The dynamics of auditory stream segregation: Effects of sudden changes in frequency, level, or modulation

Saima L. Rajasingam

Department of Vision and Hearing Sciences, Anglia Ruskin University, Cambridge CB1 1PT,

United Kingdom

Robert J. Summers and Brian Roberts[a]

School of Psychology, Aston University, Birmingham B4 7ET, United Kingdom

Running Title: Dynamics of stream segregation

Version Date: April 28, 2021

[a]Electronic mail: b.roberts@aston.ac.uk, ORCID: 0000-0002-4232-9459.

1

21 **ABSTRACT**

22   Three experiments explored the effects of abrupt changes in stimulus properties on

23 streaming dynamics. Listeners monitored 20-s-long LHL– tone sequences and reported the number

24 of streams heard throughout. Experiments 1 and 2 used pure tones and examined the effects of

25 changing triplet base frequency and level, respectively. Abrupt changes in base frequency (±3-12

26 semitones) caused significant magnitude-related falls in segregation (resetting), regardless of

27 transition direction, but an asymmetry occurred for changes in level (±12 dB). Rising-level

28 transitions usually decreased segregation significantly whereas falling-level transitions had little

29 or no effect. Experiment 3 used pure tones (unmodulated) and narrowly spaced (±25 Hz) tone pairs

30 (dyads); the two evoke similar excitation patterns but dyads are strongly modulated with a

31 distinctive timbre. Dyad-only sequences induced a strongly segregated percept, limiting scope for

32 further build-up. Alternation between groups of pure tones and dyads produced large, asymmetric

33 changes in streaming. Dyad-to-pure transitions caused substantial resetting, but pure-to-dyad

34 transitions sometimes elicited even greater segregation than for the corresponding interval in dyad-

35 only sequences (overshoot). The results indicate that abrupt changes in timbre can strongly affect

36 the likelihood of stream segregation without introducing significant peripheral-channeling cues.

37 These asymmetric effects of transition direction are reminiscent of subtractive adaptation in vision.

I. INTRODUCTION

Auditory stream segregation is the process by which sounds are grouped perceptually to form coherent representations of objects and events in the auditory scene (Bregman, 1990). The ability of the auditory system to segregate sounds into streams is commonly investigated using sequences of alternating low- (L) and high- (H) frequency pure tones, which listeners can hear either as one stream (integrated) or as two streams (segregated). It has long been known that sequences with a greater frequency separation (Δf) or faster presentation rate are more likely to be heard as segregated (Miller and Heise, 1950; Bregman and Campbell, 1971; van Noorden, 1975). Perception of these sequences is bistable, involving stochastic switching between one and two streams (Denham and Winkler, 2006; Pressnitzer and Hupé, 2006), but averaging over several trials can be used to reveal the probability of hearing a segregated percept and how this changes over time (Carlyon et al., 2001; Roberts et al., 2002). Despite decades of research on auditory stream segregation, we still know relatively little about its dynamics. One reason for this is because most studies investigating auditory streaming have used sequences whose properties remain constant throughout; another is that those studies which have introduced changes in acoustic properties have quantified their effects using one-off measures rather than measures of how the effects of a change unfold over time. The experiments reported here used tone sequences whose properties were changed at one or more time points and for which the consequences of those changes were tracked over time (cf. Haywood and Roberts, 2013; Rajasingam et al., 2018).

The mostly widely investigated aspect of the dynamics of stream segregation is an effect known as build-up, in which the tendency to segregate a repeating tone sequence of fixed rate and frequency separation into two streams increases over time (van Noorden, 1975; Bregman, 1978). Anstis and Saida (1985) investigated build-up further using long ($\geq$30 s) repeating sequences of

64   alternating pure tones (LHLH…) and discovered that build-up has two distinct stages; there is a

65   rapid increase in the tendency to hear two streams over the first 10 s followed by a slower rise

66   thereafter. Once a tone sequence ends, or is interrupted with a silent gap, the accumulated build-

67   up decays over a few seconds (Bregman, 1978; Beauvois and Meddis, 1997). A convenient way

68   to explore how the perceptual organization of later sounds is influenced by earlier sounds involves

69   a stimulus configuration in which standard test sequences (whose properties remain the same

70   across conditions) are immediately preceded by various types of induction sequence (i.e., stimuli

71   intended to cause prior build-up) or by none (control condition). Studies of this kind, or variants

72   thereof, have shown that another way in which accumulated build-up can be reduced or lost is

73   through a sudden change in the acoustic properties of the sequence, such as a change in frequency

74   region (Anstis and Saida, 1985) or in lateralization or level (Rogers and Bregman, 1998). This loss

75   may occur either because the accumulated build-up was specific to properties of the original

76   sounds, and so fails to transfer to the new sounds, or because sudden change triggers an active

77   resetting of build-up (Rogers and Bregman, 1998; Roberts et al., 2008).

78   Distinguishing experimentally between failure to transfer and resetting as accounts of the

79   loss of build-up following sudden change can be challenging, but the ability of a single deviant

80   tone at the end of an induction sequence to decrease the impact of the inducer on segregation in

81   the following test sequence suggests that there are at least some circumstances in which active

82   resetting is involved (Haywood and Roberts, 2010, 2013). In practice, the loss of build-up arising

83   from either cause is usually referred to as resetting. Unlike abrupt changes, gradual changes in

84   acoustic properties of a tone sequence such as lateralization or level have little or no effect on

85   streaming (Rogers and Bregman, 1998). Bregman (1978) proposed that build-up occurred because

86   integration was the default percept for sound sequences and that segregation emerged over several

4

87   seconds as a result of a conservative evidence-accumulation process indicating that more than one

88   source was active. The slow time course of this process was seen as serving to stabilize perception,

89   thereby preventing the auditory system from fluctuating rapidly between alternative

90   interpretations. In this functional account, the loss of build-up arising from sudden change is

91   interpreted as the resetting of this evidence-accumulation process because the change signals a

92   new auditory scene. A related idea is that a sudden correlated change applied to both subsets of

93   tones in an alternating sequence signals their common origin and so encourages their integration.

94   Most streaming studies using induction sequences have focused primarily on exploring the

95   strong segregation-promoting effect that occurs when an alternating-frequency (AF) test sequence

96   is preceded by a constant-frequency induction sequence corresponding to one subset of the test-

97   sequence tones (Rogers and Bregman, 1993; Roberts et al., 2008; Haywood and Roberts, 2010,

98   2013; Rajasingam et al., 2018). A notable exception is the study by Rogers and Bregman (1998)

99   in which the induction and test sequences both involved frequency alternation. Their study

100   included conditions exploring the effects of sudden changes in stimulus lateralization and level.

101   Also relevant is the study by Anstis and Saida (1985) in which the effects of sudden changes in

102   stimulus center frequency were explored using a variant of the induction-test configuration

103   involving alternation between an inducer (there called an adapting sequence) and a test sequence.

104   These studies, their findings, and limitations are considered in detail in Sec II and Sec III. They

105   provide evidence that sudden changes in stimulus properties can lead not only to substantial loss

106   of build-up previously accumulated but that in some cases the effects of change are directional,

107   leading to asymmetries in listeners' responses to them.

108   It is rare in everyday life to hear sequences of sounds whose properties are static; the

109   auditory scene is usually constantly changing, sometimes gradually and sometimes suddenly. The

110   aim of the current study was better characterization of the effects of acoustic change on the

111   dynamics of stream segregation. Three experiments are reported here; all used the LHL–LHL–…

112   configuration first introduced by van Noorden (1975), where the dash represents a silence equal in

113   duration to one of the tones. When this configuration is used, the one-stream percept is heard as a

114   distinctive galloping rhythm, for which the pitch of the tones is heard to move from low to high

115   and vice versa; this rhythm is lost when the L and H subsets segregate and are heard independently

116   as higher and lower-pitched streams. This way of measuring streaming is sometimes referred to as

117   the "Horse or Morse" task (Cusack et al., 2004).

118   To overcome the limitations of the one-off measures of streaming used in many previous

119   studies, all three experiments involved continuous monitoring of the perceptual organization of the

120   tone sequence throughout a trial, allowing the time course of any effects of an abrupt transition to

121   be measured and compared, including with the time course of build-up at the start of the sequence.

122   Experiments 1 and 2 extended earlier work on the effects of sudden changes in center frequency

123   (Anstis and Saida, 1985) or in level (Rogers and Bregman, 1998), and also demonstrated that build-

124   up was largely unaffected when the center frequency of a tone sequence changed smoothly and

125   progressively rather than staying constant. Experiment 3 explored the effects on stream segregation

126   of sudden changes in timbre with minimal excitation-pattern cues (Hartmann and Johnson, 1991);

127   this was achieved using abrupt transitions between unmodulated and modulated tones or vice

128   versa. Responses to these transitions showed even more marked asymmetries than those observed

129   for level changes, indicating strong directional effects. This outcome represents a major challenge

130   to Bregman's (1978) functional account of build-up and further indicates that neurophysiological

131   models of build-up based on the slow accumulation of adaptation need to account for rapid

132   direction-sensitive changes in that adaptation following sudden transitions in acoustic properties.

6

II. EXPERIMENT 1

133

134   To our knowledge, the only experiment to explore parametrically the effect of changing

135 the center frequency of an AF tone sequence on subsequent streaming is the fourth experiment

136 reported by Anstis and Saida (1985). They presented a 4-s adapting sequence of tones (LHLH…)

137 with fixed properties that alternated with a 1-s test sequence whose center frequency fell at one of

138 eleven values from the set ±12, 4, 3, 2, 1, and 0 semitones (ST) relative to the center frequency of

139 the adapting sequence (1 kHz). Switches between the adapting and test stimuli were without

140 silences and so occurred seamlessly. The adapting and test stimuli shared a fixed Δf of 2 ST (i.e.,

141 for the adapting stimulus, L tones = 944 Hz and H tones = 1060 Hz). The purpose of the adapting

142 stimulus was to induce a build-up in the tendency to hear two streams; its presentation rate was 4

143 cycles/s, corresponding to a tone repetition time (TRT) of 125 ms. Listeners had control of the

144 presentation rate of the test stimulus and their task was to adjust it as necessary to ensure that the

145 test stimulus lay at the perceptual borderline between integration and segregation. This measure,

146 known as the nulling rate because the adjustment was being made to offset the effect of build-up,

147 was taken to be the mean of the adjustment settings over the final 30 s of the 90-s trial and so only

148 one estimate was obtained per trial. Changes in nulling rate with center frequency of the test

149 stimulus were used to plot a tuning curve for the effect of the adapting stimulus—the more build-

150 up that transferred from the adapting stimulus to the test stimulus, the slower the nulling rate must

151 be set for the test stimulus to track the perceptual borderline.

152   The tuning curve obtained was relatively narrow, with a flat top and steep skirts; it was

153 also positioned asymmetrically, centered on the +1-ST test stimulus. Build-up produced by the

154 adapting stimulus led to slower nulling rates for a test stimulus within ±1 ST of the tuning

155 frequency (i.e., 0 – 2 ST), but the effect of the adapting stimulus was largely extinguished when

156   the test stimulus was ±2 ST or more away from the tuning frequency. This pattern suggests that

157   build-up transfers better when the center frequency is shifted upwards than downwards. Anstis and

158   Saida (1985) noted that the bandwidth of tuning was broadly comparable with that of the auditory

159   critical band (Scharf, 1970) but they used only one adapting stimulus and so it is not known

160   whether the bandwidth of tuning to the adapting stimulus is affected by Δf. For example, increasing

161   Δf would lead to a greater overlap in frequency range between adapting and test stimuli for a given

162   shift in center frequency. Furthermore, the combination of small Δf (2 ST) and slow rate (TRT=125

163   ms) used limited considerably the extent of build-up during the adapting stimulus compared with

164   that which would have been produced if larger values of Δf and faster rates had been used (van

165   Noorden, 1975). Anstis and Saida (1985) offered no explanation for the asymmetrical frequency

166   tuning of build-up found in their study. The procedure they used did not allow investigation of the

167   time course of the effect of frequency change on subsequent streaming.

168   There is evidence that an induction sequence which changes gradually in lateralization or

169   level towards that for a steady test sequence, giving a smooth transition between them, leads to a

170   similar effect on the build-up of stream segregation as an induction sequence whose properties

171   match those of the test sequence (Rogers and Bregman, 1998). However, to our knowledge, the

172   effects of smooth and progressive change have not been explored in the context of frequency. Our

173   pilot observations suggested that tone sequences whose triplet base frequency changed in this way

174   were broadly as effective at inducing build-up as tone sequences with constant base frequency,

175   despite the differences in peripheral channeling between the two types of stimulus. Furthermore,

176   if confirmed, this outcome would pose a challenge to neurophysiological models in which build-

177   up in the tendency to hear two streams is a result of multi-second adaptation caused by repeated

178   stimulation of central auditory neurons with the same best frequency (e.g., Micheyl et al., 2005).

8

179   The first experiment reported here investigated the effect of introducing a single sudden

180   change (transition) in base frequency (corresponding to a distinct change in pitch range) in the

181   middle of a long test sequence. The magnitude and direction of this change was varied; Δf for the

182   test sequence was also varied. The abrupt transition was presented in the context of on-going small

183   but progressive changes throughout the rest of the sequence (0.5 ST between adjacent triplets), a

184   value which fell within the narrow adapting region identified by Anstis and Saida (1985). Gradual

185   changes of this kind, at least in level or lateralization, are known to have relatively little impact on

186   the subsequent likelihood of reporting stream segregation (Rogers and Bregman, 1998).

187   **A. Method**

188   *1. Listeners*

189   Listeners were recruited mainly from the student population at Aston University, gave

190   informed consent, and received either course credit or payment for taking part. They were first

191   tested using a screening audiometer (Interacoustics AS208, Assens, Denmark) to ensure that their

192   audiometric thresholds at 0.5, 1, 2, and 4 kHz did not exceed 20 dB hearing level. All listeners

193   who passed this screening took part in a training session designed to familiarize them with the task

194   and stimuli before proceeding to the main session; exclusion criteria were predefined in relation

195   to a listener's profile of responses in the reference conditions (see Sec IIA3). Twelve listeners (6

196   males) successfully completed the experiment (mean age = 25.3 years, range = 21.8 – 29.3). This

197   research was approved by the Aston University Ethics Committee.

198   *2. Stimuli and conditions*

199   The test sequence was 20 s long and comprised 50 LHL– triplets. Each tone was 100-ms

200   long (including 10-ms raised-cosine ramps). The silence at the end of each triplet was also 100 ms

201   long, giving an onset-to-onset duration between triplets of 400 ms. This rate of presentation is

202   known to facilitate streaming based on frequency separation (e.g., Bregman and Campbell, 1971;

203   van Noorden, 1975). The base frequency of each triplet—defined as the frequency of its constituent

204   L tones—was constant in some conditions but varied in others and ranged from a maximum of 1

205   kHz to a minimum of 0.5 kHz. The frequency of the H tones was set relative to the L tones

206   according to the desired low-high frequency difference for the test sequence (Δf), which was 4, 6,

207   or 8 ST. For example, when the base frequency was 1000 Hz, the frequency of the H tones for

208   these values of Δf was 1260, 1414, and 1587 Hz, respectively. The range of frequency separations

209   used was chosen to reduce ceiling and floor effects, and to provide information on any interactions

210   that might occur between frequency separation and condition. All tones were presented at 73 dB

211   sound pressure level (SPL).

212   There were ten conditions, for which the base frequency of the triplets was manipulated. A

213   schematic depicting the test sequences is shown in Fig. 1; the left, middle, and right panels

214   illustrate conditions C1-C4, C5-C7, and C8-C10, respectively. In C1 and C2, the base frequency

215   was constant at 1 kHz and 0.5 kHz, respectively. In C3 and C4, the base frequency changed by 0.5

216   ST/triplet and followed either a linear rise-fall (C3) or fall-rise (C4) trajectory moving gradually

217   between the minimum and maximum base frequencies[1]. In C5-C7, the base frequency followed

218   the same rising path as C3 for the first half of the sequence to reach 1 kHz but at 10 s (triplet 26)

219   there was an *abrupt fall* in base frequency of 3 ST (C5), 6 ST (C6), or 12 ST (C7); thereafter, a

220   rising trajectory was resumed unless and until the maximum was reached, after which the falling

221   path was followed. In C8-C10, the base frequency followed the same falling path as C4 for the

222   first half of the sequence to reach 0.5 kHz, but at 10 s there was an *abrupt rise* in base frequency

223   of 3 ST (C8), 6 ST (C9), or 12 ST (C10); thereafter, a falling trajectory was resumed unless and

224 until the minimum was reached, after which the rising path was followed. All other properties of

225 the test sequences remained the same across conditions.

226 All stimuli were synthesized at a sampling rate of 20 kHz using MITSYN (Henke, 2005).

227 They were played back at 16-bit resolution over Sennheiser HD 480-13II earphones (Hannover,

228 Germany) via a Sound Blaster X-Fi HD sound card (Creative Technology Ltd., Singapore),

229 programmable attenuators (Tucker-Davis Technologies, TDT PA5, Alachua, FL), and a

230 headphone buffer (TDT HB7). Output levels were calibrated using a sound-level meter (Brüel and

231 Kjaer, type 2209, Nærum, Denmark) coupled to the earphones by an artificial ear (type 4153).

232 Diotic presentation was used throughout.

233
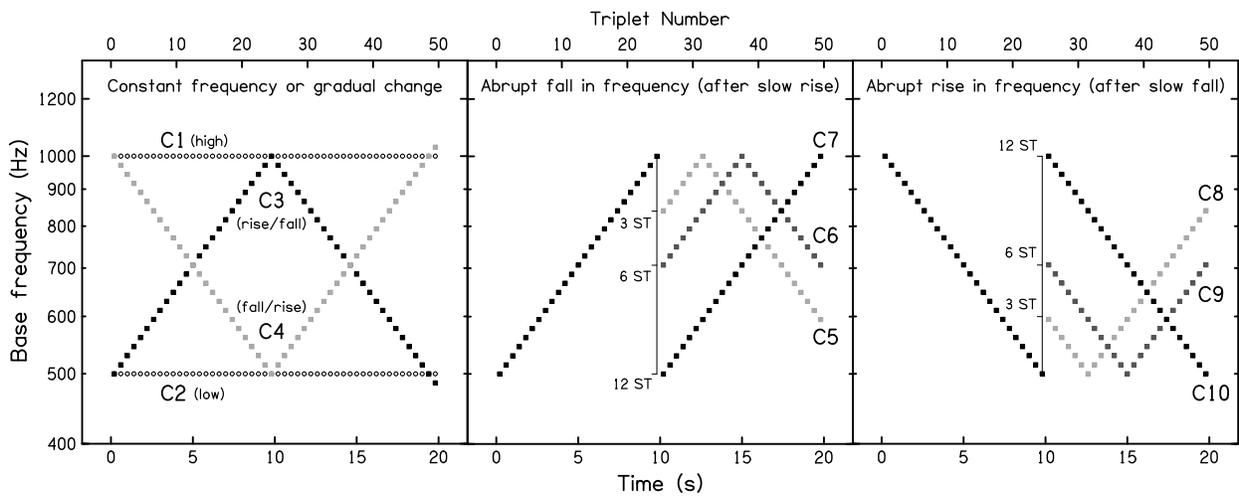
234

FIG. 1. Stimuli for experiment 1—schematic illustrating the conditions used (C1-C10). The left panel shows the base-frequency contours (i.e., L-tone frequencies) for the test sequences used in the constant (C1-C2), gradual rise/fall (C3), and gradual fall/rise (C4) conditions; each symbol (circle or square) represents one LHL– triplet. The middle panel shows the base-frequency contours for the conditions involving a slow rise over the first 10 s followed by an abrupt fall of 3, 6, or 12 ST (C5-C7). The right panel shows the base-frequency contours for the conditions involving a slow fall over the first 10 s followed by an abrupt rise of 3, 6, or 12 ST (C8-C10).

235

3. Procedure

Listeners completed the experiment in a single-walled sound-attenuating chamber (Industrial Acoustics 401A, Winchester, UK) housed within a quiet room. They were free to take breaks between trials whenever they wished. After reading the instructions, listeners completed one training block of trials identical to those used in the main experiment (see below); a second training block was offered but rarely required. During the training and main experiment, stimuli were presented in a newly randomized order in each block for each listener. Completing all stages (audiometry, training, and main experiment) usually took ~3 hours, divided into two separate sessions. The experiment was run using a program written in Visual Basic (Visual Studio, 2010, ver. 10.0); the program read from the hardware clock to record key-press timings.

Each trial was initiated 1 s after the listener pressed "Enter" on the computer keyboard. Listeners were instructed to monitor the test sequence continuously throughout; they were asked to indicate as soon as possible whether they were hearing integration (one stream) or segregation (two streams) by pressing either the "A" or "L" keys, respectively. Thereafter, listeners were asked to press the appropriate key every time their perception of the test sequence changed. They were asked to avoid listening actively for either integration or segregation, but simply to report which of the two percepts they heard at that moment; on occasions when the percept was ambiguous, listeners were asked to report the more dominant (cf. Haywood and Roberts, 2013; Rajasingam et al., 2018). At the end of each trial, there was a 5-s pause before listeners could initiate the next trial. Combined with the trial-initiation delay (1 s), this ensured a minimum silent gap of 6 s during which any prior build-up would decay before the onset of the next trial.

Each combination of condition (10 levels) and Δf (3 levels) was presented ten times in the main experiment, once in each block, giving 300 trials. Using three different Δf values also

12

259   provided a useful means of predefining criteria for excluding data. It is well established in the

260   literature that, for a given rate of presentation, an increase in the frequency separation between

261   subsets of pure tones increases the tendency to hear two streams (Miller and Heise, 1950; van

262   Noorden, 1975; Anstis and Saida, 1985). Therefore, for a listener's data to be included, the mean

263   overall extent of segregation for the conditions using steady sequences (C1 and C2) had to rise

264   when Δf was increased from 4 ST to 6 ST and rise again when Δf was increased from 6 ST to 8

265   ST. No listeners were replaced in this experiment.

266   *4. Data analysis and availability*

267   Response data from each trial were divided into twenty 1-s-long time bins (i.e., 0-1 s, 1-2

268   s, …, 19-20 s). For each time bin, the percentage of time for which the listener reported the test

269   sequence as segregated was computed from the timings of individual key presses. This value was

270   recorded only if the listener's first response had occurred before the current time bin or within the

271   first 0.5 s of that time bin. Owing to the small number of trials meeting this criterion for the 0-1 s

272   time bin, responses made during that interval were used only in the context of calculations

273   involving subsequent time bins; the 0-1 s time bin was excluded from all further analysis and

274   graphical representation (cf. Haywood and Roberts, 2013; Rajasingam et al., 2018).

275   For each listener, the data for each time bin were averaged across trial blocks separately

276   for each combination of condition and Δf. Each mean was computed only from those trials for

277   which that time bin met the acceptance criterion described above. On occasions when one of these

278   means was missing for a particular listener (13 cases, corresponding to ~0.2% of the data and all

279   occurring within the first few time bins), mean imputation was used to replace the missing value

280   with the mean of the corresponding values obtained from the other listeners. Finally, the data were

281   averaged across listeners, for each combination of condition and Δf, to yield the overall mean

282   percentage of time for which the test sequence was heard as segregated for each successive time

283   bin. This measure of the average time course of stream segregation over the test sequence is used

284   to display the results. Note that the standard errors were computed using the individual means only

285   from those listeners for whom an experimental estimate was obtained.

286   The effects of base frequency *per se* and of smooth, progressive changes in base frequency

287   were explored by comparing the extent of stream segregation across the constant and gradual-

288   change conditions for the full duration of the tone sequences used (in 1-s time bins, excluding 0-1

289   s). The effect of an abrupt transition in base frequency at 10 s in a given condition was explored

290   by comparing the extent of stream segregation following the transition with that for appropriate

291   comparator conditions during the same time interval. The reference comparison was with the test

292   sequence that followed the same base-frequency contour up to 10 s but which then changed

293   direction without discontinuity; other comparisons were between abrupt transitions of different

294   magnitude or direction. These comparisons were made using a single, longer, time interval that

295   focused on the period of peak response to the transition. This time interval was 4.0 s long and

296   began 1.2 s (3 triplets) after the transition[2], to allow sufficient time for the effect of the change to

297   be reflected in listeners' responses. It was long enough to encompass fully the peak response to

298   the transition and was chosen to correspond to the remaining time available between transitions in

299   subsequent experiments, which included conditions with transitions every 5.2 s.

300   All statistical analyses reported here were computed using R 3.6.3 (R Core Team, 2020)

301   and the ez analysis package (Lawrence, 2016). The time-series data obtained from the calculations

302   described above were analyzed using repeated-measures analysis of variance (ANOVA); the

303   measure of effect size reported here is partial eta squared (η^2_p). Comparisons among conditions

304   with constant or gradually changing base frequency were conducted using three factors—
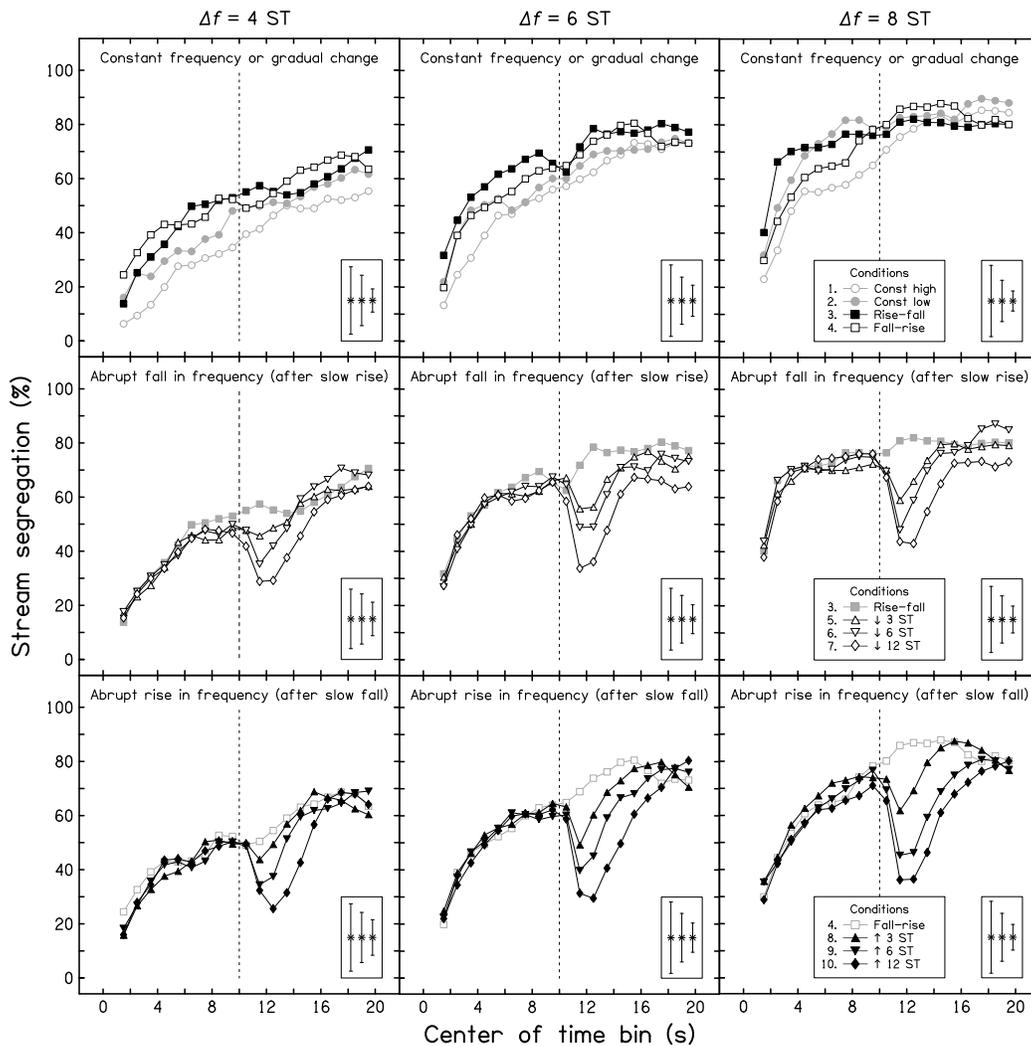
305   frequency separation between test-sequence tones (Δf), condition (C), and time interval (T, with

306   levels corresponding to time bins 1-2 s to 19-20 s). Comparisons exploring the effects of abrupt

307   transitions were based on the single time interval described above and so were conducted using

308   two factors (Δf and C). Pairwise comparisons (two tailed) were conducted using the restricted

309   least-significant-difference test (Snedecor and Cochran, 1967; Keppel and Wickens, 2004). The

310   research data underlying this publication are available on-line from a repository hosted by Aston

311   University[3].

B. Results and discussion

313   The results averaged across listeners are shown in Fig. 2. The upper panels display the

314   results for the conditions in which the base frequency of the test sequence was either constant or

315   gradually changing; the middle and lower panels display the results for the conditions in which

316   there was an abrupt fall or rise in base frequency, respectively, 10 s after the sequence began. To

317   facilitate comparisons, the middle and lower panels also reproduce the results for the gradual-

318   change reference conditions (C3 for the abrupt-fall cases and C4 for the abrupt-rise cases). These

319   results are considered in turn.

320   Although there is a suggestion in the mean data that slightly greater segregation is

321   associated with the lower base frequency for the constant conditions, and with the gradual-change

322   conditions relative to the constant conditions, this was not borne out by the analysis. An ANOVA

323   exploring the effects of the constant and gradual-change conditions is presented in Table I (data

324   for the series of 1-s time bins). Two of the three factors influenced streaming as main effects—

325   segregation was greater for larger frequency separations (means: 4 ST = 45.6%, 6 ST = 60.9%,

326   and 8 ST = 72.6%; $p < 0.001$) and tended to increase over time ($p < 0.001$)—but there was no main

327   effect of condition ($p = 0.378$) and none of the interaction terms were significant. The absence of a

328 main effect of condition remained even if the time bins included in the analysis were restricted to

329 the fast phase of build-up (1-2 s to 10-11 s; p=0.124). No difference was anticipated between the

330 two base frequencies (0.5 kHz and 1 kHz) because both fell within the frequency range for which

331 the ratio Δf over base frequency at the border between one- and two-stream percepts is roughly

332 constant (Miller and Heise, 1950). All conditions elicited broadly similar patterns of build-up—an

333 initial phase (up to ~10-12 s) that was relatively fast followed by a slower phase (Anstis and Saida,

334 1985; Haywood and Roberts, 2013; Rajasingam et al., 2018). In general, an increase in Δf tended

335 to increase both the rate and final extent of build-up.

336

16

FIG. 2. Results for experiment 1—effects of base-frequency condition (1-10, see insets in right panels) and test-sequence frequency separation (Δf) on the extent and time course of reported stream segregation (n=12). Responses for each trial were divided into 1-s time bins, for which the results were averaged across repetitions of the same type, and finally across listeners. Data for the first time bin (0-1 s) were excluded owing to the limited number of responses made during this interval (see main text). Note that the time shown on the abscissa indicates the center of the corresponding time bin. Results for each frequency separation (4, 6, and 8 ST) are displayed in separate panels from left to right. Results for the constant frequency or gradual change conditions (C1-C4), conditions involving an abrupt fall in frequency after a slow rise (C5-C7), and conditions involving an abrupt rise in frequency after a slow fall (C8-C10) are shown in the upper, middle, and lower panels, respectively. For ease of comparison, the results for the appropriate reference conditions (see upper panels) are reproduced in gray in the middle and lower panels. For clarity, the means displayed are not accompanied by individual error bars. Instead, each panel includes an inset showing summary information on the inter-subject standard errors obtained for each time bin in each condition (left = largest value, center = mean value across all time bins and conditions, right = smallest value).

337   Given that all tone sequences involving gradual change used a progressive shift in triplet

338   base frequency of 0.5 ST every 0.4 s, spanning a full octave over the first 10 s, the occurrence of

339   a similar pattern of results for the constant and gradual-change conditions indicates that the build-

340   up of stream segregation for AF tone sequences does not depend on repeated stimulation of the

341   same peripheral channels over several seconds. Furthermore, this finding suggests that the build-

342   up of stream segregation does not require extended stimulation of populations of central auditory

343   neurons with the same best frequency.

344   Inspection of Fig. 2 indicates that an abrupt fall or rise in base frequency decreased

345   subsequent stream segregation and that the extent of this grew progressively as the size of the

346   transition increased. An ANOVA exploring the effects of abrupt changes in base frequency is

347   presented in Table II (data from a single 4.0-s time bin starting 1.2 s after the transition). This

348   analysis showed significant main effects of Δf (p<0.001) and transition size (S, p<0.001). Despite

349   the suggestion in the mean data that the loss of segregation produced by an abrupt rise in frequency

350   (lower panels) was greater than that produced by an abrupt fall (middle panels), there was neither

351   a main effect of transition direction (D, p=0.812) nor a significant interaction involving it. Only

352   one interaction was significant ($\Delta f \times$ S, p=0.046) and this probably arose because the 3-ST

353   transitions had relatively little effect for the smallest Δf used (4 ST).

354   Given that there were no significant effects involving transition direction, we report here

355   the mean change in segregation for each size of transition after averaging across direction and Δf.

356   These values correspond to the difference in segregation in percentage points (% pts) produced by

357   that transition over the 4.0-s time bin relative to its reference case over the same interval (C3 and

358   C4 for falling and rising transitions, respectively). The mean changes in segregation for the 3-ST,

359   6-ST, and 12-ST transitions were -9.3% pts (p=0.002), -17.6% pts (p<0.001), and -29.9% pts

360   (p<0.001), respectively; all other pairwise comparisons were also significant (range: p=0.004–

361   p<0.001).] Although the loss of segregation associated with the 6-ST and 12-ST abrupt rises was

362   nominally a third larger (in % pts) than for their falling counterparts, these differences disappear

363   if the losses are interpreted in proportion to the extent of build-up taking place over 10 s for the

364   relevant gradual-change reference conditions. Notably, either a sudden rise or fall leads to a near-

365   complete loss of build-up for 12-ST transitions. Overall, there is no evidence to suggest an

366   asymmetrical effect of transition direction on streaming. In all cases, Fig. 2 shows that the impact

367   of the transition on streaming was greatest after ~2-3 s and thereafter the extent of segregation

368   began to grow again. The time course of this recovery from resetting was similar to that of the

369   original phase of build-up, eventually slowing as listeners' responses began to converge with those

370   for the corresponding reference cases. By the end of the sequence, segregation had mostly or

371   completely returned to where it would have been without the sudden transition.

372   The results of this experiment differ in important ways from those of its counterpart

373   reported by Anstis and Saida (1985). Their results indicated a narrow adapting region (~2 ST wide)

18

374   that was tuned asymmetrically (~1 ST above adapting-stimulus center frequency) and outside

375   which no transfer of build-up occurred. In contrast, our results indicate a much broader tuning of

376   the adapting region with shallower skirts—the resetting effect of a sudden change of 3 ST was

377   significant but partial and a change as large as 12 ST was required to cause a near-complete loss

378   of segregation. There was also no evidence of an effect of transition direction like that observed

379   by Anstis and Saida (1985). These outcomes suggest that the results of their study were strongly

380   influenced by one or more of the design factors considered above, which included the shorter

381   interval over which build-up could occur (4 s), the nulling procedure involving rate adjustment,

382   and the use of only a single small Δf (2 ST) and relatively long TRT (125 ms). Alternatively, or in

383   addition, it cannot be ruled out that our use of a gradually-changing adapting stimulus rather than

384   a constant one may have increased tolerance for a change in base frequency, widening the apparent

385   tuning. For example, in terms of Bregman's (1978) evidence-accumulation hypothesis of build-

386   up, perhaps a larger sudden change is necessary against a background of gradual change before a

387   new auditory scene is assumed and the evidence accumulation process is restarted.

388   **III. EXPERIMENT 2**

389   Rogers and Bregman (1998) explored the effects on subsequent streaming of gradual and

390   abrupt changes in stimulus lateralization—based on interaural time difference (ITD) cues,

391   interaural level difference (ILD) cues, and loudspeaker position—or in overall stimulus level. They

392   used an induction sequence (4.8 s) followed by a short test sequence (1.2 s); both tone sequences

393   were configured in the form HLH–HLH–…. Listeners were asked to provide a one-off judgment

394   of the extent of stream segregation at the end of the stimulus using a rating scale (1-8, where 1

395   corresponded to fully segregated and 8 to fully integrated). On the first trial of each condition, the

396   listener heard sequences with a 9-ST separation. On the basis of the response to a given trial, Δf

397   for the next trial was raised or lowered by 1 ST (over the range 5-14 ST) to make the percept

398   increasingly ambiguous. Through an iterative process and averaging, this provided a measure of

399   the border between segregation and integration (threshold Δf) that could be compared across

400   conditions. There were two reference cases—the no-change condition in which the induction and

401   test sequences had identical properties (maximum build-up) and the control condition in which the

402   induction sequence was replaced by white noise (no build-up).

403   Relative to an induction sequence whose properties matched those of the test sequence, an

404   induction sequence whose properties changed gradually and finished close to those of the test

405   sequence had little effect (lateralization) or no effect (level) on subsequent stream segregation. In

406   contrast, an abrupt change in lateralization at the induction/test boundary caused a large shift in

407   threshold Δf, indicating that (depending on the specific cues manipulated) between half and all the

408   build-up accumulated during the induction sequence had been lost or reset. The direction of the

409   spatial change (leftwards or rightwards) did not affect the extent of this loss. However, the effect

410   of an abrupt 12-dB change in stimulus level (from 59 to 71 dB A or vice versa) was strongly

411   directional. The sudden-louder condition (rising level) caused a loss of about two thirds in the

412   accumulated build-up, but the sudden-softer condition (falling level) had no effect. This

413   asymmetry favors an account of the loss of build-up based on an active resetting process rather

414   than a failure to transfer from the induction sequence. The second experiment reported here

415   extends the investigation of the effects of abrupt transitions in level on streaming judgments, and

416   their directional properties, by introducing occasional or frequent changes during the test sequence.

417   For conditions involving occasional transitions, the time interval between them (5.2 s) was

418   sufficiently long to observe the initial response to a transition and the main phase of recovery

419   during the new steady state.

420

A. Method

Except where described, the same method was used as for experiment 1. Twelve listeners (2 males, mean age = 20.4 years, range = 18.9 – 21.9) took part and successfully completed the experiment; no listeners were excluded and replaced. In this experiment, the L tones were always set to a constant base frequency of 1 kHz and conditions differed only in the presentation levels used for the triplets. Two levels were used, allowing sequences to be constructed involving abrupt 12-dB transitions between triplets.

There were five conditions (C1-C5). In C1 and C2, the triplet level was fixed at 73 dB SPL (high) and 61 dB SPL (low), respectively. The inclusion of the constant conditions was mainly to provide reference cases against which the effect of abrupt transitions could be determined, but also provided a test of whether there was any effect of absolute level *per se* on streaming. In C3, there were abrupt changes in triplet level between high (starting value) and low every three triplets (i.e., rapid alternation every 1.2 s). In C4 and C5, the alternation of abrupt changes in level occurred more slowly—once every 13 triplets (i.e., after 5.2 s, 10.4 s, and 15.6 s)—and stimulus level began either low (C4, LHLH) or high (C5, HLHL). The 5.2-s interval was chosen to provide sufficient scope for build-up between transitions so that any resetting arising from a particular transition would be evident. For the alternating-level conditions, the final group of triplets was truncated by one (C3) or two triplets (C4 and C5) to ensure a common test-sequence duration of 20 s. Including the low- and high-level starting cases for the slow alternation conditions ensured that each transition direction (low-to-high and high-to-low) was represented equally often and at different times in the test sequence. All other properties of the test sequences remained the same across

442   conditions. Each combination of condition (5 levels) and Δf (3 levels) was presented ten times in

443   the main experiment, once in each block, giving 150 trials.

444   Listeners completed this experiment in a single session, which typically took ~1½ hours.

445   Time-series data were computed from listeners' responses in the same way as described for

446   experiment 1. On occasions when an individual mean was missing (24 cases, all occurring within

447   the first few time bins and corresponding to ~0.7% of the data), the missing value was replaced

448   using mean imputation. Once again, the results were analyzed using repeated-measures ANOVA.

449   **B. Results and discussion**

450   The results averaged across listeners are shown in Fig. 3. The upper panels display results

451   for the conditions in which the level of the test sequence was either constant or alternated rapidly

452   (every three triplets); the lower panels display results for the conditions in which there was an

453   abrupt fall or rise in stimulus level after every thirteen triplets (i.e., three transitions per sequence—

454   T1, T2, and T3) and also reproduce the results for the constant reference cases. These results are

455   considered in turn.

456   The constant-high and constant-low conditions (C1 and C2) showed similar extents of

457   segregation and patterns of build-up over time, but the pattern for the rapid-alternation condition

458   (C3) began to diverge from the others ~5-10 s after the start of the test sequence. This divergence

459   was manifest as a suppression of build-up that appeared greatest for the largest Δf tested (8 ST).

460   The ANOVA for the conditions in which stimulus level was either constant or alternated rapidly

461   is presented in Table III. Two of the three factors influenced streaming as main effects—

462   segregation was greater for larger frequency separations (means: 4 ST = 39.1%, 6 ST = 56.9%,

463   and 8 ST = 70.4%; p<0.001) and tended to increase over time (p<0.001)—but there was no main

464   effect of condition (p=0.196). Two of the two-way interactions were also significant—condition ×

22

465 time interval (p<0.001) and $\Delta f \times$ time interval (p<0.001). The former arose mainly because the loss

466 of segregation caused by multiple changes in level was largely confined to the latter half of the

467 sequence, perhaps at least partly because this was the time period over which there was more scope

468 for loss of segregation. The latter arose mainly because the tendency for stream segregation to

469 continue increasing during the second half of the sequence was greater for smaller frequency

470 separations.
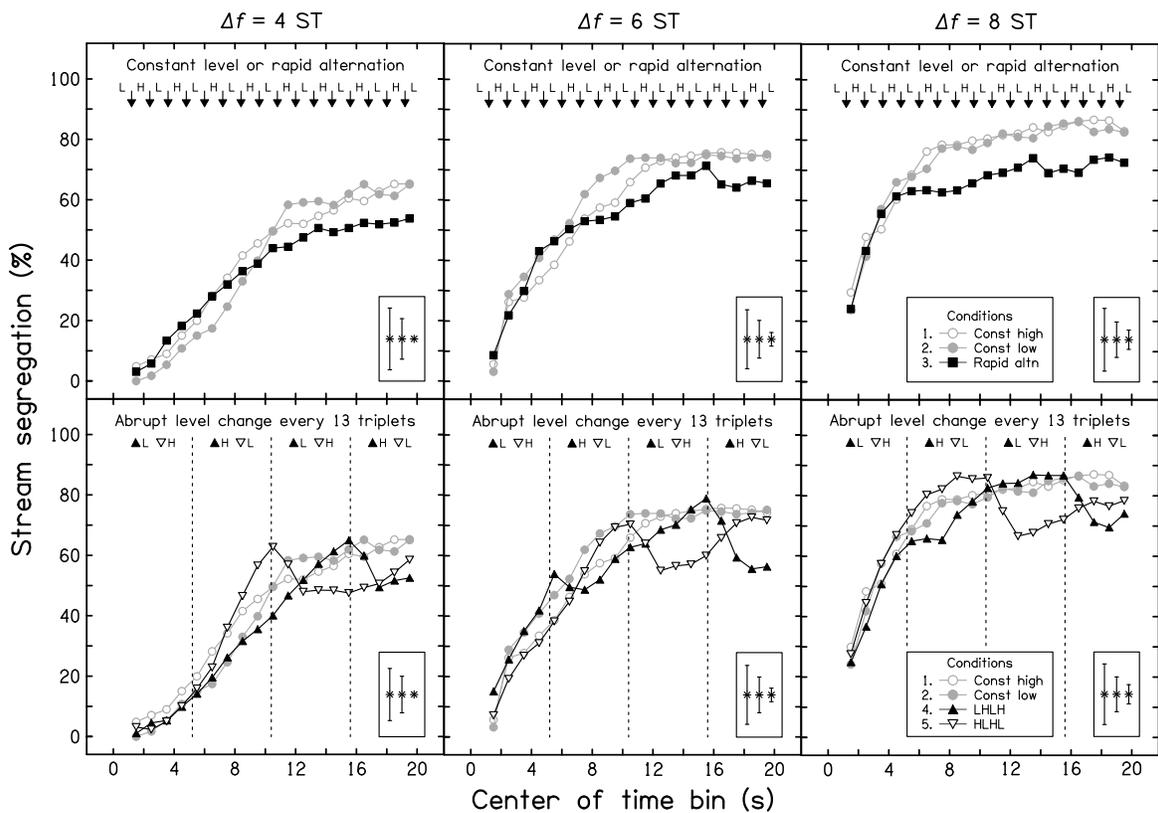
471

FIG. 3. Results for experiment 2—effects of level condition (1-5, see insets in right panels) and test-sequence frequency separation (Δf) on the extent and time course of reported stream segregation (n=12). Results for the conditions involving either constant level or rapid alternation in level (C1-C3) are shown in the upper panels. Results for the conditions involving abrupt changes in level every 13 triplets (C4-C5) are shown in the lower panels; also reproduced in gray are the results for the appropriate reference conditions. Otherwise as for Fig. 2.

472   Overall, these outcomes indicate that there was no effect of level *per se* on stream

473 segregation over the 12-dB range tested, but that the rapid alternation in level acted to reduce the

474   build-up of stream segregation. The absence of an effect of presentation level on the results for C1

475   and C2 is unsurprising. Although it has been shown that increasing presentation level can lead to

476   a fall in stream segregation for a given center frequency and Δf (Rose and Moore, 2000),

477   presumably owing to the broadening of auditory filter bandwidths (Glasberg and Moore, 1990),

478   any effect of this kind would have been modest over the range tested here. Also, that study

479   measured the fission boundary and so listeners were instructed to try to segregate one subset of

480   tones from a sequence, whereas our listeners were given neutral listening instructions. Given the

481   similar profiles for the constant high- and low-level cases, the lower segregation in the latter half

482   of the sequence associated with fast alternations implies that, when averaged, abrupt rises and falls

483   in level tend to suppress build-up. We now consider the directional effects of individual transitions.

484   Inspection of the lower panels of Fig. 3 suggests that sudden L-to-H transitions in level

485   tended to decrease subsequent stream segregation but that sudden H-to-L transitions had little or

486   no effect (with the possible exception of T1, discussed later). ANOVAs exploring the effects of

487   abrupt changes in level every 13 triplets (i.e., three transitions per test sequence) are presented in

488   Table IV. A 4.0-s time bin was used for the first two transitions, starting 1.2 s after each change

489   (cf. experiment 1), but this was reduced to 3.4 s for the third transition owing to the termination of

490   the sequence at 20 s. There was a main effect of Δf for all three transitions ($p<0.001$ in all cases),

491   reflecting the usual tendency for greater streaming with larger values of Δf. For condition, there

492   was a main effect for T2 ($p=0.002$) and T3 ($p=0.007$), respectively, but only a non-significant trend

493   for T1 ($p=0.074$) and so it was not considered further. The latter was probably a consequence of

494   the limited time available for build-up from scratch during the first 5 s of the test sequence. There

495   was no significant $\Delta f \times C$ interaction for any transition number.

496　　　　The effects of individual transitions were explored further using pairwise comparisons after

497　　the results were collapsed across Δf. For L-to-H and H-to-L transitions in level, respectively, the

498　　reference cases were the results for the constant-high (C1) and constant-low (C2) conditions during

499　　the corresponding time interval. Hence, the reference cases were matched for the stimulus

500　　properties of the test cases following the transition. Note that, between them, conditions C4

501　　(LHLH) and C5 (HLHL) provided data for one transition in each direction for each transition

502　　number. There was a significant loss of segregation associated with the L-to-H transitions (change

503　　for T2: -12.6% pts, p=0.018; change for T3: -14.8% pts, p=0.007), but for the H-to-L transitions

504　　there was no effect for T2 (change: -0.2% pts, p=0.948) and only a small loss for T3 (change: -

505　　5.7% pts, p=0.027). Taken together, these outcomes indicate an asymmetry similar to that reported

506　　by Rogers and Bregman (1998). Note that the tendency for suppression of build-up in sequences

507　　with fast alternations (C3) can be explained by the resetting effects of multiple L-to-H transitions.

508　　Rogers and Bregman (1998) interpreted the asymmetry in terms of Bregman's (1978) functional

509　　account of build-up, arguing that it reflected the greater importance of sudden increases in level,

510　　because such increases usually indicate the onset of new sound sources.

511　　　　　Finally, it merits note that one aspect of the current results motivated the development of

512　　the final experiment reported here. Following the H-to-L transition in level at T1 when $\Delta f = 4$ ST,

513　　there is a suggestion in the data that segregation *increased* (i.e., overshoot rather than resetting of

514　　segregation); indeed, if considered in isolation, this change would be significant (change: +13.3%

515　　pts, p=0.009). Further evidence that overshoot can occur is provided by experiment 3.

516　　**IV. EXPERIMENT 3**

517　　　　　Another context in which marked perceptual asymmetries have been observed is auditory

518　　search. Asemi et al. (2003) used a task in which a target and distractors were presented

519   simultaneously over loudspeakers at different positions in the frontal-horizontal plane. They found

520   that the reaction time for detecting a narrowband noise, amplitude-modulated tone, or frequency-

521   modulated tone among pure-tone distractors was largely unaffected by the number of distractors—

522   indicating "pop out"—but that the time needed to detect a pure tone among temporally fluctuating

523   distractors increased with the number of distractors. The same asymmetry was observed by Cusack

524   and Carlyon (2003) for pure and frequency-modulated tones presented sequentially. These results

525   show that the auditory system uses temporal changes in the amplitude and frequency of sound as

526   a basic feature for the detection of a sound in an acoustic scene. In other words, sounds possessing

527   this basic feature will be more salient and attention-grabbing.

528   Previous studies of stream segregation have used modulated sounds, both narrowband (e.g.,

529   Cusack and Roberts, 1999) and wideband (e.g., Grimault et al., 2002), but to our knowledge only

530   in the context of investigating the effects of introducing differences between the two subsets of

531   sounds comprising the sequence. Given the asymmetry found in auditory search tasks, and our

532   own pilot observations, we considered that introducing sudden transitions between unmodulated

533   and modulated sounds offered a potentially fruitful candidate for observing directional effects of

534   these transitions. Abrupt changes in the center frequency or overall level of a tone sequence

535   inevitably introduce differences in long-term excitation pattern between corresponding subsets of

536   tones. Given the potential impact of peripheral-channeling cues on stream segregation (Hartmann

537   and Johnson, 1991; Roberts et al., 2002; Moore and Gockel, 2012), we wished to minimize them

538   by transitioning between pure tones (unmodulated) and narrowly spaced two-tone complexes

539   (modulated, cf. Cusack and Roberts, 1999). These two types of sound differ markedly in timbre.

540   The sequence configurations used and the timing of the transitions corresponded closely with their

541   counterparts in experiment 2.

A. Method

Except where described, the same method was used as for experiments 1 and 2. Twelve listeners (2 males, mean age = 20.9 years, range = 19.4 – 25.3) took part and successfully completed the experiment; three listeners were excluded and replaced. In this experiment, the L tones were always set to a constant base frequency of 1 kHz and conditions differed only in the nature of the tones used to construct the triplets. Two types of tone were used—pure tones (T) and narrowly spaced pairs of pure tones known as dyads (D)—which allowed sequences to be constructed involving abrupt transitions in timbre between groups of triplets (from unmodulated to modulated or vice versa) without introducing excitation-pattern cues. Our informal observations with sequences of this kind suggested a marked asymmetry in the effect of transition direction on subsequent judgments of stream segregation.

Tone dyads were constructed by adding two pure tones of equal level and centered (±25 Hz) on the frequency of their pure-tone counterparts. Each constituent tone was attenuated by 3 dB relative to its pure-tone counterpart, such that the root-mean-square (RMS) power of each pure tone and corresponding dyad was the same. One constituent tone began in sine phase and the other in negative sine phase and their addition with 50-Hz separation gave exactly 5 cycles of full-depth amplitude modulation over 100 ms. Given that the center frequency of the L-tone dyads was 1000 Hz (H tones = 4, 6, or 8 ST above), the two components were always unresolved (equivalent rectangular bandwidth of the auditory filter at 1000 Hz ≈ 132 Hz; Glasberg and Moore, 1990) and the average excitation pattern of each dyad and its pure-tone counterpart was almost identical. Note that a pair of unresolved components also creates a frequency modulation at the output of the cochlear filters, which depends on the relative amplitude of each component (Hartmann, 1998). To our knowledge, the effect of a correlated transition in timbre for both subsets of sounds without

565 peripheral-channeling cues has not previously been investigated. The strong modulation of the

566 dyads resulting from the interaction of the two components within the same auditory filter gave

567 them a distinctive timbre; indeed, the quality of a sequence of triplets composed of dyads was

568 reminiscent of the sound produced by stridulating crickets.

569   There were five conditions (C1-C5). In C1 and C2, the triplets were always composed of

570 pure tones or dyads, respectively. In C3, there were abrupt changes in triplet timbre between pure

571 tones (starting value) and dyads every three triplets (i.e., rapid alternation every 1.2 s). If the

572 dominant effect of the timbre transitions was to cause resetting, then this should be manifest as an

573 overall suppression of the build-up of stream segregation (cf. experiment 2). In C4 and C5, the

574 alternation of abrupt changes in modulation occurred more slowly—once every 13 triplets (i.e., at

575 5.2 s, 10.4 s, and 15.6 s)—and the sequence began with either pure tones (C4, TDTD) or dyads

576 (C5, DTDT). For the alternating-timbre conditions, the final group of triplets was truncated by one

577 (C3) or two triplets (C4 and C5) to ensure a common test-sequence duration of 20 s. Including the

578 pure-tone and dyad-starting cases for the slow alternation conditions ensured that each transition

579 direction (pure-to-dyad and dyad-to-pure) was represented equally often and at different times in

580 the test sequence. All other properties of the test sequences remained the same across conditions.

581 Pure tones and dyads were presented at 73 dB SPL.

582   Each combination of condition (5 levels) and Δf (3 levels) was presented ten times in the

583 main experiment, once in each block, giving 150 trials. Listeners completed this experiment in a

584 single session, which typically took ~1½ hours. Time-series data were computed from listeners'

585 responses in the same way as described for experiment 1. On occasions when an individual mean

586 was missing (44 cases, corresponding to ~1.3% of the data and all occurring within the first few

587 time bins), mean imputation was used to replace the missing value.

588

589

590 **B. Results and discussion**

591   The results averaged across listeners are shown in Fig. 4. The upper panels show results

592   for the conditions in which the timbre of the sounds comprising the test sequence was either

593   constant (unmodulated or modulated) or alternated rapidly (every three triplets); the lower panels

594   show results for the conditions in which there was an abrupt change from a pure tone to a dyad or

595   vice versa every thirteen triplets (transitions T1, T2, and T3) and also reproduce the results for the

596   reference cases. These results are considered in turn.
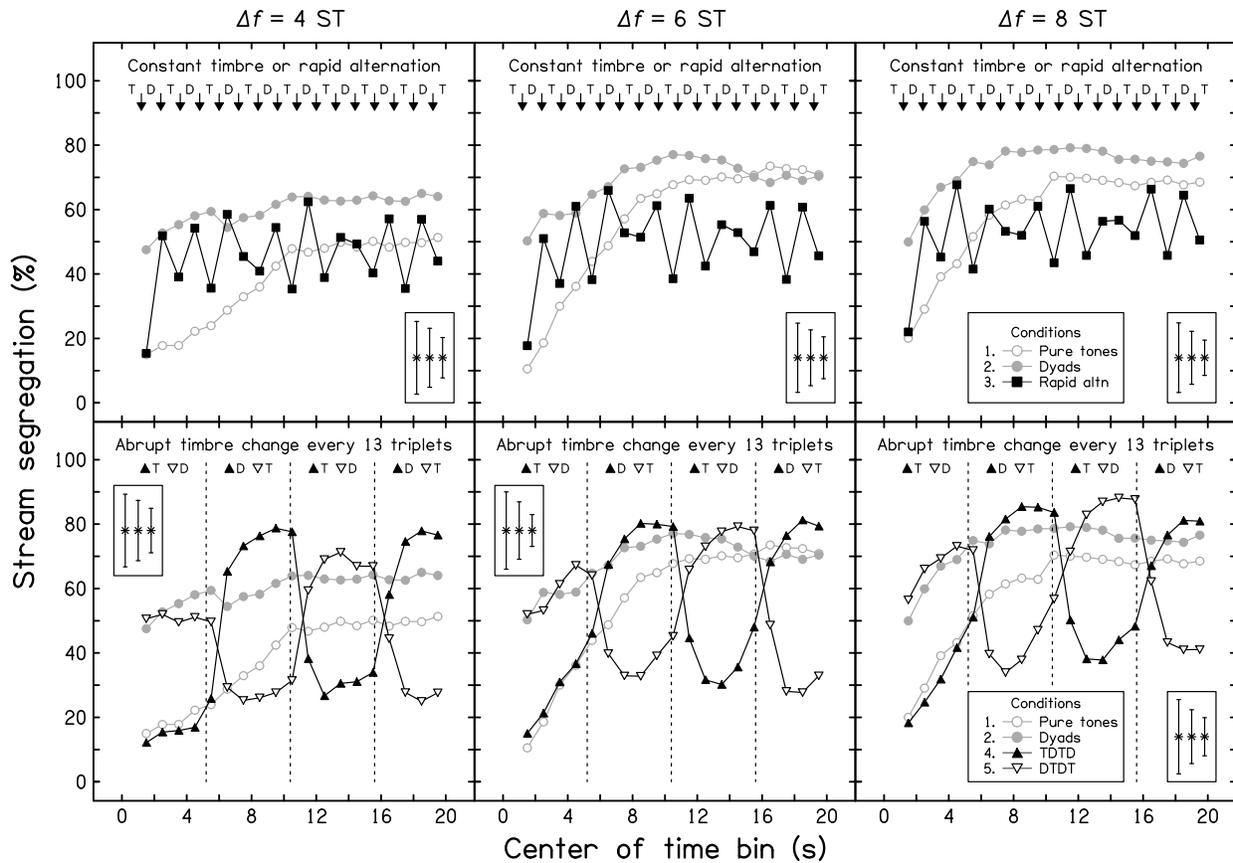
FIG. 4. Results for experiment 3—effects of modulation condition (1-5, see insets in right panels) and test-sequence frequency separation (Δf) on the extent and time course of reported stream segregation (n=12). Results for the conditions involving either constant timbre or rapid alternation in timbre (C1-C3) are shown in the upper panels. Results for the conditions involving abrupt changes in timbre every 13 triplets (C4-C5) are shown in the lower panels; also reproduced in gray are the results for the appropriate reference conditions. Otherwise as for Fig. 2.

The pure-tone-only and dyad-only conditions (C1 and C2) differed in that the extent of stream segregation was greater for the latter at the start of the sequence, but the difference between them tended to decline over time. The somewhat ragged profile seen for the rapid-alternation condition (C3) is a consequence of aliasing; a regular sawtooth pattern is observed if the results are plotted using 1.2-s time bins (corresponding exactly to 3 triplets). This pattern represents an oscillation between lesser and greater tendencies to give two-stream judgments in response to the pure-tone and dyad components of the sequence, respectively. However, superimposed on this

30

605   pattern is an overall loss of segregation that emerged ~5-10 s after the start of the sequence (cf. the

606   corresponding condition in experiment 2). This divergence was manifest as a suppression of build-

607   up that appeared to be greater for the larger values of Δf tested (6 ST and 8 ST).

608   The ANOVA for the conditions in which the stimulus modulation for the test sequence was

609   either constant or alternated rapidly is presented in Table V. Two of the three factors influenced

610   streaming as main effects—segregation was greater for larger frequency separations (means: 4 ST

611   = 48.0%, 6 ST = 58.4%, and 8 ST = 61.8%; p<0.001) and tended to increase over time (p<0.001).

612   The main effect of condition was not significant overall (p=0.117), but became so if the time bins

613   included in the analysis were restricted to the fast phase of build-up (1-2 s to 10-11 s; p=0.043),

614   reflecting the considerably greater segregation for dyad than pure-tone sequences during this

615   phase. All of the two-way interactions were also significant—condition × time interval (p<0.001),

616   Δf × condition (p=0.020) and Δf × time interval (p=0.001). Once again, the C × T interaction arose

617   mainly because the loss of segregation caused by multiple changes was largely confined to the

618   latter half of the sequence, for which there was more scope for loss of segregation, and the Δf × T

619   interaction arose mainly because the tendency for stream segregation to continue increasing during

620   the second half of the sequence was greater for smaller frequency separations. The Δf × C

621   interaction probably reflects the smaller suppression of segregation observed for the rapid-

622   alternation condition for the smallest Δf tested.

623   One possible explanation for the greater stream segregation observed here for dyad

624   sequences than for pure-tone sequences with the same Δf is suggested by the results of Cusack and

625   Roberts (1999). They used repeating LHL–LHL–… sequences of two-tone complexes in which

626   the L stimuli (center frequency = 1000 Hz) had a fixed component separation of 100 Hz in all

627   conditions whereas the H stimuli (center frequency = 1200 Hz) had a component separation

31

628 corresponding to one of seven values (80 Hz to 140 Hz, in 10-Hz steps) across conditions. Least

629 segregation was reported when the H stimuli had the same relative bandwidth as the L stimuli

630 (match = 120 Hz) rather than the same modulation rate (100 Hz). Notwithstanding the use of a

631 smaller fixed component separation of 50 Hz in the current experiment, it seems likely that using

632 the same component separation for the H and L dyads introduced an additional factor supporting

633 the build-up of stream segregation to that provided by Δf.

634   Inspection of the lower panels of Fig. 4 shows that slow alternations of the sudden changes

635 in timbre caused dramatic changes in perception between integrated and segregated. The overall

636 pattern suggests that sudden D-to-T transitions (i.e., modulated to unmodulated) decreased

637 subsequent streaming but that sudden T-to-D (i.e., unmodulated to modulated) transitions tended

638 to have the opposite effect. ANOVAs exploring the effects of abrupt changes in level every 13

639 triplets are presented in Table VI; the same time bins were used for transitions T1-T3 as for their

640 counterparts in experiment 2. For all three transitions, there was a main effect of Δf (range:

641 p=0.017–p<0.001), reflecting the usual tendency for streaming to increase with Δf, and of

642 condition (range: p=0.007–p=0.001), reflecting the evident differences between conditions during

643 the observation interval. Since there was a significant $\Delta f \times C$ interaction for two of the three

644 transitions, pairwise comparisons were made separately for each Δf. For D-to-T and T-to-D

645 transitions, respectively, the reference cases were the results for the pure-tone-only (C1) and dyad-

646 only (C2) conditions during the corresponding time interval, again matching the reference cases

647 to the stimulus properties of the test cases following the transition. Between them, conditions C4

648 (TDTD) and C5 (DTDT) provided data for one transition in each direction at each Δf for T1-T3.

649   For each direction of change, the results of these pairwise comparisons are summarized in

650 Table VII for all nine combinations of transition number and Δf. For the D-to-T transitions (i.e.,

651 modulated to unmodulated), all nine combinations were associated with a fall in segregation

652 (overall mean difference = -25.6% pts); seven cases were significant and the losses were often

653 substantial. For the T-to-D transitions (i.e., unmodulated to modulated), all nine combinations were

654 associated with an *increase* in subsequent segregation, but the overall mean difference was

655 considerably smaller (+7.5% pts); only two cases showed significant overshoot (T1 for Δf = 4 ST,

656 +16.8% pts, p=0.005; T2 for Δf = 8 ST, +8.1% pts, p=0.004) and a third approached significance

657 (T3 for Δf = 4 ST, +12.1% pts, p=0.057). Note, however, that using the dyad-only segregation

658 score during the corresponding interval represents an exceptionally conservative reference case

659 for estimating overshoot following T2 and T3. This is because the previous D-to-T transition will

660 have reset almost all prior build-up.

661   For T2 and T3, using the mean segregation score averaged over the fast phase of build-up

662 (time bins 1-2 s to 10-11 s) is arguably a more reasonable reference case. For example, if this

663 reference were used instead, the marginal case noted above would become significant (T3 for Δf

664 = 4 ST, +19.6% pts, p=0.013). It is also the case that the "headroom" available to demonstrate

665 overshoot following T-to-D transitions was quite limited for larger Δfs because of the high

666 segregation scores associated with dyad sequences. These issues suggest that further research with

667 stimuli of this kind would benefit from two changes in the experimental design. First, to include

668 only one transition per trial (as was done here in experiment 1), to avoid the difficulties of choosing

669 an appropriate reference case for subsequent transitions and also to allow more time to observe

670 streaming before and after. Second, to include smaller Δfs than were tested here to allow greater

671 headroom for overshoot effects to be manifest. Nonetheless, it seems reasonable overall to

672 conclude that D-to-T transitions do not lead to resetting but instead tend to increase subsequent

673   streaming. The implications of these results for accounts of build-up and the kinds of mechanism

674   that might explain the observed asymmetry are considered below.

675   Finally, it should be acknowledged that a contribution to the results from audible distortion

676   products generated by the dyads cannot be ruled out entirely, given the relatively high presentation

677   level and the absence of background noise. The most prominent combination tone generated from

678   a pair of primaries (f_1 and f_2) is usually the cubic difference tone ($2f_1$-f_2), particularly for f_2/f_1 ratios

679   ≤1.10 (Goldstein, 1967). For the tone pair constituting the L dyads, f_1 = 975 Hz and f_2 = 1025 Hz,

680   giving a ratio of ~1.05 and generating a cubic difference tone at 925 Hz. Although it would have

681   been lower in level than the primaries, this distortion product may have been sufficient to increase

682   the level of excitation on the lower skirt of the excitation pattern evoked by the dyads, leading to

683   greater than anticipated differences in peripheral channeling between corresponding pure tones

684   and dyads. Even if this were the case, however, it is not clear how this could account for the strong

685   directional effects observed for sudden transitions between pure tones and dyads. Rather, we argue

686   that the critical factor is most probably the sudden changes in modulation.

687   **V. CONCLUDING DISCUSSION**

688   For tone sequences involving one or more correlated transitions in acoustic properties—

689   i.e., where the high- and low-frequency subsets change together to the same extent on the same

690   dimension—the effect of a sudden change can be influenced not only by the property being altered

691   but also by the direction of that change. Experiment 1 explored the effects of sudden changes in

692   triplet base frequency and found that part of the build-up of stream segregation prior to a transition

693   can transfer over a wider frequency region (more than half an octave), and more equally for sudden

694   rises and falls, than had been suggested by the results for the particular set of values tested by

695   Anstis and Saida (1985). Relative to maintaining a constant base frequency, the progressive

696   accumulation of gradual changes in base frequency made little or no difference to the build-up of

697   stream segregation for a tone sequence. This outcome casts doubt on models of auditory streaming

698   in which build-up depends on extended stimulation of populations of central auditory neurons with

699   the same best frequency (e.g., Micheyl et al., 2005; Pressnitzer et al., 2008; Bee et al., 2010).

700   Rather, this outcome suggests a mechanism in which accumulated build-up in the tendency for

701   stream segregation (adaptation) can be transferred between neurons with different best

702   frequencies, so long as there are no abrupt changes in base frequency as the tone sequence unfolds.

703   Experiments 2 and 3 explored the effects of sudden changes in level and modulation,

704   respectively. Sudden transitions in level (±12 dB) produced smaller changes in segregation than

705   those associated with sudden transitions in base frequency and, in accord with the findings of

706   Rogers and Bregman (1998), there was a clear asymmetry in the effect of transition direction.

707   Rising transitions (softer-to-louder) caused significant loss of build-up (resetting) but falling

708   transitions (louder-to-softer) had little or no effect. The effects of sudden changes in tone

709   modulation on stream segregation were larger, with the losses for D-to-T transitions (i.e., from

710   modulated to unmodulated) approaching the size of those for changes in base frequency, and the

711   effect of direction was even more marked. Specifically, T-to-D transitions (i.e., from unmodulated

712   to modulated) in some cases led to even greater segregation than that for dyad-only sequences

713   during the corresponding time interval (i.e., overshoot).

714   Rogers and Bregman (1998) interpreted the asymmetry they observed for sudden level

715   changes in terms of Bregman's (1978) functional account of build-up, arguing that a sudden rise

716   in level causes a loss of build-up but a sudden fall does not because only the former can signal the

717   activation of a new sound source. However, it is hard to see how this argument might be extended

718   to account for the directional effects of changes in tone modulation, particularly given the evidence

719   that sudden changes from unmodulated to modulated tones (T-to-D transitions) sometimes result

720   in greater stream segregation. Although speculative, three plausible accounts merit discussion.

721   First, the overshoot sometimes observed after T-to-D transitions may be a short-term contrast

722   effect arising from the greater tendency for dyad-only sequences to be heard as segregated relative

723   to pure-tone sequences. Longer-term contrast effects, occurring across trials, have previously been

724   reported for AF tone sequences following changes in Δf (Snyder et al., 2008, 2009). Second, it

725   may be possible to extend attention-switching accounts of the loss of build-up after an abrupt

726   transition in stimulus properties (Rajasingam et al., 2018; see also Cusack et al., 2004; Thompson

727   et al., 2011) to explain the overshoot that can occur after a T-to-D transition, on the basis that the

728   switch in attention is to modulated sounds, which have primary attention-grabbing properties (cf.

729   Asemi et al., 2003; Cusack and Carlyon, 2003). Third, the occurrence of overshoot may be

730   indicative of the operation of some inhibitory or suppressive process whose accumulation and

731   release affects the extent of stream segregation.

732   The phenomenon of subtractive adaptation has long been known in the visual system (e.g.,

733   Geisler, 1983; Hayhoe et al., 1992) and mechanisms of this kind have since been proposed to

734   account for the multi-second build-up of stream segregation for a repeating but unchanging

735   sequence of tones (e.g., Micheyl et al., 2005; Pressnitzer et al., 2008; Bee et al., 2010). The basis

736   of these accounts is that the intensity of the response of tonotopically tuned neurons in the central

737   auditory system to a repeating tone sequence gradually declines through the slow accumulation of

738   inhibition or suppression, leading to a progressive narrowing of their receptive fields. As a result,

739   the receptive fields of neurons best tuned to the H and L subsets initially overlap but over time two

740   distinct subpopulations emerge, leading to the perception of separate streams. A sudden change of

741   sufficient magnitude in the base frequency of the tone sequence resets this process because a

36

742 different population of neurons is activated. In this regard, note that the changes in stimulus level

743 or modulation used in experiments 2 and 3 did not involve changes in base frequency and so the

744 transitions were not anticipated to change the tuned populations of neurons responding to these

745 sequences. Presumably, rising-level transitions and transitions from modulated to unmodulated

746 tones (D-to-T) led quickly to partial or complete release, respectively, of accumulated inhibition

747 resulting in a loss of build-up. By this account, neither falling-level transitions nor transitions from

748 unmodulated to modulated tones (T-to-D) led to release of inhibition. Given that a 12-dB

749 difference in level *per se* has little or no effect on the build-up of stream segregation, a falling-

750 level transition therefore produces little or no effect on subsequent streaming. Presumably,

751 overshoot sometimes arises for T-to-D transitions because the tendency to hear two streams is

752 greater for dyad sequences than for pure-tone sequences, leading to increased inhibition rather than

753 a release from it following the transition. Future research might investigate the responses of

754 auditory cortical neurons to tone sequences involving sudden changes in level or modulation.

755   In conclusion, the experiments reported here have extended our knowledge of the dynamics

756 of auditory stream segregation. Most notably, we have demonstrated that the effects of sudden

757 correlated transitions in stimulus modulation are strongly directional, including instances in which

758 a sudden change from unmodulated (pure tones) to modulated (dyads) sounds leads to greater

759 segregation (overshoot) rather than a loss of build-up. It is not obvious how Bregman's (1978)

760 functional account of build-up might be adapted to explain these findings, but there are plausible

761 accounts based on stimulus contrast effects, attention switching, or neural mechanisms involving

762 the accumulation and release of inhibition or suppression. These accounts might be investigated

763 by extending the perceptual experiments reported here to include transitions between modulated

764 sounds with different modulation rates and by exploring the effects of transitions of this kind on

765   the responses of auditory cortical neurons. This approach should help elucidate further how stream

766   segregation functions in changing auditory scenes.

776   **FOOTNOTES**

777   (1) Rather than changing direction, the base frequency of the final triplet in C3 and C4 continued

778   on its established trajectory and consequently fell 0.5 ST below the nominal minimum (486 Hz)

779   and above the nominal maximum (1029 Hz), respectively. Note that the final triplet occurred too

780   late to have any appreciable effect on listeners' responses.

781   (2) In experiment 1, the start time for the 4.0-s time interval used to explore the effect of abrupt

782   transitions on streaming was 11.2 s (i.e., 1.2 s after the transition at 10.0 s), which did not align

783   exactly with the set of 1-s time bins plotted in Fig. 2. Similar issues of alignment arose in relation

784   to the transition times used in experiments 2 and 3 (5.2 s, 10.4 s, and 15.6 s). To create longer time

785   intervals for statistical analysis with start times that did not correspond to an integer number of

786 seconds, the response data for each trial were also divided into finer-grained (0.2 s) time bins.

787 Longer time intervals were constructed by combining the appropriate set of 0.2-s time bins.

788 (3) See https://doi.org/10.17036/researchdata.aston.ac.uk.00000496 (Last viewed April 27, 2021).

789

790   **REFERENCES**

791   Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones,"

792   J. Exp. Psychol. Hum. Percept. Perform. **11,** 257-271.

793   Asemi, N., Sugita, Y., and Suzuki, Y. (**2003**). "Auditory search asymmetry between pure tone and

794   temporal fluctuating sounds distributed on the frontal-horizontal plane," Acta Acust. Acust. **89,**

795   346-354.

796   Beauvois, M.W., and Meddis, R. (**1997**). "Time decay of auditory stream biasing," Percept.

797   Psychophys. **59,** 81-86.

798   Bee, M.A., Micheyl, C., Oxenham, A.J., and Klump, G.M. (**2010**). "Neural adaptation to tone

799   sequences in the songbird forebrain: Patterns, determinants, and relation to the build-up of auditory

800   streaming," J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol. **196,** 543-557.

801   Bregman, A.S. (**1978**). "Auditory streaming is cumulative," J. Exp. Psychol. Hum. Percept.

802   Perform. **4,** 380-387.

803   Bregman, A.S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT

804   Press, Cambridge, MA), pp. 1-790.

805   Bregman, A.S., and Campbell, J. (**1971**). "Primary auditory stream segregation and perception of

806   order in rapid sequences of tones," J. Exp. Psychol. **89,** 244-249.

807   Carlyon, R.P, Cusack, R., Foxton, J.M., and Robertson, I.H. (**2001**). "Effects of attention and

808   unilateral neglect on auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **27,**

809   115-127.

810   Cusack, R., and Carlyon, R.P. (**2003**). "Perceptual asymmetries in audition," J. Exp. Psychol. Hum.

811   Percept. Perform. **29,** 713-725.

812   Cusack, R., Deeks, J., Aikman, G., and Carlyon, R.P. (**2004**). "Effects of location, frequency

813   region and time course of selective attention on auditory scene analysis," J. Exp. Psychol. Hum.

814   Percept. Perform. **30,** 643-656.

815   Cusack, R., and Roberts, B. (**1999**). "Effects of similarity in bandwidth on the auditory sequential

816   streaming of two-tone complexes," Percept. **28,** 1281-1289.

Denham, S.L., and Winkler, I. (**2006**). "The role of predictive models in the formation of auditory streams," J. Physiol. Paris **100,** 154-170.

Geisler, W.S. (**1983**). "Mechanism of visual sensitivity: Backgrounds and early dark adaptation," Vision Res. **23,** 1423-1432.

Glasberg, B.R., and Moore, B.C.J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47,** 103-138.

Goldstein, J.L. (**1967**). "Auditory nonlinearity," J. Acoust. Soc. Am. **41,** 676-689.

Grimault, N., Bacon, S.P., and Micheyl, C. (**2002**). "Auditory stream segregation on the basis of amplitude-modulation rate," J. Acoust. Soc. Am. **111,** 1340-1348.

Hartmann, W.M. (**1998**). *Signals, Sound, and Sensation* (Springer, New York), Chapter 17, pp. 393-398.

Hartmann, W.M., and Johnson, D. (**1991**). "Stream segregation and peripheral channeling," Music Percept. **9,** 155-183.

Hayhoe, M.M., Levin, M.E., and Koshel, R.J. (**1992**). "Subtractive processes in light adaptation," Vision Res. **32,** 323-333.

Haywood, N.R., and Roberts, B. (**2010**). "Build-up of the tendency to segregate auditory streams: Resetting effects of single deviant tones," J. Acoust. Soc. Am. **128,** 3019-3031.

Haywood, N.R., and Roberts, B. (**2013**). "Build-up of auditory stream segregation induced by tone sequences of constant or alternating frequency and the resetting effects of single deviants," J. Exp. Psychol. Hum. Percept. Perform. **39,** 1652-1666.

Henke, W.L. (**2005**). "MITSYN: A coherent family of high-level languages for time signal processing [software package]," (WLH, Belmont, MA).

Keppel, G., and Wickens, T.D. (**2004**). *Design and Analysis: A Researcher's Handbook, 4th ed.* (Pearson Prentice Hall, Englewood Cliffs, NJ), pp. 1-611.

Lawrence, M.A. (**2016**). "ez: Easy analysis and visualization of factorial experiments (R package version 4.4-0) [software]," available at https://cran.r-project.org/package=ez (Last viewed December 6, 2019).

844   Micheyl, C., Tian, B., Carlyon, R.P., and Rauschecker, J.P. (**2005**). "Perceptual organization of

845   tone sequences in the auditory cortex of awake macaques," Neuron **48,** 139-148.

846   Miller, G.A., and Heise, G.A. (**1950**). "The trill threshold," J. Acoust. Soc. Am. **22,** 637-638.

847   Moore, B.C.J., and Gockel, H. (**2012**). "Properties of auditory stream formation," Phil. Trans. R.

848   Soc. B **367,** 919-931.

849   Pressnitzer, D., and Hupé, J.M. (**2006**). "Temporal dynamics of auditory and visual bistability

850   reveal common principles of perceptual organization," Current Biol. **16,** 1351-1357.

851   Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I.M. (**2008**). "Perceptual organization of

852   sound begins in the auditory periphery," Current Biol. **18,** 1124-1128.

853   R Core Team (**2020**). "R: A language and environment for statistical computing," R Foundation

854   for Statistical Computing, Vienna, Austria, available at http://www.r-project.org/ (Last viewed

855   March 19, 2021).

856   Rajasingam, S.L. (**2016**). "The dynamics of auditory stream segregation for tone sequences with

857   gradually and abruptly varying stimulus properties," Doctoral thesis, Aston University,

858   Birmingham, UK.

859   Rajasingam, S.L., Summers, R.J., and Roberts, B. (**2018**). "Stream biasing by different induction

860   sequences: Evaluating stream capture as an account of the segregation-promoting effects of

861   constant-frequency inducers," J. Acoust. Soc. Am. **144,** 3409-3420.

862   Roberts, B., Glasberg, B.R., and Moore, B.C.J. (**2002**). "Primitive stream segregation of tone

863   sequences without differences in fundamental frequency or passband," J. Acoust. Soc. Am. **112,**

864   2074-2085.

865   Roberts, B., Glasberg, B.R., and Moore, B.C.J. (**2008**). "Effects of the build-up and resetting of

866   auditory stream segregation on temporal discrimination," J. Exp. Psychol. Hum. Percept. Perform.

867   **34,** 992-1006.

868   Rogers, W.L., and Bregman, A.S. (**1993**). "An experimental evaluation of three theories of

869   auditory stream segregation," Percept. Psychophys. **53,** 179-189.

870   Rogers, W.L., and Bregman, A.S. (**1998**). "Cumulation of the tendency to segregate auditory

871   streams: Resetting by changes in location and loudness," Percept. Psychophys. **60,** 1216-1227.

872 Rose, M.M., and Moore, B.C.J. (**2000**). "Effects of frequency and level on auditory stream

873 segregation," J. Acoust. Soc. Am. **108,** 1209-1214.

874 Scharf, B. (**1970**). "Critical Bands," in Foundations of Modern Auditory Theory, Vol. I, edited by

875 J.V. Tobias. (Academic Press, New York), pp. 159-202.

876 Snedecor, G.W., and Cochran, W.G. (**1967**). *Statistical Methods*, 6th ed. (Iowa U.P., Ames, Iowa),

877 pp. 1-310.

878 Snyder, J.S., Carter, O.L., Hannon, E.E., and Alain, C. (**2009**). "Adaptation reveals multiple levels

879 of representation in auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **35,**

880 1232-1244.

881 Snyder, J.S., Carter, O.L., Lee, S.K., Hannon, E.E., and Alain, C. (**2008**). "Effects of context on

882 auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **34,** 1007-1016.

883 Thompson, S.K., Carlyon, R.P., and Cusack, R. (**2011**). "An objective measurement of the build-

884 up of auditory streaming and of its modulation by attention," J. Exp. Psychol. Hum. Percept.

885 Perform. **37,** 1253-1262.

886 van Noorden, L.P.A.S. (**1975**). "Temporal coherence in the perception of tone sequences,"

887 Doctoral thesis, Eindhoven University of Technology, the Netherlands.

888

889 **TABLE I**

890 Results for experiment 1—effects of stimulus base frequency (*constant or gradual-change*

891 *conditions*). Summary of the three-way repeated-measures ANOVA for all time bins (1-2 s to 19-

892 20 s) across conditions C1-C4. All significant terms are shown in bold.

Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**39.693**	**<0.001**	**0.783**
Base-frequency condition (C)	(3, 33)	1.062	0.378	0.088
Time interval (T)	**(18, 198)**	**22.927**	**<0.001**	**0.676**
$\Delta f \times$ C	(6, 66)	1.505	0.190	0.120
$\Delta f \times$ T	(36, 396)	1.118	0.299	0.092
C × T	(54, 594)	0.595	0.990	0.051
$\Delta f \times$ C × T	(108, 1188)	1.195	0.093	0.098

893

894 **TABLE II**

895 Results for experiment 1—effects of an *abrupt change* in stimulus base frequency. Summary of

896 the three-way repeated-measures ANOVA (single 4.0-s time interval, beginning 1.2 s after the

897 transition) across conditions C3-C10. All significant terms are shown in bold.

Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**33.909**	**<0.001**	**0.755**
Direction of change (D)	(1, 11)	0.059	0.812	0.005
Size of change (S)	**(3, 33)**	**28.790**	**<0.001**	**0.724**
$\Delta f \times$ D	(2, 22)	0.379	0.689	0.033
$\Delta f \times$ S	**(6, 66)**	**2.278**	**0.046**	**0.172**
D × S	(3, 33)	0.402	0.752	0.035
$\Delta f \times$ D × S	(6, 66)	0.574	0.749	0.050

898

44

899 **TABLE III**

900 Results for experiment 2—effects of stimulus level (*constant or rapid-alternation conditions*).

901 Summary of the three-way repeated-measures ANOVA for all time bins (1-2 s to 19-20 s) across

902 conditions C1-C3. All significant terms are shown in bold.

Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**30.001**	**<0.001**	**0.732**
Level condition (C)	(2, 22)	1.755	0.196	0.138
Time interval (T)	**(18, 198)**	**65.515**	**<0.001**	**0.856**
$\Delta f \times$ C	(4, 44)	1.242	0.307	0.101
$\Delta f \times$ T	**(36, 396)**	**5.057**	**<0.001**	**0.315**
C $\times$ T	**(36, 396)**	**1.985**	**<0.001**	**0.153**
$\Delta f \times$ C $\times$ T	(72, 792)	1.064	0.342	0.088

903

904 **TABLE IV**

905 Results for experiment 2—effects of *abrupt changes* in stimulus level. Summary of the two-way

906 repeated-measures ANOVA for each transition (single time intervals) across conditions C1, C2,

907 C4, and C5. All significant terms are shown in bold.

Part (a): Results for first transition (T1)				
Factor	*df*	*F*	*p*	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**48.230**	**<0.001**	**0.814**
Level condition (C)	(3, 33)	2.532	0.074	0.187
$\Delta f \times$ C	(6, 66)	1.631	0.153	0.129
Part (b): Results for second transition (T2)				
Factor	*df*	*F*	*p*	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**13.247**	**<0.001**	**0.546**
Level condition (C)	**(3, 33)**	**6.303**	**0.002**	**0.364**
$\Delta f \times$ C	(6, 66)	0.593	0.735	0.051
Part (c): Results for third transition (T3)				
Factor	*df*	*F*	*p*	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**10.916**	**<0.001**	**0.498**
Level condition (C)	**(3, 33)**	**4.838**	**0.007**	**0.305**
$\Delta f \times$ C	(6, 66)	0.615	0.718	0.053

908

909 **TABLE V**

910 Results for experiment 3—effects of stimulus modulation (timbre, *constant or rapid-alternation*

911 *conditions*). Summary of the three-way repeated-measures ANOVA for all time bins (1-2 s to 19-

912 20 s) across conditions C1-C3. All significant terms are shown in bold.

Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**11.444**	**<0.001**	**0.510**
Modulation condition (C)	(2, 22)	2.374	0.117	0.178
Time interval (T)	**(18, 198)**	**10.395**	**<0.001**	**0.486**
$\Delta f \times$ C	**(4, 44)**	**3.257**	**0.020**	**0.228**
$\Delta f \times$ T	**(36, 396)**	**1.945**	**0.001**	**0.150**
C $\times$ T	**(36, 396)**	**4.929**	**<0.001**	**0.309**
$\Delta f \times$ C $\times$ T	(72, 792)	1.134	0.217	0.093

913

914 **TABLE VI**

915 Results for experiment 3—effects of *abrupt changes* in stimulus modulation (timbre). Summary

916 of the two-way repeated-measures ANOVA for each transition (single time intervals) across

917 conditions C1, C2, C4, and C5. All significant terms are shown in bold.

Part (a): Results for first transition (T1)				
Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**10.553**	**<0.001**	**0.490**
Modulation condition (C)	**(3, 33)**	**6.790**	**0.001**	**0.382**
$\Delta f \times$ C	**(6, 66)**	**2.472**	**0.032**	**0.183**
Part (b): Results for second transition (T2)				
Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**8.941**	**<0.001**	**0.448**
Modulation condition (C)	**(3, 33)**	**4.847**	**0.007**	**0.306**
$\Delta f \times$ C	(6, 66)	1.410	0.224	0.114
Part (c): Results for third transition (T3)				
Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf)	**(2, 22)**	**4.933**	**0.017**	**0.310**
Modulation condition (C)	**(3, 33)**	**5.079**	**0.005**	**0.316**
$\Delta f \times$ C	**(6, 66)**	**2.869**	**0.015**	**0.207**

918

919 **TABLE VII**

920 Results for experiment 3—effects of *abrupt changes* in stimulus modulation (timbre). Summary

921 of the pairwise comparisons exploring the change in stream segregation following D-to-T and T-

922 to-D transitions, relative to their reference cases, for each transition number and Δf. All significant

923 cases are shown in bold.

Part (a): Results for D-to-T transitions (difference scores)			
Transition number	Δf=4 ST (mean, p)	Δf=6 ST (mean, p)	Δf=8 ST (mean, p)
T1	-10.5% pts, 0.108	**-25.1% pts, 0.001**	**-22.1% pts, 0.040**
T2	**-18.6% pts, 0.042**	**-34.9% pts, 0.004**	**-27.4% pts, 0.028**
T3	**-23.1% pts, 0.030**	**-42.2% pts, 0.003**	-26.2% pts, 0.068
Part (b): Results for T-to-D transitions (difference scores)			
Transition number	Δf=4 ST (mean, p)	Δf=6 ST (mean, p)	Δf=8 ST (mean, p)
T1	**+16.8% pts, 0.005**	+4.7% pts, 0.519	+6.0% pts, 0.357
T2	+5.1% pts, 0.493	+2.1% pts, 0.796	**+8.1% pts, 0.004**
T3	+12.1% pts, 0.057	+8.7% pts, 0.255	+3.9% pts, 0.343

924