

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

Saima L. Rajasingam, Robert J. Summers, and Brian Roberts

Citation: [The Journal of the Acoustical Society of America](#) **149**, 3769 (2021); doi: 10.1121/10.0005049

View online: <https://doi.org/10.1121/10.0005049>

View Table of Contents: <https://asa.scitation.org/toc/jas/149/6>

Published by the [Acoustical Society of America](#)

---

## ARTICLES YOU MAY BE INTERESTED IN

[The influence of envelope shape on the lateralization of amplitude-modulated, low-frequency sound](#)

[The Journal of the Acoustical Society of America](#) **149**, 3133 (2021); <https://doi.org/10.1121/10.0004788>

[Spectral envelope position and shape in sustained musical instrument sounds](#)

[The Journal of the Acoustical Society of America](#) **149**, 3715 (2021); <https://doi.org/10.1121/10.0005088>

[Perceptual similarity between piano notes: Simulations with a template-based perception model](#)

[The Journal of the Acoustical Society of America](#) **149**, 3534 (2021); <https://doi.org/10.1121/10.0004818>

## REVIEWS OF ACOUSTICAL PATENTS

[The Journal of the Acoustical Society of America](#) **149**, 3689 (2021); <https://doi.org/10.1121/10.0005067>

[Absolute pitch is disrupted by a memory illusion](#)

[The Journal of the Acoustical Society of America](#) **149**, 2829 (2021); <https://doi.org/10.1121/10.0004776>

[Near real-time detection of low-frequency baleen whale calls from an autonomous surface vehicle: Implementation, evaluation, and remaining challenges](#)

[The Journal of the Acoustical Society of America](#) **149**, 2950 (2021); <https://doi.org/10.1121/10.0004817>

---



# Across Acoustics

The official podcast highlighting authors' research from our publications

## The dynamics of auditory stream segregation: Effects of sudden changes in frequency, level, or modulation<sup>a)</sup>

Saima L. Rajasingam,<sup>1</sup> Robert J. Summers,<sup>2,b)</sup> and Brian Roberts<sup>2,c)</sup>

<sup>1</sup>Department of Vision and Hearing Sciences, Anglia Ruskin University, Cambridge CB1 1PT, United Kingdom

<sup>2</sup>School of Psychology, Aston University, Birmingham B4 7ET, United Kingdom

### ABSTRACT:

Three experiments explored the effects of abrupt changes in stimulus properties on streaming dynamics. Listeners monitored 20-s-long low- and high-frequency (LHL–) tone sequences and reported the number of streams heard throughout. Experiments 1 and 2 used pure tones and examined the effects of changing triplet base frequency and level, respectively. Abrupt changes in base frequency ( $\pm 3$ –12 semitones) caused significant magnitude-related falls in segregation (resetting), regardless of transition direction, but an asymmetry occurred for changes in level ( $\pm 12$  dB). Rising-level transitions usually decreased segregation significantly, whereas falling-level transitions had little or no effect. Experiment 3 used pure tones (unmodulated) and narrowly spaced ( $\pm 25$  Hz) tone pairs (dyads); the two evoke similar excitation patterns, but dyads are strongly modulated with a distinctive timbre. Dyad-only sequences induced a strongly segregated percept, limiting scope for further build-up. Alternation between groups of pure tones and dyads produced large, asymmetric changes in streaming. Dyad-to-pure transitions caused substantial resetting, but pure-to-dyad transitions sometimes elicited even greater segregation than for the corresponding interval in dyad-only sequences (overshoot). The results indicate that abrupt changes in timbre can strongly affect the likelihood of stream segregation without introducing significant peripheral-channeling cues. These asymmetric effects of transition direction are reminiscent of subtractive adaptation in vision.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0005049>

(Received 19 September 2020; revised 28 April 2021; accepted 3 May 2021; published online 2 June 2021)

[Editor: Christian Lorenzi]

Pages: 3769–3784

### 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 ( $\Delta f$ ) 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 that 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 most 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 alternating pure tones (LHLH...) and discovered that build-up has two distinct stages; there is a rapid increase in the tendency to hear two streams over the first 10 s followed by a slower rise thereafter. Once a tone sequence ends or is interrupted with a silent gap, the accumulated build-up decays over a few seconds (Bregman, 1978; Beauvois and Meddis, 1997). A convenient way to explore how the perceptual organization of later

<sup>a)</sup>Preliminary poster presentations on parts of this research were given at the Annual Meeting of the British Society of Audiology, Keele, UK, September 2014, and the 169th Meeting of the Acoustical Society of America, Pittsburgh, PA, USA, May 2015.

<sup>b)</sup>ORCID: 0000-0003-4857-7354.

<sup>c)</sup>Electronic mail: b.roberts@aston.ac.uk, ORCID: 0000-0002-4232-9459.

sounds is influenced by earlier sounds involves a stimulus configuration in which standard test sequences (whose properties remain the same across conditions) are immediately preceded by various types of induction sequence (i.e., stimuli intended to cause prior build-up) or by none (control condition). Studies of this kind, or variants thereof, have shown that another way in which accumulated build-up can be reduced or lost is through a sudden change in the acoustic properties of the sequence, such as a change in frequency region (Anstis and Saida, 1985) or in lateralization or level (Rogers and Bregman, 1998). This loss may occur either because the accumulated build-up was specific to properties of the original sounds, and so fails to transfer to the new sounds, or because sudden change triggers an active resetting of build-up (Rogers and Bregman, 1998; Roberts *et al.*, 2008).

Distinguishing experimentally between failure to transfer and resetting as accounts of the loss of build-up following sudden change can be challenging, but the ability of a single deviant tone at the end of an induction sequence to decrease the impact of the inducer on segregation in the following test sequence suggests that there are at least some circumstances in which active resetting is involved (Haywood and Roberts, 2010, 2013). In practice, the loss of build-up arising from either cause is usually referred to as resetting. Unlike abrupt changes, gradual changes in acoustic properties of a tone sequence, such as lateralization or level, have little or no effect on streaming (Rogers and Bregman, 1998). Bregman (1978) proposed that build-up occurred because integration was the default percept for sound sequences and that segregation emerged over several seconds as a result of a conservative evidence-accumulation process indicating that more than one source was active. The slow time course of this process was seen as serving to stabilize perception, thereby preventing the auditory system from fluctuating rapidly between alternative interpretations. In this functional account, the loss of build-up arising from sudden change is interpreted as the resetting of this evidence-accumulation process because the change signals a new auditory scene. A related idea is that a sudden correlated change applied to both subsets of tones in an alternating sequence signals their common origin and so encourages their integration.

Most streaming studies using induction sequences have focused primarily on exploring the strong segregation-promoting effect that occurs when an alternating-frequency (AF) test sequence is preceded by a constant-frequency induction sequence corresponding to one subset of the test-sequence tones (Rogers and Bregman, 1993; Roberts *et al.*, 2008; Haywood and Roberts, 2010, 2013; Rajasingam *et al.*, 2018). A notable exception is the study by Rogers and Bregman (1998) in which the induction and test sequences both involved frequency alternation. Their study included conditions exploring the effects of sudden changes in stimulus lateralization and level. Also relevant is the study by Anstis and Saida (1985) in which the effects of sudden changes in stimulus center frequency were explored using a variant of the induction-test configuration involving alternation

between an inducer (there called an adapting sequence) and a test sequence. These studies, their findings, and their limitations are considered in detail in Secs. II and III. They provide evidence that sudden changes in stimulus properties can lead not only to substantial loss of build-up previously accumulated but that in some cases the effects of change are directional, leading to asymmetries in listeners' responses to them.

It is rare in everyday life to hear sequences of sounds whose properties are static; the auditory scene is usually constantly changing, sometimes gradually and sometimes suddenly. The aim of the current study was better characterization of the effects of acoustic change on the dynamics of stream segregation. Three experiments are reported here; all used the LHL-LHL-... configuration first introduced by van Noorden (1975), where the dash represents a silence equal in duration to one of the tones. When this configuration is used, the one-stream percept is heard as a distinctive galloping rhythm, for which the pitch of the tones is heard to move from low to high and vice versa; this rhythm is lost when the L and H subsets segregate and are heard independently as higher- and lower-pitched streams. This way of measuring streaming is sometimes referred to as the "horse or Morse" task (Cusack *et al.*, 2004).

To overcome the limitations of the one-off measures of streaming used in many previous studies, all three experiments involved continuous monitoring of the perceptual organization of the tone sequence throughout a trial, allowing the time course of any effects of an abrupt transition to be measured and compared, including with the time course of build-up at the start of the sequence. Experiments 1 and 2 extended earlier work on the effects of sudden changes in center frequency (Anstis and Saida, 1985) or in level (Rogers and Bregman, 1998) and also demonstrated that build-up was largely unaffected when the center frequency of a tone sequence changed smoothly and progressively rather than staying constant. Experiment 3 explored the effects on stream segregation of sudden changes in timbre with minimal excitation-pattern cues (Hartmann and Johnson, 1991); this was achieved using abrupt transitions between unmodulated and modulated tones or vice versa. Responses to these transitions showed even more marked asymmetries than those observed for level changes, indicating strong directional effects. This outcome represents a major challenge to Bregman's (1978) functional account of build-up and further indicates that neurophysiological models of build-up based on the slow accumulation of adaptation need to account for rapid direction-sensitive changes in that adaptation following sudden transitions in acoustic properties.

## II. EXPERIMENT 1

To our knowledge, the only experiment to explore parametrically the effect of changing the center frequency of an AF tone sequence on subsequent streaming is the fourth experiment reported by Anstis and Saida (1985). They presented a 4-s adapting sequence of tones (LHLH...) with fixed properties that alternated with a 1-s test sequence



whose center frequency fell at one of 11 values from the set  $\pm 12, 4, 3, 2, 1$ , and 0 semitones (ST) relative to the center frequency of the adapting sequence (1 kHz). Switches between the adapting and test stimuli were without silences and so occurred seamlessly. The adapting and test stimuli shared a fixed  $\Delta f$  of 2 ST (i.e., for the adapting stimulus, L tones = 944 Hz and H tones = 1060 Hz). The purpose of the adapting stimulus was to induce a build-up in the tendency to hear two streams; its presentation rate was 4 cycles/s, corresponding to a tone repetition time (TRT) of 125 ms. Listeners had control of the presentation rate of the test stimulus, and their task was to adjust it as necessary to ensure that the test stimulus lay at the perceptual borderline between integration and segregation. This measure, known as the nulling rate because the adjustment was being made to offset the effect of build-up, was taken to be the mean of the adjustment settings over the final 30 s of the 90-s trial, so only one estimate was obtained per trial. Changes in nulling rate with center frequency of the test stimulus were used to plot a tuning curve for the effect of the adapting stimulus—the more build-up that transferred from the adapting stimulus to the test stimulus, the slower the nulling rate must be set for the test stimulus to track the perceptual borderline.

The tuning curve obtained was relatively narrow, with a flat top and steep skirts; it was also positioned asymmetrically, centered on the +1-ST test stimulus. Build-up produced by the adapting stimulus led to slower nulling rates for a test stimulus within  $\pm 1$  ST of the tuning frequency (i.e., 0–2 ST), but the effect of the adapting stimulus was largely extinguished when the test stimulus was  $\pm 2$  ST or more away from the tuning frequency. This pattern suggests that build-up transfers better when the center frequency is shifted upward than downward. Anstis and Saida (1985) noted that the bandwidth of tuning was broadly comparable with that of the auditory critical band (Scharf, 1970), but they used only one adapting stimulus, so it is not known whether the bandwidth of tuning to the adapting stimulus is affected by  $\Delta f$ . For example, increasing  $\Delta f$  would lead to a greater overlap in frequency range between adapting and test stimuli for a given shift in center frequency. Furthermore, the combination of small  $\Delta f$  (2 ST) and slow rate (TRT = 125 ms) used limited considerably the extent of build-up during the adapting stimulus compared with that which would have been produced if larger values of  $\Delta f$  and faster rates had been used (van Noorden, 1975). Anstis and Saida (1985) offered no explanation for the asymmetrical frequency tuning of build-up found in their study. The procedure they used did not allow investigation of the time course of the effect of frequency change on subsequent streaming.

There is evidence that an induction sequence that changes gradually in lateralization or level toward that for a steady test sequence, giving a smooth transition between them, leads to a similar effect on the build-up of stream segregation as that of an induction sequence whose properties match those of the test sequence (Rogers and Bregman, 1998). However, to our knowledge, the effects of smooth and progressive change have not been explored in the

context of frequency. Our pilot observations suggested that tone sequences whose triplet base frequency changed in this way were broadly as effective at inducing build-up as tone sequences with constant base frequency, despite the differences in peripheral channeling between the two types of stimulus. Furthermore, if confirmed, this outcome would pose a challenge to neurophysiological models in which build-up in the tendency to hear two streams is a result of multi-second adaptation caused by repeated stimulation of central auditory neurons with the same best frequency (e.g., Micheyl *et al.*, 2005).

The first experiment reported here investigated the effect of introducing a single sudden change (transition) in base frequency (corresponding to a distinct change in pitch range) in the middle of a long test sequence. The magnitude and direction of this change was varied;  $\Delta f$  for the test sequence was also varied. The abrupt transition was presented in the context of on-going small but progressive changes throughout the rest of the sequence (0.5 ST between adjacent triplets), a value that fell within the narrow adapting region identified by Anstis and Saida (1985). Gradual changes of this kind, at least in level or lateralization, are known to have relatively little impact on the subsequent likelihood of reporting stream segregation (Rogers and Bregman, 1998).

## A. Method

### 1. Listeners

Listeners were recruited mainly from the student population at Aston University, gave informed consent, and received either course credit or payment for taking part. They were first tested using a screening audiometer (Interacoustics AS208, Assens, Denmark) to ensure that their audiometric thresholds at 0.5, 1, 2, and 4 kHz did not exceed 20 dB hearing level. All listeners who passed this screening took part in a training session designed to familiarize them with the task and stimuli before proceeding to the main session; exclusion criteria were predefined in relation to a listener's profile of responses in the reference conditions (see Sec. II A 3). Twelve listeners (six males) successfully completed the experiment (mean age = 25.3 yr, range = 21.8–29.3). This research was approved by the Aston University Ethics Committee.

### 2. Stimuli and conditions

The test sequence was 20 s long and comprised 50 LHL– triplets. Each tone was 100-ms long (including 10-ms raised-cosine ramps). The silence at the end of each triplet was also 100 ms long, giving an onset-to-onset duration between triplets of 400 ms. This rate of presentation is known to facilitate streaming based on frequency separation (e.g., Bregman and Campbell, 1971; van Noorden, 1975). The base frequency of each triplet—defined as the frequency of its constituent L tones—was constant in some conditions but varied in others and ranged from a maximum of 1 kHz to a minimum of 0.5 kHz. The frequency of the H

tones was set relative to the L tones according to the desired low-high frequency difference for the test sequence ( $\Delta f$ ), which was 4, 6, or 8 ST. For example, when the base frequency was 1000 Hz, the frequency of the H tones for these values of  $\Delta f$  was 1260, 1414, and 1587 Hz, respectively. The range of frequency separations used was chosen to reduce ceiling and floor effects and to provide information on any interactions that might occur between frequency separation and condition. All tones were presented at 73 dB sound pressure level (SPL).

There were ten conditions for which the base frequency of the triplets was manipulated. A schematic depicting the test sequences is shown in Fig. 1; the left, middle, and right panels illustrate conditions C1–C4, C5–C7, and C8–C10, respectively. In C1 and C2, the base frequency was constant at 1 kHz and 0.5 kHz, respectively. In C3 and C4, the base frequency changed by 0.5 ST/triplet and followed either a linear rise-fall (C3) or fall-rise (C4) trajectory, moving gradually between the minimum and maximum base frequencies.<sup>1</sup> In C5–C7, the base frequency followed the same rising path as C3 for the first half of the sequence to reach 1 kHz, but at 10 s (triplet 26) there was an *abrupt fall* in base frequency of 3 ST (C5), 6 ST (C6), or 12 ST (C7); thereafter, a rising trajectory was resumed unless and until the maximum was reached, after which the falling path was followed. In C8–C10, the base frequency followed the same falling path as C4 for the first half of the sequence to reach 0.5 kHz, but at 10 s there was an *abrupt rise* in base frequency of 3 ST (C8), 6 ST (C9), or 12 ST (C10); thereafter, a falling trajectory was resumed unless and until the minimum was reached, after which the rising path was followed. All other properties of the test sequences remained the same across conditions.

All stimuli were synthesized at a sampling rate of 20 kHz using MITSYN (Henke, 2005). They were played back at 16-bit resolution over Sennheiser HD 480-13II earphones (Hannover, Germany) via a Sound Blaster X-Fi HD sound

card (Creative Technology Ltd., Singapore), programmable attenuators (Tucker-Davis Technologies, TDT PA5, Alachua, FL), and a headphone buffer (TDT HB7). Output levels were calibrated using a sound-level meter (Brüel & Kjaer, type 2209, Nærum, Denmark) coupled to the earphones by an artificial ear (type 4153). Diotic presentation was used throughout.

### 3. Procedure

Listeners completed the experiment in a single-walled sound-attenuating chamber (Industrial Acoustics 401 A, 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 h, divided into two separate sessions. The experiment was run using a program written in VISUAL BASIC (Visual Studio 2010, version 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

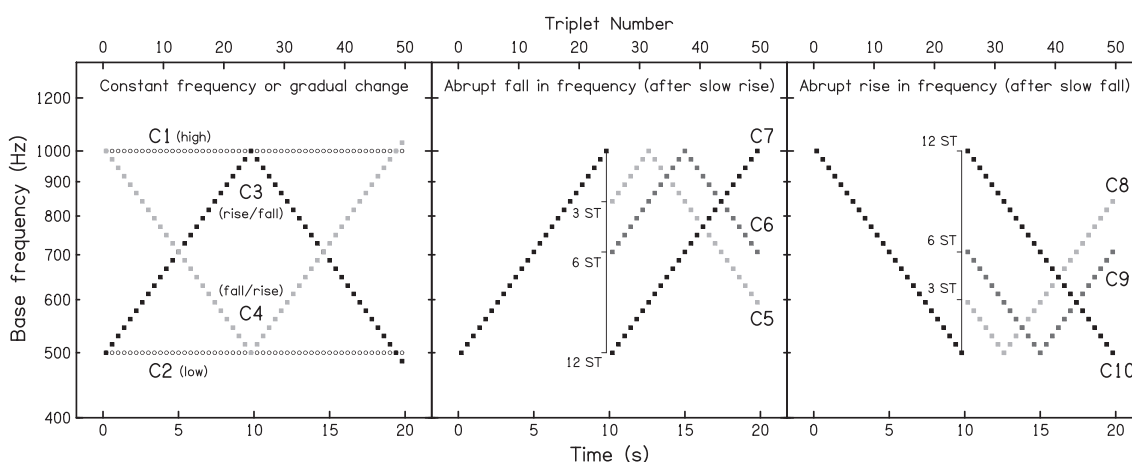


FIG. 1. Stimuli for experiment 1—schematic illustrating the conditions used (C1–C10). The left panel shows the base-frequency contours (i.e., L-tone frequencies) for the test sequences used in the constant (C1–C2), gradual rise/fall (C3), and gradual fall/rise (C4) conditions; each symbol (circle or 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).

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 (ten levels) and  $\Delta f$  (three levels) was presented ten times in the main experiment, once in each block, giving 300 trials. Using three different  $\Delta f$  values also provided a useful means of predefining criteria for excluding data. It is well established in the literature that, for a given rate of presentation, an increase in the frequency separation between subsets of pure tones increases the tendency to hear two streams (Miller and Heise, 1950; van Noorden, 1975; Anstis and Saida, 1985). Therefore, for a listener's data to be included, the mean overall extent of segregation for the conditions using steady sequences (C1 and C2) had to rise when  $\Delta f$  was increased from 4 ST to 6 ST and rise again when  $\Delta f$  was increased from 6 ST to 8 ST. No listeners were replaced in this experiment.

#### 4. Data analysis and availability

Response data from each trial were divided into twenty 1-s-long time bins (i.e., 0–1 s, 1–2 s, ..., 19–20 s). For each time bin, the percentage of time for which the listener reported the test sequence as segregated was computed from the timings of individual key presses. This value was recorded only if the listener's first response had occurred before the current time bin or within the first 0.5 s of that time bin. Owing to the small number of trials meeting this criterion for the 0–1 s time bin, responses made during that interval were used only in the context of calculations involving subsequent time bins; the 0–1 s time bin was excluded from all further analysis and graphical representation (cf. Haywood and Roberts, 2013; Rajasingam *et al.*, 2018).

For each listener, the data for each time bin were averaged across trial blocks separately for each combination of condition and  $\Delta f$ . Each mean was computed only from those trials for which that time bin met the acceptance criterion described above. On occasions when one of these means was missing for a particular listener (13 cases, corresponding to  $\sim 0.2\%$  of the data and all occurring within the first few time bins), mean imputation was used to replace the missing value with the mean of the corresponding values obtained from the other listeners. Finally, the data were averaged across listeners, for each combination of condition and  $\Delta f$ , to yield the overall mean percentage of time for which the test sequence was heard as segregated for each successive time bin. This measure of the average time course of stream segregation over the test sequence is used to display the results. Note that the standard errors were computed using the individual means only from those listeners for whom an experimental estimate was obtained.

The effects of base frequency *per se* and of smooth, progressive changes in base frequency were explored by comparing the extent of stream segregation across the

constant and gradual-change conditions for the full duration of the tone sequences used (in 1-s time bins, excluding 0–1 s). The effect of an abrupt transition in base frequency at 10 s in a given condition was explored by comparing the extent of stream segregation following the transition with that for appropriate comparator conditions during the same time interval. The reference comparison was with the test sequence that followed the same base-frequency contour up to 10 s but which then changed direction without discontinuity; other comparisons were between abrupt transitions of different magnitude or direction. These comparisons were made using a single, longer, time interval that focused on the period of peak response to the transition. This time interval was 4.0 s long and began 1.2 s (three triplets) after the transition,<sup>2</sup> to allow sufficient time for the effect of the change to be reflected in listeners' responses. It was long enough to encompass fully the peak response to the transition and was chosen to correspond to the remaining time available between transitions in subsequent experiments, which included conditions with transitions every 5.2 s.

All statistical analyses reported here were computed using R 3.6.3 (R Core Team, 2020) and the ez analysis package (Lawrence, 2016). The time-series data obtained from the calculations described above were analyzed using repeated-measures analysis of variance (ANOVA); the measure of effect size reported here is partial eta squared ( $\eta^2_p$ ). Comparisons among conditions with constant or gradually changing base frequency were conducted using three factors—frequency separation between test-sequence tones ( $\Delta f$ ), condition (C), and time interval (T, with levels corresponding to time bins 1–2 s to 19–20 s). Comparisons exploring the effects of abrupt transitions were based on the single time interval described above and so were conducted using two factors ( $\Delta f$  and C). Pairwise comparisons (two tailed) were conducted using the restricted least-significant-difference test (Snedecor and Cochran, 1967; Keppel and Wickens, 2004). The research data underlying this publication are available on-line from a repository hosted by Aston University.<sup>3</sup>

## B. Results and discussion

The results averaged across listeners are shown in Fig. 2. The upper panels display the results for the conditions in which the base frequency of the test sequence was either constant or gradually changing; the middle and lower panels display the results for the conditions in which there was an abrupt fall or rise in base frequency, respectively, 10 s after the sequence began. To facilitate comparisons, the middle and lower panels also reproduce the results for the gradual-change reference conditions (C3 for the abrupt-fall cases and C4 for the abrupt-rise cases). These results are considered in turn.

Although there is a suggestion in the mean data that slightly greater segregation is associated with the lower base frequency for the constant conditions, and with the gradual-change conditions relative to the constant conditions, this

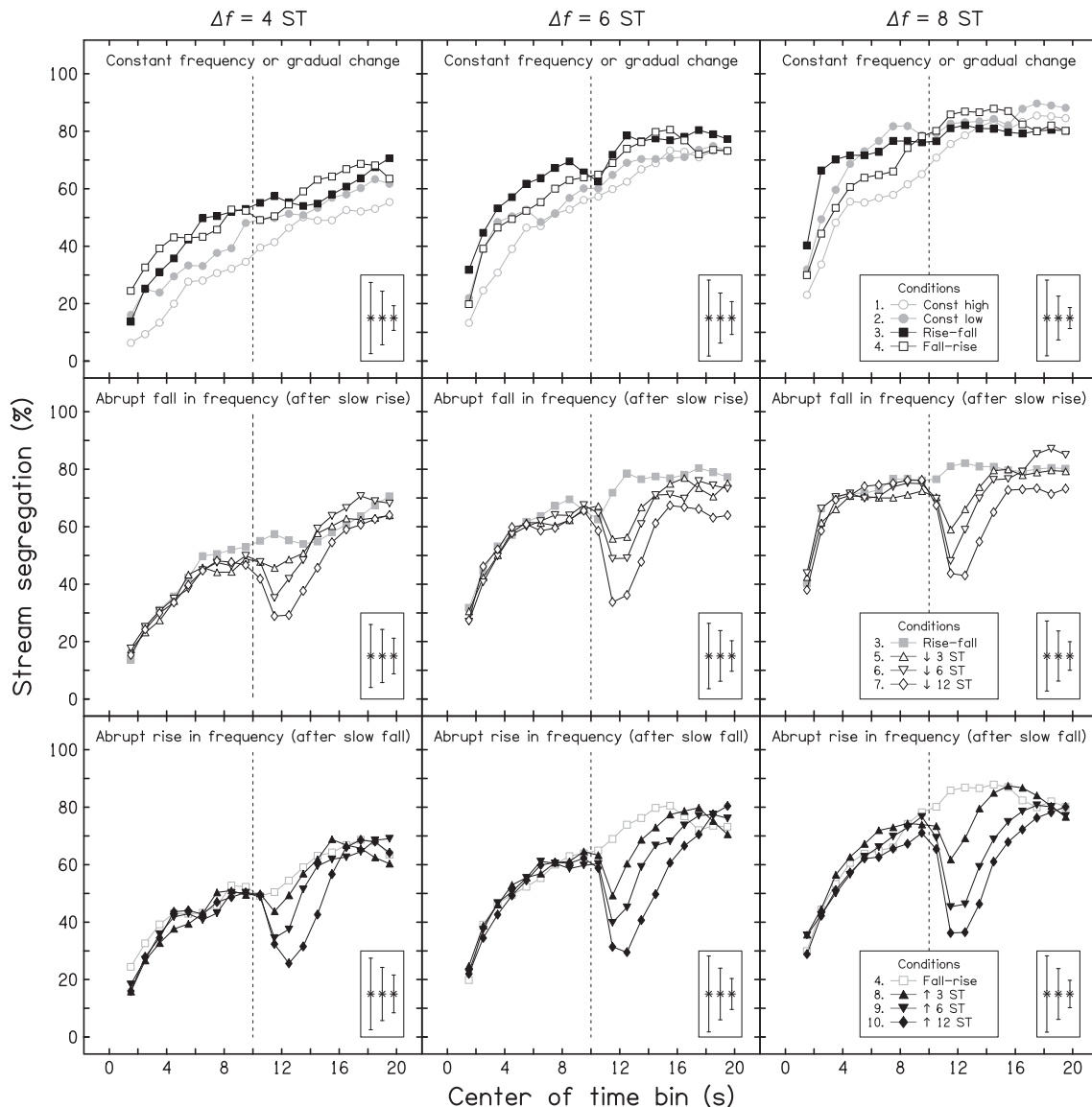


FIG. 2. Results for experiment 1—effects of base-frequency condition (1–10, see insets in right panels) and test-sequence frequency separation ( $\Delta f$ ) on the extent and time course of reported stream segregation ( $n = 12$ ). Responses for each trial were divided into 1-s time bins, for which 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).

was not borne out by the analysis. An ANOVA exploring the effects of the constant and gradual-change conditions is presented in Table I (data for the series of 1-s time bins). Two of the three factors influenced streaming as main effects—segregation was greater for larger frequency separations (means: 4 ST = 45.6%, 6 ST = 60.9%, and 8 ST = 72.6%;  $p < 0.001$ ) and tended to increase over time ( $p < 0.001$ )—but there was no main effect of condition ( $p = 0.378$ ), and none of the interaction terms were significant. The absence of a main effect of condition remained even if the time bins included in the analysis were restricted to the

fast phase of build-up (1–2 s to 10–11 s;  $p = 0.124$ ). No difference was anticipated between the two base frequencies (0.5 and 1 kHz) because both fell within the frequency range for which the ratio  $\Delta f$  over base frequency at the border between one- and two-stream percepts is roughly constant (Miller and Heise, 1950). All conditions elicited broadly similar patterns of build-up—an initial phase (up to ~10–12 s) that was relatively fast followed by a slower phase (Anstis and Saida, 1985; Haywood and Roberts, 2013; Rajasingam *et al.*, 2018). In general, an increase in  $\Delta f$  tended to increase both the rate and final extent of build-up.



TABLE I. Results for experiment 1—effects of stimulus base frequency (*constant or gradual-change conditions*). Summary of the three-way repeated-measures ANOVA for all time bins (1–2 s to 19–20 s) across conditions C1–C4. All significant terms are shown in bold.

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

Given that all tone sequences involving gradual change used a progressive shift in triplet base frequency of 0.5 ST every 0.4 s, spanning a full octave over the first 10 s, the occurrence of a similar pattern of results for the constant and gradual-change conditions indicates that the build-up of stream segregation for AF tone sequences does not depend on repeated stimulation of the same peripheral channels over several seconds. Furthermore, this finding suggests that the build-up of stream segregation does not require extended stimulation of populations of central auditory neurons with the same best frequency.

Inspection of Fig. 2 indicates that an abrupt fall or rise in base frequency decreased subsequent stream segregation and that the extent of this grew progressively as the size of the transition increased. An ANOVA exploring the effects of abrupt changes in base frequency is presented in Table II (data from a single 4.0-s time bin starting 1.2 s after the transition). This analysis showed significant main effects of  $\Delta f$  ( $p < 0.001$ ) and transition size (S,  $p < 0.001$ ). Despite the suggestion in the mean data that the loss of segregation produced by an abrupt rise in frequency (lower panels) was greater than that produced by an abrupt fall (middle panels), there was neither a main effect of transition direction (D,  $p = 0.812$ ) nor a significant interaction involving it. Only one interaction was significant ( $\Delta f \times S$ ,  $p = 0.046$ ), and this probably arose because the 3-ST transitions had relatively little effect for the smallest  $\Delta f$  used (4 ST).

TABLE II. Results for experiment 1—effects of an *abrupt change* in stimulus base frequency. Summary of the three-way repeated-measures ANOVA (single 4.0-s time interval, beginning 1.2 s after the transition) across conditions C3–C10. All significant terms are shown in bold.

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

Given that there were no significant effects involving transition direction, we report here the mean change in segregation for each size of transition after averaging across direction and  $\Delta f$ . These values correspond to the difference in segregation in percentage points (% pts) produced by that transition over the 4.0-s time bin relative to its reference case over the same interval (C3 and C4 for falling and rising transitions, respectively). The mean changes in segregation for the 3-ST, 6-ST, and 12-ST transitions were  $-9.3\%$  pts ( $p = 0.002$ ),  $-17.6\%$  pts ( $p < 0.001$ ), and  $-29.9\%$  pts ( $p < 0.001$ ), respectively; all other pairwise comparisons were also significant (range:  $p = 0.004$  to  $p < 0.001$ ). Although the loss of segregation associated with the 6-ST and 12-ST abrupt rises was nominally a third larger (in % pts) than for their falling counterparts, these differences disappear if the losses are interpreted in proportion to the extent of build-up taking place over 10 s for the relevant gradual-change reference conditions. Notably, either a sudden rise or fall leads to a near-complete loss of build-up for 12-ST transitions. Overall, there is no evidence to suggest an asymmetrical effect of transition direction on streaming. In all cases, Fig. 2 shows that the impact of the transition on streaming was greatest after  $\sim 2$ – $3$  s, and thereafter the extent of segregation began to grow again. The time course of this recovery from resetting was similar to that of the original phase of build-up, eventually slowing as listeners' responses began to converge with those for the corresponding reference cases. By the end of the sequence, segregation had mostly or completely returned to where it would have been without the sudden transition.

The results of this experiment differ in important ways from those of its counterpart reported by Anstis and Saida (1985). Their results indicated a narrow adapting region ( $\sim 2$  ST wide) that was tuned asymmetrically ( $\sim 1$  ST above adapting-stimulus center frequency) and outside which no transfer of build-up occurred. In contrast, our results indicate a much broader tuning of the adapting region with shallower skirts—the resetting effect of a sudden change of 3 ST was significant but partial, and a change as large as 12 ST was required to cause a near-complete loss of segregation. There was also no evidence of an effect of transition direction like that observed by Anstis and Saida (1985). These outcomes suggest that the results of their study were strongly influenced by one or more of the design factors considered above, which included the shorter interval over which build-up could occur (4 s), the nulling procedure involving rate adjustment, and the use of only a single small  $\Delta f$  (2 ST) and a relatively long TRT (125 ms). Alternatively, or in addition, it cannot be ruled out that our use of a gradually changing adapting stimulus rather than a constant one may have increased tolerance for a change in base frequency, widening the observed tuning. For example, in terms of Bregman's (1978) evidence-accumulation hypothesis of build-up, perhaps a larger sudden change is necessary against a background of gradual change before a new auditory scene is assumed and the evidence accumulation process is restarted.



### III. EXPERIMENT 2

Rogers and Bregman (1998) explored the effects on subsequent streaming of gradual and abrupt changes in stimulus lateralization—based on interaural time difference (ITD) cues, interaural level difference (ILD) cues, and loudspeaker position—or in overall stimulus level. They used an induction sequence (4.8 s) followed by a short test sequence (1.2 s); both tone sequences were configured in the form HLH–HLH–.... Listeners were asked to provide a one-off judgment of the extent of stream segregation at the end of the stimulus using a rating scale (1–8, where 1 corresponded to fully segregated and 8 to fully integrated). On the first trial of each condition, the listener heard sequences with a 9-ST separation. On the basis of the response to a given trial,  $\Delta f$  for the next trial was raised or lowered by 1 ST (over the range 5–14 ST) to make the percept increasingly ambiguous. Through an iterative process and averaging, this provided a measure of the border between segregation and integration (threshold  $\Delta f$ ) that could be compared across conditions. There were two reference cases—the no-change condition in which the induction and test sequences had identical properties (maximum build-up) and the control condition in which the induction sequence was replaced by white noise (no build-up).

Relative to an induction sequence whose properties matched those of the test sequence, an induction sequence whose properties changed gradually and finished close to those of the test sequence had little effect (lateralization) or no effect (level) on subsequent stream segregation. In contrast, an abrupt change in lateralization at the induction/test boundary caused a large shift in threshold  $\Delta f$ , indicating that (depending on the specific cues manipulated) between half and all the build-up accumulated during the induction sequence had been lost or reset. The direction of the spatial change (leftward or rightward) did not affect the extent of this loss. However, the effect of an abrupt 12-dB change in stimulus level (from 59 to 71 dB A or vice versa) was strongly directional. The sudden-louder condition (rising level) caused a loss of about two-thirds in the accumulated build-up, but the sudden-softener condition (falling level) had no effect. This asymmetry favors an account of the loss of build-up based on an active resetting process rather than a failure to transfer from the induction sequence. The second experiment reported here extends the investigation of the effects of abrupt transitions in level on streaming judgments, and their directional properties, by introducing occasional or frequent changes during the test sequence. For conditions involving occasional transitions, the time interval between them (5.2 s) was sufficiently long to observe the initial response to a transition and the main phase of recovery during the new steady state.

#### A. Method

Except where described, the same method was used as for experiment 1. Twelve listeners (two males, mean age = 20.4 yr, 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 it 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, 10.4, 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 conditions. Each combination of condition (five levels) and  $\Delta f$  (three levels) was presented ten times in the main experiment, once in each block, giving 150 trials.

Listeners completed this experiment in a single session, which typically took  $\sim 1\frac{1}{2}$  h. Time-series data were computed from listeners' responses in the same way as described for experiment 1. On occasions when an individual mean was missing (24 cases, all occurring within the first few time bins and corresponding to  $\sim 0.7\%$  of the data), the missing value was replaced using mean imputation. Once again, the results were analyzed using repeated-measures ANOVA.

#### B. Results and discussion

The results averaged across listeners are shown in Fig. 3. The upper panels display results for the conditions in which the level of the test sequence was either constant or alternated rapidly (every three triplets); the lower panels display results for the conditions in which there was an abrupt fall or rise in stimulus level after every 13 triplets (i.e., three transitions per sequence—T1, T2, and T3) and also reproduce the results for the constant reference cases. These results are considered in turn.

The constant-high and constant-low conditions (C1 and C2) showed similar extents of segregation and patterns of build-up over time, but the pattern for the rapid-alternation condition (C3) began to diverge from the others  $\sim 5$ – $10$  s after the start of the test sequence. This divergence was

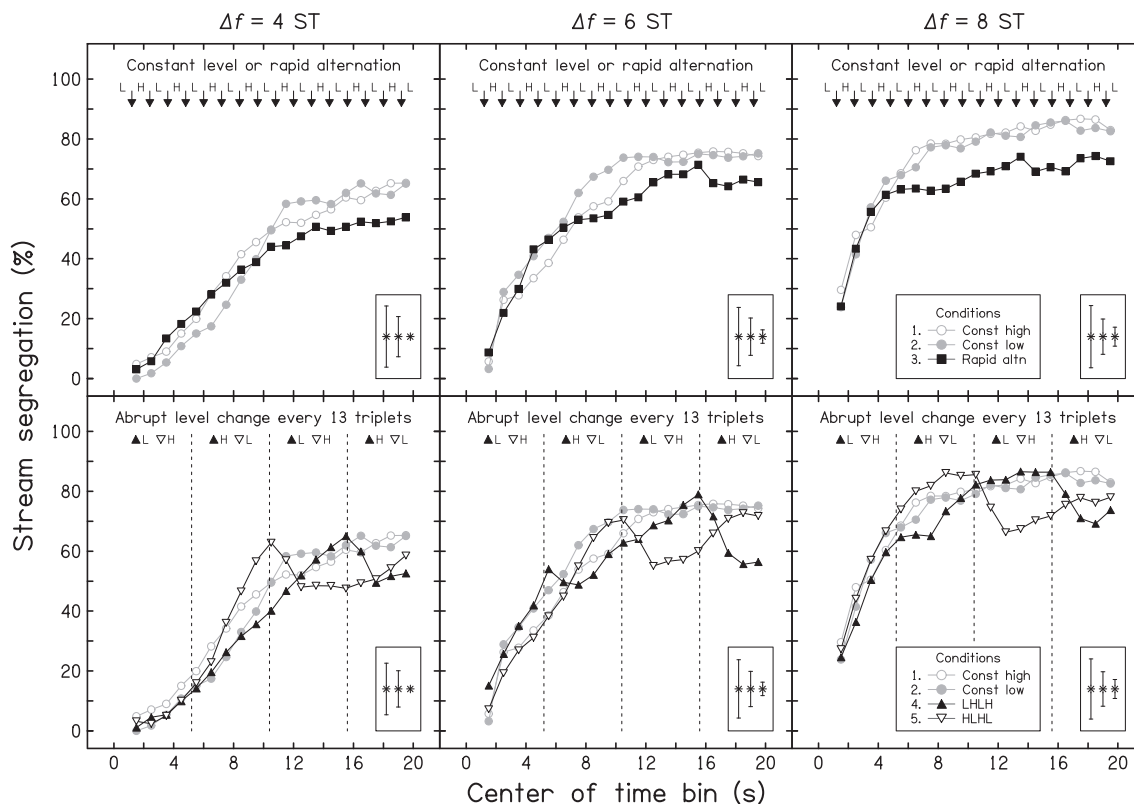


FIG. 3. Results for experiment 2—effects of level condition (1–5, see insets in right panels) and test-sequence frequency separation ( $\Delta f$ ) on the extent and time course of reported stream segregation ( $n = 12$ ). Results for the conditions involving either constant level or rapid alternation 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 are the results for the appropriate reference conditions. Otherwise as for Fig. 2.

manifest as a suppression of build-up that appeared greatest for the largest  $\Delta f$  tested (8 ST). The ANOVA for the conditions in which stimulus level was either constant or alternated rapidly is presented in Table III. Two of the three factors influenced streaming as main effects—segregation was greater for larger frequency separations (means: 4 ST = 39.1%, 6 ST = 56.9%, and 8 ST = 70.4%;  $p < 0.001$ ) and tended to increase over time ( $p < 0.001$ )—but there was no main effect of condition ( $p = 0.196$ ). Two of the two-way interactions were also significant—condition  $\times$  time interval ( $p < 0.001$ ) and  $\Delta f \times$  time interval ( $p < 0.001$ ). The former arose mainly because the loss of segregation caused

by multiple changes in level was largely confined to the latter half of the sequence, perhaps at least partly because this was the time period over which there was more scope for loss of segregation. The latter arose mainly because the tendency for stream segregation to continue increasing during the second half of the sequence was greater for smaller frequency separations.

Overall, these outcomes indicate that there was no effect of level *per se* on stream segregation over the 12-dB range tested but that the rapid alternation in level acted to reduce the build-up of stream segregation. The absence of an effect of presentation level on the results for C1 and C2 is unsurprising. Although it has been shown that increasing presentation level can lead to a fall in stream segregation for a given center frequency and  $\Delta f$  (Rose and Moore, 2000), presumably owing to the broadening of auditory filter bandwidths (Glasberg and Moore, 1990), any effect of this kind would have been modest over the range tested here. Also, that study measured the fission boundary, so listeners were instructed to try to segregate one subset of tones from a sequence, whereas our listeners were given neutral listening instructions. Given the similar profiles for the constant high- and low-level cases, the lower segregation in the latter half of the sequence associated with fast alternations implies that, when averaged, abrupt rises and falls in level tend to suppress build-up. We now consider the directional effects of individual transitions.

TABLE III. Results for experiment 2—effects of stimulus level (constant or rapid-alternation conditions). Summary of the three-way repeated-measures ANOVA for all time bins (1–2 s to 19–20 s) across conditions C1–C3. All significant terms are shown in bold.

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

Inspection of the lower panels of Fig. 3 suggests that sudden L-to-H transitions in level tended to decrease subsequent stream segregation but that sudden H-to-L transitions had little or no effect (with the possible exception of T1, discussed later). ANOVAs exploring the effects of abrupt changes in level every 13 triplets (i.e., three transitions per test sequence) are presented in Table IV. A 4.0-s time bin was used for the first two transitions, starting 1.2 s after each change (cf. experiment 1), but this was reduced to 3.2 s for the third transition owing to the termination of the sequence at 20 s. There was a main effect of  $\Delta f$  for all three transitions ( $p < 0.001$  in all cases), reflecting the usual tendency for greater streaming with larger values of  $\Delta f$ . For condition, there was a main effect for T2 ( $p = 0.002$ ) and T3 ( $p = 0.007$ ), respectively, but only a non-significant trend for T1 ( $p = 0.074$ ), so it was not considered further. The latter was probably a consequence of the limited time available for build-up from scratch during the first 5 s of the test sequence. There was no significant  $\Delta f \times C$  interaction for any transition number.

The effects of individual transitions were explored further using pairwise comparisons after the results were collapsed across  $\Delta f$ . For L-to-H and H-to-L transitions in level, respectively, the reference cases were the results for the constant-high (C1) and constant-low (C2) conditions during the corresponding time interval. Hence, the reference cases were matched for the stimulus properties of the test cases following the transition. Note that, between them, conditions C4 (level: LHLH) and C5 (level: HLHL) provided data for one transition in each direction for each transition number.

TABLE IV. Results for experiment 2—effects of abrupt changes in stimulus level. Summary of the two-way repeated-measures ANOVA for each transition (single time intervals) across conditions C1, C2, C4, and C5. All significant terms are shown in bold.

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

There was a significant loss of segregation associated with the L-to-H transitions (change for T2:  $-12.6\%$  pts,  $p = 0.018$ ; change for T3:  $-14.8\%$  pts,  $p = 0.007$ ), but for the H-to-L transitions there was no effect for T2 (change:  $-0.2\%$  pts,  $p = 0.948$ ) and only a small loss for T3 (change:  $-5.7\%$  pts,  $p = 0.027$ ). Taken together, these outcomes indicate an asymmetry similar to that reported by Rogers and Bregman (1998). Note that the tendency for suppression of build-up in sequences with fast alternations (C3) can be explained by the resetting effects of multiple L-to-H transitions. Rogers and Bregman (1998) interpreted the asymmetry in terms of Bregman's (1978) functional account of build-up, arguing that it reflected the greater importance of sudden increases in level, because such increases usually indicate the onset of new sound sources.

Finally, it merits note that one aspect of the current results motivated the development of the final experiment reported here. Following the H-to-L transition in level at T1 when  $\Delta f = 4$  ST, there is a suggestion in the data that segregation *increased* (i.e., overshoot rather than resetting of segregation); indeed, if considered in isolation, this change would be significant (change:  $+13.3\%$  pts,  $p = 0.009$ ). Further evidence that overshoot can occur is provided by experiment 3.

#### IV. EXPERIMENT 3

Another context in which marked perceptual asymmetries have been observed is auditory search. Asemi *et al.* (2003) used a task in which a target and distractors were presented simultaneously over loudspeakers at different positions in the frontal-horizontal plane. They found that the reaction time for detecting a narrowband noise, amplitude-modulated tone, or frequency-modulated tone among pure-tone distractors was largely unaffected by the number of distractors—indicating “pop out”—but that the time needed to detect a pure tone among temporally fluctuating distractors increased with the number of distractors. The same asymmetry was observed by Cusack and Carlyon (2003) for pure and frequency-modulated tones presented sequentially. These results show that the auditory system uses temporal changes in the amplitude and frequency of sound as basic features for the detection of a sound in an acoustic scene. In other words, sounds possessing these basic features will be more salient and attention-grabbing.

Previous studies of stream segregation have used modulated sounds, both narrowband (e.g., Cusack and Roberts, 1999) and wideband (e.g., Grimault *et al.*, 2002), but to our knowledge only in the context of investigating the effects of introducing differences between the two subsets of sounds comprising the sequence. Given the asymmetry found in auditory search tasks and our own pilot observations, we considered that introducing sudden transitions between unmodulated and modulated sounds offered a potentially fruitful candidate for observing directional effects of these transitions. Abrupt changes in the center frequency or overall level of a tone sequence inevitably introduce differences

in long-term excitation pattern between corresponding subsets of tones. Given the potential impact of peripheral-channeling cues on stream segregation (Hartmann and Johnson, 1991; Roberts *et al.*, 2002; Moore and Gockel, 2012), we wished to minimize them by transitioning between pure tones (unmodulated) and narrowly spaced two-tone complexes (modulated; cf. Cusack and Roberts, 1999). These two types of sound differ markedly in timbre. The sequence configurations used and the timing of the transitions corresponded closely with their counterparts in experiment 2.

## A. Method

Except where described, the same method was used as for experiments 1 and 2. Twelve listeners (two males, mean age = 20.9 yr, 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 ( $\pm 25$  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 five 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  $\approx$  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 peripheral-channeling cues has not previously been investigated. The strong modulation of the dyads resulting from the interaction of the two components within the same auditory filter gave them a distinctive timbre; indeed, the quality of a sequence of triplets composed of dyads was reminiscent of the sound produced by stridulating crickets.

There were five conditions (C1–C5). In C1 and C2, the triplets were always composed of pure tones or dyads, respectively. In C3, there were abrupt changes in triplet timbre between pure tones (starting value) and dyads every three triplets (i.e., rapid alternation every 1.2 s). If the dominant effect of the timbre transitions was to cause resetting, then this should be manifest as an overall suppression of the build-up of stream segregation (cf. experiment 2). In C4 and C5, the alternation of abrupt changes in modulation occurred more slowly—once every 13 triplets (i.e., at 5.2, 10.4, and 15.6 s)—and the sequence began with either pure tones (C4, TDTD) or dyads (C5, DTDT). For the alternating-timbre 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 pure-tone and dyad-starting cases for the slow alternation conditions ensured that each transition direction (pure-to-dyad and dyad-to-pure) was represented equally often and at different times in the test sequence. All other properties of the test sequences remained the same across conditions. Pure tones and dyads were presented at 73 dB SPL.

Each combination of condition (five levels) and  $\Delta f$  (three levels) was presented ten times in the main experiment, once in each block, giving 150 trials. Listeners completed this experiment in a single session, which typically took  $\sim 1\frac{1}{2}$  h. Time-series data were computed from listeners' responses in the same way as described for experiment 1. On occasions when an individual mean was missing (44 cases, corresponding to  $\sim 1.3\%$  of the data and all occurring within the first few time bins), mean imputation was used to replace the missing value.

## B. Results and discussion

The results averaged across listeners are shown in Fig. 4. The upper panels show results for the conditions in which the timbre of the sounds comprising the test sequence was either constant (unmodulated or modulated) or alternated rapidly (every three triplets); the lower panels show results for the conditions in which there was an abrupt change from a pure tone to a dyad or vice versa every 13 triplets (transitions T1, T2, and T3) and also reproduce the results for the reference cases. These results are considered in turn.

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 three 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 pattern is an overall loss of segregation that emerged  $\sim 5$ –10 s after the start of the



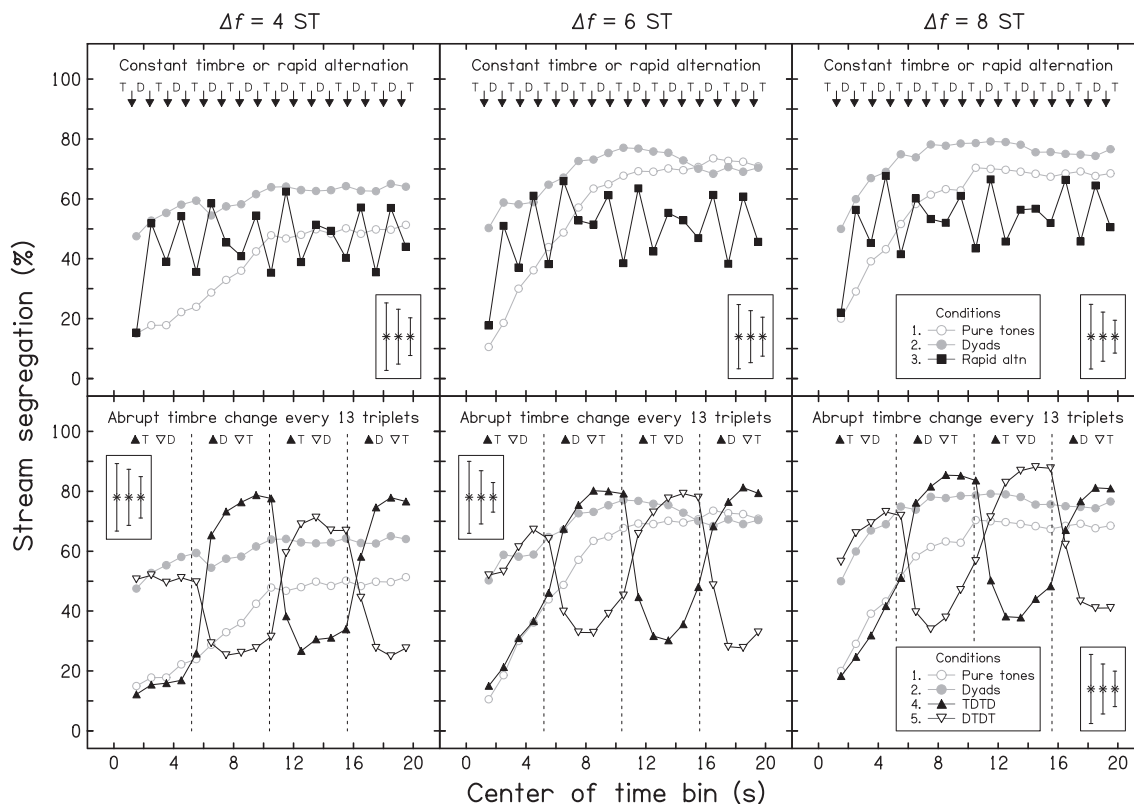


FIG. 4. Results for experiment 3—effects of modulation condition (1–5, see insets in right panels) and test-sequence frequency separation ( $\Delta f$ ) on the extent and time course of reported stream segregation ( $n = 12$ ). Results for the conditions involving either constant timbre or rapid alternation 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 are the results for the appropriate reference conditions. Otherwise as for Fig. 2.

sequence (cf. the corresponding condition in experiment 2). This divergence was manifest as a suppression of build-up that appeared to be greater for the larger values of  $\Delta f$  tested (6 and 8 ST).

The ANOVA for the conditions in which the stimulus modulation for the test sequence was either constant or alternated rapidly is presented in Table V. Two of the three factors influenced streaming as main effects—segregation was greater for larger frequency separations (means: 4 ST = 48.0%, 6 ST = 58.4%, and 8 ST = 61.8%;  $p < 0.001$ ) and tended to increase over time ( $p < 0.001$ ). The main effect of condition was not significant overall ( $p = 0.117$ )

but became so if the time bins included in the analysis were restricted to the fast phase of build-up (1–2 s to 10–11 s;  $p = 0.043$ ), reflecting the considerably greater segregation for dyad than pure-tone sequences during this phase. All of the two-way interactions were also significant—condition  $\times$  time interval ( $p < 0.001$ ),  $\Delta f \times$  condition ( $p = 0.020$ ), and  $\Delta f \times$  time interval ( $p = 0.001$ ). Once again, the  $C \times T$  interaction arose mainly because the loss of segregation caused by multiple changes was largely confined to the latter half of the sequence, for which there was more scope for loss of segregation, and the  $\Delta f \times T$  interaction arose mainly because the tendency for stream segregation to continue increasing during the second half of the sequence was greater for smaller frequency separations. The  $\Delta f \times C$  interaction probably reflects the smaller suppression of segregation observed for the rapid-alternation condition for the smallest  $\Delta f$  tested.

One possible explanation for the greater stream segregation observed here for dyad sequences than for pure-tone sequences with the same  $\Delta f$  is suggested by the results of Cusack and Roberts (1999). They used repeating LHL–LHL–... sequences of two-tone complexes in which the L stimuli (center frequency = 1000 Hz) had a fixed component separation of 100 Hz in all conditions, whereas the H stimuli (center frequency = 1200 Hz) had a component separation corresponding to one of seven values (80–140 Hz, in 10-Hz steps) across conditions. Least segregation was

TABLE V. Results for experiment 3—effects of stimulus modulation (timbre, constant or rapid-alternation conditions). Summary of the three-way repeated-measures ANOVA for all time bins (1–2 s to 19–20 s) across conditions C1–C3. All significant terms are shown in bold.

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

reported when the H stimuli had the same relative bandwidth as the L stimuli (match = 120 Hz) rather than the same modulation rate (100 Hz). Notwithstanding the use of a smaller fixed component separation of 50 Hz in the current experiment, it seems likely that using the same component separation for the H and L dyads introduced an additional factor supporting the build-up of stream segregation to that provided by  $\Delta f$ .

Inspection of the lower panels of Fig. 4 shows that slow alternations of the sudden changes in timbre caused dramatic changes in perception between integrated and segregated. The overall pattern suggests that sudden D-to-T transitions (i.e., modulated to unmodulated) decreased subsequent streaming but that sudden T-to-D (i.e., unmodulated to modulated) transitions tended to have the opposite effect. ANOVAs exploring the effects of abrupt changes in level every 13 triplets are presented in Table VI; the same time bins were used for transitions T1–T3 as for their counterparts in experiment 2. For all three transitions, there was a main effect of  $\Delta f$  (range:  $p = 0.017$  to  $p < 0.001$ ), reflecting the usual tendency for streaming to increase with  $\Delta f$ , and of condition (range:  $p = 0.007$  to  $p = 0.001$ ), reflecting the evident differences between conditions during the observation interval. Since there was a significant  $\Delta f \times C$  interaction for two of the three transitions, pairwise comparisons were made separately for each  $\Delta f$ . For D-to-T and T-to-D transitions, respectively, the reference cases were the results for the pure-tone-only (C1) and dyad-only (C2) conditions during the corresponding time interval, again matching the reference cases to the stimulus properties of the test cases

following the transition. Between them, conditions C4 (TDTD) and C5 (DTDT) provided data for one transition in each direction at each  $\Delta f$  for T1–T3.

For each direction of change, the results of these pairwise comparisons are summarized in Table VII for all nine combinations of transition number and  $\Delta f$ . For the D-to-T transitions (i.e., modulated to unmodulated), all nine combinations were associated with a fall in segregation (overall mean difference =  $-25.6\%$  pts); seven cases were significant, and the losses were often substantial. For the T-to-D transitions (i.e., unmodulated to modulated), all nine combinations were associated with an *increase* in subsequent segregation, but the overall mean difference was considerably smaller ( $+7.5\%$  pts); only two cases showed significant overshoot (T1 for  $\Delta f = 4$  ST,  $+16.8\%$  pts,  $p = 0.005$ ; T2 for  $\Delta f = 8$  ST,  $+8.1\%$  pts,  $p = 0.004$ ), and a third approached significance (T3 for  $\Delta f = 4$  ST,  $+12.1\%$  pts,  $p = 0.057$ ). Note, however, that using the dyad-only segregation score during the corresponding interval represents an exceptionally conservative reference case for estimating overshoot following T2 and T3. This is because the previous D-to-T transition will have reset almost all prior build-up.

For T2 and T3, using the mean segregation score averaged over the fast phase of build-up (time bins 1–2 s to 10–11 s) is arguably a more reasonable reference case. For example, if this reference were used instead, the marginal case noted above would become significant (T3 for  $\Delta f = 4$  ST,  $+19.6\%$  pts,  $p = 0.013$ ). It is also the case that the “headroom” available to demonstrate overshoot following T-to-D transitions was quite limited for larger  $\Delta f$ s because of the high segregation scores associated with dyad sequences. These issues suggest that further research with stimuli of this kind would benefit from two changes in the experimental design: first, to include only one transition per trial (as was done here in experiment 1), to avoid the difficulties of choosing an appropriate reference case for subsequent transitions and also to allow more time to observe streaming before and after; second, to include smaller  $\Delta f$ s than were tested here to allow greater headroom for overshoot effects to be manifest. Nonetheless, it seems reasonable overall to conclude that D-to-T transitions do not lead to resetting but instead tend to increase subsequent streaming. The implications of these results for accounts of build-up and the kinds of mechanism that might explain the observed asymmetry are considered below.

Finally, it should be acknowledged that a contribution to the results from audible distortion products generated by the dyads cannot be ruled out entirely, given the relatively high presentation level and the absence of background noise. The most prominent combination tone generated from 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  $\leq 1.10$  (Goldstein, 1967). For the tone pair constituting the L dyads,  $f_1 = 975$  Hz and  $f_2 = 1025$  Hz, giving a ratio of  $\sim 1.05$  and generating a cubic difference tone at 925 Hz. Although it would have been lower in level than the primaries, this distortion product may have been sufficient to increase the

TABLE VI. Results for experiment 3—effects of abrupt changes in stimulus modulation (timbre). Summary of the two-way repeated-measures ANOVA for each transition (single time intervals) across conditions C1, C2, C4, and C5. All significant terms are shown in bold.

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

TABLE VII. Results for experiment 3—effects of *abrupt changes* in stimulus modulation (timbre). Summary of the pairwise comparisons exploring the change in stream segregation following D-to-T and T-to-D transitions, relative to their reference cases, for each transition number and  $\Delta f$ . All significant cases are shown in bold.

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

level of excitation on the lower skirt of the excitation pattern evoked by the dyads, leading to greater than anticipated differences in peripheral channeling between corresponding pure tones and dyads. Even if this were the case, however, it is not clear how this could account for the strong directional effects observed for sudden transitions between pure tones and dyads. Rather, we argue that the critical factor is most probably the sudden changes in modulation.

## V. CONCLUDING DISCUSSION

For tone sequences involving one or more correlated transitions in acoustic properties—i.e., where the high- and low-frequency subsets change together to the same extent on the same dimension—the effect of a sudden change can be influenced not only by the property being altered but also by the direction of that change. Experiment 1 explored the effects of sudden changes in triplet base frequency and found that part of the build-up of stream segregation prior to a transition can transfer over a wider frequency region (more than half an octave), and more equally for sudden rises and falls, than had been suggested by the results for the particular set of values tested by Anstis and Saida (1985). Relative to maintaining a constant base frequency, the progressive accumulation of gradual changes in base frequency made little or no difference to the build-up of stream segregation for a pure-tone sequence. This outcome casts doubt on models of auditory streaming in which build-up depends on extended stimulation of populations of central auditory neurons with the same best frequency (e.g., Micheyl *et al.*, 2005; Pressnitzer *et al.*, 2008; Bee *et al.*, 2010). Rather, this outcome suggests a mechanism in which accumulated build-up in the tendency for stream segregation (adaptation) can be transferred between neurons with different best frequencies, so long as there are no abrupt changes in base frequency as the tone sequence unfolds.

Experiments 2 and 3 explored the effects of sudden changes in level and modulation, respectively. Sudden transitions in level ( $\pm 12$  dB) produced smaller changes in segregation than those associated with sudden transitions in base frequency, and, in accord with the findings of Rogers and

Bregman (1998), there was a clear asymmetry in the effect of transition direction. Rising transitions (softer-to-louder) caused significant loss of build-up (resetting), but falling transitions (louder-to-softer) had little or no effect. The effects of sudden changes in tone modulation on stream segregation were larger, with the losses for D-to-T transitions (i.e., from modulated to unmodulated) approaching the size of those for changes in base frequency, and the effect of direction was even more marked. Specifically, T-to-D transitions (i.e., from unmodulated to modulated) in some cases led to even greater segregation than that for dyad-only sequences during the corresponding time interval (i.e., overshoot).

Rogers and Bregman (1998) interpreted the asymmetry they observed for sudden level changes in terms of Bregman's (1978) functional account of build-up, arguing that a sudden rise in level causes a loss of build-up but a sudden fall does not because only the former can signal the activation of a new sound source. However, it is hard to see how this argument might be extended to account for the directional effects of changes in tone modulation, particularly given the evidence that sudden changes from unmodulated to modulated tones (T-to-D transitions) sometimes result in greater stream segregation. Although speculative, three plausible accounts merit discussion. First, the overshoot sometimes observed after T-to-D transitions may be a short-term contrast effect arising from the greater tendency for dyad-only sequences to be heard as segregated relative to pure-tone sequences. Longer-term contrast effects, occurring across trials, have previously been reported for AF tone sequences following changes in  $\Delta f$  (Snyder *et al.*, 2008; Snyder *et al.*, 2009). Second, it may be possible to extend attention-switching accounts of the loss of build-up after an abrupt transition in stimulus properties (Rajasingam *et al.*, 2018; see also Cusack *et al.*, 2004; Thompson *et al.*, 2011) to explain the overshoot that can occur after a T-to-D transition, on the basis that the switch in attention is to modulated sounds, which have primary attention-grabbing properties (cf. Asemi *et al.*, 2003; Cusack and Carlyon, 2003). Third, the occurrence of overshoot may be indicative of the operation of some inhibitory or suppressive process whose accumulation and release affect the extent of stream segregation.



The phenomenon of subtractive adaptation has long been known in the visual system (e.g., Geisler, 1983; Hayhoe *et al.*, 1992), and mechanisms of this kind have since been proposed to account for the multi-second build-up of stream segregation for a repeating but unchanging sequence of tones (e.g., Micheyl *et al.*, 2005; Pressnitzer *et al.*, 2008; Bee *et al.*, 2010). The basis of these accounts is that the intensity of the response of tonotopically tuned neurons in the central auditory system to a repeating tone sequence gradually declines through the slow accumulation of inhibition or suppression, leading to a progressive narrowing of their receptive fields. As a result, the receptive fields of neurons best tuned to the H and L subsets initially overlap, but over time two distinct subpopulations emerge, leading to the perception of separate streams. A sudden change of sufficient magnitude in the base frequency of the tone sequence resets this process because a different population of neurons is activated. In this regard, note that the changes in stimulus level or modulation used in experiments 2 and 3 did not involve changes in base frequency, so the transitions were not anticipated to change the frequency-tuned populations of neurons responding to these sequences. Presumably, rising-level transitions and transitions from modulated to unmodulated tones (D-to-T) led quickly to partial or complete release, respectively, of accumulated inhibition, resulting in a loss of build-up. By this account, neither falling-level transitions nor transitions from unmodulated to modulated tones (T-to-D) led to release of inhibition. Given that a 12-dB difference in level *per se* has little or no effect on the build-up of stream segregation, a falling-level transition therefore produces little or no effect on subsequent streaming. Presumably, overshoot sometimes arises for T-to-D transitions because the tendency to hear two streams is greater for dyad sequences than for pure-tone sequences, leading to increased inhibition rather than a release from it following the transition. Future research might investigate the responses of auditory cortical neurons to tone sequences involving sudden changes in level or modulation.

In conclusion, the experiments reported here have extended our knowledge of the dynamics of auditory stream segregation. Most notably, we have demonstrated that the effects of sudden correlated transitions in stimulus modulation are strongly directional, including instances in which a sudden change from unmodulated (pure tones) to modulated (dyads) sounds leads to greater segregation (overshoot) rather than a loss of build-up. It is not obvious how Bregman's (1978) functional account of build-up might be adapted to explain these findings, but there are plausible accounts based on stimulus contrast effects, attention switching, or neural mechanisms involving the accumulation and release of inhibition or suppression. These accounts might be investigated by extending the perceptual experiments reported here to include transitions between modulated sounds with different modulation rates and by exploring the effects of transitions of this kind on the responses of auditory cortical neurons. This approach should help elucidate further how stream segregation functions in changing auditory scenes.

## ACKNOWLEDGMENTS

This research was by supported Aston University, which funded a Ph.D. studentship for S.R. under the supervision of B.R. We thank Nick Haywood and Brian Moore for their comments on an earlier version of this manuscript and Mark Georgeson for drawing our attention to the literature on subtractive adaptation in vision. The experiments reported here correspond to reanalyzed versions of experiments 1, 3, and 4 in the doctoral thesis of S.R. (Rajasingam, 2016).

<sup>1</sup>Rather than changing direction, the base frequency of the final triplet in C3 and C4 continued on its established trajectory and consequently fell 0.5 ST below the nominal minimum (486 Hz) and above the nominal maximum (1029 Hz), respectively. Note that the final triplet occurred too late to have any appreciable effect on listeners' responses.

<sup>2</sup>In experiment 1, the start time for the 4.0-s time interval used to explore the effect of abrupt transitions on streaming was 11.2 s (i.e., 1.2 s after the transition at 10.0 s), which did not align exactly with the set of 1-s time bins plotted in Fig. 2. Similar issues of alignment arose in relation to the transition times used in experiments 2 and 3 (5.2, 10.4, and 15.6 s). To create longer time intervals for statistical analysis with start times that did not correspond to an integer number of seconds, the response data for each trial were also divided into finer-grained (0.2 s) time bins. Longer time intervals were constructed by combining the appropriate set of 0.2-s time bins.

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

- Anstis, S., and Saida, S. (1985). "Adaptation to auditory streaming of frequency-modulated tones," *J. Exp. Psychol. Hum. Percept. Perform.* **11**, 257–271.
- Asemi, N., Sugita, Y., and Suzuki, Y. (2003). "Auditory search asymmetry between pure tone and temporal fluctuating sounds distributed on the frontal-horizontal plane," *Acta Acust. united Acust.* **89**, 346–354.
- Beauvois, M. W., and Meddis, R. (1997). "Time decay of auditory stream biasing," *Percept. Psychophys.* **59**, 81–86.
- Bee, M. A., Micheyl, C., Oxenham, A. J., and Klump, G. M. (2010). "Neural adaptation to tone sequences in the songbird forebrain: Patterns, determinants, and relation to the build-up of auditory streaming," *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **196**, 543–557.
- Bregman, A. S. (1978). "Auditory streaming is cumulative," *J. Exp. Psychol. Hum. Percept. Perform.* **4**, 380–387.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA), pp. 1–790.
- Bregman, A. S., and Campbell, J. (1971). "Primary auditory stream segregation and perception of order in rapid sequences of tones," *J. Exp. Psychol.* **89**, 244–249.
- Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (2001). "Effects of attention and unilateral neglect on auditory stream segregation," *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 115–127.
- Cusack, R., and Carlyon, R. P. (2003). "Perceptual asymmetries in audition," *J. Exp. Psychol. Hum. Percept. Perform.* **29**, 713–725.
- Cusack, R., Deeks, J., Aikman, G., and Carlyon, R. P. (2004). "Effects of location, frequency region and time course of selective attention on auditory scene analysis," *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 643–656.
- Cusack, R., and Roberts, B. (1999). "Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes," *Perception* **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), Chap. 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]," <https://cran.r-project.org/package=ez> (Last viewed December 6, 2019).
- Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (2005). "Perceptual organization of tone sequences in the auditory cortex of awake macaques," *Neuron* **48**, 139–148.
- Miller, G. A., and Heise, G. A. (1950). "The trill threshold," *J. Acoust. Soc. Am.* **22**, 637–638.
- Moore, B. C. J., and Gockel, H. (2012). "Properties of auditory stream formation," *Phil. Trans. R. Soc. B* **367**, 919–931.
- Pressnitzer, D., and Hupé, J. M. (2006). "Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization," *Curr. Biol.* **16**, 1351–1357.
- Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). "Perceptual organization of sound begins in the auditory periphery," *Curr. Biol.* **18**, 1124–1128.
- R Core Team (2020). "R: A language and environment for statistical computing" (R Foundation for Statistical Computing, Vienna, Austria), <http://www.r-project.org/> (Last viewed March 19, 2021).
- Rajasingam, S. L. (2016). "The dynamics of auditory stream segregation for tone sequences with gradually and abruptly varying stimulus properties," Ph.D. thesis, Aston University, Birmingham, UK.
- Rajasingam, S. L., Summers, R. J., and Roberts, B. (2018). "Stream biasing by different induction sequences: Evaluating stream capture as an account of the segregation-promoting effects of constant-frequency inducers," *J. Acoust. Soc. Am.* **144**, 3409–3420.
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," *J. Acoust. Soc. Am.* **112**, 2074–2085.
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2008). "Effects of the build-up and resetting of auditory stream segregation on temporal discrimination," *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 992–1006.
- Rogers, W. L., and Bregman, A. S. (1993). "An experimental evaluation of three theories of auditory stream segregation," *Percept. Psychophys.* **53**, 179–189.
- Rogers, W. L., and Bregman, A. S. (1998). "Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness," *Percept. Psychophys.* **60**, 1216–1227.
- Rose, M. M., and Moore, B. C. J. (2000). "Effects of frequency and level on auditory stream segregation," *J. Acoust. Soc. Am.* **108**, 1209–1214.
- Scharf, B. (1970). "Critical bands," in *Foundations of Modern Auditory Theory*, Vol. I, edited by J. V. Tobias (Academic, New York), pp. 159–202.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods*, 6th ed. (Iowa State University, Ames, IA), pp. 1–310.
- Snyder, J. S., Carter, O. L., Hannon, E. E., and Alain, C. (2009). "Adaptation reveals multiple levels of representation in auditory stream segregation," *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 1232–1244.
- Snyder, J. S., Carter, O. L., Lee, S. K., Hannon, E. E., and Alain, C. (2008). "Effects of context on auditory stream segregation," *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 1007–1016.
- Thompson, S. K., Carlyon, R. P., and Cusack, R. (2011). "An objective measurement of the build-up of auditory streaming and of its modulation by attention," *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 1253–1262.
- van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Eindhoven University of Technology, Eindhoven, the Netherlands.