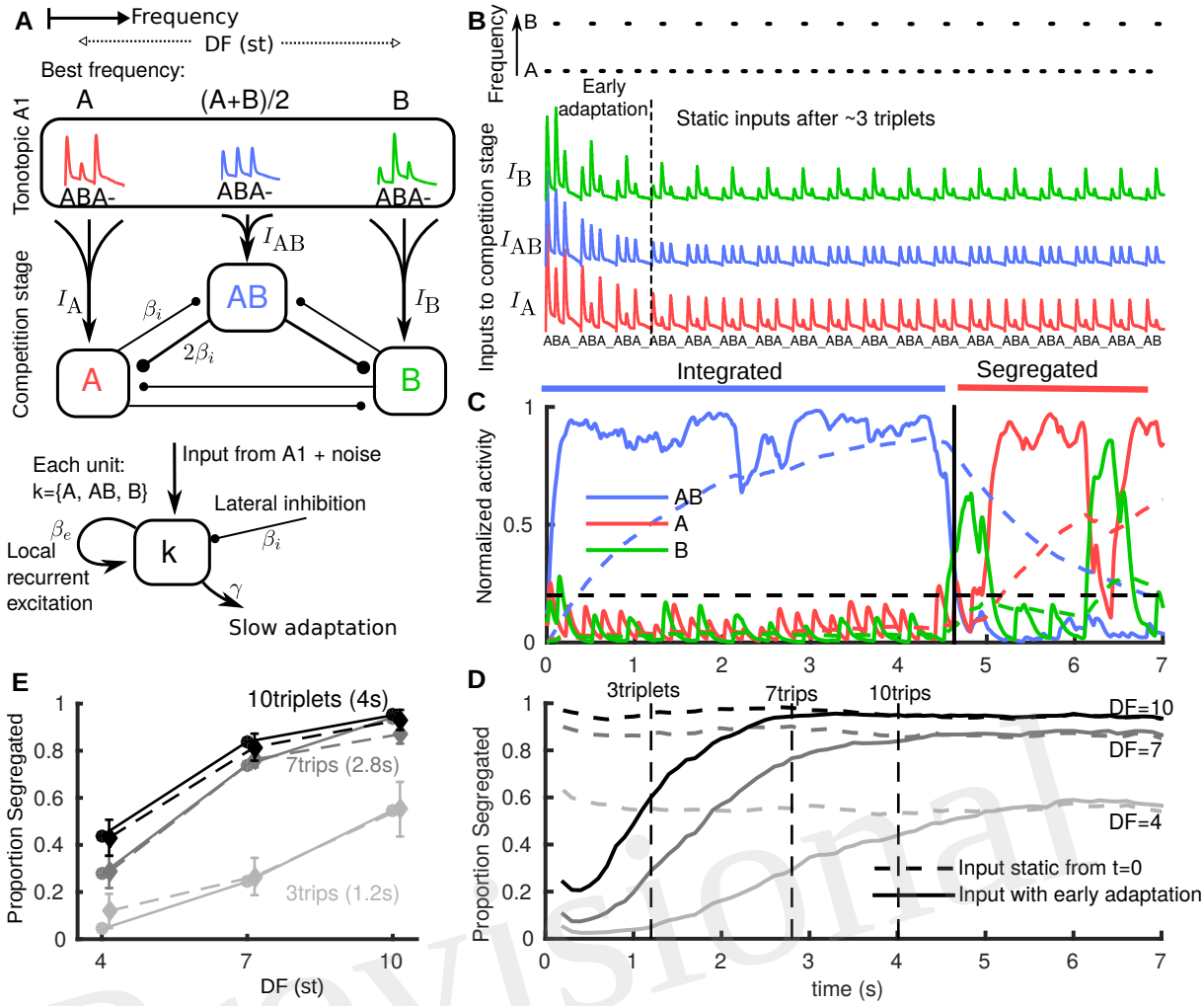

Fig. 1 Neuromechanistic model captures initial bias to integration and build-up of stream segregation. (A) Model schematic with two stages: tonotopic A1 and a competition stage (downstream of and pooling inputs from A1). A1 encodes only stimulus features, while the downstream competition stage encodes percepts. Inputs from lower frequency A and higher B tones generate onset-plateau responses in A1 dependent on difference in frequency (DF) in semitones (st). In the competition stage three units encode the integrated percept (AB), the segregated A stream, and the segregated B stream. Units are in competition through mutual inhibition, pool inputs from A1, have recurrent NMDA excitation (timescale 70 ms) and undergo adaptation on a slow timescale (timescale 1.4 s). (B) (top) Stimulus paradigm where low A tones, high B tones and silences (_) each of 100 ms repeat in an ABA_ triplet pattern. (below) A1 responses to tones adapt rapidly (timescale 500 ms) with tonotopic dependence emerging and overall amplitude reducing during first 2–3 triplets. Vertical offset for visualisation only. (C) One model simulation showing the activation threshold (horizontal dashed), and each population's NMDA variable (solid, pulsatile inputs appear smoothed in sub-threshold activity) and adaptation variable (dashed). When the central AB unit is active (integrated), the peripheral units are suppressed through mutual inhibition. Increasing adaptation for AB increases the probability of noise inducing a switch; when units A or B become active and dominant after ~ 4.5 s (segregated), the integrated (AB) unit is suppressed. (D) Build-up function computed as time-binned trial-averaged proportion segregated computed from $N = 500$ model simulations. With no early adaptation of inputs from A1 (input static), there is no build-up and stable proportion segregation from long-term alternations is reflected at onset. Early adaptation of inputs from A1 gives initial bias toward integrated and proportion segregated gradually builds up to DF-dependent value of long-term alternations. (E) Snapshots from build-up after 3, 7 and 10 triplets from model (each solid curve in E corresponds to a dashed vertical line in D) are compared with psychoacoustic data (N=8 normal hearing subjects) with percept reported at the end of presentation (dashed curves; errorbars show s.e.m., same for all plots)

DF the receptive fields overlap, leading to a common drive for an intermediate population encoding integration (Fig. 1A). Our model mimics the periodic, pulsatile responses and stimulus feature dependence from A1[11], which are pooled as inputs to a competition stage residing downstream (A1 encodes only stimulus features, not the percepts). At the competition stage peripheral units A and B encode segregation and a central unit AB encodes integrated. The competition network incorporates the mechanisms of mutual inhibition, slow adaptation and additive noise shown to play an important role in perceptual bistability[18,19]. Recurrent excitation with an NMDA-like timescale links responses and thereby percepts across silent gaps between tones and triplets (Fig. 1C). Our model captures the complex dynamics of perceptual alternations,

reproducing characteristic features such as the log-normal distribution of perceptual durations as well as dependence of perceptual durations on parameters such as DF [17]. We focused previously on the alternations after the first perceptual switch; the initial bias to integrated and build-up were not addressed.

Here, we propose that the initial integration bias is determined by early broad tonotopic tuning of neuronal responses in A1, while the multisecond timescale of build-up is due to slow adaptation downstream of A1. Recovery of early adaptation, say after a stimulus pause, can further explain the reset to the integrated percept. Furthermore, we find in new experiments, a class of transient perturbations (single unexpected tones in the ongoing stimulus) that subsequently promote segregation, in contrast to the widely reported resets to integrated. Our model, motivated from neurophysiological studies, provides a mechanistic explanation for build-up and resetting whilst also accounting for new experimental findings.

2 Results

2.1 Neuromechanistic model explains initial bias to integration and build-up of stream segregation

In order to study build-up in our existing model, we made one change to the inputs based on further observations about the early responses to triplets in A1 [11]. We introduced rapid adaptation (timescale 500 ms) for both input amplitude and DF dependence (Fig. 1B). During the first 2–3 triplets input evolves as if driven by a DF that is effectively small but gradually increasing to a static value. The AB unit receives enough input bias to become active, suppress the peripheral units and become dominant first (Fig. 1C). Time-binned build-up functions (three DF and two input cases) were computed by averaging across simulations. In the input static case (Fig. 1D dashed) the inputs are assumed post fast-adaptation (Fig. 1B after 3 triplets) and the time-course only reflects the static probability of post build-up alternations. In the input adapting case (Fig. 1D solid) responses are initially biased to integrated and gradually build-up to the static probability of later alternations. The slower timescale of this build-up arises from the mechanisms already established in [17] for the competition stage downstream of A1. In particular, there is a slower adaptation process at the competition stage. In psychoacoustic experiments, the build-up process can be sampled with short stimulus presentations of different lengths with percepts reported at the end. Vertical lines in Fig. 1D show three such snapshots from the model (Fig. 1E solid). These are compared with psychoacoustic data (Fig. 1E dashed) for three DF and two presentation length conditions. A repeated measures ANOVA showed a significant effect of DF ($F(2,14) = 37.49, P < 0.001$), of presentation length ($F(2,14) = 19.49, P < 0.005$) and their interaction ($F(4,28) = 4.34, P < 0.05$), see App. A. The close match with these data show that the model accurately captures build-up (increasing segregation with both DF and presentation length). Our model is the first to produce the bias to integrated in a manner directly motivated from neurophysiology data [9,11] (fast adaptation in A1) and to produce gradual build-up due to a slower adaptation timescale downstream of A1 (at the competition stage in our model).

2.2 Promotion of segregation by distractor and deviant tones

In psychoacoustic experiments we reproduced a previously reported reset toward integration for a brief pause between triplets (paradigm, Fig. 2A; data Fig. 2D). In all experiments reported here, the stimulus ends in three normal triplets with the last triplet reported as integrated or segregated [20,21]. In Fig. 2D, if the test conditions (300 ms or 600 ms pause) showed no effect, the orange and red curves would align with the black ten triplet control. For a full reset to integrated the test conditions would align with the three triplet grey curve. Our results show, consistent with existing studies [22,23], that brief stimulus pauses can result in a partial reset back toward integrated. The pause conditions had a significant effect on proportion segregated ($F(2,14) = 5.126, P < 0.05$), see App. A. The reset is of a similar magnitude for all pause duration and DF conditions.

In a new experiment six triplet presentations are used with a perturbation in the third triplet (full details in Sec. 4 and App. A). In the *distractor case* (Fig. 2B), an additional tone is inserted in the normal gap between the third and fourth triplet: ... ABA_ABAdABA_ ..., where 'd' is 2 semitones (st) higher than B. In the *deviant* case (Fig. 2C), the B-tone in the third triplet is a deviant: ... ABA_ADA_ABA_ ..., where 'D' is 2 st higher than B. A shorter presentation length was used relative to the pause experiment to avoid ceiling effects (saturation at proportion segregated=1). See Fig. 2E, where again, for no effect the test conditions would align with the black control case and for a reset to integrated, move down toward the grey three-triplet case. We found an opposite effect from pauses for a deviant or distractor tone during the ongoing triplet sequence: promotion of segregation. The increase in proportion segregated is significant

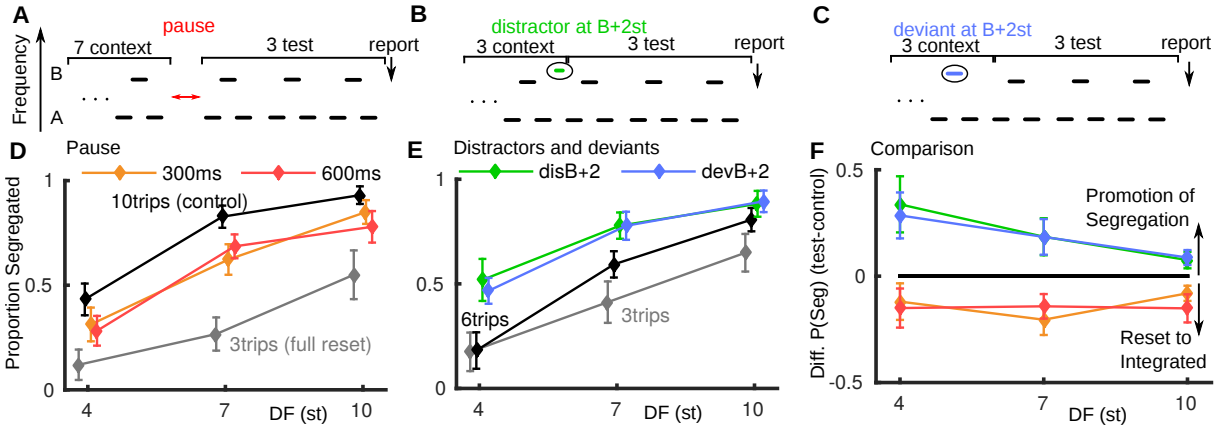

Fig. 2 Promotion of segregation by distractor and deviant tones, behavioral data (normal hearing subjects). Psychoacoustic data from N=8 normal hearing subjects. (A) Paradigm for pause of 300 or 600 ms after 7 context triplets, followed by 3 test triplets, where subjects report percept of final triplet. (B) Paradigm for distractor falling in the normal gap between last of 3 context triplets and first of 3 test triplets. (C) Paradigm for deviant B tone in last of 3 context triplets. (D) Brief pauses in stimulus presentation result in a partial reset to integrated. The test conditions (red, orange) would align with the ten-triplet control (black) if the pause had no effect and align with the three-triplet control (grey) for a full reset to integrated. (E) Both a distractor tone in the gap between triplets or a deviant tone within a triplet can promote segregation. Proportion segregation increased for all test conditions (green, blue) relative to the control condition black. (F) Direct comparison between stimulus pauses and distractor or deviant tones shows an opposite effect. The difference in proportion segregated between test and control conditions in A and B is plotted; when the difference is positive there is promotion of segregation, when negative a reset to integrated.

for these test conditions ($F(3,21) = 5.80, P < 0.05$). There is a similar effect for the deviant and distractor cases (largest for small DF). A distractor at 15 st above B showed no effect (not shown); see App. A.

For each experiment, by calculating the difference in proportion segregated between the test cases (colored curves) and control cases (black curves) in Fig. 2D–E, we can make a direct comparison between the two types of perturbation (Fig. 2F). A negative (positive) difference indicates a reset toward integrated (promotion of segregation). The promotion of segregation by a single-tone perturbation during triplets is a new and unexpected finding, opposite to the effect of a pauses and other perturbations previously reported. To better understand this phenomenon, we focused on the distractor tones and further investigated their relative frequency to the triplet tones (Fig. 4F). Before reporting these data we explore perturbations with the model.

2.3 Rapid recovery of adapted A1 responses explains reset to integration for pauses

In the model we assume that when the stimulus resumes after even a brief pause, it will be partially recovered from adaptation (to a state similar to stimulus onset) (Fig. 3A). Figure 3B shows a simulation-averaged build-up function comparing a case without a stimulus pause (input Fig. 1B) to a case with a pause input (input Fig. 3A). When the stimulus turns off the proportion segregated decreases (increases for DF=4) toward 0.5. When the stimulus resumes the amplitude and effective DF of inputs from A1 have partially recovered; the proportion segregated continues to decrease (starts decreasing for DF=4) before resuming gradual build-up. In this way, the model accounts for the partial reset toward integration across all DF conditions, compare red/orange curves in Fig. 3E (model) with Fig. 2F (experiments).

2.4 Model hypothesis on differential processing of non-triplet tones

For a distractor tone in the model, in order to compute an input amplitude, we first assumed the same rules as for the standard A and B tones. One modification was to assume a reduced response in A1 at the A-location due to higher repetition rate and the distractor immediately following an A-tone offset (stimulus-specific adaptation[24,25]). Until now the responses in A1 were taken directly as inputs to the competition stage, without modification. However, in initial simulations we found almost no effect of introducing a single new tone. A further assumption is that a distractor tone, arriving in a window where silence was expected, would be detected as a new event, and boosted (approximately to the level of an un-adapted tone) as input to the competition stage. Figure 3D shows inputs for a distractor 2 st above a normal B (B+2), see App. B. Still, only a small reset toward integrated is observed (Fig. 3E). Using the same assumptions for a deviant

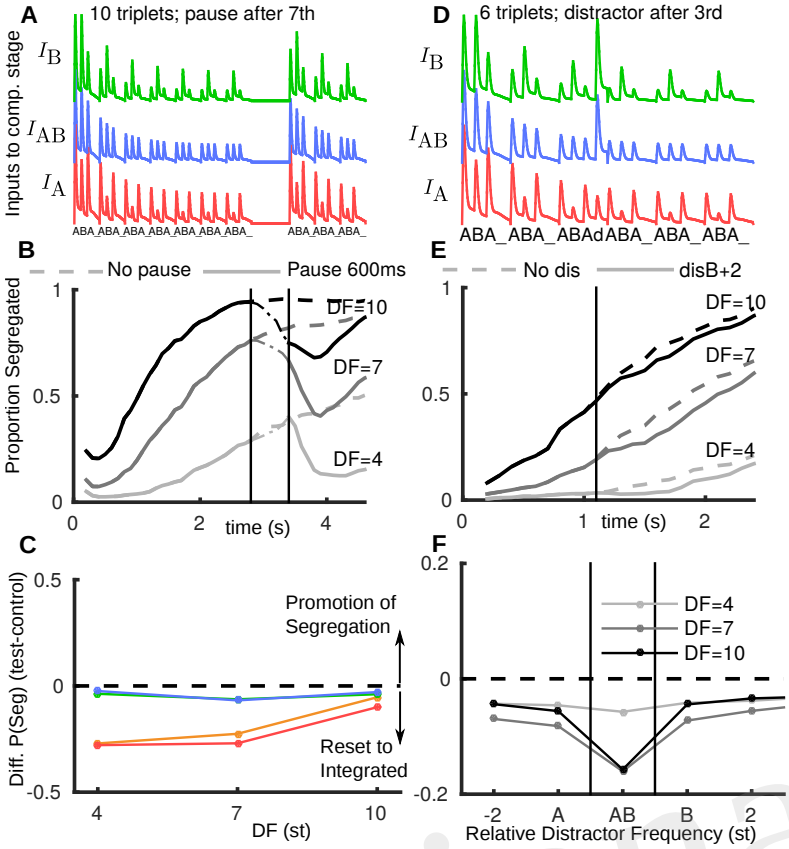

Fig. 3 Rapid recovery of adapted A1 neural responses explains reset to integration after pauses in the model. (A) A1 responses fed as inputs to competition stage with a pause in presentation; after pause inputs are unadapted. (B) Build-up functions from model with stimulus as shown in A (solid). Dashed curves without pause same as Fig 1D (solid). (C) The model captures the effect of stimulus pauses but not distractor or deviant tones (compare with Fig 2C, same color conventions). (D) Inputs to competition stage, with distractor d after third triplet at B+2 (2st above a normal B). Distractor tone response is assumed to have a normal tonotopic representation in A1, but be relatively more adapted at the A-location due to higher repetition rate and immediately following an A-tone offset. Distractor tone response in A1 is boosted as input to the competition stage, so the response to d is larger than for preceding tones. (E) Build-up function with distractor tone (solid) shows slight reset to integrated in comparison with no distractor case (dashed, as Fig 1D solid). (F) Across a range of tonotopic locations for the distractor tone, the model would predict a modest reset to integrated. Effect is largest when the distractor is at (A+B)/2 (labeled AB) and the DF is large, as the AB unit in the model would receive more input than peripheral units. Note x-axis does not have fixed spacing and distance between A and B changes with DF.

B tone at B+2 we find a similar effect (Fig. 3C). A comparison with the experimental data from Fig. 2F shows that the model has not captured the effects of deviants and distractors. A further exploration varying relative frequency of the distractor tone (Fig. 3F) shows that the model would predict a large reset toward integrated when it is at a frequency $(A+B)/2$, in which case the AB unit receives the most additional input from the distractor tone. However, this prediction was not borne out in later experiments.

Using the model, we tested a new hypothesis for how novel inputs, tones that are saliently not part of a triplet, propagate from A1 to the competition stage. These include tones not fitting the temporal pattern of a regular triplet (e.g. the distractor tone) or not matching the frequency of the tones in a regular triplet (e.g. a deviant tone); in informal listening either case is saliently different from a normal triplet. We suppose that the AB unit, encoding the integrated percept, will only receive inputs matching a normal triplet, while as before, the unexpected event results in boosted input to the competition stage (Fig. 4A). For example, a distractor tone B+2 leads to a larger than expected input at B, but no input to AB (Fig 4B). The build-up function shows an increase in segregation due to the peripheral units receiving more input. In both the distractor and deviant cases segregation is promoted, recapitulating the behavior with the reported experimental data, compare Fig. 4B (model) with Fig. 2F (experiments). Note that the model captures the largest promotion of segregation occurring for small DF.

We further applied the model to predict the dependence of change in proportion segregated on the frequency of a distractor tone (Fig. 4E). Predictions: 1) the promotion of segregation occurs for a range of relative frequencies for the distractor tone, 2) the effect is strongest when the distractor tone is close

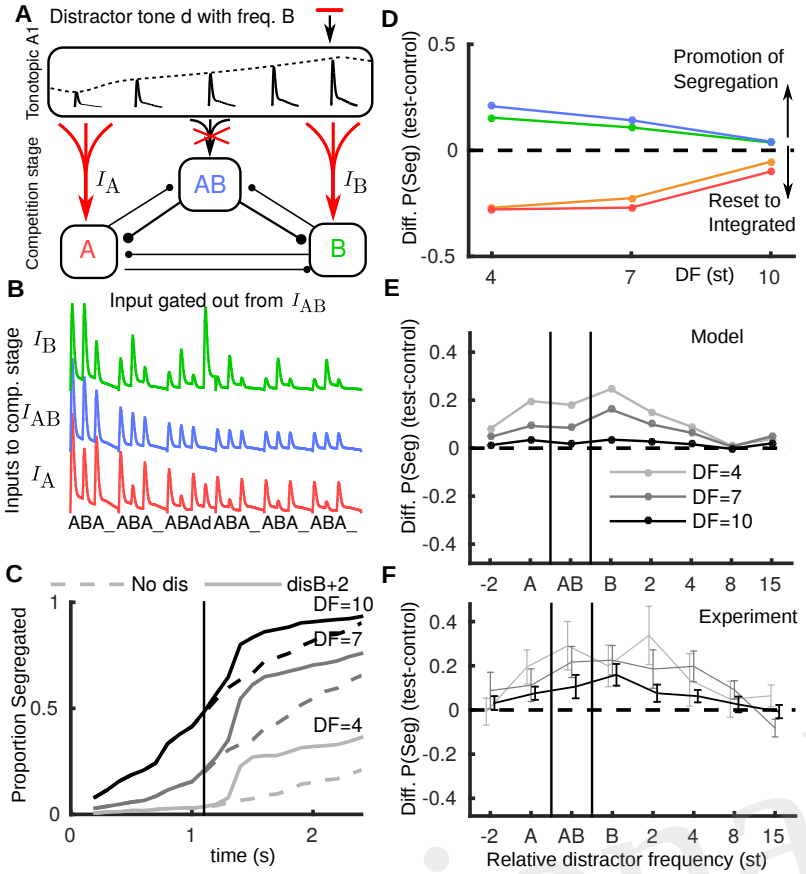

Fig. 4 Non-triplet (deviant or distractor) tones are gated out from AB unit. (A) Schematic showing how a distractor d with, e.g. the frequency of a B tone, propagates in the model when boosted to A and B units and gated out from the AB unit. (B) Model inputs from A1 with a distractor tone (at B+2) after the third triplet where it is not seen by the AB unit, contrast with Fig 3D. (C) Build-up function in this case shows that the distractor tone results in an immediate increase in segregation, contrast with Fig 3E. (D) Based on the new assumption the model captures, along with the resetting effects of pauses, the promotion of segregation for distractor and deviant tones, compare with Fig 2C (same color conventions). (E) The model predicts the largest effect for the distractor tone when it is close to the B location, that the effect is largest for small DF and that the effect diminishes if the distractor tone is too far above B or below A. Note x-axis does not have fixed spacing and distance between A and B changes with DF. (F) Experimental data showing promotion of segregation with respect to the tonotopic location of the distractor tone.

to the A and B tones, 3) there is no effect if the distractor is too far in tonotopy from the A and B tones, and 4) asymmetry, e.g. that the effect is more prominent when the distractor is near or above the B tone than when it is near or below the A tone. Further experiments confirmed these general trends for distractor tones at more frequencies (total 8) relative to the A and B tones (Fig. 4F). One experiment tested distractors aligned with the A (disA), the B (disB) or directly between (disAB). These conditions showed a significant effect on proportion segregated ($F(3,21) = 5.00, P < 0.05$). Another experiment tested distractors above B (disB+4, disB+8) and below A (disA-2). These conditions did not show a significant effect ($F(3,21) = 2.145, P = 0.125$), which is indicative of the diminishing effect of the distractors away from the A and B tones.

3 Discussion

We report new insights on the dynamics of build-up in perceptual segregation, including the initial bias toward integration, and the effects of pauses, distractor tones and deviant tones. In audition the initial percept is typically integration with segregation developing over seconds [2,3]. But such biasing toward integration has eluded neuronally-based explanation. We suggest that the initial bias is determined by broad onset activation in neurons selective to low-level features (e.g. tone frequency [11,15]) in or even below primary sensory cortex, prior to early adaptation and emergence of strong feature dependence. This property at onset biases the initial conditions that are propagated downstream of A1 to areas where identification of perceptual patterns, competition between them and build-up develops more slowly. Our study focused on auditory

streaming, but the principle could generalize to motion plaid displays consisting of two gratings moving in different directions, also showing an initial bias toward integrated pattern motion [26]. Neural responses in visual areas representing the relevant low-level feature (motion direction) show a broader initial activation and a bias toward the vector average, i.e. integrated, direction [27]. Our experiments and modeling demonstrate that the bias in the auditory case is partially restored during pauses that allow some recovery from early and fast adaptation (as brief as sub-sec), thereby allowing a refresh of the biased initial conditions. While various changes in stimuli can also interrupt build-up and reset toward integration we have discovered a class of perturbations that promote segregation rather than integration. In auditory streaming a transient perturbation that disrupts the triplet pattern (e.g., replacing B by a deviant D or adding a distractor tone d between triplets) promotes segregation. Intuitively, these events could briefly make one of the streams more salient and cause a switch. Our model provided an opportunity to seek a more mechanistic explanation. Based on our experiments and modeling we propose that non-triplet tones are processed differently downstream from primary auditory cortex. Furthermore, our results support the notion of auditory streaming being bistable between perceptual states, where a pause or aberrant tone can flip the percept in a specific direction and the perturbation's effect is still evident several seconds later.

3.1 Timescale of the reset to integrated

Using short stimulus presentations, we confirmed a partial reset to integrated for pauses of 300 or 600 ms, but did not find an increasing trend between the two conditions. A reset to integrated has been shown with pauses longer than 1 s using short stimulus presentations [28,29,30] and with briefer pauses (<1 s) using long stimulus presentations (during bistable alternations) [23]. Ref. 31 showed a reset for multi-second pauses, using EEG recordings and a mismatch negativity paradigm.

In our model, initial A1 responses had a large amplitude and broad tonotopic tuning; fast adaptation on a common timescale led to static responses with lower amplitude and tightened tuning. The tonotopic component is key for the initial bias toward integration. A rapid recovery of the fast adaptation following a stimulus pause led to a partial reset to integrated, consistent with our data. For a long enough pause there must be a full reset to integrated, as if hearing the stimulus for the first time. Ref. 32 suggested biasing from previous stimuli would have recovered within 6 s and this was confirmed by later studies [22,29]. Our results suggest that although rapid recovery of adaptation in A1 may explain the partial reset to integrated (even for very brief stimulus pauses), the multi-second timescale of longer term recovery may also be related to processes downstream of A1.

3.2 Link between context and perturbations

A sudden change after a sequence of context triplets causes at least partial resetting of build-up back toward integration, as shown for a change in ear of presentation [2], a shift in perceived loudness and/or location [5,6], a switch in attention [4,33] and a pause in presentation (as described above); see review ref. 34. Like a pause, a switch in attention could allow recovery from adaptation. Otherwise a one-time shift of the entire stimulus in location or intensity (an increase, but not decrease [6]) could recruit previously unstimulated and, therefore, unadapted neurons. We may view the triplets preceding the deviant and distractor tones during build-up as context. Different types of context can bias perception toward (i.e. prime for) integration or segregation [35,29,30,36]. Even for a context of a single stream of tones, say A_A_A_A_, that would alone promote segregation for subsequent test triplets ABA_ ABA_..., similar disruptions as above at the end of the context sequence lead to integration, as if the effect of the context was undone [5,6,22]. Also, a single deviant A' at the end of an A_A_A_... context can reduce or eliminate the expected bias toward segregation [37,21,38]. So while these various disruptions favor integration and may a priori lead one to a generalized expectation that a single transient distractor tone (between triplets) or a single deviant tone (within a triplet) should also promote integration, we found the opposite — promotion of segregation in the subsequent triplets. Nevertheless our results do not contradict these previous studies. Studies looking at the effects of deviant tones did so by placing these at the end of a single stream context [37,21,38]; in our study we placed the deviant or distractor at the end of context triplets. Ref. 33 included an experiment with a single delayed-onset deviant B tone, but did not report promotion of segregation or resetting. Nevertheless, other stimulus perturbations may promote segregation. Further experiments should consider whether single-tone deviants in features other than frequency (e.g. lateralization or loudness) can promote segregation.

3.3 Promotion of segregation and differential processing of non-triplet events

Our model accounts for the observed segregation-promoting effects by assuming that inputs propagate to the competition stage in a differential manner, A1 responses to a deviant or a distractor tone do not reach the "integration (AB) unit". It encodes a non-trivial rhythm and can be viewed as more sensitive to, effectively selective against, sounds that break the triplet pattern. Our implicit assumption is that the aberrant tone is identified as a mismatch and is deflected from reaching AB. Viewed differently, an incoming sound inconsistent with the integrated percept might result in the integration unit being briefly suppressed, allowing the peripheral units to take over. The crucial aspect is that the incoming tones have a differential effect on the integrated and segregated units. The effects of distractor tones also show a dependence on tonotopy, which led us to favor an input-based explanation.

Our results allow us to rule out some other potential explanations for the effects of non-triplet perturbations. Suppose that such perturbations indiscriminately cause a switch in perception away from the current percept. One might argue that we saw switches only from integrated to segregated since we considered perturbations only during build-up, when integration is thought to be dominant. However, our data do not support the idea of switches from segregated to integrated. At DF=10, where $\sim$50% of trials are already segregated after 3 triplets, we saw no evidence of a reset or switch back toward integration (Fig. 2E), either in the group data, or in individual subjects (not shown). This facet of the data is consistent with the proposed notion that input propagates from A1 to the segregated units, but not the integrated unit. Another hypothesis could be that any transient, salient perturbation distinct from standard triplets promotes segregation. However, our data showed no effect for distractor tones sufficiently far in frequency from the A or B tones. Our modeling work shows that this interaction could be through input from the distractor tones still propagating to segregated units with tonotopic dependence.

Ref. 38 showed that hearing a single A tone before the triplets could make that stream more salient. Could there be a similar effect in our data, where the distractor tone makes one of the streams more salient, or briefly direct attention toward it? The range of conditions for which we found promotion of segregation includes several cases where the perturbation is not an A or a B tone. The distractor tone d appears in a sequence ... ABA_ABAdABA_.... It could be that the d is being grouped into a new triplet (AdA or B-d-B), thus making the A or B stream more salient (or highlighting their separation) ahead of the upcoming test triplets. For a distractor or deviant tone, the proposed mechanism in our model boosts inputs to the competition stage for the segregated units whilst gating out input to the integrated unit. This selective transient modulation of input gains could be viewed as a brief top-down attentional effect. However, for an attention mechanism, the selective gain would likely act in response to the perturbation mismatch with some delay. In our current model we have idealized the transmission of input from A1 to competition stage without a delay.

3.4 Build-up and bistability in models

Most existing computational models of auditory streaming have focused on reproducing the dependence[1] of perceptual bias on DF and presentation rate[39,40,41], the dynamics of build-up[42,43] or both[44]. A complete theoretical framework for streaming should account for build-up as well as the later alternations, given that the probability of perceiving segregation converges to the long-term probability of bistable alternations. Some recent models focused on post-build-up alternations (auditory bistability)[45,46,17,47]. The initial bias to integration is set by specifying a priori initial conditions[46,47]. In ref. 45 the bias is emergent through an early stage of an algorithmic pattern discovery. Our model that accounts for alternations, and was further developed here to describe build-up, is the first treatment to explain the initial bias for integration through a direct link to observed neurophysiological responses[9,11]. To the best of our knowledge, no other model has been used to investigate resetting effects, or the effects of perturbations in general.

3.5 Future directions for our model

Our current neuromechanistic model relies on a lumped version of a distributed network, a few discrete units pool inputs from different tonotopic locations in A1. Although this view allows the model to account for many phenomena (stimulus parameter dependence, build-up, alternations, resetting for pauses), the notion of differential processing introduced to account for promotion of segregation approaches the limit of our modeling framework, and suggests the need for a richer description. One avenue for extension would be to consider a continuous feature space in DF, as proposed in ref. 39, at least at the A1 stage of the

model. Although the rules for the tonotopic spread of A1 responses allowed us to considered, for example, distractor tones away from the three locations A, B and (A+B)/2, a more refined description would define how A1 responds in time to any combination of tones across DF (and consider other paradigms, for example, involving frequency-banded maskers[48]). As a further extension we could introduce an additional dimension to the feature space, e.g., selectivity to different repetition rates of the streams and the relative timing (phase) of the inputs. A first step in this direction was made in[44], with the use of delay lines to introduce a temporal feature space. Beyond this, a suitable theoretical framework to study might be a coupled oscillator network sensitive to frequencies in the range of the repetition rates of the tone sequences, not just the frequency of the tones (like tonotopy in A1). fMRI studies have implicated a broader network involved in streaming, including areas associated with rhythm and timing[49]. The design of such a network and the necessary mechanisms for competition could build directly on our present model. Such networks have been used in studies of rhythm perception[50] and in phenomenological studies of perceptual bistability[51]. Such a richer description would allow one to pursue the origin of the differential processing we propose here and to investigate the effects of temporal coherence, a strong cue in auditory stream segregation[52].

3.6 Conclusion

Our model with the developments presented here is the first grounded in neurophysiological detail to account for build-up and subsequent bistable alternations. We propose that the initial bias to integrated arises naturally from the rapid but delayed emergence of low-level feature dependence and that the more gradual timescale of build-up comes from competition mechanisms downstream of A1. This is the first explanation of integration bias and build-up motivated directly from neurophysical data (responses to triplet sequences in A1[11]).

New findings presented here challenge the current understanding of how the segregation of auditory objects is affected by interruptions and perturbations. A reset of the build-up process results from an established class of perturbations that shift the entire triplet stimulus in location, loudness or timing. We illustrate that the rapid recovery of responses in A1 can explain resetting for stimulus pauses. We demonstrated a new and opposite effect, promotion of segregation, by a complementary class of perturbations that transiently alter a single triplet or introduce a new non-triplet element. Our modeling in conjunction with confirmed experimental predictions led to a new hypothesis: that new non-triplet events (deviant or distractor tones) are gated out from the neural population encoding the complex integrated rhythm.

4 Materials and methods

4.1 Neuromechanistic model

The neuronal circuits for competition and perceptual encoding are assumed to be downstream and receiving inputs from A1. The periodic inputs mimic the A1-responses to ABA- sequences reported in ref. 11. Neuronal activity is described by mean firings rates and competitive interactions emerge through a combination of excitatory and inhibitory connections, slow adaptation and intrinsic noise. We provide a brief outline of the model architecture, mechanisms and inputs here; the full model equations and further details in the App. B.

The schematic in Fig. 1A shows downstream units A, B and AB that respectively pool inputs from regions of A1 centered at locations with best frequencies A, B and the midpoint between (A + B)/2. We associate a variable r_k ($k = \{A, AB, B\}$) with each unit representing the mean firing rate of a population of neurons in the competition network. For each unit r_k the intrinsic dynamics are illustrated in Fig. 1A and described by a differential equation like the following,

$$\tau_r \dot{r}_{\mathrm{AB}} = -r_{\mathrm{AB}} + F\Big(\beta_e e_{\mathrm{AB}} - \beta_i(r_{\mathrm{AB}} + r_{\mathrm{A}} + r_{\mathrm{B}}) - g a_{\mathrm{AB}} + I_{\mathrm{AB}} + \chi_{\mathrm{AB}}\Big). \quad (1)$$

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 cortical timescale is $\tau_r = 10\,\mathrm{ms}$. A sigmoidal firing rate (smooth threshold) function F (see App. B) process all inputs to the unit. Local excitation e_{AB} has strength $\beta_e = 0.65$ and evolves on an NMDA-like timescale $\tau_{\mathrm{e}} = 70\,\mathrm{ms}$. Global inhibition (assumed instantaneous and so proportional to the cortical variables r_k) has strength $\beta_i = 0.3$. Note $\beta_e > \beta_i$ so there is net local excitation. Linear spike frequency adaptation (slow negative feedback) a_{AB} has strength $g = 0.045$ and a timescale of 1.4 s.

The input I_{AB} mimics A1 cortical responses to triplet tone sequences; full details are given in App. B. There are two components to the early adaptation of these responses, both consistent with observations from ref. 11 and sharing a common timescale $\tau_{\mathrm{A1}} = 500\,\mathrm{ms}$ (Fig. 1B). Firstly, the overall amplitude of

responses decays. Secondly, the effective DF is initially small i.e. the DF dependence of the responses takes time to emerge. After a stimulus pause, the A1 adaptation is assumed to rapidly recover ($\tau_{\rm rec} = 100$ ms), such that when the stimulus resumes after a an adequate pause (say $2 \times \tau_{\rm rec}$) the model inputs resemble those after initial stimulus onset (Fig. 3B). For a distractor tone (or a deviant tone) input amplitude and tonotopic spread are consistent with a partially recovered response to the tone. At the tonotopic location A, responses to a distractor are reduced, because the distractor immediately follows the offset of a normal A tones (referred to as temporal forward masking in ref. 9). Intrinsic additive noise $\chi_{\rm AB}$ is an independent Ornstien-Uhlenbeck process for each r_k.

Numerical simulations were implemented with an Euler-Murayama scheme with a stepsize of $0.5\tau_r$. Build-up functions were computed as time-binned averages across 500 simulations. For each time bin the fraction of trials with more activity in the AB unit than the summed activities of the A and B units was taken as the measure of proportion integrated. Computations were implemented in `Matlab` and batch processed using the function `parfor`; no special computing hardware was required. In all computations, the same set of 500 randomized initial conditions and the same 500 instantiations of the noise process (i.e. frozen noise) was used for each r_k. This ensures that any differences between conditions is entirely due to changes to model parameters (e.g. reflecting different stimulus properties). For example, in Fig. 3B and E, the control (No pause, No dis) curves only deviate from the test simulations (Pause 600ms, disB+2) from the time point where the perturbation is introduced.

4.2 Psychoacoustic experiments

Our experimental paradigm is well suited for studying the effects of perturbations on how the subsequent triplets are perceived. In all experiments presented here (with pauses, distractors or deviants) the perturbation was followed by three normal triplets and subjects reported their perception of the final triplet, roughly 1 s after the perturbation. Three triplets provides enough stimulus duration to make a reliable perceptual judgement [20,21]. This design precludes the possibility of subjects reporting, say a distractor tone, as being its own segregated stream, as the distractor occurs well before the final triplet. If continuous perceptual reports were used, confusion might arise about classifying an unexpected tone into it's own stream at the moment the distractor is detected. A final possibility would be to use an objective measure of streaming [37,33]. An appropriate paradigm could be the one used in ref. 33, where performance in a deviant detection task functioned as an objective measure for streaming and showed qualitative agreement with subjective perceptual reports. In the objective task, subjects had to detect a single delayed-onset B tone and performance was best during integration. Given that the objective task relies on the detection of a delayed-onset deviant and that some trials would need to involve another deviant tone (the perturbations studied here), it could become rather confusing for a subject. It would be challenging for a subject to distinguish between multiple types of aberrant tone, ignoring some and reporting the presence of others.

Procedure. Subjects sat in an acoustically shielded chamber and pressed keys on a keyboard to indicate their perceptual response. In each task, a short ABA- sequence ranging between three and 10 triplets was played, and the subjects reported with button presses whether the last triplet of the sequence sounded most like the integrated percept or the segregated percept and guessed if unsure. The integrated percept was defined as hearing the A and the B tones together in a galloping rhythm, and the segregated percept was defined as hearing the A tones and B tones separately in two distinct streams. Subjects were instructed to respond as quickly as possible and had up to 5 s — the length of the inter-stimulus interval (ISI) — to respond.

Stimuli. The repeating ABA_ triplet consist of 100 ms pure tones with 10 ms linear ramps, where the '_' indicates a silence also lasting 100 ms; in total, the duration of each ABA_ triplet is 400 ms. An inter-trial interval of 5 s was included between all trials. The higher frequency B tones are a variable DF semitones (st) above the lower frequency A tones. Cosine squared ramps with 10 ms rise and fall times were used. Each tone sequence was played binaurally through Etymotic headphones at 65 dB SPL. Three DF conditions were used for all experiments: DF $\in \{4, 7, 10\}$ st. From trial to trial the A-tone base-frequencies were roved between 420 Hz and 1060 Hz, separated by intervals of 4 st; correspondingly, the B tone frequencies ranged between 530 Hz and 1888 Hz. The roving of base frequencies and ISI of 5 s were chosen to avoid any latent adaptation from one trial to the next [22,30,53].

Subjects. Seventeen subjects in total, including one of the authors, took part in the experiments (10 female, 7 male), aged 20-51, mean age 27.9. Subjects were reimbursed for their participation and all experimental procedures complied with human subject research guidelines as approved by the University Committee on Activities Involving Human Subjects at New York University (IRB-FY2016-310). All subjects provided written informed consent and were required to pass a hearing screening.

Conditions. The stimulus paradigm for the pause experiment is shown in Fig. 2A. A total of 15 conditions (3 DF conditions crossed with 5 stimulus length/pause combinations) were tested with 20 repetitions of each condition (total of 300 trials for each of 8 subjects). Two test conditions consistent of 7 context triplets, followed by a pause of 300 or 600 ms followed by 3 test triplets 8 blocks of 15 trials. Three control conditions of 3 (test only), 7 (context only) and 10 (no pause control) triplets were tested in 9 blocks of 20 trials and the test conditions. Control and test conditions were run in separate block sections to avoid confusion about timing of perceptual reports.

Schematics of the stimulus paradigm for the distractor and deviant experiments are shown in Figs. 2B and C. Three different experimental sessions, with eight subjects each, were conducted for different experimental conditions. Subjects performed 20 blocks of 15 trials each, where the length of each trial ranged from 1.2 s to 2.4 s in length. In each experiment, two control conditions included a 3 triplet and a 6 triplet condition with no deviants or distractors. Along with the two control conditions, each experiment included three distractor or deviant conditions, 6 triplets in length. Distractor tones were 50 ms in length and were inserted symmetrically in the 100ms inter-triplet gap between the third and fourth triplets of the sequence, so that there was 25ms of silence on either side of the distractor. Across the three experimental sessions, the following frequencies (in st, relative to the A and B tones of the triplets) of distractor tones were tested: A-2, A, (A+B)/2, B, B+2, B+4, B+8, B+15. Deviant tones involved a change in frequency to the B tone of the third triplet. In the one deviant tone condition tested, the B-tone was increased in frequency by 2 st.

Acknowledgements

The authors express their gratitude to Rodica Curtu, Yonatan Fishman, David Heeger, Antje Ihlefeld, Andrew Oxenham and Elyse Sussman for helpful feedback on earlier versions of this manuscript. J Rankin was supported by a Swartz Foundation postdoc grant. POP received support from the NYU DURF undergraduate research program. Partial support for J Rinzel comes from an NIH K-18 award, K18 DC011602.

Appendix

A Statistical analyses for experimental results

All statistical analyses presented here utilized the software `R`[54] with the package `ez`[55], which produces repeated measures analysis for variance (ANOVA) while handling sphericity tests and appropriate corrections to p-values where necessary[56].

For Experiment 1 the aim was to reproduce a known result, that brief stimulus pauses result in a partial reset towards the integrated percept. In test conditions a 300 or 600 ms pause in a ten triplet presentation was inserted after the seventh triplet, leaving three test triplets at the end. Reference control conditions (3 or 7 triplets) and the main control condition (10 triplets) reveal the behaviour with no pause (10), before the pause (7) and for the test triplets on their own (3). Data from all conditions in this experiment are shown in Fig. 5A. A first analysis shows that the build-up is occurring for the control conditions, that is, increasing proportion segregated with DF and presentation length. An ANOVA table for repeated measures ($N = 8$, as in all experiments) within-subject factors DF and *cond* (presentation length) for the three control conditions is labelled Experiment 1A in Table 1. In this section the term *cond* represents the relevant set of conditions in each experiment, refer to the headings for each experiment in Table 1. In general, we report significant effects at the standard $\alpha = 0.05$-level and, where a Mauchly Spheriricity test reached significance for the given factor, we report the Greenhouse-Geiser corrected p-value p_{GG}. The factors DF, *cond* and their interaction showed significant effects (Exp. 1A Table 1). Next, we compare the relevant control condition with the test conditions (dashed black and red/orange curves Fig. 5A). The effect *cond* for these conditions is tested in Experiment 1B in Table 1; we found a significant effect of DF and *cond*, but not their interaction. The observed reset to integrated for short stimulus pauses is significant.

In Experiments 2–4 the effect of eight distractor cases and one deviant case were tested across three experiments. Each experiment had the same design with control conditions of 3 and 6 triplets and three test conditions (Fig. 5B–D). In each figure the relevant comparison is between the main control condition (6 triplets) and the test conditions (color). In general the test conditions promote segregation and we report whether these effects were significant for each experiment (Table 1). For Experiment 2 there was a significant effect of DF, *cond* and their interaction. Visual inspection of Fig. 5B suggests the significant effect of *cond* comes from increases in promotion segregated for the disB+2 and devB+2 conditions. For Experiment 3 there was a significant effect of DF, but not significant effect for *cond* or their interaction. Visual inspection of Fig. 5B shows that the disB+4 had the largest effect. These data support the notion that distractor tones far from the As and Bs have less of an effect than those close to A and B. For Experiment 4 there was a significant effect of DF and *cond*, but not their interaction. Visual inspection of Fig. 5B shows that the disAB condition (distractor at (A+B)/2) had the largest effect.

There was some but not complete overlap in the subjects participating in each Experiments 2–4. We therefore wanted to check for consistency in the control conditions across the experiments to ensure making comparisons for test conditions across the experiments is relevant (Fig 4F). Table 1 (bottom) shows an ANOVA table including exp (experiment number) as a between subjects factor. We found no significant effects for exp, or its interactions with other factors, confirming that comparison across these experiments is appropriate.

One might wish to apply post-hoc tests to further explore the significant effects for the variable *cond* in the ANOVAs reported above. Visual comparison of test conditions (colored curves) with the relevant control condition (dashed black

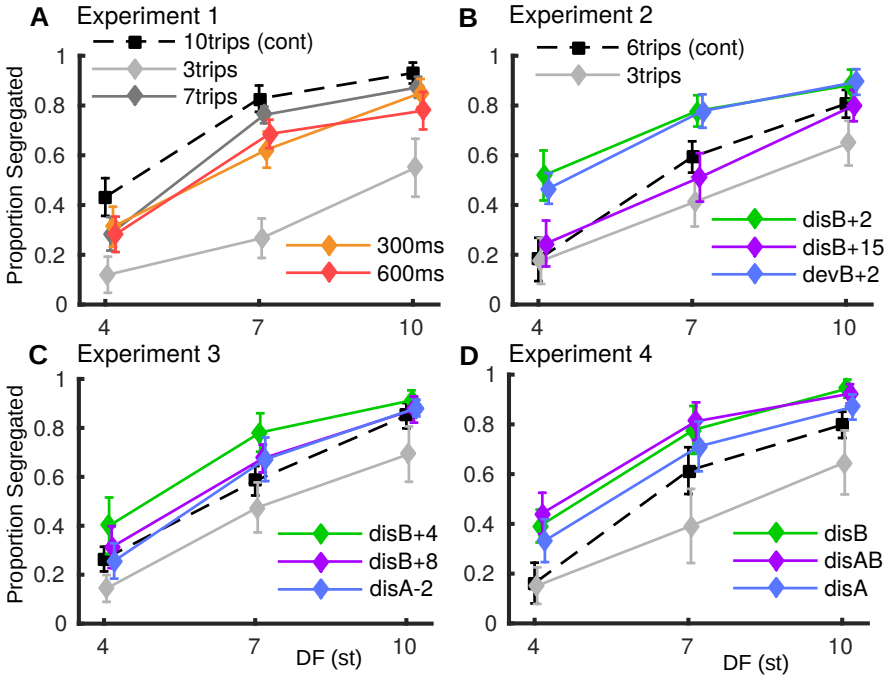

Fig. 5 Experiments 1–4. (A) Pause experiment with three control conditions for fixed-length presentations with indicated number of triplets (black/grey curves). Control conditions were plotted and compared with the model in Fig. 1E. Test conditions with pause duration indicated (orange/red). Test conditions plotted with two control conditions in Fig. 2A. (B) Distractor and deviant experiment with two control conditions with indicated number of triplets (black/grey). One deviant and two distractor cases were tested (blue/green/purple curves). One deviant and one distractor condition were plotted in Fig. 2B. (C) As B for additional distractor cases tested in Experiment 3. (D) As B for additional distractor cases tested in Experiment 4. All distractor conditions from Experiments 2–4 were plotted in Fig. 4F.

curve) for each experiment represented in Fig. 5 shows multiple cases that might be significant if tested in post-hoc comparisons. Indeed, for many of the conditions paired t-tests between the control and test conditions do reach significance at the $p < 0.05$ level. However, to rule out the chance of making Type I errors due to multiple comparisons being made (6 comparisons for Experiment 1, 9 comparisons for Experiments 2–4), it is appropriate to apply a Bonferroni adjustment to the significance levels. No conditions reach significance with the conservative Bonferroni adjustments. We note that applying a Tukey Honest Significant Differences analysis is not appropriate with our repeated measures experimental design [56].

B Details of the model

The network structure and neural mechanisms forming the basis of our model were originally motivated in [17]. In this section we give a complete description of the model, specifying the exact formulation used in the present study. The firing rate variables r_k are indexed by $k = \{AB, A, B\}$ for each population shown in Fig. 1A with the associated adaptation a_k and recurrent excitation e_k variables (note that 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:

$$\begin{aligned}
\tau_r \dot{r}_{\mathrm{AB}} &= -r_{\mathrm{AB}} + F\Big(\beta_e e_{\mathrm{AB}} - \beta_i(r_{\mathrm{AB}} + r_{\mathrm{A}} + r_{\mathrm{B}}) - g a_{\mathrm{AB}} + I_{\mathrm{AB}} + \chi_{\mathrm{AB}}\Big),\\
\tau_r \dot{r}_{\mathrm{A}} &= -r_{\mathrm{A}} + F\Big(\beta_e e_{\mathrm{A}} - \beta_i(2r_{\mathrm{AB}} + r_{\mathrm{A}} + r_{\mathrm{B}}) - g a_{\mathrm{A}} + I_{\mathrm{A}} + \chi_{\mathrm{A}}\Big),\\
\tau_r \dot{r}_{\mathrm{B}} &= -r_{\mathrm{B}} + F\Big(\beta_e e_{\mathrm{B}} - \beta_i(2r_{\mathrm{AB}} + r_{\mathrm{A}} + r_{\mathrm{B}}) - g a_{\mathrm{B}} + I_{\mathrm{B}} + \chi_{\mathrm{B}}\Big),\\
\tau_a \dot{a}_{\mathrm{AB}} &= -a_{\mathrm{AB}} + r_{\mathrm{AB}},\\
\tau_a \dot{a}_{\mathrm{A}} &= -a_{\mathrm{A}} + r_{\mathrm{A}},\\
\tau_a \dot{a}_{\mathrm{B}} &= -a_{\mathrm{B}} + r_{\mathrm{B}},\\
\tau_{\mathrm{e}} \dot{e}_{\mathrm{AB}} &= -e_{\mathrm{AB}} + r_{\mathrm{AB}},\\
\tau_{\mathrm{e}} \dot{e}_{\mathrm{A}} &= -e_{\mathrm{A}} + r_{\mathrm{A}},\\
\tau_{\mathrm{e}} \dot{e}_{\mathrm{B}} &= -e_{\mathrm{B}} + r_{\mathrm{B}},
\end{aligned} \tag{2}$$

with time constants $\tau_r = 10\,\mathrm{ms}$ (cortical), $\tau_a = 1.4\,\mathrm{s}$ (spike frequency adaptation), $\tau_{\mathrm{e}} = 70\,\mathrm{ms}$ (NMDA-excitation). The strength of recurrent excitation is given by $\beta_e = 0.65$, lateral inhibition $\beta_i = 0.3$ and adaptation $g = 0.045$. Note that the profile of inhibition used here, with non-uniform synaptic weights and independent of DF, was determined after fitting the model to behavioural data [17]. Note that although within-unit inhibition is included, $\beta_e > \beta_i$, so there is always net

Experiment 1A: cond = {3 trip; 7 trip; 10 trip control}					
Effect	dfn	dfd	F	p [† if GG-corrected]	ges
DF	2	14	37.486	**0.000**	0.571
cond	2	14	19.486	**0.001**†	0.492
DF:cond	4	28	4.336	**0.007**	0.094
Experiment 1B: cond = {10 trip control; 300 ms gap; 600 ms gap}					
Effect	dfn	dfd	F	p [† if GG-corrected]	ges
DF	2	14	22.945	**0.001**†	0.603
cond	2	14	5.126	**0.021**	0.127
DF:cond	4	28	1.457	0.242	0.021
Experiment 2: cond = {6 trip control; disB+2; disB+15; devB+2}					
Effect	dfn	dfd	F	p [† if GG-corrected]	ges
DF	2	14	19.511	**0.000**	0.521
cond	3	21	5.796	**0.038**†	0.203
DF:cond	6	42	2.414	**0.043**	0.055
Experiment 3: cond = {6 trip control; disB+4; disB+8; disA-2 }					
Effect	dfn	dfd	F	p [† if GG-corrected]	ges
df	2	14	29.461	**0.000**†	0.624
cond	3	21	2.145	0.125	0.064
df:cond	6	42	0.873	0.523	0.017
Experiment 4: cond = {6 trip control; disB; disAB; disA}					
Effect	dfn	dfd	F	p [† if GG-corrected]	ges
DF	2	14	53.899	**0.000**	0.604
cond	3	21	5.002	**0.035**†	0.167
DF:cond	6	42	1.632	0.162	0.029
Experiments 2-4 compare controls: cond = {3 trip; 6 trip control}, exp = {2; 3; 4}					
Effect	dfn	dfd	F	p [† if GG-corrected]	ges
exp	2	18	0.057	0.945	0.003
DF	2	36	49.041	**0.001**†	0.491
cond	1	18	12.191	**0.003**	0.067
exp:DF	4	36	0.029	0.998	0.001
exp:cond	2	18	0.018	0.982	0.000
DF:cond	2	36	4.372	**0.020**	0.016
exp:DF:cond	4	36	0.906	0.471	0.007

Table 1 Analysis of Variance (ANOVA) tables for repeated measures experiments (N=8 subjects) shown in Fig. 5. Columns are effect degrees of freedom (dfn), error degrees of freedom (dfd), F-value, p-value, generalized eta-squared effect size (ges). Significant p-values (0.05 significance level) are bold. A star indicates that the Greenhouse-Geiser corrected p-value was used due to Mauchly's sphericity test reaching significance at the $\alpha = 0.05$. In all experiments frequency difference conditions were DF ={4,7,10}. The first table (Experiment 1A) compares the control conditions of different lengths for Experiment 1 (Fig. 5A). The second (Experiment 1B) compares the main 10 triplet control with the test conditions. Similarly for Experiments 2–4 comparing the 6 triplet control with the test conditions. The last table compares the 3 and 6 triplet control conditions across Experiments 2–4. Each experiment had a different set of N=8 subjects but we found no effect for exp (subject group), i.e. the subject groups gave similar results for the controls. This demonstrates that it is relevant to compare data from the test conditions in Experiments 2–4, as done in Fig. 4F.

within-unit excitation. The firing rate function F is given by

$$F(u) = \frac{1}{1 + \exp\left(k_F(-u + \theta_F)\right)}, \tag{3}$$

where $\theta_F = 0.2$ is a threshold parameter and $k_F = 12$ is a slope parameter.

Additive noise is introduced with independent stochastic processes χ_{A}, χ_{B} and χ_{AB} added to the inputs of each population. Input noise is modeled as an Ornstien-Uhlenbeck process:

$$\dot{\chi_k} = -\frac{\chi_k}{\tau_d} + \gamma\sqrt{\frac{2}{\tau_{\mathrm{X}}}}\xi_k(t), \tag{4}$$

where $\tau_{\mathrm{X}} = 100\,\mathrm{ms}$ (a standard choice[19,57]) is the timescale, the strength γ equals 0.0875 and $\xi(t)$ is 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.

B.1 Model inputs and early adaptation

The particular form of the periodic inputs are based on recorded responses from A1 with ABA_ triplets [11]. We capture the basic form of these responses to tones (TR) 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:

$$\mathrm{TR}(t) = H(t)\left[\frac{\exp(2)}{\alpha_1^2}t^2\exp\left(\frac{-2t}{\alpha_1}\right) + \Lambda_2\frac{\exp(2)}{\alpha_2^2}t^2\exp\left(\frac{-2t}{\alpha_2}\right)\right], \tag{5}$$

with plateau amplitude fraction $\Lambda_2 < 1$ and rise times $\alpha_1 < \alpha_2$. The constant terms $\frac{\exp(2)}{\alpha_{\{1,2\}}}$ terms normalise the amplitude at $t = \alpha_{\{1,2\}}$ of the individual onset functions to 1. A standard Heaviside function H (step function where $H(t) = 0$ zero for $t < 0$ and $H(t) = 1$ for $t > 0$) 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 [11].

The spread of input is defined via the weighting function

$$w_p(\mathrm{DF}, t) = Q(t)I_p\exp\left(\frac{-R(t)\mathrm{DF}}{\sigma_p}\right), \tag{6}$$

where the tonotopic decay constant is $\sigma_p = 9.7$ st, the input amplitude is $I_p = 0.6$, $R(t)$ represents effective DF adaptation (increasing with time) and $Q(t)$ represents amplitude adaptation (decreasing with time). These are the two components of the early fast-adaptation in A1 sharing a common timescale $\tau_{\mathrm{A1}} = 500$ ms. The tonotopic spread of inputs in A1 evolves with time according to

$$R(t) = 1 - (1-p)\exp(-t/\tau_{\mathrm{A1}}), \tag{7}$$

where the initial DF fraction is $p = 0.1$ ($R(t)$ rises from 0.1 to 1; effective DF rises from 0.1DF to DF). The input amplitude evolves according to

$$Q(t) = 1 + m\exp(-t/\tau_{\mathrm{A1}}), \tag{8}$$

where the $1 + m$ ($m = 2.5$) is the initial input amplitude factor ($Q(t)$ decays exponentially from 3.5 to 1; input amplitude decays from $3.5I_p$ to I_p).

In order to specify the amount of input received by each unit, I_{AB}, I_{A} and I_{B}, in (2), we first construct sequences of tone responses $\mathrm{TR}_A(t)$ (A_A_...) and $\mathrm{TR}_B(t)$ (_B_...) where the tones and silences ("_") each have a duration of 100 ms. Inputs for a repeating ABA_... sequence are given by

$$\begin{aligned} I_{\mathrm{AB}}(t) &= w(DF/2, t)(\mathrm{TR}_A(t) + \mathrm{TR}_B(t)),\\ I_{\mathrm{A}}(t) &= w(0, t)\mathrm{TR}_A(t) + w(DF, t)\mathrm{TR}_B(t),\\ I_{\mathrm{B}}(t) &= w(DF, t)\mathrm{TR}_A(t) + w(0, t)\mathrm{TR}_B(t), \end{aligned} \tag{9}$$

and plotted in Fig 1B. Respectively, equations (7) and (8) describe the early adaptation on the timescale τ_{A1} of the effective DF and the amplitude of responses in A1; see Fig 1B "Early adaption". After this initial adaptation during ~ 3 triplets, $w(DF, t_{\mathrm{late}}) = I_p\exp(-DF/\sigma_p)$ is independent of time; see Fig 1B "Static inputs". After a stimulus pause, both components recover on a timescale $\tau_{\mathrm{rec}} = 100$ ms. The amplitude component can recover completely (8) and the tonotopic spread partially recovers ($p = 0.325$ in (7), rather than 0.1); see Fig 3A.

We now specify how the formulation of the model in the present study relates to the one in [17]. In our previous study a slow synaptic depression on the recurrent excitation was introduced, but here we assume this does not play a role in the build up phase, i.e. we use static excitation (denoted e_{fix} in [17]). To maintain a match to our experimental data under this assumption g, β_e, γ and I_p were adjusted relative to the values used in [17]. In the present study we use global, rather than DF-dependent, inhibition (denoted i_{gbl} in [17]), see our previous paper for further discussion on this point. The input terms in (2) given by I_{AB}, I_{A}, I_{B} refer to the input to the competition stage, which may be different to the A1 responses, e.g. particular when inputs from distractor tones are gated out; see Fig 4A.

B.2 Inputs from distractor and deviant tones, simple implementation of SSA

For a distractor tone at tonotopic location d, or a deviant tone at tonotopic location D, the amplitude response in A1 can be computed in terms of the frequency difference $\mathrm{DF_d}$ (or $\mathrm{DF_D}$) between d (or D) and the tonotopic locations A, (A+B)/2 and B. The weighting function for a distractor (similarly for a deviant) is given by

$$w_{\mathrm{d}}(\mathrm{DF_d}, t) = I_{\mathrm{d}}Q(t)\exp\left(\frac{-\mathrm{DF_d}}{\sigma_{\mathrm{d}}}\right), \tag{10}$$

where, the amplitude adapts through $Q(t)$ and the tonotopic spread is assumed broad $\sigma_{\mathrm{d}} = 2.7\sigma_p$ (for example when above or below the A and B tones). In the presence of the ABA_ triplets, the A location is hit by more tones and, if a distractor immediately follows at A, it will be significantly adapted due to stimulus specific adaptation (SSA) [24,25] in A1. As such, a relatively smaller response is assumed at the A location (factor 0.5 in (11)). This ad hoc, straightforward implementation of SSA is illustrated in Fig. 6B. We provide a more general implementation of SSA below. We now let $\mathrm{TR_d}(t)$ (..._____d_____...) represent an impulse (5) at the specific time of the additional distractor tone. A distractor tone, as a salient new event, is assumed boosted ($I_{\mathrm{d}} = 2.8I_p$) when it is it integrated as input to the competition stage (see Fig

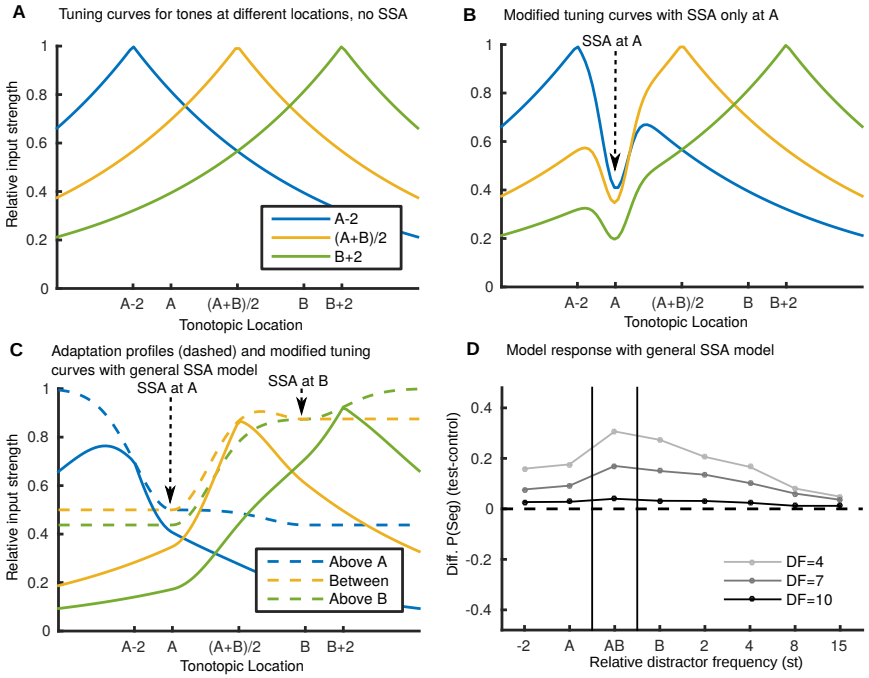

Fig. 6 Amplitude of A1 responses for distractor tones at different locations relative to A and B tones. (A) Tonotopic tuning of responses to tones in A1 at locations A-2, (A+B)/2 and B+2 without SSA (e.g. responses to isolated tones with no prior input). With no SSA, the tuned response is translated horizontally depending on the location of the tone. (B) Representative tuning curves with SSA only at the A location. (C) Tuning curves with general SSA model. More tones arrive at the A location and it will be more adapted than the B location. The profile of adaptation is shown for tones below (dashed blue), between (dashed yellow) or above (dashed green) the A and B tones. Solid curves show the tonotopic tuning of responses for tones at different location (legend in A); these are computed by multiplying the tuning curves in panel A with the adaptation profiles (dashed curves) in B. (D) As Fig 4E with general SSA model rather than SSA only at A location. Shows change in proportion segregated as a function of distractor location relative to A and B tones. Note x-axis does not have fixed spacing and distance between A and B changes with DF. Apart than the SSA model, the same assumptions are used (boost of inputs to A-unit and B-unit, no input to AB-unit (Fig 4A).

3D, where the distractor tone d gives larger amplitude input to the competition stage than preceding tones). For, say, a distractor tone at tonotopic location B+2 the modified inputs would be

$$\begin{aligned}\widehat{I_{\mathrm{AB}}}(t) &= I_{\mathrm{AB}}(t) + w_{\mathrm{d}}(\mathrm{DF}/2+2)\mathrm{TR}_{\mathrm{d}}(t),\\ \widehat{I_{\mathrm{A}}(t)} &= I_{\mathrm{A}}(t) + 0.5w_{\mathrm{d}}(\mathrm{DF}+2)\mathrm{TR}_{\mathrm{d}}(t),\\ \widehat{I_{\mathrm{B}}(t)} &= I_{\mathrm{B}}(t) + w_{\mathrm{d}}(2)\mathrm{TR}_{\mathrm{d}}(t),\end{aligned} \tag{11}$$

see Fig. 3D. For a deviant tone D we use the same rules ($w_{\mathrm{D}}(\mathrm{DF}_{\mathrm{D}},t) = w_{\mathrm{d}}(\mathrm{DF}_{\mathrm{d}},t)$), but the impulse $\mathrm{TR}_{\mathrm{D}}(t)$ (. . . _ _ _ _ D_ _ _ _ _ . . .) would replace a B tone in $\mathrm{TR}_{\mathrm{B}}(t)$. Incorporating the assumption illustrated in Fig 4A, that distractor tone responses in A1 do not propagate to the integrated unit, $\widehat{I_{\mathrm{AB}}(t)} = I_{\mathrm{AB}}(t)$ in (11); see Fig. 4B.

B.3 General model for stimulus specific adaptation in A1

Here we provide a more general description of how neuronal responses in A1 depend on the tonotopic location of a new tone subject due to SSA from preceding tones. Our implementation of SSA is based primarily on feedforward effects. In SSA a location that has received a sustained input will be adapted in response to further input at the same tonotopic location, with a bandwidth of around 3–4 st in A1[24,25]. We provide a plausible, general implementation of SSA in our model, that could describe A1 responses and be used to determine the inputs from distractor tones to the model's competition stage. Then general schema described below for computing the relative amplitude of responses to new tones, additional to the ABA_ triplets yields very similar results to the ad hoc description above, compare Fig 4E with Fig. 6D.

The general principal is to determine how the tuning curve for a new tone might be modified, based on previous inputs from the regular triplet tones. Example tuning curves for new tones (shown unadapted in Fig. 6A), are modified by the adaptation profiles (dashed curves in Fig. 6B), dependent on the relative location of the new tone to preceding inputs. The adaptation profiles show the most adaptation close to the A tones (fast repetition rate), and less adaptation close to the B tones (slow repetition rate). For a new tone below A, the tuning curves (blue solid curve in Fig. 6B) is carved out on the right hand side. For a new tone above B, the tuning curves (green solid curve in Fig. 6B) is carved out on the left hand side. For a tone in between the tuning curve is carved out on either side (yellow solid curve in Fig. 6B). Below we give a more complete, mathematical description of how the modified tuning curves are calculated.

In this more general formulation, functions will be defined in terms of a tonotopic coordinate y, rather than in terms of a frequency difference $\mathrm{DF_d}$, as used above in (10). In the absence of any prior input, an isolated tone will elicit a response in A1, largest at the tonotopic location of the tone, and decaying on either side (Fig. 6A). In[17], the tuning of these responses was assumed to have a symmetric exponential decay and, for a tone at a location N, this can be described by

$$TC(y,N) = \exp\left(\frac{-|y-N|}{\sigma_{\mathrm{tc}}}\right), \tag{12}$$

where $\sigma_{\mathrm{tc}} = 4\sigma_p$ is broad relative to the post-adaptation tuning width for the A and B tones in (6). In the presence of repeating ABA_ triplets that precede a new tone, the tuned responses will depend on the location of the new tone relative to the As or the Bs. In general, if a series of tones has been arriving at a specific tonotopic location L (either A or B) then the tuning curve of any subsequent tones will be altered. For a new tone N_+ above L the left side of its tuning curve will be reduced. For a new tone N_- below L the right side of its tuning curve will be reduced. The following equation describes the Gaussian *adaptation profile* AP around the L location

$$AP_+(y,L) = \begin{cases} 1 - c_L \exp\left(\frac{-(y-L)^2}{2(BW/2)^2}\right), & y < L \\ 1 - c_L, & y \geq L, \end{cases} \tag{13}$$

where $BW = 4$ is the bandwidth of adaptation and c_L is the amplitude of adaptation, which will be larger when, for example, the preceding sequence of L tones has a higher repetition rate. Equation 13 is 1 for $y \ll L$, decreases with Gaussian decay to $1 - c_L$ as y approaches L from below and is $1 - c_L$ for $y \geq L$. We similarly define AP for a tone below L

$$AP_-(y,L) = \begin{cases} 1 - c_L, & y \leq L \\ 1 - c_L \exp\left(\frac{-(y-L)^2}{2(BW/2)^2}\right) & y > L. \end{cases} \tag{14}$$

In this way the modified tuning curve $\widehat{TC}$ for a tone N_+ above L is given by multiplying the tuning curve with the appropriate adaptation profile

$$\widehat{TC}(y,N_+,L) = TC(y,N_+)AP_+(y,L), \tag{15}$$

and for a tone N_- below L is similarly given by

$$\widehat{TC}(y,N_-,L) = TC(y,N_-)AP_-(y,L). \tag{16}$$

If a tuning curve will be modulated by two sequences of tones L_1 and L_2, an additional argument in (15) or (16) can signify further modulation of the tuning curve by a second adaptation profile, e.g. $\widehat{TC}(y,N_-,L_1,L_2) = TC(y,N_-)AP_-(y,L_1)AP_-(y,L_2)$. These functions can now be used to work out the tuning curves for responses to deviant tones d, relative to the locations of the A and B tones featured in the ABA_ triplet sequence. Assuming significantly more adaptation at the A location due to the higher repetition rate, we set the adaptation strengths associated respectively with the A and B locations to be $c_A = 0.5$ and $c_B = 0.125$. The adaptation profile for a tone below A (which is also below B) will be

$$AP_{A-}(y,A,B) = AP_-(y,A)AP_-(y,B), \tag{17}$$

and is plotted dashed blue in Fig. 6B. For a tone between A and B (above A and below B), we have

$$AP_{AB}(y,A,B) = AP_+(y,A)AP_-(y,B), \tag{18}$$

plotted dashed yellow in Fig. 6B. For a tone above B (also above A), we have

$$AP_{B+}(y,A,B) = AP_+(y,B)AP_+(y,A), \tag{19}$$

plotted dashed green in Fig. 6B. For example, the tuning curve for a new tone (e.g. distractor tone) arriving at a location A-2 (Fig. 6B solid blue) is given by

$$\widehat{TC}_{A-}(y,A-2,A,B) = TC(y,A-2)AP_{A-}(y,A,B), \tag{20}$$

at a location (A+B)/2 (Fig. 6B solid yellow) is given by

$$\widehat{TC}_{AB}(y,(A+B)/2,A,B) = TC(y,(A+B)/2)AP_{AB}(y,A,B), \tag{21}$$

and at a location B+2 (Fig. 6B solid green) is given by

$$\widehat{TC}_{B+}(y,B+2,A,B) = TC(y,B+2)/2)AP_{B+}(y,A,B). \tag{22}$$

To summarise, for $\widehat{TC}$, the first argument is tonotopic location, the second argument the location of a new tone. The subscript $A-$, AB or $B+$ indicates whether the new tone is below, between, or above the A and B tones. The third and fourth arguments are the adapted locations for preceding tones (here A and B from the ABA_ triplets). Having defined the relative amplitude across tonotopy in A1, we now describe the final steps to determine the inputs to the model's competition stage. Similar to (11), the inputs for, say, a distractor tone d above B

$$\begin{aligned} \widehat{I_{\mathrm{AB}}}(t) &= I_{\mathrm{AB}}(t) + I_{\mathrm{ssa}}Q(t)\widehat{TC}_{B+}((A+B)/2,d,A,B), \\ \widehat{I_{\mathrm{A}}(t)} &= I_{\mathrm{A}}(t) + I_{\mathrm{ssa}}Q(t)\widehat{TC}_{B+}(A,d,A,B), \\ \widehat{I_{\mathrm{B}}(t)} &= I_{\mathrm{B}}(t) + I_{\mathrm{ssa}}Q(t)\widehat{TC}_{B+}(B,d,A,B), \end{aligned} \tag{23}$$

where $Q(t)$ describes early onset adaptation and $I_{\mathrm{ssa}} = 3I_p$ is the boosted amplitude for a salient new tone. Again, if we were to incorporate the assumption illustrated in Fig 4A, that no input from a distractor tone reaches in AB-unit, we set $\widehat{I_{\mathrm{AB}}(t)} = I_{\mathrm{AB}}(t)$. Fig. 6D shows the effect on proportion segregated of distractor tones at different tonotopic locations with the general model for SSA presented here. The general model for SSA captures the same features as show in Fig. 4E, also based on the same assumptions illustrated in Fig. 4A, but with a different implementation of SSA.

References

1. LPAS van Noorden. *Temporal coherence in the perception of tone sequences*. Ph.D. Thesis, Eindhoven University of Technology, 1975.
2. SM Anstis and S Saida. Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, 11(3):257–271, 1985.
3. AS Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.
4. RP Carlyon, R Cusack, JM Foxton, and IH Robertson. Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1):115–127, 2001.
5. WL Rogers and AS Bregman. An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics*, 53(2):179–189, 1993.
6. WL Rogers and AS Bregman. Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception & Psychophysics*, 60(7):1216–1227, 1998.
7. D Pressnitzer and JM Hupé. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, 16(13):1351–1357, 2006.
8. SL Denham, K Gyimesi, G Stefanics, and I Winkler. Perceptual bistability in auditory streaming: How much do stimulus features matter? *Learning & Perception*, 5(2):73–100, 2013.
9. YI Fishman, DH Reser, JC Arezzo, and M Steinschneider. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research*, 151(1):167–187, 2001.
10. YI Fishman, JC Arezzo, and M Steinschneider. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *The Journal of the Acoustical Society of America*, 116(3): 1656–1670, 2004.
11. C Micheyl, B Tian, RP Carlyon, and JP Rauschecker. Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, 48(1):139–148, 2005.
12. MA Bee and GM Klump. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *Journal of Neurophysiology*, 92(2):1088–1104, 2004.
13. MA Bee, C Micheyl, AJ Oxenham, and GM Klump. Neural adaptation to tone sequences in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory streaming. *Journal of Comparative Physiology A*, 196 (8):543–557, 2010. ISSN 0340-7594, 1432-1351. doi: 10.1007/s00359-010-0542-4.
14. N Itatani and GM Klump. Neural correlates of auditory streaming in an objective behavioral task. *Proceedings of the National Academy of Sciences*, 111(29):10738–10743, 2014.
15. D Pressnitzer, M Sayles, C Micheyl, and IM Winter. Perceptual organization of sound begins in the auditory periphery. *Current Biology*, 18(15):1124–1128, 2008.
16. C Scholes, AR Palmer, and CJ Sumner. Stream segregation in the anesthetized auditory cortex. *Hearing research*, 328: 48–58, 2015.
17. J Rankin, E Sussman, and J Rinzel. Neuromechanistic model of auditory bistability. *PLOS Computational Biology*, 11 (11):e1004555, 2015.
18. R Moreno-Bote, J Rinzel, and N Rubin. Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology*, 98(3):1125–1139, 2007.
19. A Shpiro, R Moreno-Bote, N Rubin, and J Rinzel. Balance between noise and adaptation in competition models of perceptual bistability. *Journal of Computational Neuroscience*, 27:37–54, 2009. ISSN 0929-5313.
20. RP Carlyon, CJ Plack, DA Fantini, and R Cusack. Cross-modal and non-sensory influences on auditory streaming. *Perception*, 32(11):1393–1402, 2003.
21. NR Haywood and B Roberts. Build-up of the tendency to segregate auditory streams: resetting effects evoked by a single deviant tone. *The Journal of the Acoustical Society of America*, 128(5):3019–3031, 2010.
22. MW Beauvois and R Meddis. Time decay of auditory stream biasing. *Perception & Psychophysics*, 59(1):81–86, 1997.
23. SL Denham, K Gyimesi, G Stefanics, and I Winkler. Stability of perceptual organisation in auditory streaming. In *The neurophysiological bases of auditory perception*, pages 477–487. Springer, 2010.
24. N Ulanovsky, L Las, D Farkas, and I Nelken. Multiple time scales of adaptation in auditory cortex neurons. *The Journal of Neuroscience*, 24(46):10440–10453, 2004.
25. N Taaseh, A Yaron, and I Nelken. Stimulus-specific adaptation and deviance detection in the rat auditory cortex. *PLOS ONE*, 6(8):e23369–e23369, 2011.
26. J-M Hupé and D Pressnitzer. The initial phase of auditory and visual scene analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):942–953, 2012.
27. GH Recanzone and RH Wurtz. Shift in smooth pursuit initiation and MT and MST neuronal activity under different stimulus conditions. *Journal of Neurophysiology*, 82(4):1710–1727, 1999. ISSN 0022-3077, 1522-1598.
28. R Cusack, J Decks, G Aikman, and RP Carlyon. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4): 643, 2004.
29. JS Snyder, OL Carter, S-K Lee, EE Hannon, and C Alain. Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4):1007, 2008.
30. Joel S Snyder, Olivia L Carter, Erin E Hannon, and Claude Alain. Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4):1232, 2009.
31. ES Sussman, J Horváth, I Winkler, and M Orr. The role of attention in the formation of auditory streams. *Perception & psychophysics*, 69(1):136–152, 2007.
32. AS Bregman. Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, 4(3):380, 1978.
33. SK Thompson, RP Carlyon, and R Cusack. An objective measurement of the build-up of auditory streaming and of its modulation by attention. *Journal of Experimental Psychology: Human Perception and Performance*, 37(4):1253–1262, 2011. ISSN 1939-1277 0096-1523. doi: 10.1037/a0021925.
34. BCJ Moore and HE Gockel. Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):919–931, 2012.

35. E Sussman and M Steinschneider. Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Research*, 1075(1):165–174, 2006. ISSN 0006-8993. doi: 10.1016/j.brainres.2005.12.074.
36. T Rahne and E Sussman. Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. *European Journal of Neuroscience*, 29(1):205–211, 2009.
37. B Roberts, BR Glasberg, and BCJ Moore. Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4):992, 2008.
38. NR Haywood and B Roberts. Build-up of auditory stream segregation induced by tone sequences of constant or alternating frequency and the resetting effects of single deviants. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6):1652–1666, 2013.
39. F Almonte, VK Jirsa, EW Large, and B Tuller. Integration and segregation in auditory streaming. *Physica D: Nonlinear Phenomena*, 212(1):137–159, 2005.
40. DL Wang and P Chang. An oscillatory correlation model of auditory streaming. *Cognitive Neurodynamics*, 2(1):7–19, 2008.
41. M Elhilali and SA Shamma. A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *The Journal of the Acoustical Society of America*, 124(6):3751–3771, 2008.
42. SL McCabe and MJ Denham. A model of auditory streaming. *Advances in Neural Information Processing Systems*, 101(3):1611–1621, 1996.
43. MW Beauvois and R Meddis. A computer model of auditory stream segregation. *The Quarterly Journal of Experimental Psychology*, 43(3):517–541, 1991.
44. MW Beauvois and R Meddis. Computer simulation of auditory stream segregation in alternating-tone sequences. *The Journal of the Acoustical Society of America*, 99(4):2270–2280, 1996.
45. RW Mill, TM Bőhm, A Bendixen, I Winkler, and SL Denham. Modelling the emergence and dynamics of perceptual organisation in auditory streaming. *PLOS Computational Biology*, 9(3):e1002925, 2013.
46. SA Steele, D Tranchina, and J Rinzel. An alternating renewal process describes the buildup of perceptual segregation. *Frontiers in Computational Neuroscience*, 8(166):1–13, 2015.
47. D Barniv and I Nelken. Auditory streaming as an online classification process with evidence accumulation. *PLOS ONE*, 10(12):e0144788, 2015.
48. M Elhilali, J Xiang, SA Shamma, and JZ Simon. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLOS Biology*, 7(6):e1000129, 2009. ISSN 1545-7885. doi: 10.1371/journal.pbio.1000129.
49. M Kashino and HM Kondo. Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):977–987, 2012.
50. EW Large, JA Herrera, and MJ Velasco. Neural networks for beat perception in musical rhythm. *Frontiers in Systems Neuroscience*, page 159, 2015. doi: 10.3389/fnsys.2015.00159.
51. R Borisyuk, D Chik, and Y Kazanovich. Visual perception of ambiguous figures: synchronization based neural models. *Biological cybernetics*, 100(6):491–504, 2009. ISSN 0340-1200.
52. M Elhilali, L Ma, C Micheyl, AJ Oxenham, and SA Shamma. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*, 61(2):317–329, 2009.
53. J Sussman-Fort and E Sussman. The effect of stimulus context on the buildup to stream segregation. *Frontiers in Neuroscience*, 8:93, 2014. doi: 10.3389/fnins.2014.00093.
54. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `https://www.R-project.org/`.
55. Michael A. Lawrence. *ez: Easy analysis and visualization of factorial experiments.*, 2013. URL `http://CRAN.R-project.org/package=ez`. R package version 4.2-2.
56. David C. Howell. *Fundamental Statistics for the Behavioral Sciences*. Cengage Learning, 2013. ISBN 978-1-285-53183-0.
57. Jeffrey Seely and Carson C Chow. Role of mutual inhibition in binocular rivalry. *Journal of Neurophysiology*, 106(5): 2136–2150, 2011.