Attentional control via synaptic gain mechanisms in auditory streaming

James Rankin a,*, John Rinzel b,c

a Department of Mathematics, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Rd, Exeter EX4 4QF, UK
b Center for Neural Science, New York University, 4 Washington Place, 10003 New York, NY, United States
c Courant Institute of Mathematical Sciences, New York University, 251 Mercer St, 10012 New York, NY, United States

ARTICLE INFO

Keywords:
Auditory streaming
Bistability
Attention
Computational model

ABSTRACT

Attention is a crucial component in sound source segregation allowing auditory objects of interest to be both singled out and held in focus. Our study utilizes a fundamental paradigm for sound source segregation: a sequence of interleaved tones, A and B, of different frequencies that can be heard as a single integrated stream or segregated into two streams (auditory streaming paradigm). We focus on the irregular alternations between integrated and segregated that occur for long presentations, so-called auditory bistability. Psychoacoustic experiments demonstrate how attentional control, a listener’s intention to experience integrated or segregated, biases perception in favour of different perceptual interpretations. Our data show that this is achieved by prolonging the dominance times of the attended percept and, to a lesser extent, by curtailing the dominance times of the unattended percept, an effect that remains consistent across a range of values for the difference in frequency between A and B. An existing neuromechanistic model describes the neural dynamics of perceptual competition downstream of primary auditory cortex (A1). The model allows us to propose plausible neural mechanisms for attentional control, as linked to different attentional strategies, in a direct comparison with behavioural data. A mechanism based on a percept-specific input gain best accounts for the effects of attentional control.

1. Introduction

In a dynamic auditory world the brain must isolate objects of interest and resolve ambiguity between sounds that agree with different interpretations of the environment (Bregman, 1994). Intriguingly, even stationary, ambiguous stimuli can induce spontaneous alternations between competing perceptual interpretations, i.e. perceptual bistability (van Noorden, 1975). The dominant percept depends on stimulus parameters (Moreno-Bote et al., 2010; Brascamp et al., 2015; Rankin et al., 2015) and on a listener’s ability to manipulate what they hear through attentional control (Pressnitzer and Hupé, 2006), i.e. through their intention to hear a specific perceptual interpretation. A common psychoacoustic stimulus used in experiments on auditory object separation consists of alternating pure tone sequences organized in repeating ABA_ triplets, where A and B are separated by a difference in frequency DF and “_” is a silent gap (van Noorden, 1975). The sequence can be perceived as integrated (Int) into one, or as segregated (Seg) into two streams (Fig. 1A). The initial percept is typically integrated, but as time proceeds a perceptual transition to segregated becomes more likely, a phenomenon called build-up (Bregman, 1994; Anstis and Saida, 1985). For short duration presentations (10–30s), behavioral studies of build-up have characterized the effects of attention (Carlyon et al., 2003; Macken et al., 2003; Beauvois and Meddis, 1997; Snyder et al., 2006), context (Rahne and Sussman, 2009; Snyder et al., 2008), temporal coherence (Shamma et al., 2011) and sudden stimulus changes (Rogers and Bregman, 1998; Roberts et al., 2008; Haywood and Roberts, 2010). For extended presentations (minutes long) of ABA_ sequences, the build-up phase is followed by persistent irregular alternations between integrated and segregated (Fig. 1B), a phenomenon called bistability (Pressnitzer and Hupé, 2006; Kondo and Kashino, 2009; Winkler et al., 2012; Rankin et al., 2015; Denham et al., 2018).

Attention is a fundamental aspect of sensory processing that mediates sensory selection and the content of cognitive awareness in general (e.g. Gilbert and Sigman, 2007; Harris and Thele, 2011). In the auditory streaming paradigm, the build-up to segregation depends on participants attending to the stimulus (Carlyon et al., 2001), although build-up can occur without attention in some cases (Macken et al., 2003; Sussman et al., 2007). In auditory bistability, one may study percept-ideal: attentional control by an instruction such as ‘hold the integrated percept’. Indeed, it has been shown that participants perceive the
attended percept a larger proportion of the time (Pressnitzer and Hupé, 2006; Billig et al., 2018; Kondo et al., 2018). While behavioural (Carlcyon et al., 2001), single-cell electrophysiology (Micheyl et al., 2005; Itatani and Klump, 2014), brain imaging (Gutschalk et al., 2005; Kashino and Kondo, 2012) and modelling studies (Rankin et al., 2015; Little et al., 2020) support the involvement of a hierarchy of brain regions including cortical areas in auditory bistability, the mechanisms of attention are unknown.

We focus on attentional mechanisms that lead to a bias in the proportion of time hearing the attended percept, as reported previously in behavioural experiments (Pressnitzer and Hupé, 2006; Billig et al., 2018; Kondo et al., 2018). In the literature, the influence of attentional control (Klink et al., 2008; Van Ee et al., 2006) is also referred to as volitional control (Pressnitzer and Hupé, 2006; Kondo et al., 2018) or intention (Billig et al., 2018); for the remainder of the manuscript we use the term attention to refer to attentional control in terms of a subject’s intention to hear the integrated or segregated percept. In new behavioural experiments we extend the investigation of attention to interactions with stimulus-driven biases (via different DF values; see Fig. 1C). Our neuromechanistic model (Rankin et al., 2015), which accurately reproduces the dynamics of auditory bistability (Fig. 1C-E), will be used to investigate several potential mechanisms for attentional control. We consider three mechanisms implemented via task-dependent adjustments to synaptic weights at the input targets in the model’s competition stage, downstream of primary auditory cortex (A1). A qualitative and quantitative comparison with experimental data allows for the likely validity of different mechanisms to be explored.

The manuscript is organised as follows. The Materials and methods sections introduce our previously published computational model and describes procedures for psychoacoustic experiments. Results sections report data from new behaviour experiments on attention in auditory bistability over a range of DF-values. Several attentional mechanisms proposed in the model are compared with behavioural data. Discussion sections cover literature on attention for auditory streaming and a comparison of findings with the literature on studies of bistable visual perception.

2. Materials and methods

2.1. Neuromechanistic model of auditory streaming

The neuromechanistic model presented in Rankin et al. (2015) is based on plausible neurocomputational mechanisms found throughout cortex, as motivated from neurophysiological studies of auditory streaming (Fishman et al., 2001; Fishman et al., 2004; Micheyl et al., 2005) and general models of perceptual bistability (Taylor et al., 2002; Laing and Chow, 2002; Wilson, 2003; Shpiro et al., 2009). The model captures the dynamics of alternations after the build-up phase, i.e. after the first perceptual switch (Fig. 1B). It uses dynamic inputs that directly link to sensory features as represented by the neuronal responses of pre-competition stages: inputs based on electrophysiologically-recorded A1 responses to interleaved A and B tones. It features a tonotopic organization with three units assumed to be downstream and receiving input from A1 (Fig. 1D). Two peripheral units receive input from regions of A1 and experiment (dashed; error bars show SEM). Equidominance for DF between integrated and segregated (bistability) if DF is not too large or small (Rankin et al., 2015).

Fig. 1. A: The triplet stimulus is perceived as either one integrated stream ABA,ABA,... or two segregated streams A_A_A_A_... and _B___B___... B: Perceptual reports for 120s of a single trial. Initial percept is integrated followed by a switch to segregated within first ~10s (build-up phase). Subsequently perception alternates every ~5s between integrated and segregated (bistability) if DF is not too large or small (Rankin et al., 2015). C: Dependence of mean percept duration on DF in model (solid) and experiment (dashed; error bars show SEM). Equidominance for DF=5st. Duration integrated decreases and duration segregated increases with DF. This cross-diagram demonstrated a generalisation of the effect of stimulus strength manipulations from visual bistability to auditory bistability (Rankin et al., 2015). D: 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 the difference in frequency (DF) given 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). E: One model simulation showing the activation threshold (horizontal black), and each population’s NMDA variable (solid, pulsatile inputs appear smoothed in sub-threshold activity). When the central AB unit is active (integrated), the peripheral units are suppressed through mutual inhibition.
centered at locations with best frequencies \(f_a\) or \(f_b\), and a third unit receives input from tonotopically intermediate A1 locations, say centered at \((f_a + f_b)/2\). The perceptual interpretations are classified through criteria on a units’ firing patterns, for example, the central unit (receiving A and B inputs via A1) encodes integrated. We use a dynamical systems framework with firing rate based neuronal competition mediated by mutual inhibition, adaptation and noise; neural mechanisms that proved successful in accounting for many of the characteristic behaviors of perceptual bistability (Shpiro et al., 2009). Our model incorporates in its inputs the onset and transient dynamics of A1 responses with gaps between triplets (Fishman et al., 2001; Fishman et al., 2004; Michely et al., 2005) (Fig. 1D “Tonotopic A1”). The inclusion of recurrent NMDA-like excitation allows for some neuronal memory that links each tone and each triplet to the next. This neuro-mechanistic approach with neuro-based time- and feature-dependence of inputs is distinct from recent modeling studies of perceptual bistability that involve percept-based inputs and competition (Taylor et al., 2002; Laing and Chow, 2002; Wilson, 2003; Shpiro et al., 2009).

A recent paper extended the model to capture build-up (Rankin et al., 2017) and provided the first mechanistic explanations for several phenomena: (1) build-up results from adaptation in A1 that biases competition downstream, (2) resetting to an integrated percept occurring after stimulus pauses are due to rapid recovery of this adaptation, and (3) our new finding, that unexpected additional tones promote segregation, is explained by differential input gating from A1 to downstream populations. The model has been adapted to account for performance in streaming tasks for cochlear implant users (Paredes-Gallardo et al., 2019) and has further been used to explore entrainment in dynamic auditory environments (Byrne et al., 2019). It provides an ideal framework to explore the mechanisms of attention for auditory bistability.

### 2.2. Attention mechanisms overview

The model will be used to investigate three attention mechanisms that are linked to different potential outcomes for the behavioural data. These outcomes are described briefly to provide context for the data. Mechanism M1- ImpAtt-Mixed involves a task-related static boost to inputs delivered from low level processing and leads to a bias towards the attended percept via an increase to durations of the attended percept and a decrease to the unattended percept. Mechanism M2- Recur-IncAtt involves a static boost to local recurrent excitation, respectively and leads to an increase to the durations of the attended percept, but with no change to the unattended percept. Finally, for M3- StDep-DecUn, Inputs to the units representing the attended percept are boosted, but only when the unattended unit is active, which leads to a decrease in the unattended percept but with no change to the attended percept. Further details are given as part of the results in Sections 3.2-3.5.

### 2.3. Participants

Pilot data following the procedure of Pressnitzer and Hupé (2006) (see Appendix B) guided the design, in particular the required subject numbers. Nine subjects (four female) with a mean age 25 years took part in the experiment and were reimbursed for their time at a $10 hourly rate. 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 (study IRB-FY2016-310). All subjects provided written informed consent. Each subject completed two repetitions of a 9-trial block covering each combination of the three attention tasks (Passive, Attend Seg, Attend Int) and the DF conditions (3,5,8), giving a total of 33 four-minute trials per subject. A 9 × 9 latin square design was used to determine the order of conditions for each subject. Two participants who did not switch at least three times during every trial were excluded from the analysis.

### 2.4. Stimuli

The stimuli consist of repeating ABA triplets, where each tone is 100 ms long and where ‘-’ indicates a silence also lasting 100 ms; each ABA triplet is therefore 0.4 s in duration. The higher frequency B tones are a variable DF semitones (st) above the lower frequency A tones. Cosine squared ramps are used with 5 ms rise and fall times. During 4 min trials the tone sequence is played binaurally through Etymotic headphones at 62 dB SPL. There was a minimum of 20 s gap between trials and B tone base frequencies were roved over six frequencies in the range 554–987 Hz (giving corresponding A tone frequencies in the range 740–1,319 Hz, depending on DF). Each tone frequency pair occurred an equal number of times for each condition. Roving of base frequencies and breaks between trials reduced the possibility for latent adaptation carrying over from one trial to the next.

### 2.5. Experimental procedures

Subjects sat in an acoustically shielded chamber and indicated their perceptual responses with button presses on a keyboard. In a continuous two alternative forced choice 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 with 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. In the Passive condition, subjects were instructed to passively report their percepts without attempting to hear one perceptual organization over another. In the Attend Int task subjects were asked to “try to hear the one stream interpretation with galloping rhythm more,” and in the Attend Seg task to “try to hear the two streams more.” Subjects reported their percepts by holding specific keys associated with each percept. 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 continuous two-alternative forced choice 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 (Pressnitzer and Hupé, 2006; Mill et al., 2013; Rankin et al., 2017). Here no instruction for “indeterminate” responses was provided. Durations shorter than 0.5 s (one triplet) were excluded from the analysis. Given the two choice 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.

### 2.6. Statistical analyses

The factors tested in this study, attention condition (task = {Passive, Attend Seg, Attend Int}) and frequency difference (DF = {3,5,8}) have been explored in earlier studies (Pressnitzer and Hupé, 2006; Rankin et al., 2015). Here, the measures of interest are the durations of integrated and segregated perceptual phases in the bistable auditory streaming paradigm. As in these earlier studies, the perceptual interpretation (percept = {integrated, segregated}) is a within-subject factor and the main interactions task x percept and df x percept are tested. An earlier study from the authors (Rankin et al., 2015) showed an effect size of \(\eta^2 = 0.21\) for the interaction df x percept, which would require a sample size of 8 participants to detect with 90% power. In-house data reproducing the experiments from Pressnitzer and Hupé (2006) (Fig. 5) had an effect size of \(\eta^2 = 0.33\) for the interaction task x percept, which would also require a sample size of 8 participants to detect with 90% power. The present study, including both these factors, used 9 participants.
Statistical analyses were carried out in the statistical package R and subject number estimates were calculated with GPower (Faul et al., 2009). Greenhouse-Geisser (GG) corrected $P$-values are reported if a Mauchly sphericity (MS) test reached significance. In ANOVA tables the GG-corrected $P$-values are highlighted in italics if the MS test reached significance.

3. Results

3.1. Behavioural experiments on DF-dependent attentional biasing

In behavioural experiments a stimulus parameter DF was varied for three different attention conditions: Passive, Attend Int and Attend Seg. The proportion of time that participants perceived the integrated percept is shown in Fig. 2A. For the passive case at DF = 5st, there is approximate equidominance between the integrated (perceived just over 50% of the time) and segregated percepts. For smaller DF integrated is perceived more (around 60% of the time) and for larger DF segregated is perceived more (also around 60% of the time). Instructing participants to attend to the integrated percept (Attend Int; see Methods for instructions to participants) results in that percept being perceived more of the time (an increase of around 15 percentage points) relative to the Passive case (compare black and blue curves); similarly for Attend Seg (compare black and red curves). The results of a repeated measures ANOVA (Table 1 in Appendix A) show that the attention condition task: df ($F(2, 12) = 31.93; P < .0001; \eta^2 = 0.53$) and difference in frequency stimulus parameter df ($F(2, 12) = 133.60; P < .0001; \eta^2 = 0.83$) both have a highly significant, large effect on the proportion of time spent perceiving integrated (their interaction task: df was not significant). This demonstrates that DF significantly affects the predominance of the each percept (as reported previously in Rankin et al. (2015)), as does the attention condition (as reported in Pressnitzer and Hupé (2006)) over a range of DF-values.

The archetypal cross-diagram (Fig. 1C) shows that perception is biased (with longer durations) towards integrated at low DF, towards segregated at large DF and approximately equidominant at DF = 5 (Fig. 2B). In the Attend Seg condition (panel C) the segregated durations are lengthened significantly (compare dashed curves between panels B and panel C), whilst the integrated durations decrease slightly (compare solid curves). As one might expect, the opposite trend is observed for the Attend Int condition (panel D) where integrated durations increase (compare solid curves between panels B and D) and segregated durations decrease slightly (compare dashed curves). The reader may wish to refer forward to Fig. 3 in which the change in duration is plotted directly and where the increase/decrease relative to the passive case can be seen more clearly. Treating the perceptual interpretation as a within-subject factor (percept), a significant, small effect was found for the interaction task: df ($F(4, 24) = 3.67; P = .018; \eta^2 = 0.37$). Other interactions involving the percept were strongly significant with a large effect size task: percept ($F(2, 12) = 17.37; P < .001; \eta^2 = 0.64$) and df: percept ($F(2, 12) = 31.33; P < .001; \eta^2 = 0.40$), however the individual factors and the interaction of all three did not reach significance. Treating the durations for each percept as a dependent variable we found a significant, large effect for the attention condition task ($F(2, 12) = 6.49; P = .037; \eta^2 = 0.17$) and a highly significant, large effect for the stimulus parameter df ($F(2, 12) = 17.55; P < .001; \eta^2 = 0.38$) on integrated durations (but no significant interaction). For segregated durations we found a highly significant, large effect for the attention condition task ($F(2, 12) = 25.18; P < .0001; \eta^2 = 0.18$) and a significant, large effect for the stimulus parameter df ($F(2, 12) = 16.84; P = .0045; \eta^2 = 0.52$) but no significant interaction. Overall we can conclude that attention significantly biases perceptual durations in the direction of the attended percept (Fig. 2A) and that this bias is primarily mediated via an increase in the durations of the attended percept (Fig. 2C-D).

3.2. Modelling attention

We investigate gain control mechanisms, implemented as task-selective synaptic gains at units in the model’s competition stage, downstream of A1’s sensory feature extraction. We will introduce the naming convention summarising how the mechanism is implemented (e. g. via inputs or recurrent excitation) and the effect that gain adjustments

---

Fig. 2. A: Proportion of time for integrated percept against DF for three conditions: Passive, Attend Seg and Attend Int (see legend). Dark curves averaged across subjects (error bars show SEM; N = 7) and faint curves show individual-subject data. B-D: Experimental cross diagrams in DF under three conditions showing integrated and segregated durations (see legend). Averaged (dark) and individual-subject (faint) data indicated as in panel A.
Table 1
Repeated measures ANOVA tables for Proportion Integrated, durations with the percept (integrated or segregated) as a within-subject factor and durations broken down into integrated or segregated. Other within-subject factors are task (Passive Attend Int, Attend Seg) and the stimulus difference in frequency df (DF = 2,5,8). Degrees of Freedom (DoF), F-statistic, P-values and generalized η² are given. Factors for which Mauchly’s test for non-sphericity reached significance the Greenhouse-Geisser-corrected P-value is indicated in italics. In the text effect sizes with η² > 0.14 reported as large, with 0.06 < η² < 0.14 as medium and with η² < 0.06 as small. For significant P-values, * indicates P < 0.05, ** P < 0.01, *** P < 0.001, and **** P < 0.0001.

| Proportion Integrated | Effect | DoFn | DoFd | F     | P or P(GG) | η²    |
|-----------------------|--------|------|------|-------|------------|-------|
| task                  | 2      | 12   | 31.973 | < .0001 | 0.533 **   |       |
| df                    | 2      | 12   | 133.599 | < .0001 | 0.826 ***  |       |
| tank+df               | 4      | 24   | 0.505 | .732  | 0.024      |       |

| Durations (percept as within-sub) | Effect | DoFn | DoFd | F     | P or P(GG) | η² |
|-----------------------------------|--------|------|------|-------|------------|----|
| task                              | 2      | 12   | 1.250 | .307  | 0.015      |    |
| df                                | 2      | 12   | 3.868 | .051  | 0.081      |    |
| percept                           | 1      | 6    | 0.411 | .545  | 0.005      |    |
| tank+df                           | 4      | 24   | 3.668 | 0.018 | 0.037      |    |
| tank+percept                      | 2      | 12   | 17.369 | < .001 *** | 0.159    |    |
| df+percept                        | 2      | 12   | 31.327 | < .001 *** | 0.401    |    |
| tank+df+percept                   | 4      | 24   | 0.409 | .800  | 0.006      |    |

| Integrated durations              | Effect | DoFn | DoFd | F     | P or P(GG) | η² |
|-----------------------------------|--------|------|------|-------|------------|----|
| task                              | 2      | 12   | 6.488 | .0367 * | 0.166      |    |
| df                                | 2      | 12   | 17.545 | < .001 *** | 0.378    |    |
| tank+df                           | 4      | 24   | 1.477 | 0.240 | 0.041      |    |

| Segregated durations              | Effect | DoFn | DoFd | F     | P or P(GG) | η² |
|-----------------------------------|--------|------|------|-------|------------|----|
| task                              | 2      | 12   | 25.177 | < .0001 | 0.178      |    |
| df                                | 2      | 12   | 16.837 | .00447 ** | 0.515    |    |
| tank+df                           | 4      | 24   | 2.645 | .0584 | 0.047      |    |

have on durations (e.g. increasing attended durations) below. Mechanisms M1-InpAtt-Mixed and M2-Recur-IncAtt involve a task-related static boost to inputs delivered from low level processing and a static boost to local recurrent excitation, respectively. M3-StDep-DecUn, on the other hand, is state-dependent. Inputs to the units representing the attended percept (e.g. unit AB for Attend Int) are boosted, but only when the unattended unit is active. This restorative drive acts as an effective disinhibition for the attended unit(s). The effect of each mechanism is compared to our behavioural data below.

3.3. Mechanism M1-InpAtt-Mixed: static input-based attention

The mechanism M1-InpAtt-Mixed statically boosts the inputs to the unit associated with the attended percept (-InpAtt) resulting in a mix of increased attended percept durations and decreased unattended durations (-Mixed).

As discussed in Section 2.1, the integrated percept is encoded by a single unit downstream of A1 that pools inputs from A and B tonotopic locations, whereas segregated is encoded by two units that receive their inputs from either the A location or from the B location. Partly inspired by experiments in binocular rivalry that explored input-strength (i.e. image contrast) based biasing as being equivalent to attentional gain (Mueller and Blake, 1989; Chong et al., 2005; Chong and Blake, 2006), we implemented an input-based mechanism for attention in our auditory streaming model (Fig. 3A and C). The mechanism M1-InpAtt-Mixed is straightforward: during the Attend Seg task there is a static increase of the A and B unit’s input with amplitude λI, while the AB unit’s inputs are unchanged (panel A). For Attend Seg (panel B left), as λI is increased, the segregated durations are extended (upward sloping dashed curves; black is Passive condition) and the integrated durations are reduced to a lesser extent (downward sloping solid curves; black is Passive condition). Conversely, during Attend Int (mechanism illustrated in panel C), inputs to the AB unit receive a static boost of amplitude λI, while the A and B units’ inputs are unchanged. For Attend Int durations (panel B right), the result is an increase to integrated durations with λI and concurrent slight decrease to segregated durations.

A comparison between the model mechanism M1-InpAtt-Mixed and subject-averaged experimental data is shown for best-fit values of λI and λS in Fig. 3D–F. Panels D and E show experimental data from Fig. 2 with model responses overlaid. The model accurately captures the effect of attention on the proportion of time spent in the integrated percept (Fig. 3D), with a slight difference in the slope across DF-values for Attend Int. The model captures the qualitative effect on durations for the two attention conditions, showing a significant increase in segregated durations for Attend Seg (and slight decrease in integrated durations) along with a significant increase in integrated durations for Attend Int (and slight decrease in segregated durations). There is a quantitative mismatch for the Attend Int condition, most notably at large DF. The same data from panel E are re-plotted in panel F, where the increase or decrease of durations in the attended conditions are plotted relative to the Passive condition. The decrease in durations for the unattended percept is seen more clearly in this projection of the data, which will be used to compare different attention mechanisms in Fig. 4 (where panels A and F from Fig. 3 are repeated).

3.4. Mechanism M2-Recur-IncAtt: boost to recurrent excitation

This mechanism statically boosts the recurrent excitation in the unit (s) encoding the attended percept (-Recur) resulting in an increase to attended percept durations (-IncAtt).

Under mechanism M2-Recur-IncAtt for Attend Seg, the recurrent excitation strength in the A and B units is increased by a factor μS (Fig. 4B); the equivalent for Attend Int, boosting the AB unit, is not shown. This mechanism qualitatively captures the effect of attention on proportion of time integrated (not shown) and a comparison for the change in duration relative to the Passive case is given in panel E. The increase in durations for the attended percept (upper curves) is captured, but the slope with respect to DF is too shallow for Attend Seg (dashed darker curve on left) and too steep for Attend Int (darker solid curve on right). The modest decrease in the durations of the unattended percept is not captured by this mechanism.

Mechanism M2-Recur-IncAtt, based on a boost to recurrent excitation, is effectively state dependent, i.e. only active when the attended unit (units) is (are) on. This results in a boost to the attended percept but with no change to the unattended percept durations as the mechanism shuts down when the attended unit is inactive (with no activity to drive excitation recurrently). In summary, M2-Recur-IncAtt misses key qualitative features of the data that were successfully captured by M1-InpAtt-Mixed (compare panels D and E).

3.5. Mechanism M3-StDep-DecUn: state-dependent bias to switch away from unattended percept

This mechanism boosts the inputs to the unit(s) encoding the attended percept in a state-dependent way (-StDep) resulting in an decrease to unattended percept durations (-DecUn).

The third mechanism explored here is inspired by the observation, made prior to conducting our own experiments, that in one study of attention in auditory streaming (Pressnitzer and Hupé, 2006), and some studies of attention in binocular rivalry (Hancock and Andrews, 2007;
Lack, 1978), the effect is to reduce the duration of the unattended percept (whilst not changing durations of the attended percept). As such we implemented a mechanism that acts to shorten the unattended durations whilst leaving the attended percept durations unchanged. The mechanism operates by boosting the inputs to the attended percept but only during dominance of unattended units (Fig. 4C). The result is to shorten the unattended durations, while leaving the durations for the attend percept unchanged. This mechanism can qualitatively capture the effect of attention on proportion of time integrated as reported in Fig. 2 (not shown) but fails to adequately capture the effect of attention on percept durations in our data (Fig. 2 and in other studies (Billig et al., 2018; Kondo et al., 2018)). The attentional effects reported in Pressnitzer and Hupé (2006) (which we, see Fig. 5, and others have failed to reproduce) would be captured by this mechanism as discussed further below.

4. Discussion

In psychacoustic experiments we investigated how attentional control can bias the perceptual organisation of the auditory streaming paradigm parameterised via DF for bistability between integrated and segregated. Our study builds on earlier work (Pressnitzer and Hupé, 2006; Billig et al., 2018; Kondo et al., 2018) by considering a range of parameter values where perception favoured integrated (lower DF) or segregated (higher DF), along with a near-equidominant case (DF = 5). We found that relative to passive listening conditions, participants were able to bias their perception toward integrated or segregated with attentional focus and that the strength of this bias was consistent across a range of DF conditions. Interestingly, this bias was achieved by a combination of lengthening of durations for the attended percept and, to a lesser extent, shortening of durations for the unattended percept. Our findings were consistent when attention was directed toward either integrated or segregated, and across a range of DF-values.

A computational model developed to account for the dynamics of auditory bistability across a range of DF-values allowed an exploration of different attention mechanisms. Inspired by a range of effects for attention observed in visual rivalry experiments (Lack, 1978; Mueller and Blake, 1989; Chong et al., 2005; Van Ee et al., 2005; Pressnitzer and Hupé, 2006; Hancock and Andrews, 2007) and in auditory bistability (Pressnitzer and Hupé, 2006; Billig et al., 2018; Kondo et al., 2018) (see sections below), we proposed three different mechanisms. As a bias in the proportion of time spent holding one percept can be achieved by increasing the dominant durations, decreasing the non-dominant...
durations, or by a mix of these two effects, we proposed mechanisms to capture each scenario. The first mechanism M1-InpAtt-Mixed (Fig. 3) gives a static boost to synaptic weights at targets of inputs from A1 at the model’s competition stage; this captures the mixed scenario, where attended durations increase and unattended durations decrease. A second mechanism M2-Recur-IncAtt (Fig. 4B) gives a boost to recurrent synaptic weights within the competition stage, which captures the scenario where durations of the attended percept are lengthened but the unattended durations are unchanged. We note that state-dependence is emergent for this mechanism: the recurrent excitation is only active when the units encoded the attended percept are active. Finally, a third mechanism M3-StDep-DecUn (Fig. 4C) is explicitly state dependent, whereby synaptic weights for the attended unit(s) increase, but only when the unattended percept is active, which captures a decrease in only the unattended durations. From a participant’s point of view this could be a strategy along the lines of “try to switch away from the unattended percept”, whereas M2-Recur-IncAtt would be along the lines of “when the attended percept is heard, try to maintain that percept”. We note that these three mechanisms lead to different effects by changing synaptic weights in the competition stage, downstream of A1 (rather than modulating activity in A1, say via descending feedback).

A comparison between each mechanism and our experimental data revealed that the input-based mechanism (M1-InpAtt-Mixed) provides the best match with qualitative features of our data, i.e. a mixed effect on durations of the attended and unattended. However, given subtle differences in the effect on perceptual durations observed in other studies on attentional control, we suspect that other mechanisms (or strategies from a participant’s point of view) may be at play in different

Fig. 4. Comparison between mechanisms M1-InpAtt-Mixed, M2-Recur-IncAtt and M3-StDep-DecUn. A–C: Mechanisms illustrated for Attend Seg. In M1-InpAtt-Mixed and M2-Recur-IncAtt boost to inputs \( \lambda_S \) or to recurrent excitation \( \mu_S \) is static, whereas for M3-StDep-DecUn the input boost is state dependent and therefore a function of time \( \lambda_S(t) \). D–F: Change in duration relative to Passive case for M1-InpAtt-Mixed, M2-Recur-IncAtt and M3-StDep-DecUn.

Fig. 5. Psychacoustic experiments \((N = 8)\) with fixed \(DF = 5\), following similar protocol to Pressnitzer and Hupé (2006); integrated durations as blue bars, segregated durations as red bars. Error bars show SEM.
situations as discussed below.

4.1. Comparison with auditory behavioural studies

Several earlier studies have looked at attentional effects for short stimulus presentations (Carlyon et al., 2003; Macken et al., 2003; Beauvois and Meddis, 1997; Snyder et al., 2006; Brace and Sussman, 2021), with a few studies focusing on attentional control during bistable alternations during long stimulus presentations (Pressnitzer and Hupé, 2006; Billig et al., 2018; Kondo et al., 2018). These studies, and our presented data, are consistent in showing a longer proportion of time spent hearing the attended percept, increasing by around 15%. Whilst Pressnitzer and Hupé (2006) reported that bias towards the attended percept results exclusively from a reduction of durations for the unattended percept, our results here (Fig. 1C-D) and more recent studies have not reproduced this finding (Billig et al., 2018; Kondo et al., 2018). Indeed, consistent with our results, a mixed effect where the durations of the attended percept increase and the unattended durations decrease were reported in Kondo et al. (2018) (and our pilot data in Fig. 5). Billig et al. (2018) found an asymmetry between the two tasks: for the Attend Int task, a larger decrease in segregated durations relative to small increase in integrated durations and for the Attend Seg task, a mix of increased segregated and decreased integrated durations (though we note that the subject’s attention was directed to just one of the streams A or B rather than the segregated percept in general). Our investigation also considered DF-values that favoured one of the percepts in the passive case and found that the strength of attentional modulation on the proportion of time spent in either percept to be quite consistent across DF and across attentional task (Attend Int or Attend Seg). We found consistency across tasks with a mixed effect on durations with a large increase for the attended percept’s durations (also in line with Levelt’s proposition II for DF manipulations (Rankin et al., 2015); see Section 4.3). Differences across studies could result from differences in stimulus parameters (e.g. Pressnitzer and Hupé, 2006; Billig et al., 2018 used slower presentation rates) or from different strategies used by subjects, which may be affected by the specific instructions used on the attention tasks. Indeed, the mechanisms proposed in the present study could be linked to different strategies that could account for our data and contrasting results across studies.

4.2. Neural recordings

Here we review literature on the neural basis of auditory stream segregation (where the involvement of A1 remains unclear) whilst providing context for a recent studies also investigating attentional control for auditory bistability (Billig et al., 2018; Kondo et al., 2018).

In humans, neural correlates of auditory bistability have been studied with EEG (Szalárdy et al., 2013; Higgins et al., 2020) and a number of studies have focused on attention during build-up (Snyder et al., 2006; Sussman et al., 2007; Brace and Sussman, 2021). An MEG study (Gutschalk et al., 2005) localized to auditory cortex showed differences in activation between the two percepts and implicated non-primary auditory areas in maintaining percepts rather than only stimulus features (as in A1), which is further supported by single-cell recordings in macaques (Fishman et al., 2004; Michieley et al., 2005). These results were further illuminated by intracranial recordings showing that percept maintaining and switching responses are represented in auditory cortex (Curtu et al., 2019), with bistability-associated responses localised predominantly in non-primary areas. Imaging studies (fMRI) have shown activation of broader thalamocortical (Kondo and Kashino, 2009) and cerebellar (Kashino and Kondo, 2012) networks around the time of perceptual switches. We note that the perceptual interpretations are not equal in auditory streaming (in contrast with e.g. binocular rivalry) and that different sequences of activation along thalamocortical pathways are associated with switches in different directions (Kondo and Kashino, 2009). A recent EEG study (Higgins et al., 2020) showed differences in activity in the ventral auditory pathway (associated with object recognition) across perceptual interpretations and in epochs associated with different switch directions. Attentional effects for auditory bistability have recently been studied with magnetoencephalography (MEG) recordings (Billig et al., 2018) and with magnetic resonance spectroscopy (MRS) (Kondo et al., 2018), where the ratio of inhibitory-excitatory concentrations was found to correlate with strength of volitional modulation.

4.3. Comparison with non-auditory rivalry literature

A wealth of literature on perceptual rivalry, the most widely-studied example being binocular rivalry, provides valuable context that has motivated several aspects of this study and offers valuable points for comparison.

Binocular rivalry, and other bistable visual paradigms, including dynamic plauds (Pressnitzer and Hupé, 2006), share common characteristics (Leopold and Logothetis, 1999) with bistable auditory perception: exclusivity (one percept dominates at a time), inevitability (perception eventually switches) and randomness (variability in the timing of switches). This generalisation between visual and auditory bistability was also found to extend to stimulus strength manipulations (Rankin et al., 2015), which predominantly lead to longer durations for the dominant percept (rather than shorter durations for the non-dominant percept), a property known as Levelt’s proposition II in the rivalry literature (Moreno-Bote et al., 2010; Brascamp et al., 2015). Attentional effects on dominance durations have been widely explored in binocular rivalry (e.g. Lack (1978); Mueller and Blake, 1989; Meng and Tong, 2004; Van Ee et al., 2005; Chong and Blake, 2006; Chong et al., 2005; Hancock and Andrews, 2007). Different mechanisms that involve excitation or inhibition of the dominant or non-dominant eye (e.g. percept) were hypothesised and linked to the dominant durations increasing or non-dominant durations decreasing in Lack (1978, Chapter IV). Lack’s data supported a mixed mechanism where the dominant durations increased and the non-dominant durations decreased. However, other studies report a range of effects in binocular rivalry (and other types of rivalry including ambiguous figures) including a reduction of the non-dominant durations (Pressnitzer and Hupé, 2006; Hancock and Andrews, 2007), prolonging of dominant durations (Chong et al., 2005; Van Ee et al., 2005) or a mix of the two effects (Meng and Tong, 2004; Van Ee et al., 2005). Experimentalists have hypothesised that attentional control in binocular rivalry could be equivalent to a low-level input gain, i.e. contrast changes for the stimulus in one eye when the corresponding percept is active (Mueller and Blake, 1989; Chong et al., 2005) or inactive (Mueller and Blake, 1989). The effect of attention control on dominance durations depends on a range of factors including the bistable paradigm considered, choice of stimulus parameters (e.g. the baseline contrast in binocular rivalry) and the type of attention tested (percept-based or via a sub-task that requires deployment of attentional resources). We suggest that other factors could be at play at the individual level including general inter-subject variability, a given subject’s experience with the stimulus, or deployment of different strategies that may be influenced by instructions.

4.4. Computational models

Models of auditory stream segregation have been developed with a range of approaches, with a few addressing auditory bistability, but none looking at attentional mechanisms. Below we highlight some modelling studies from non-auditory bistability. Most existing computational models of auditory streaming focused on reproducing the dependence of perceptual bias, and/or the dynamics of build-up, on DF and presentation rate (recent reviews: Szabó et al., 2016; Snyder and Elhilali, 2017; Rankin and Rinzel, 2019; Rankin and Rinzel, 2019). Several recent models focused on auditory bistability with competition dynamics (Mill et al., 2013; Rankin et al., 2015) or probabilistic switching schemes (Steele et al., 2015; Barnel and Nelken,
A recent paper implemented a three-stage hierarchical model including computation of auditory features, objects and typical competition mechanisms (Little et al., 2020). Competition-based models proposed for visual bistability incorporate mutual inhibition, slow adaptation and noise (Laing and Chow, 2002). Our neuromechanistic approach with neuro-based time- and feature-dependence of inputs is distinct from recent modeling studies of auditory bistability that involve percept-based inputs and competition (Mill et al., 2013; Barniv and Nelken, 2015); it is an ideal framework to investigate attention as common mechanisms used in the model are found throughout cortex (cross-inhibition, recurrent excitation, neural adaptation) and likely part of attentional architecture.

Some models for visual rivalry have multiple stages and also include some feature representation (Wilson, 2003; Said and Heeger, 2013; Klink et al., 2008; Li et al., 2017). Klink et al. (2008) modelled attentional control via a top-down mechanism interacting with early stages of processing. A recent study features an attentional mechanism reminiscent of the recurrent excitation implemented here, although the study focused on attending to the stimulus as a whole, rather than the effect of attentional control on percept durations (Li et al., 2017).

4.5. Perspectives

How might we distinguish different attentional strategies or mechanisms in future experimental and modelling work? In the visual literature, attempts have been made to find an equivalent manipulation of the stimulus to capture the effects of attentional control, e.g. in binocular rivalry, by finding an equivalent adjustment to stimulus contrast that mimics attention that is either fixed (Chong and Blake, 2006; Klink et al., 2008) or dependent on the currently-reported perceptual state (Mueller and Blake, 1989; Chong et al., 2005). In binocular rivalry the two percepts are directly linked with the stimulus delivered to each eye, however in auditory bistability the two percepts are not equivalent and involve different organisations of the same set of tones. There is not a parameter that can be tweaked to, say explicitly boost the integrated percept (like contrast for one eye). However, we know that DF (and other parameters) do shift perceptual bias towards one percept (toward integrated at low DF and segregated at high DF). A perceptual-state-dependent bias could be introduced via a dynamic shift in DF when a given percept is active (and shift back when inactive). However, sharp changes in DF can elicit resetting effects (Rogers and Bregman, 1998; Roberts et al., 2008; Haywood and Roberts, 2010; Rankin et al., 2017; Higgins et al., 2021), but this could be avoided by using spectrally broader, non-pure-tone stimuli with less salient changes to DF. Streaming paradigms that probe time-varying stimulus elements with feature variation in streams (Rahne and Sussman, 2009; Bendixen et al., 2010), or with slowly varying stimulus parameters (Byrne et al., 2019), move towards more realistic listening environments. Studies that incorporate speech elements into streaming and bistability paradigms (Kashino and Kondo, 2012; Billig et al., 2013; David et al., 2017) build further links to investigating attention in multi-talker environments (Zion Golumbic et al., 2013).

Another avenue to explore attentional strategies would be to give instructions to subjects that encourage attentional bias, but only in certain epochs of perceptual dominance (e.g. for Attend Seg: “switch away from integrated if you hear it, but listen passively to segregated” versus “hold onto segregated if you hear it, but listen passively to integrated”). This may encourage different strategies that relate to different theoretical mechanisms, allowing for the effects seen across this and other studies to be teased apart (Pressnitzer and Hupé, 2006; Billig et al., 2018; Kondo et al., 2018). Transient perturbations (e.g. deviant or extra tones (Rankin et al., 2017; Higgins et al., 2021)) may further have different attention-mechanism-dependent effects that could be explored with experiments and modelling.

4.6. Conclusions

We contribute to a renewed interest in the affects of attentional control in the auditory streaming (Gutschalk et al., 2015; Mehta et al., 2016; Kaya and Elhilali, 2017; Billig et al., 2018; Kondo et al., 2018; Brace and Sussman, 2021) with an exploration of how different mechanisms may be distinguished based on attentional effects during dominance of attended or unattended perceptual organisations. A percept-specific input gain provides the best match with qualitative features of our data; the target is a synaptic gain at the model’s competition stage, downstream of A1. Bottom up manipulations of input strengths at an earlier stage than the units encoding specific percepts would not be able to capture the effects of attentional control. More generally, the proposed modelling mechanisms open up avenues to understand how and when attentional control is active during auditory bistability. The locus of attentional control, and the neural mechanisms through which it acts, could be revealed via perceptual-state-dependent stimulus design or via nuanced instructions, complemented by non-invasive imaging approaches (Billig et al., 2018; Higgins et al., 2020), by intracranial studies (Curtu et al., 2019) and by theoretical developments.

Data availability

All experimental data and model code are available in the github repository james-rankin/auditory-streaming: https://github.com/james-rankin/auditory-streaming

Funding

Rankin acknowledge support from an Engineering and Physical Sciences Research Council (EPSRC) New Investigator Award (EP/R03124X/1) and from the EPSRC Hub for Quantitative Modelling in Healthcare (EP/T017856/1). Rankin was supported by the Swartz Foundation whilst a postdoc at NYU. Rinzel acknowledges an NYU, University Research Challenge Fund grant in support of the behavioral experiments.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. ANOVA tables

Repeated-measures ANOVA tables are reported in Table 1.

Appendix B. Attention at equidominance to reproduce results from Pressnitzer and Hupé (2006)

A pilot experiment used stimulus parameters (DF = 5 and the 125 ms tones) in line with Pressnitzer and Hupé (2006) and used a similar procedure as outlined above in Section 2.5 (we note that in our continuous two choice task there was no specific instruction about times when participants were confused or unable to classify the percept as integrated or segregated). The results in terms of percept durations are shown in Fig. 5. Whilst the
proportion of time spent hearing integrated for each condition was qualitatively similar to Pressnitzer and Hupé (2006) (not shown), this was achieved by both increasing durations of the attended percept and decreasing durations of the unattended percept (rather than only decreasing unattended durations as in Pressnitzer and Hupé (2006)). These pilot results were qualitatively closer to more recent studies also finding a mixed effect on attended and unattended durations (Billig et al., 2018; Kondo et al., 2018).

Appendix C. Model equations

We note two errors in the description of the model in Rankin et al. (2015): inhibition from the $r_{AB}$ unit to the $r_A$ and $r_B$ units is assumed stronger than other inhibitory connections by a factor of 2 (incorrectly reported as a factor of 1 in our earlier study) and the tonotopic decay of input amplitudes is equal $\sigma_p = 4.25$ (incorrectly reported as twice this value in our earlier study). Here the adaptation strength $g = 0.125$ and the tonotopic decay $\sigma_p = 5.5$ were tuned to match subject-average data in the Passive condition (as shown in Fig. 3E left) and these values were kept fixed for the attention conditions.

The model equations are given by

$$
\begin{align*}
\tau_r \dot{r}_{AB} &= -r_{AB} + F \left( \beta_r + \mu_i \right) d_{AB} e_{AB} - \beta_r r_{AB} - \beta_i (r_A + r_B) - \gamma \dot{e}_{AB} + \left[ 1 + \lambda_i(t) \right] \left[ I_A + I_B \right] \frac{DF}{2} + \chi_{AB}, \\
\tau_r \dot{r}_A &= -r_A + F \left( \beta_r + \mu_i \right) d_{A} e_{A} - \beta_r r_{A} - 2\beta_i r_{AB} + \gamma \dot{e}_{A} + \left[ 1 + \lambda_i(t) \right] \left[ I_A + \frac{DF}{2} I_B \right] + \chi_{A}, \\
\tau_r \dot{r}_B &= -r_B + F \left( \beta_r + \mu_i \right) d_{B} e_{B} - \beta_r r_{B} - 2\beta_i r_{AB} + \gamma \dot{e}_{B} + \left[ 1 + \lambda_i(t) \right] \left[ I_B + \frac{DF}{2} I_A \right] + \chi_{B}, \\
\tau_d \dot{d}_{AB} &= -d_{AB} + \left( 1 - k r_{AB} \right), \\
\tau_d \dot{d}_A &= -d_A + \left( 1 - k r_{A} \right), \\
\tau_d \dot{d}_B &= -d_B + \left( 1 - k r_{B} \right).
\end{align*}
$$

(1)

Terms involved in the attention mechanisms $\mu_i$, $\mu_x$, $\lambda_i(t)$ and $\lambda_x(t)$ are defined at the end of this section. The synaptic time constant for each unit is $\tau_r = 10$ ms. The function $F$ translates the synaptic inputs to each population into a firing rate and takes a sigmoidal form

$$
F(u) = \frac{1}{1 + \exp(\theta_F(\theta_f - u + \theta_j))},
$$

with threshold $\theta_F = 0.2$ and slope $k_F = 12$. Excitation with strength $\beta_r = 0.85$ is assumed to be local, to work on an intermediate NMDA-like timescale of $\tau_r = 70$ ms and undergo slow synaptic depression on a timescale $\tau_d = 3$ s with strength $\kappa = 0.25$. Inhibition with strength $\beta_i = 0.3$ is assumed global and to act instantaneously. Inhibition from the $r_{AB}$ unit to the $r_A$ and $r_B$ units is stronger by a factor of 2 as in Huguet et al. (2014). Spike-frequency adaptation has strength $g = 0.0525$ and a slow timescale $\tau_a = 1.4$s. Inputs to the model mimic the onset-plateau responses to pure tones in A1 with onset timescale $\tau_i = 15$ ms, plateau timescale $\tau_2 = 82.5$ ms and peak to plateau ratio $\lambda_2 = 1/6$. Inputs are given by the following double $\alpha$-function where $H(t)$ is the heaviside function.

$$
I(t) = H(t) \left[ \frac{\exp(2) - \exp(2)}{\alpha_1^2} + \lambda_2 \exp(2) - \exp(2) \right].
$$

Input amplitudes depend on DF with peak amplitude $I_p = 0.47$, decaying away over tonotopy with a spatial scale $\sigma_p = 4$st as defined by

$$
\frac{w(DF)}{DF} = I_p \frac{DF}{\sigma_p^2}.
$$

Additive noise is introduced with independent stochastic processes $\chi_{AB}, \chi_{A}$, and $\chi_{B}$ and added to the inputs of each population as in Shpiro et al. (2009) and Seely and Chow (2011). Input noise is modeled as an Ornstein-Uhlenbeck process:

$$
\dot{\chi}_t = -\frac{\dot{\chi}_t}{\tau_d} + \gamma \sqrt{\frac{\dot{\chi}_t}{\tau_A}}(t).
$$

where $\tau_k = 100$ ms (a standard choice Shpiro et al., 2009; Seely and Chow, 2011) is the timescale, $\gamma = 0.075$ the strength and $\dot{\chi}(t)$ a white noise process with zero mean. Note these terms appear inside the firing rate function $F$ such that firing rates $r_i$ remain positive and do not exceed 1. Simulations were run in Matlab using a standard Euler-Murayama time stepping scheme with a stepsize of 5 ms (half the value of the fastest timescale in our equations $\tau_r = 10$ ms).Reducing this timestep by a factor of 10 did not change the results.

New terms involving a boost to recurrent excitation via $\mu_i$ and $\mu_x$ or a boost to the attended units’ inputs via $\lambda_i(t)$ and $\lambda_x(t)$ (possibly time-dependent) are defined as follows for the three attention mechanisms:
• M1-ImpAtt-Mixed: The additional input is fixed as a constant and $\mu_S = \mu_A = 0$. For Attend Int $\lambda_I(t) = \lambda_I$ (the attentional gain) and $\lambda_S(t) = 0$. For Attend Seg $\lambda_I(t) = \lambda_I$ and $\lambda_S(t) = 0$.

• M2-Recur-IncAtt: Additional input terms are off $\lambda_I(t) = \lambda_S(t) = 0$. For Attend Int $\mu_I$ gives the attentional gain and $\mu_S = 0$. For Attend Seg $\mu_S$ gives the attentional gain and $\mu_I = 0$.

• M3-StDep-DecUn: Additional input terms are time dependent and $\mu_I = \mu_S = 0$. For Attend Int $\lambda_I$ in the attentional gain has sigmoidal activation with threshold $\theta_{an}$ and slope $k_{an}$ depending on the activity of the units encoding segregated ($r_A$ and $r_B$):

$$
\hat{\lambda}_I(t) = \frac{\lambda_I}{1 + \exp(-((r_A + r_B)/k_{an} - \theta_{an}))}
$$

and $\hat{\lambda}_S(t) = 0$. For Attend Seg $\hat{\lambda}_S$ is the attentional gain depending on the activity of the unit encoding integrated ($r_A$):

$$
\hat{\lambda}_S(t) = \frac{\lambda_S}{1 + \exp(-((r_A + r_B)/k_{an} + \theta_{an}))}
$$

and $\hat{\lambda}_I(t) = 0$.

All model code is available in the following GitHub repository: james-rankin/auditory-streaming.

References

Amirs, S., Saida, S., 1985. Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol. Hum. Percept. Perform. 11, 257–271.

Barniv, D., Nelken, I., 2015. Auditory streaming as an online classification process with evidence accumulation. PLoS One 10, e0147891.

Beauvois, M., Meddis, R., 1997. Time decay of auditory stream biasing. Perception Psychophys. 59, 81–86.

Brascamp, J., Klink, P., Levelt, W.J., 2015. The ‘laws of binocularity’. Trends Cogn. Sci. 19, 1406–1412.

Bregman, A., 1994. Auditory scene analysis: The perceptual organization of sound. MIT press.

Byrne, A., Rinzel, J., Rankin, J., 2019. Auditory streaming and bistability paradigm extended to a dynamic environment. Hearing Res. 383, 107807.

Carson, R., Foxton, J., Robertson, I., 2007. Effects of attention and unilateral neglect on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. 257, 115–127.

Carson, R., Plack, C., Fantini, D., Carson, R., 2003. Cross-modal and non-sensory influences on auditory streaming. Perception 32, 1393–1402.

Chong, S., Tadin, D., Blake, R., 2005. Endogenous attention prolongs dominance in binocular rivalry. J. Vis. 5, 1004-1012.

Chong, S.C., Blake, R., 2006. Exogenous attention and endogenous attention influence initial dominance in binocular rivalry. Vision Res. 46, 1794-1803.

Curto, R., Wang, X., Brunton, B.W., Nouraki, K.V., 2019. Neural signatures of auditory perceptual bistability revealed by large-scale human intracranial recordings. J. Neurosci. 39, 6482-6497.

David, M., Lavandier, M., Grimaud, N., Osenham, A.J., 2017. Sequential stream segregation of voiced and unvoiced speech sounds based on fundamental frequency. Hearing Res. 344, 235-243. https://doi.org/10.1016/j.heares.2016.11.016.

Denham, S.L., Farkas, D., Van Ee, R., Tarana, M., Kocsis, Z., Wimmer, M., Carmel, D., Winkler, I., 2018. Similar but separate systems underlie perceptual bistability in human auditory cortex. PLOS One 10.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.G., 2009. Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. Behavior. Res. Methods 41, 1149-1160.

Fishman, Y., Arezzo, J., Steinmeidner, M., 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J. Acoust. Soc. Am. 116, 1656-1670.

Fishman, Y., Reser, D., Arezzo, J., Steinmeidner, M., 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. 151, 167–187.

Gilbert, C.D., Sigman, M., 2007. Brain States: Top-Down Influences in Sensory Processing. Neuron 54, 677-696. https://doi.org/10.1016/j.neuron.2007.05.019.

Hancock, S., Andrews, T.J., 2007. The role of voluntary and involuntary attention in selecting perceptual dominance during binocular rivalry. Perception 36, 288.

Harris, K.D., Thiele, A., 2011. Cortical state and attention. Nature Rev. Neurosci. 12, 509–522.

Haywood, N., Roberts, B., 2010. Build-up of the tendency to segregate auditory streams: resetting effects evoked by a single deviant tone. J. Acoust. Soc. Am. 128, 2039-2041.

Higgins, N.C., Little, D.F., Verkes, B.D., Nave, K.M., Kuruvilla-Mythath, A., Elhilali, M., Snyder, J.S., 2020. Neural correlates of perceptual switching while listening to bistable auditory streaming stimuli. Neuroimage 204, 116220.

Higgins, N.C., Monjaras, A.G., Verkes, B., Little, D.F., Nave-Bledgett, J.E., Elhilali, M., Snyder, J.S., 2021. Resettting of auditory and visual segregation occurs after transient stimuli of the same modality. Front. Psychol. 12 https://doi.org/10.3389/fpsyg.2021.72031.

Huguet, G., Rinzel, J., Hupé, J.M., 2014. Noise and adaptation in multistable perception: Noise drives when to switch, adaptation determines percept choice. J. Vis. 14, 19. Itatani, N., Klump, G., 2014. Neural correlates of auditory streaming in an objective behavioral task. Proc. Natl. Acad. Sci. USA 111, 10738-10743.

Kashino, M., Kondo, H., 2012. Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. Philos. T Roy. Soc. B 367, 977-987.

Kaya, E.M., Elhilali, M., 2017. Modelling auditory attention. Philos. Trans. R. Soci. B: Biol. Sci. 372, 20160101. https://doi.org/10.1098/rstb.2016.0101.

Klinic, P.C., Ee, R., Nijs, M.M., Brouwer, G.J., Noest, A.J., Wessel, R.I.A.V., 2008. Early interactions between neuronal adaptation and voluntary control determine perceptual choices in bistable vision. J. Vis. 8, 16–16. https://doi.org/10.1167/8.5.1.

Kondo, H.M., Kashino, M., 2009. Involvement of the thalamocortical loop in the spontaneous switching of percept in auditory streaming. J. Neurosci. 29, 12695-12701.

Kondo, H.M., Pressnitzer, D., Shimada, V., Kochiyama, T., Kashino, M., 2018. Inhibition-excitation balance in the parietal cortex modulates volitional control for auditory and visual multistability. Sci. Rep. 8, 14548.

Lack, L.C., 1978. Selective attention and the control of binocular rivalry. volume 11 De Gruyter Mouton.

Laing, C., Chow, C., 2002. A spiking neuron model for binocular rivalry. J. Comput. Neurosci. 12, 39–53.

Leopold, D., Logothetis, N., 1999. Multistable phenomena: changing views in perception. Trends Cogn. Sci. 3, 254–264.

Li, H.H., Rankin, J., Rinzel, J., Carrasco, M., Heeger, D., 2017. Attention model of binocular rivalry. Proc. Natl. Acad. Sci. USA 114, E6192-E6201.

Little, D.F., Snyder, J.S., Elhilali, M., 2020. Ensemble modeling of auditory streaming reveals potential sources of bistability across the perceptual hierarchy. PLoS Comput. Biol. 16, e1007746.

Macken, W.J., Tremblay, S., Houghton, R.J., Nicholls, A.P., Jones, D.M., 2003. Does auditory streaming require attention? evidence from attentional selectivity in short-term memory. J. Exp. Psychol. Hum. Perform. 29, 43–51.

Mehta, A.H., Vasin, I., Osenham, A.J., Shamma, S., 2016. Neural correlates of attention and streaming in a perceptually multistable auditory illusion. J. Acoust. Soc. Am. 140, 2225–2233.

Meneghetti, T., Teng, F., 2004. Can attention selectively bias bistable perception? differences between binocular rivalry and ambiguous figures. J. Vis. 4, 2.

Meyrich, C., Tian, B., Carlon, R., Rauschecker, J., 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139–148.

Mill, R., Boin, T., Bendiren, A., Winkler, I., Denham, S., 2013. Modelling the emergence and dynamics of perceptual organisation in auditory streaming. PLOS Comput. Biol. 9, e1002925.
Said, C.P., Heeger, D.J., 2013. A model of binocular rivalry and cross-orientation suppression. PLOS Comput. Biol. 9, e1002991.

Seely, J., Chow, C.C., 2011. Role of mutual inhibition in binocular rivalry. J. Neurophysiol. 106, 2136–2150.

Shamma, S., Elhilali, M., McIlvay, C., 2011. Temporal coherence and attention in auditory scene analysis. Trends Neurosci. 34, 114–123.

Shapiro, A., Moreno-Bote, R., Rubin, N., Rinzel, J., 2009. Balance between noise and adaptation in competition models of perceptual bistability. J. Comput. Neurosci. 27, 37–54.

Snyder, J., Carter, O., Lee, S.K., Hannon, E., Alain, C., 2008. Effects of context on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. 34, 1007.

Snyder, J.S., Alain, C., Picton, T.W., 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. J. Cognit. Neurosci. 18, 1–12.

Snyder, J.S., Elhilali, M., 2017. Recent advances in exploring the neural underpinnings of auditory scene perception. Ann. N.Y. Acad. Sci. https://doi.org/10.1111/nyas.13317.

Steele, S., Tranchina, D., Rinzel, J., 2015. An alternating renewal process describes the buildup of perceptual segregation. Front. Comput. Neurosci. 8, 1–13.

Sussman, E., Horvath, J., Winkler, I., Orr, M., 2007. The role of attention in the formation of auditory streams. Perception Psychophysics 69, 136–152.

Szabó, B.T., Denham, S.L., Winkler, I., 2016. Computational Models of Auditory Scene Analysis: A Review. Front. Neurosci. 10 https://doi.org/10.3389/fnins.2016.00524.

Szalardy, O., Böhm, T., Bendixen, A., Winkler, I., 2013. Event-related potential correlates of sound organization: Early sensory and late cognitive effects. Biol. Psychol. 93, 97–104.

Taylor, A.L., Cottrell, G.W., Kristan Jr, W.B., 2002. Analysis of oscillations in a reciprocally inhibitory network with synaptic depression. Neural Comput. 14, 561–581.

Van Er, R., Noest, A., Brascamp, J., van den Berg, A., 2006. Attentional control over either of the two competing percepts of ambiguous stimuli revealed by a two-parameter analysis: Means do not make the difference. Vision Res. 46, 3129–3141.

Van Er, R., Van Dam, L., Broower, G., 2005. Voluntary control and the dynamics of perceptual bistability. Vision Res. 45, 41–55.

Wilson, H., 2003. Computational evidence for a rivalry hierarchy in vision. Proc. Natl. Acad. Sci. USA 100, 14499–14503.

Winkler, I., Denham, S.L., Mill, R., Böhm, T.M., Bendixen, A., 2012. Multistability in auditory stream segregation: a predictive coding view. Philos. Trans. R. Soci. B: Biol. Sci. 367, 1001–1012.