

JULY 18 2023

# Asymmetric effects of sudden changes in timbre on auditory stream segregation

Brian Roberts  ; Nicholas R. Haywood



*J Acoust Soc Am* 154, 363–378 (2023)

<https://doi.org/10.1121/10.0020172>



View  
Online



Export  
Citation

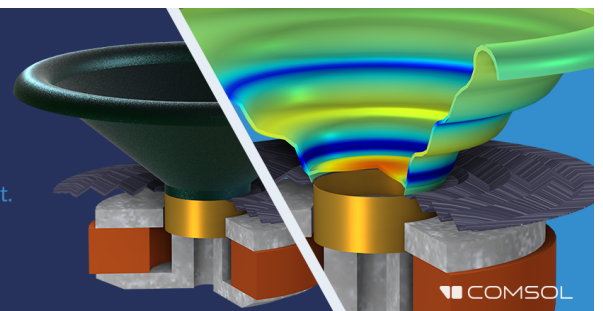
CrossMark

19 July 2023 07:35:38

## Take the Lead in Acoustics


The ability to account for coupled physics phenomena lets you predict, optimize, and virtually test a design under real-world conditions – even before a first prototype is built.

» Learn more about COMSOL Multiphysics®



COMSOL

## Asymmetric effects of sudden changes in timbre on auditory stream segregation

Brian Roberts<sup>1,a)</sup>  and Nicholas R. Haywood<sup>2</sup>

<sup>1</sup>*School of Psychology, Aston University, Birmingham B4 7ET, United Kingdom*

<sup>2</sup>*Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0SZ, United Kingdom*

### ABSTRACT:

Two experiments explored the effects of abrupt transitions in timbral properties [amplitude modulation (AM), pure tones vs narrow-band noises, and attack/decay envelope] on streaming. Listeners reported continuously the number of streams heard during 18-s-long alternating low- and high-frequency (LHL-) sequences (frequency separation: 2–6 semitones) that underwent a coherent transition at 6 s or remained unchanged. In experiment 1, triplets comprised unmodulated pure tones or 100%-depth AM was created using narrowly spaced tone pairs (dyads: 30- or 50-Hz modulation). In experiment 2, triplets comprised narrow-band noises, dyads, or pure tones with quasi-trapezoidal envelopes (10/80/10 ms), fast attacks and slow decays (10/90 ms), or vice versa (90/10 ms). Abrupt transitions led to direction-dependent changes in stream segregation. Transitions from modulated to unmodulated (or slower-modulated) tones, from noise bands to pure tones, or from slow- to fast-attack tones typically caused substantial loss of segregation (resetting), whereas transitions in the opposite direction mostly caused less or no resetting. Furthermore, for the smallest frequency separation, transitions in the latter direction usually led to increased segregation (overshoot). Overall, the results are reminiscent of the perceptual asymmetries found in auditory search for targets with or without a salient additional feature (or greater activation of that feature).

© 2023 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/10.1121/10.0020172>

(Received 8 February 2023; revised 2 June 2023; accepted 28 June 2023; published online 18 July 2023)

[Editor: Jonas Braasch]

Pages: 363–378

### I. INTRODUCTION

The sound reaching our ears is often a mixture arising from more than one source in the environment, and so the auditory system is commonly faced with the task of inferring which parts of the incoming stimulation come from a given source. The perceptual process of grouping together those acoustic elements assumed to have a common origin and separating them from other elements is known as auditory scene analysis (Bregman, 1990). An important aspect of solving this scene analysis problem concerns how sound elements are grouped and separated over time, a process called auditory stream formation and segregation (Bregman and Campbell, 1971). Most studies of stream segregation have used rapid sequences of sounds that alternate between two subsets (A and B). Sequences of this kind can be heard either as one stream of sounds moving back and forth in whatever perceptual property distinguishes them (e.g., pitch) or as two independent streams, each containing only the elements from one or the other subset. The organization heard from moment-to-moment is bistable, with spontaneous switches occurring between integrated and segregated percepts (e.g., Pressnitzer and Hupé, 2006), but the proportion of time for which stream segregation is heard varies with sequence properties. For sequences whose acoustic and

rhythmic properties remain constant throughout, the tendency for stream segregation also builds up over many seconds (e.g., Bregman, 1978; Anstis and Saida, 1985). The aim of the experiments reported here is to extend our understanding of the impact of sudden change on prior accumulated build-up, with a focus on acoustic properties for which the direction of change has asymmetric effects on subsequent streaming (Rogers and Bregman, 1998; Rajasingam *et al.*, 2021).

For a given rate of presentation, the overall likelihood of stream segregation depends on the perceived similarity of the A and B subsets (Moore and Gockel, 2002, 2012). The more dissimilar the sounds are perceived to be, across a range of acoustic properties, the more likely they are to be segregated into separate streams. The extent to which different acoustic properties determine the likelihood of stream segregation has usually been explored by manipulating the magnitude of the difference between the A and B subsets on one or more acoustic dimensions and measuring its impact on streaming. Studies of this kind have shown that introducing peripheral channeling cues—differences between subsets in which auditory filters are activated and to what extent (Hartmann and Johnson, 1991)—promotes stream segregation. Indeed, most streaming experiments have used sequences in which the A and B subsets are pure tones that have different frequencies or occupy different frequency ranges (e.g., Bregman and Campbell, 1971; van Noorden, 1975), or

<sup>a)</sup>Electronic mail: b.roberts@aston.ac.uk

are complex tones evoking different excitation patterns through differences in spectral shape or in the fundamental frequency of their resolved harmonics (e.g., Bregman *et al.*, 1990). Differences of this kind provide the most robust cues for stream segregation. Nonetheless, salient differences in other acoustic properties that produce small or negligible peripheral channeling cues can increase, or even induce, stream segregation. These include differences between the A and B subsets in their attack and decay characteristics (e.g., Singh and Bregman, 1997), envelope type (steady pure tones vs fluctuating narrow-band noises; e.g., Cusack and Roberts, 2000), or their pitch value and strength for a set of unresolved harmonics with a common passband (e.g., Vliegen *et al.*, 1999; Roberts *et al.*, 2002).

The time course of build-up is very different from that of the strong segregation-promoting effect of preceding a sequence of alternating-frequency tones with either the A or B subset alone, which occurs much more rapidly (Rogers and Bregman, 1993; Roberts *et al.*, 2008; Haywood and Roberts, 2013) and appears to have a different origin (see, e.g., Rajasingam *et al.*, 2018). Build-up for alternating sequences decays over a few seconds if the sequence is interrupted by silence (Bregman, 1978; Beauvois and Meddis, 1997). Accounts of this slow build-up have been offered based on its proposed functional role as a conservative process of evidence accumulation that two different sources of sound are active rather than one (Bregman, 1978) or on its proposed physiological basis—e.g., slow adaptation leading to a narrowing of the receptive fields of frequency-tuned neurons in auditory cortex (Micheyl *et al.*, 2005; see also Rankin *et al.*, 2017).

Outside the realm of fire and car alarms, it is unusual to hear long sequences of unchanging sounds. Nonetheless, the impact on subsequent stream segregation of sudden coherent changes to an ongoing sequence—i.e., the imposition of the same change on the properties of the A and B subsets—has so far received relatively little attention. Few acoustic properties have been explored in this context, but the results to date reveal how little is known about the dynamics of stream segregation and its functional role in everyday listening. Most experiments on the effects of abrupt transitions on alternating sequences of lower- (L) and higher- (H) frequency pure tones have examined the impact of changing the center frequency of the sequence while preserving the frequency separation of the subsets (Anstis and Saida, 1985; Rajasingam *et al.*, 2021) or of changes in lateralization of the subsets as manipulated via ear of presentation, interaural level difference, or interaural time difference cues (Rogers and Bregman, 1998). The result of sudden changes in these properties is a rapid loss of the build-up accumulated prior to the transition, either partial or complete. Elucidating whether this loss results from a passive failure of build-up to transfer to the different properties of the post-transition sounds or from an active resetting process can be hard to determine (see, e.g., Haywood and Roberts, 2010), but whatever its cause this loss is usually referred to as resetting.

For sudden changes in center frequency or lateralization, the effect of the direction of the transition on subsequent segregation is broadly symmetrical. There is, however, one property for which a clear asymmetry has long been established—the effect of transitions in level (Rogers and Bregman, 1998; Rajasingam *et al.*, 2021). Sudden increases in level (softer-to-louder transitions) lead to partial resetting but decreases (louder-to-softer transitions) have little or no effect. Rogers and Bregman (1998) proposed that this asymmetry occurred because abrupt increases in level can indicate that a new source has become active, encouraging a one-stream interpretation from which evidence for segregation must build-up anew, whereas abrupt decreases do not.

It has recently been shown that abrupt changes in another acoustic property—the presence or absence of amplitude modulation (AM)—can lead to an even more striking asymmetry (Rajasingam *et al.*, 2021, experiment 3). Specifically, changes from amplitude-modulated to steady (unmodulated) pure tones led to substantial resetting but changes from unmodulated to modulated tones led to little or no resetting and sometimes, particularly for the smallest high-low frequency difference tested [ $\Delta f = 4$  semitones (ST)], to *increased* segregation compared with a sequence that was modulated from the start. To our knowledge, such an *overshoot* effect had not previously been reported. There is also no obvious equivalent of the functional account offered by Rogers and Bregman (1998) for the asymmetry observed for sudden transitions in level—why should a change from modulated to unmodulated sounds be interpreted as a new event when a change in the opposite direction is not? Other contexts in which a similar asymmetry has been observed include auditory search, in which a modulated target is easier to detect among unmodulated distractors than the reverse (e.g., Asemi *et al.*, 2003; Cusack and Carlyon, 2003), modulation detection interference, in which the detection or discrimination of modulation on one carrier is impaired by the presence of another modulated carrier but is largely unaffected when the additional carrier is unmodulated (e.g., Yost and Sheft, 1989; Wilson *et al.*, 1990; Carlyon, 2000), and masking, in which it is easier to detect a fluctuating noise added to a steady pure or complex tone than vice versa (e.g., Hellman, 1972; Gockel *et al.*, 2002).

An advantage of the long sequences (20 s) and continuous reporting of perception used by Rajasingam *et al.* (2021) is that the evolution of listeners' responses to a sudden transition could be observed (cf. the short test sequences and one judgment per trial method used by Rogers and Bregman, 1998). However, it was evident from the results that the inclusion of three abrupt transitions over the course of the tone sequence partly curtailed the measurement of the unfolding changes in the perceptual organization following each transition. Moreover, as acknowledged by Rajasingam *et al.* (2021), the most appropriate reference condition was obvious only for the first transition, because accumulated build-up at later transitions would already have been influenced by earlier ones. This proved problematic when

attempting to establish whether any increase in stream segregation observed after a later transition represented significant overshoot. Also, the  $\Delta f$  range used (4–8 ST) meant that there was limited scope for overshoot effects to be revealed for the larger values tested, owing to the extent to which segregation would already have built up before the transition, leading to ceiling effects.

The two experiments reported here adapted the approach taken by [Rajasingam et al. \(2021\)](#) by using stimuli with only one transition in a long sequence and a less segregation-promoting  $\Delta f$  range (2–6 ST). These experiments extend the investigation of the directional effects of transitions on judgments of the number of streams heard by introducing sudden changes in a wider range of acoustic properties. These properties were selected to allow salient changes in timbre without introducing substantial peripheral-channeling cues and included transitions in AM rate, envelope type (steady vs fluctuating), and envelope shape (attack/decay). All these properties have been examined previously in the context of differences between the L and H subsets of sound sequences but not, to our knowledge, in the context of sudden coherent changes in both subsets. Implications of the results for our understanding of the dynamics of stream segregation are considered.

## II. EXPERIMENT 1

This experiment used the adapted methodology outlined above to extend the exploration of the effects on subsequent streaming of sudden transitions between modulated and unmodulated tones ([Rajasingam et al., 2021](#)). In that study, the AM rate of the modulated tones was always 50 Hz, whereas here AM rates of 30 and 50 Hz were used. This allowed the creation not only of sequences involving transitions between unmodulated and modulated tones but also sequences involving transitions between tones with different AM rates. There is some evidence that differences in AM rate between the A and B subsets can initiate or increase stream segregation. For example, [Grimault et al. \(2002\)](#) used sequences of ABA– triplets of broadband noise for which the AM rate of the A subset was fixed at 100 Hz and the AM rate of the B subset was varied across sequences from 100 to 800 Hz. They found that two streams were almost always heard when the AM rate of the B subset was more than double that of the A subset, indicating segregation based on differences in AM rate. Similar findings have since been reported by [Dolležal et al. \(2012\)](#), using ABA– triplets of sinusoidal carriers, and by [Nie and Nelson \(2015\)](#), using alternating pairs of bandpass-filtered noise (A and B). [Dolležal et al. \(2012\)](#) demonstrated that the effects of differences in AM rate on stream segregation arose primarily from temporal cues. To our knowledge, however, the effects on subsequent streaming of sudden coherent changes in AM rate applied simultaneously to both A and B subsets have not previously been investigated.

Exploring the effects of transitions between tones with different AM rates is important because [Rajasingam et al.](#)

(2021) proposed that the asymmetries they observed may have a similar origin to those seen in perceptual search tasks. In particular, the time needed to detect a modulated-tone target among pure-tone distractors is largely unaffected by the number of distractors, indicating “pop out,” whereas the time needed to detect a pure-tone target among modulated distractors increases with the number of distractors (e.g., [Asemi et al., 2003](#); [Cusack and Carlyon, 2003](#)). This account of the asymmetry in perceptual search relies on whether the target has a salient additional “feature” not present in the distractors—in this case, modulation. By analogy, the directional effect of a transition between unmodulated and modulated tones depends on whether a new salient feature is present after the transition. Given that a transition between different rates of AM does not involve the introduction or removal of an additional feature, exploring the effects of these transitions on streaming can be used to assess the proposal of [Rajasingam et al. \(2021\)](#).

## A. Method

### 1. Listeners

Listeners, mostly students, were recruited from Macquarie University and the University of Cambridge. All gave informed consent and received payment for taking part. They were first tested using a screening audiometer (Macquarie: AS208, Cambridge: Infinity 2.0; Interacoustics, Middelfart, Denmark) to ensure that their audiometric thresholds at 0.5, 1, 2, and 4 kHz did not exceed 20 dB hearing level. All listeners passed this screening. They next 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, ten from Macquarie) successfully completed the experiment (mean age = 21.8 years, range = 19–28). This research was approved by the Macquarie University Human Research Ethics Committee (Reference No. 5201700786) and the Cambridge Psychology Research Ethics Committee (Application No. PRE.2019.093).

### 2. Stimuli and conditions

Each stimulus sequence was 18 s long (45 LHL– triplets) and comprised an induction sequence (6 s, 15 triplets) followed by a longer test sequence (12 s, 30 triplets). The acoustic properties of the two sequences were identical in the reference conditions, leading to a seamless whole, but were different in the experimental conditions, leading to a sudden transition at 6 s. The test sequences were sufficiently long to explore the unfolding consequences of an abrupt change in tone properties at the induction-test boundary. Each constituent tone and the silence at the end of each triplet was 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 center frequency of the L tones was kept constant at 1 kHz and that of the H tones was set according to the desired value of  $\Delta f$ , which was 2, 4, or 6 ST. The center frequency of the H tones for these  $\Delta f$ s was 1125, 1259, and 1414 Hz, respectively. This range was chosen to enhance the likelihood of demonstrating increases in segregation (overshoot), as well as decreases (resetting), following a sudden transition in properties.

Two types of tone were used—pure tones and narrowly spaced pairs of pure tones known as dyads—which allowed sequences to be constructed involving abrupt transitions in the presence or rate of AM, and hence in timbre, between the induction and test sequences without introducing appreciable excitation-pattern cues. Pure tones were shaped using 10-ms raised-cosine ramps. Tone dyads were constructed by adding two pure tones of equal level and centered (either  $\pm 15$  or  $\pm 25$  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; their addition with 30-Hz or 50-Hz separation gave exactly 3 cycles (slower-modulation dyads) of full-depth AM or 5 cycles (faster-modulation dyads), respectively, over 100 ms. No further envelope shaping was required for the dyads; the rising half of the first cycle of modulation and the falling half of the last cycle acted as onset and offset ramps. These half cycles corresponded to ramp durations of 10.0 ms and 16.7 ms, respectively, for the faster- and slower-modulation stimuli. Given that the center frequency of the L-tone dyads was 1000 Hz (H tones = 2, 4, or 6 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. The strong sinusoidal AM of the dyads resulting from the interaction of the two components within the same auditory filter gave them a distinctive timbre and the difference in AM rate (30 Hz or 50 Hz) for the two types of dyad (3 or 5 cycles) was easily discernible.<sup>1</sup>

There were nine conditions; three were reference cases (C1–C3) and six were experimental (C4–C9). In C1–C3, all triplets throughout were composed of pure tones (PT), slower-modulation dyads (SD), or faster-modulation dyads (FD), respectively. All permutations of the three types of tone were represented in the experimental conditions; the composition of the induction and test sequences, respectively, was PT and SD (C4), SD and PT (C5), PT and FD (C6), FD and PT (C7), SD and FD (C8), and FD and SD (C9). Stimuli were synthesized in MATLAB 2018 b (Mathworks, Natick, MA) at a sampling rate of 48 kHz; sequences for each combination of condition and  $\Delta f$  were pre-assembled, stored as WAV files, and played back at 16-bit resolution. Diotic presentation was used throughout, and all tones were presented at 70 dB sound pressure level (SPL). At Macquarie University, sounds were presented

over Sennheiser 380 Pro headphones (Hannover, Germany) via an Audio Express sound card (MOTU, Cambridge, MA). Output levels were calibrated using a type 2250 sound-level meter (Brüel and Kjaer, Naerum, Denmark) and an RA0045 microphone (GRAS, Holte, Denmark) coupled to the earphones with a type 43AG ear simulator (GRAS). At the University of Cambridge, sounds were presented over Sennheiser HD-600 headphones via an RME Fireface UCX audio system (Haimhausen, Germany). Output levels were calibrated with reference to a 1-kHz full-scale signal and headphone sensitivity data using an MDO3024 oscilloscope (Tektronix, Beaverton, OR).

### 3. Procedure

At both sites, listeners completed the experiment in a custom-built double-walled sound-attenuating chamber (IAC Acoustics, Naperville, IL). They were free to take breaks between blocks of trials if desired. Completing all stages (audiometry, training, and main experiment) usually took  $\sim 2\frac{1}{2}$  hours, divided into two sessions. The experiment was run using a program written in Python 3.8 (Peirce *et al.*, 2019) using the PsychoPy toolbox (version 2020.2.3), a software package designed for precise stimulus presentation and key press recording (see Bridges *et al.*, 2020). Each trial was initiated 0.5 s after the listener pressed the space bar on the keyboard. Listeners were instructed to monitor the 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 “Q” or “P” keys, respectively (the meaning of these keys was explained during training and supported by on-screen visual feedback). Thereafter, listeners were asked to press the appropriate key every time their perception of the 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 impression (Haywood and Roberts, 2013; Rajasingam *et al.*, 2018, 2021). 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 (0.5 s), this ensured a minimum silent gap of 5.5 s during which any prior build-up would decay before the start of the next trial (Bregman, 1978).

Each combination of condition (nine levels) and  $\Delta f$  (three levels) was presented ten times in the main experiment, once in each block, giving 270 trials. Stimuli were presented in a newly randomized order in each block for each listener. Training consisted of completing a single block of trials just prior to starting the main experiment; the same approach was used as practice to begin the second session. Using three different  $\Delta f$ s 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 reference conditions (C1–C3) had to rise when  $\Delta f$  was increased from 2 ST to 4 ST and rise again when  $\Delta f$  was increased from 4 ST to 6 ST. Two listeners were excluded and replaced in this experiment.

#### 4. Data analysis and availability

Once a response key was first pressed on a given trial, the associated value (one stream or two) was held until the other key was pressed or until the end of that trial. Response data from each trial were divided into eighteen 1-s-long time bins (i.e., 0–1 s, 1–2 s, ..., 17–18 s).<sup>2</sup> 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 a given 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 (Haywood and Roberts, 2013; Rajasingam *et al.*, 2018, 2021).

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. Following the approach of our previous studies, on occasions when one of these means was missing for a particular listener (ten cases, corresponding to  $\sim 0.2\%$  of all time bins, and all occurring within the 1–2 s time bin), the missing value was replaced with the mean 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 sequence was heard as segregated for each time bin. This measure of the average time course of stream segregation over the sequence is used to display the results.

The effects of stimulus type *per se* were explored by comparing the extent of stream segregation across the reference conditions for the full duration of the sequences (in 1-s time bins, excluding 0–1 s). The time-series data obtained from the calculations described above were analyzed using repeated-measures analysis of variance (ANOVA); departures from sphericity were addressed using the Greenhouse-Geisser correction. The measure of effect size reported here is partial eta squared ( $\eta_p^2$ ). Comparisons among conditions were conducted using three factors—frequency separation between the L and H sounds ( $\Delta f$ ), stimulus type (S), and time interval (T, with levels corresponding to time bins 1–2 s to 17–18 s). All ANOVAs reported here were computed using SPSS (SPSS statistics version 21, IBM Corp., Armonk, NY).

The effects of different abrupt transitions in acoustic properties at 6 s were explored by comparing the extent of

stream segregation in the test sequence for each experimental condition with that for the appropriate reference case (i.e., the same stimulus type). Two approaches were used to make these comparisons. The main approach was to estimate the overall impact of different transitions on judgments of the test sequence. To do this, segregation scores were collapsed across time bins 7–8 s to 14–15 s to create a single 8-s-long interval; the last three time bins were omitted because any effects of the transition had mostly been lost by then and time bin 6–7 s was omitted owing to the usual short delay in the initial response to the transition at 6 s. Comparisons among the corresponding difference scores for this interval were conducted using a repeated-measures ANOVA with three factors— $\Delta f$ , stimulus pair (P), and direction of change (D)—where stimulus pair specifies the stimulus types defining the transition and direction specifies the order of those two types (i.e., which was the inducer and which the test sequence). The latter factor was organized according to which direction of change for a given pair was predicted to be more integration-promoting for subsequent streaming judgments (i.e., to cause more resetting); these predictions were based on a combination of previous research (Rajasingam *et al.*, 2021), pilot observations, and theoretical principles. The effects of individual transitions at different  $\Delta f$ s were explored further with planned pairwise comparisons (two-tailed) conducted using the restricted least-significant-difference test (Snedecor and Cochran, 1967; Keppel and Wickens, 2004).

In the second approach, comparisons were made between the experimental and reference cases for successive 1-s time bins throughout the sequence, using a bootstrapping analysis in MATLAB 2018 b to explore the evolution of the impact of the transition on subsequent judgments of streaming. This analysis estimated for each time bin the probability that the measured difference in segregation between the experimental condition and the relevant reference case was significantly different from zero. Note that the reference case used for comparison changed at 6 s, to reflect the transition in the experimental condition. The data were resampled 5000 times to ensure a stable estimate of the confidence intervals, and the Holm-Bonferroni method was used to correct the significance level for multiple comparisons. The research data and stimuli underlying this publication are available online from a repository hosted by Aston University.<sup>3</sup>

## B. Results and discussion

The results averaged across listeners are shown in Fig. 1 (see supplementary material for alternative versions<sup>4</sup>). The top panels display the results for the reference (no transition) conditions, in which stimulus type remained constant throughout the 18-s sequence; the remaining panels display the results for the experimental (transition) conditions, in which there was an abrupt change in stimulus properties 6 s after the sequence began. The upper middle, lower middle, and bottom panels show the effects of abrupt transitions (in both directions) between stimulus types SD and PT, FD and

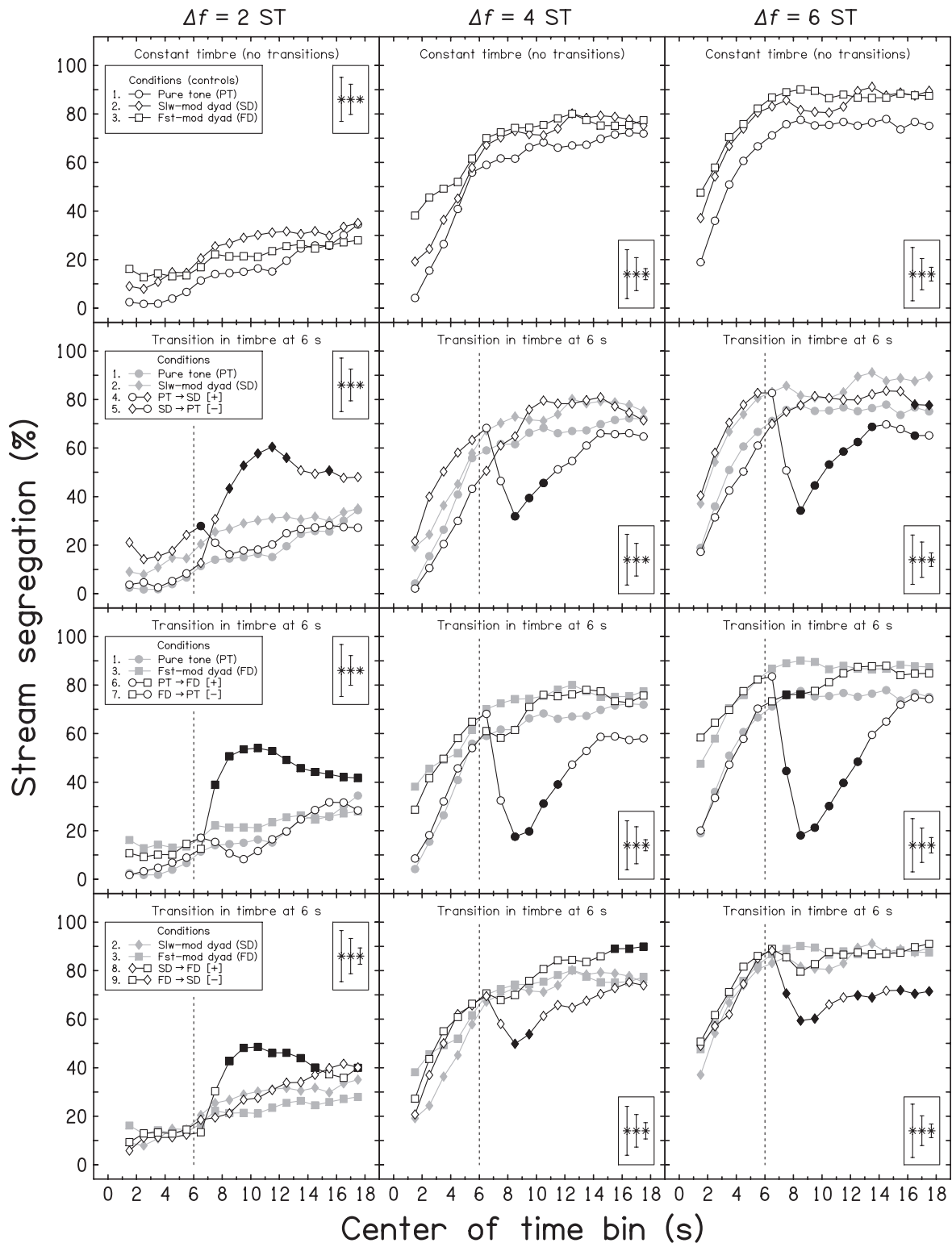


FIG. 1. Results for experiment 1—effects of condition (reference = 1–3, experimental = 4–9; see insets in left panels) and frequency separation ( $\Delta f$ ) on the extent and time course of reported stream segregation ( $n = 12$ ). Responses for each trial were divided into 1-s time bins; the time shown on the abscissa indicates the center of the corresponding time bin. Results for each  $\Delta f$  (2, 4, and 6 ST) are displayed in separate panels from left to right. Results for the reference conditions (no transition) are shown in the top panels; results for the conditions involving transitions between stimulus types PT and SD, PT and FD, or SD and FD are displayed in the upper middle, lower middle, and bottom panels, respectively. Black-filled symbols indicate time bins for which the bootstrapping analysis showed a significant difference from the corresponding reference case. For ease of comparison, the results for the appropriate reference conditions are reproduced in gray in the lower panels. For clarity, the means 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).

19 July 2023 07:35:38

PT, and SD and FD, respectively. For each stimulus pair, the transition directions predicted to be less or more integration-promoting are indicated in the insets as [+]  
(higher segregation) and [-] (lower segregation), respectively. These panels also reproduce the results for the corresponding reference cases. Time bins for which the bootstrapping analysis showed a significant difference (see Sec. II A 4) from the relevant reference case are indicated by filled black symbols; there were no cases in which this happened prior to the transition. Results for the reference and experimental conditions are considered in turn.

Inspection of the top panels of Fig. 1 shows clear evidence that stream segregation built up over time in the reference conditions, and was greater in rate of increase and final extent for larger  $\Delta f$ s. The observed patterns are in accord with research suggesting a faster initial phase of build-up over the first 10 s of a sequence followed by a slower phase thereafter (e.g., Anstis and Saida, 1985; Haywood and Roberts, 2013). There is also some indication that the dyad reference conditions (C2 and C3) were associated with more stream segregation than the pure-tone condition (C1) earlier on in the sequence, but that the difference between them tended to decline over time; a similar pattern was observed by Rajasingam *et al.* (2021). No consistent effect of AM rate was apparent. The ANOVA for the reference conditions is presented in Table I. Two of the three factors influenced streaming as highly significant main effects—stream segregation was greater for larger frequency separations (means: 2 ST = 20.2%, 4 ST = 62.3%, and 6 ST = 75.9%;  $p < 0.001$ ) and tended to increase over time ( $p < 0.001$ ). There was no main effect of stimulus type ( $p = 0.148$ ), but the stimulus type  $\times$  time interval interaction was significant ( $p = 0.034$ ), which is in accord with the difference between the response profiles for dyad and pure-tone sequences described above. The initial tendency towards greater segregation for the dyad sequences compared with otherwise-matched pure-tone sequences may have been a consequence of the modulated L and H tones sharing a common AM rate rather than a common relative bandwidth—the former is associated with more stream segregation than the latter for sequences of this kind (Cusack and Roberts, 1999) and so the

TABLE I. Results for experiment 1—effects of stimulus type (no-transition reference cases). Summary of the three-way repeated-measures ANOVA for the segregation scores from all time bins (1–2 s to 17–18 s) across conditions C1–C3. Factor  $\Delta f$  has 3 levels (2, 4, and 6 ST) and stimulus type has 3 levels—PT, SD, and FD. All significant terms are shown in bold.

Factor	$df$	$F$	$p$	$\eta_p^2$
<b>Frequency separation in sequence (<math>\Delta f</math>)</b>	<b>(1.069, 11.761)</b>	<b>54.056</b>	<b>&lt;0.001</b>	<b>0.831</b>
Stimulus type (S)	(1.419, 15.606)	2.252	0.148	0.170
<b>Time interval (T)</b>	<b>(1.692, 18.614)</b>	<b>35.105</b>	<b>&lt;0.001</b>	<b>0.761</b>
$\Delta f \times S$	(2.339, 25.731)	0.791	0.482	0.067
$\Delta f \times T$	<b>(5.014, 55.153)</b>	<b>9.166</b>	<b>&lt;0.001</b>	<b>0.455</b>
$S \times T$	<b>(5.262, 57.883)</b>	<b>2.571</b>	<b>0.034</b>	<b>0.189</b>
$\Delta f \times S \times T$	(4.880, 53.682)	0.608	0.690	0.052

difference in relative bandwidth may have acted as an additional segregation cue. The only other significant interaction term was  $\Delta f \times$  time interval ( $p < 0.001$ ), which arose mainly from the greater rate of change of stream segregation for larger  $\Delta f$ s during the initial phase of build-up.

Inspection of the other panels of Fig. 1 shows that sudden changes from modulated to unmodulated tones, or tones modulated at a slower rate, typically led to decreased subsequent segregation (i.e., to resetting of build-up), particularly for the larger values of  $\Delta f$ . Sudden changes in the opposite direction were less integration-promoting—the transition typically had little effect on subsequent streaming for the larger  $\Delta f$ s, and for a 2-ST  $\Delta f$  there was clear evidence of a tendency for stream segregation to *increase*, rather than to fall, following the transition (i.e., overshoot). The effect on streaming was typically greatest after  $\sim 3$  s for transitions causing resetting, but not until after  $\sim 5$  s for transitions causing overshoot. The time constant for decay of the effect of transition was considerably longer, but that effect was usually diminished, and sometimes lost altogether, by the last few seconds of the test sequence. Most of the 1-s time bins in the transition conditions identified in the bootstrapping analysis as significantly different from their references align with the resetting and overshoot effects described above, but the 7–8 s and 8–9 s time bins following the PT-to-FD transition in the 6-ST case are exceptions. Those cases can be accounted for in terms of the higher segregation for FD stimuli established prior to the transition, with which the sequence changing from PT to FD needs a few seconds to catch up. The last two time bins following the PT-to-SD transition in the 6-ST case are an exception without obvious explanation.

The ANOVA exploring the effects of abrupt changes in stimulus type is presented in Table II. The main effects of frequency separation and direction of change were highly significant ( $p < 0.001$ ), and the main effect of the stimulus pair approached but did not quite reach significance ( $p = 0.061$ ). None of the interaction terms were significant. The effects of individual transitions were explored further using pairwise

TABLE II. Results for experiment 1—effects of abrupt changes in stimulus properties. Summary of the three-way repeated-measures ANOVA for scores corresponding to the difference in stream segregation between experimental (transition) and reference (no-transition) cases over a single 8.0-s time interval (collapsed across time bins 7–8 s to 14–15 s). Factor  $\Delta f$  has 3 levels (2, 4, and 6 ST), stimulus pair has 3 levels (stimulus pairs: SD & PT, FD & PT, and FD & SD), and direction of change has 2 levels (forward and reverse within-pair order, corresponding to transitions predicted to be more and less integration-promoting, respectively). All significant terms are shown in bold.

Factor	$df$	$F$	$p$	$\eta_p^2$
<b>Frequency separation in sequence (<math>\Delta f</math>)</b>	<b>(1.964, 21.605)</b>	<b>33.651</b>	<b>&lt;0.001</b>	<b>0.754</b>
Stimulus pair (P)	(1.499, 16.491)	3.614	0.061	0.247
<b>Direction of change (D)</b>	<b>(1.000, 11.000)</b>	<b>28.770</b>	<b>&lt;0.001</b>	<b>0.723</b>
$\Delta f \times P$	(3.111, 34.226)	1.804	0.163	0.141
$\Delta f \times D$	(1.547, 17.012)	0.068	0.892	0.006
$P \times D$	(1.472, 16.191)	1.820	0.197	0.142
$\Delta f \times P \times D$	(2.243, 24.671)	0.121	0.907	0.011



TABLE III. Results for experiment 1—effects of abrupt changes in stimulus properties. Parts (a) and (b) show a summary of pairwise comparisons (t tests) exploring the change in stream segregation following more and less integration-promoting transitions, respectively, relative to their reference cases, for each transition type and  $\Delta f$ . All significant cases are shown in bold.

Part (a): Results for more integration-promoting transitions (differences re reference cases)			
Transition type	$\Delta f = 2$ ST (mean, $p$ )	$\Delta f = 4$ ST (mean, $p$ )	$\Delta f = 6$ ST (mean, $p$ )
SD to PT	+3.4% pts, 0.569	<b>-16.5% pts, 0.035</b>	<b>-21.0% pts, &lt;0.001</b>
FD to PT	-1.2% pts, 0.810	<b>-28.7% pts, &lt;0.001</b>	<b>-35.5% pts, &lt;0.001</b>
FD to SD	-0.7% pts, 0.929	<b>-13.3% pts, 0.026</b>	<b>-18.0% pts, 0.008</b>
Part (b): Results for less integration-promoting transitions (differences re reference cases)			
Transition type	$\Delta f = 2$ ST (mean, $p$ )	$\Delta f = 4$ ST (mean, $p$ )	$\Delta f = 6$ ST (mean, $p$ )
PT to SD	<b>+20.6% pts, 0.003</b>	+0.1% pts, 0.992	-5.0% pts, 0.177
PT to FD	<b>+25.4% pts, &lt;0.001</b>	-4.1% pts, 0.552	-5.5% pts, 0.081
SD to FD	<b>+20.0% pts, 0.010</b>	+3.2% pts, 0.650	-2.4% pts, 0.466

comparisons that are summarized in Table III (parts a and b) for all 18 combinations of transition type (i.e., two directions per stimulus pair) and  $\Delta f$ . The mean values quoted correspond to the change in segregation scores, in percentage points (% pts), produced by the transition case relative to its reference over the selected 8.0-s time interval. Figure 2 displays these means and corresponding inter-subject standard errors across  $\Delta f$  for each stimulus pair in a separate panel; open and filled symbols indicate results for the transition directions predicted to be less and more integration-promoting, respectively. The top axis summarizes the corresponding mean difference scores between the two directions.

All transitions predicted to be more integration-promoting (part a) led to substantial and significant resetting for 4-ST and 6-ST  $\Delta f$ s (range: -13.3% pts to -35.5% pts,

$p = 0.035 - p < 0.001$ ), and all transitions predicted to be less integration-promoting (part b) led to substantial and significant overshoot for the 2-ST  $\Delta f$  (range: +20.0% pts to +25.4% pts,  $p = 0.010 - p < 0.001$ ). The remaining cases shown in parts a and b were associated with only small and non-significant effects. Importantly, sudden changes in the rate of AM produced directional effects like those observed for transitions between unmodulated and modulated tones. When collapsed across  $\Delta f$  and stimulus pair, the transitions predicted to be more integration-promoting were associated on average with 20.4% pts less segregation than their counterparts. Transitions between PT and FD stimuli were associated with the largest mean difference scores. The overall pattern of response profiles suggests a summation of the effects of three factors—one based on the direction of the transition (with greater resetting following transitions from more to less modulated tones), one based on  $\Delta f$  (less resetting is possible for small  $\Delta f$ s) and one that biases subsequent streaming judgments towards segregation. This proposal is considered further in Sec. IV.

Two other factors that might have affected the results merit brief comment. The tone dyads may have generated audible distortion products (e.g., Goldstein, 1967) that introduced greater than anticipated differences in peripheral channeling between the dyads and corresponding pure tones. However, as noted by Rajasingam *et al.* (2021), there is no obvious mechanism by which these distortion products might account for the strong directional effects observed for transitions between these stimulus types. Also, the loudness of corresponding modulated and unmodulated tones with the same rms power may not have been identical (Moore *et al.*, 1999). However, any difference between them would have been small compared with that resulting from the 12-dB transitions in level used previously (Rogers and Bregman, 1998; Rajasingam *et al.*, 2021), and so seems unlikely to account for the pronounced asymmetries on streaming judgments observed here.

One aspect of the current results that differs somewhat from the nearest equivalent data (Rajasingam *et al.*, 2021, experiment 3) concerns the effect of  $\Delta f$ . That experiment included two conditions in which each tone sequence

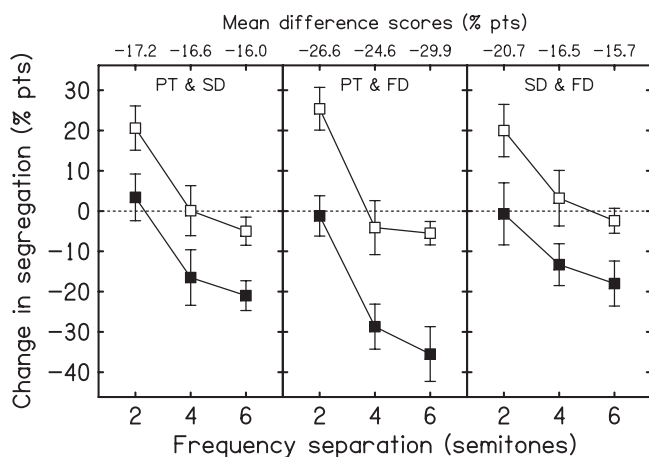


FIG. 2. Results for experiment 1—summary of the effects of abrupt changes in stimulus properties over the selected 8.0-s interval (collapsed across time bins 7–8 s to 14–15 s). Changes in segregation (% pts) are shown for the experimental conditions relative to their corresponding reference cases. Results for each stimulus pair are displayed in separate panels from left to right. Within each panel, the results (means and inter-subject standard errors) are displayed from left to right for each  $\Delta f$  (2, 4, and 6 ST), using open and filled symbols for the transition direction predicted to be less integration-promoting (labels in order shown) and more integration-promoting (labels in reverse order), respectively. The dashed line in each panel indicates the division between overshoot (above) and resetting (below). The top axis shows the mean difference scores between corresponding cases for the two transition directions.

contained three transitions, rather than just one as here, between exactly corresponding FD and PT stimuli. It is not straightforward to identify a suitable reference case against which to assess the effects of second and subsequent transitions, but for one of those conditions, there was an initial transition from PT to FD stimuli 5.2 s after the sequence began, which is only 0.8 s earlier than for the corresponding case in the current study. This first transition was associated with significant overshoot for a  $\Delta f$  of 4 ST (the smallest value tested), whereas overshoot in the experiment reported here required a smaller  $\Delta f$  (2 ST) to be significant, albeit with the greater size anticipated for the inclusion of the 2-ST case. Notwithstanding the other contextual differences between the two experiments, this discrepancy suggests that the extent of the tendency towards more segregated judgments for smaller  $\Delta f$ s may have been influenced by the  $\Delta f$  range to which the listeners were exposed.

The finding that transitions in AM rate produced asymmetric effects on streaming like those seen for transitions between unmodulated and modulated tones casts some doubt on the notion that these asymmetries are akin to those found in perceptual search tasks (e.g., Asemi *et al.*, 2003; Cusack and Carlyon, 2003). This is because it seems highly plausible that a modulated target would stand out among unmodulated distractors, owing to the presence of an additional feature (AM), whereas a change in AM rate does not constitute an additional feature. Nonetheless, a modified version of this account remains possible in which the asymmetry can arise from one set of sounds causing greater activation on a given perceptual dimension, rather than necessarily possessing an extra feature. Cusack and Carlyon (2003) came to this conclusion in the context of their finding that longer sounds were easier to select from short distractors than vice versa in an auditory search task. In the context of the current streaming study, this implies that higher rates of AM lead to greater activation on some perceptual dimensions than do lower rates. Presumably, greater activation on a given dimension equates to higher salience, and hence to easier attentional selection. Counterintuitively, however, applying the same logic to the effects of sudden changes in level on subsequent streaming implies that louder sounds lead to less activation than softer ones. This conundrum is considered further in Sec. IV.

### III. EXPERIMENT 2

Given that sudden changes in level and in the presence or rate of AM are the only changes thus far known to be associated with clear directional effects on subsequent stream segregation, this asymmetry was explored further by extending the range of stimulus types tested. Specifically, we examined the effects of transitions between fluctuating sounds (narrow-band noises) and non-fluctuating sounds (pure tones), and between pure tones with different envelope shapes (damped vs ramped). Transitions between these stimulus types do not involve substantial peripheral-channeling cues. The effects of transitions between PT and FD stimuli

were also measured in the same context to act as a benchmark for comparison with the effects of the other transitions. The inclusion of transitions between narrow-band noises and pure tones was motivated by evidence of a strong asymmetry between these stimuli in perceptual search tasks (narrow-band noises are more salient; Asemi *et al.*, 2003) and pilot observations made with our streaming task. There is also a marked difference in timbre between these stimuli, which is known to facilitate stream segregation when the L and H subsets are distinguished from one another in this way (e.g., Cusack and Roberts, 2000).

The inclusion of transitions between damped and ramped tones was motivated by evidence that the attack duration of a sound is an important determinant of its timbre (e.g., Iverson, 1995) and that differences in attack/decay profile across the L and H subsets of a pure-tone sequence increase their tendency to segregate (Singh and Bregman, 1997). Our initial hypothesis was that the percussive character of damped sounds, with their more definite perceived onset, might act as an additional feature, such that transitions from damped to ramped (i.e., loss of that feature) might be more integration-promoting than vice versa. However, pilot observations indicated the opposite pattern. Perhaps relevant here is the finding of Cusack and Carlyon (2003) for auditory search, noted above, that longer sounds are easier to select from short distractors than vice versa. This is because ramped sounds are known to be perceived as longer than corresponding damped sounds when they have equal intensity and physical duration (e.g., Stecker and Hafer, 2000; Wang *et al.*, 2014).

### A. Method

Except where described, the same method was used as for experiment 1. Twelve listeners (four males, nine from Macquarie, mean age = 23.9 years, range = 18–32) took part and successfully completed the experiment; no listeners were excluded and replaced. Eight listeners had previously completed experiment 1. In this experiment, triplets composed of pure tones or faster-modulation dyads were retained in the same form as for experiment 1, but triplets composed of narrow-band noises (NB) with 100-ms quasi-trapezoidal amplitude envelopes (including  $\pm 10$  ms raised cosine ramps) or pure tones with asymmetric envelopes were also used. The latter sounds had linear ramps and were either “damped” with a fast attack (FA, 10 ms) and slow decay (90 ms) or “ramped” with a slow attack (SA, 90 ms) and fast decay (10 ms). Narrow-band noises were created by the addition of sine waves spaced 1 Hz apart with a random starting phase for each component; each token was generated using a different set of random phases. The equal-amplitude passband was  $\pm 1$  ST relative to the center frequency—e.g., the lower and upper cutoff frequencies for a band with a 1000-Hz center frequency were 943.9 and 1059.5 Hz. To avoid generating edge pitches (von Békésy, 1963), components were also present outside the passband. These components were attenuated by 80 dB/octave and

those below  $-60$  dB were excluded. For each condition involving noise bursts, a hundred unique sequences were generated and stored as WAV files, from which one was selected at random on any given trial with the constraint that no unique sequence could be presented twice to the same listener. All stimuli had the same rms power and presentation level (70 dB SPL).

There were eleven conditions, five of which were reference cases (C1–C5) and six were experimental (C6–C11). In C1–C5, all triplets throughout were composed of stimulus types PT, SA, FA, NB, or FD, respectively. Only a subset of the possible permutations was represented in the experimental conditions; the composition of the induction and test sequences, respectively, was PT and FD (C6), FD and PT (C7), PT and NB (C8), NB and PT (C9), SA and FA (C10), and FA and SA (C11). Each combination of condition (11 levels) and  $\Delta f$  (three levels) was presented ten times in the main experiment, once in each block, giving 330 trials. Listeners usually took  $\sim 3$  h to complete the experiment over two sessions. Time-series data were computed from listeners' responses as for experiment 1. When an individual mean was missing (five cases,  $\sim 0.1\%$  of the data and all within the 1–2 s time bin), the missing value was replaced as before using mean imputation. Again, the results were analyzed using repeated-measures ANOVAs and bootstrapping.

## B. Results and discussion

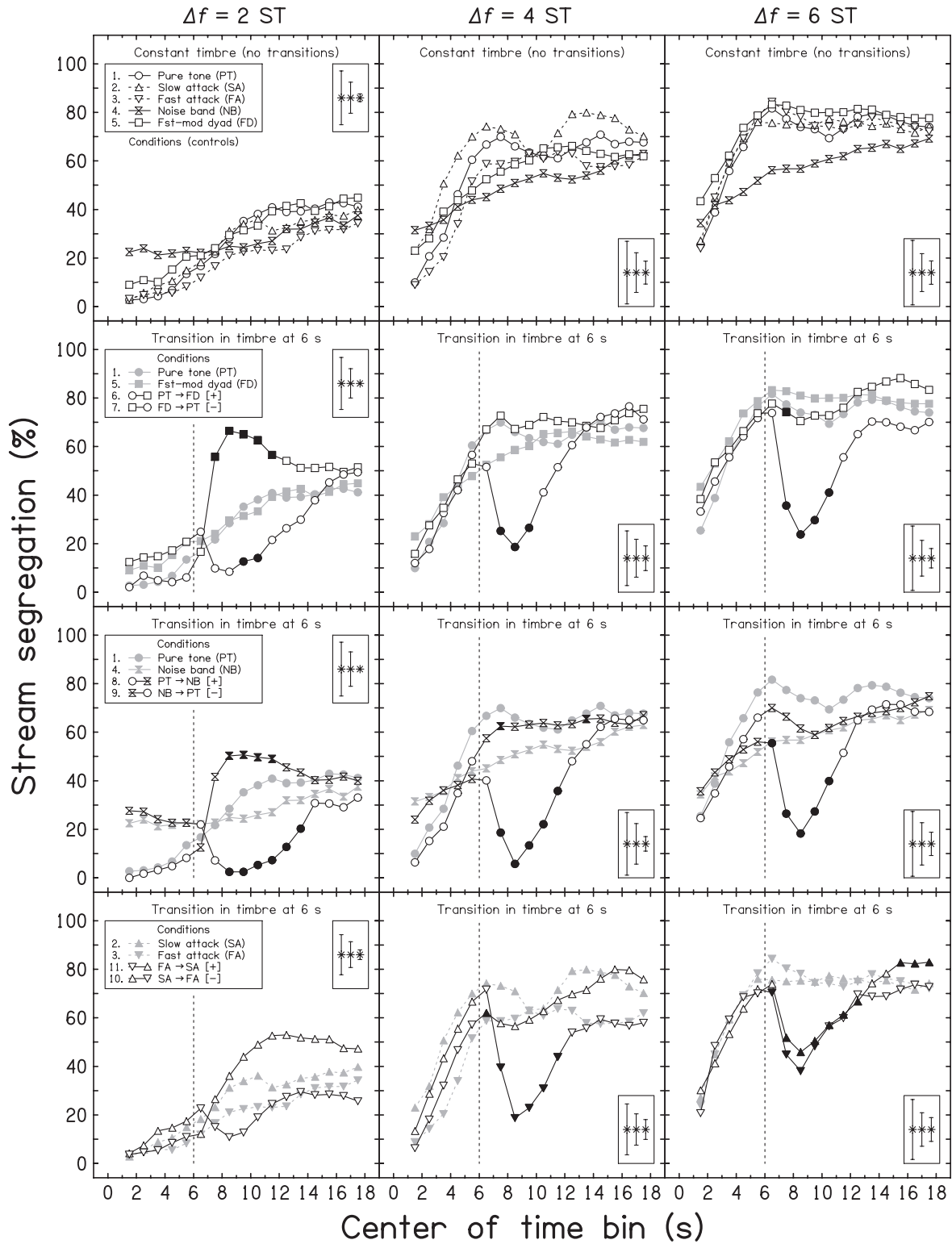
The results averaged across listeners are shown in Fig. 3 (see supplementary material for alternative versions<sup>4</sup>). As before, the top panels display the results for the reference conditions and the remaining panels display the results for the experimental conditions. The upper middle, lower middle, and bottom panels show the effects of abrupt transitions (in both directions) between stimulus types FD and PT, NB and PT, and SA and FA, respectively, relative to their reference cases. Experimental time bins for which there was a significant difference from the reference case in the bootstrapping analysis are shown by filled black symbols. Results for the reference and experimental conditions are considered in turn.

Figure 3 again shows clear evidence of build-up in the reference conditions, and for that build-up to be greater for larger  $\Delta f$ s. There is also an indication that, compared with the other reference conditions, the NB case (C4) caused a slower rate of build-up and a smaller increase in stream segregation with increasing  $\Delta f$ . The differences between dyad and pure-tone sequences observed in experiment 1 were largely absent here. Consistent with Singh and Bregman (1997), there was little evidence to suggest systematic differences in stream segregation between sequences of FA and SA tones. The apparent dip in reported segregation around 10 s for the SA condition (C2) when  $\Delta f = 4$  ST does not have any obvious cause—it cannot easily be attributed to specific listeners and is assumed to have occurred by chance.

The ANOVA for the reference conditions is presented in Table IV. Stream segregation was greater for larger  $\Delta f$ s (means: 2 ST = 25.7%, 4 ST = 55.1%, and 6 ST = 68.0%;  $p < 0.001$ ) and tended to increase over time ( $p < 0.001$ ). There was no main effect of stimulus type ( $p = 0.468$ ), but both two-way interaction terms involving it were significant. The main cause of the stimulus type  $\times$  time interval ( $p = 0.004$ ) and  $\Delta f \times$  stimulus type ( $p = 0.017$ ) interactions is probably the pattern of results for the NB condition. The former probably reflects the shallower response profiles seen for the NB condition and the latter probably reflects the slower rise in segregation with increasing  $\Delta f$  for that condition. Although PT and NB sequences have occasionally been included in the same experiments (e.g., Dannenbring and Bregman, 1976), we are not aware of any studies that have compared the build-up of stream segregation for alternating sequences of pure tones with that for otherwise equivalent narrow-band noises. In addition, the  $\Delta f \times$  time interval interaction was highly significant ( $p < 0.001$ ); as for experiment 1, this arose mainly from the greater rate of change of stream segregation for larger  $\Delta f$ s during the initial phase of build-up. The three-way interaction term was not significant.

The effects of transitions are shown in the other panels of Fig. 3. Sudden changes from FD to PT, from NB to PT, or from SA to FA (ramped to damped) typically led to decreased subsequent segregation (i.e., resetting of build-up). This was especially true for the larger  $\Delta f$ s tested, but there was also more evidence of resetting overall when  $\Delta f$  was 2 ST than was apparent in experiment 1. The latter was mainly attributable to the NB-to-PT transition, for which there was more build-up prior to the change and hence more scope for resetting. Sudden changes in the opposite direction were less integration-promoting for all transitions tested, but the exact pattern varied across stimulus pairs and  $\Delta f$ . For transitions from PT to FD or NB, overshoot was characteristic for the 2-ST case, but the effect on subsequent streaming was typically small for larger values of  $\Delta f$ . The outcome was somewhat different for FA-to-SA transitions. There was a hint of overshoot for the 2-ST case, little effect at 4 ST, but clear evidence of resetting at 6 ST. Overall, the time constants for the development and decay of the impact of a transition on streaming appear broadly similar to those observed in experiment 1, except perhaps for a slightly faster and more complete decay by the last few seconds of the sequence. Again, most time bins in the transition conditions identified in the bootstrapping analysis as significantly different from their references align with the resetting and overshoot effects described earlier, but the last three time bins following the FA-to-SA transition in the 6-ST case are an exception without obvious explanation.

The ANOVA exploring abrupt changes in stimulus type is presented in Table V. The main effects of frequency separation and direction of change were highly significant ( $p < 0.001$ ), but there was no significant effect of stimulus pair ( $p = 0.619$ ). The only significant interaction term was stimulus pair  $\times$  direction of change ( $p = 0.005$ ). This



19 July 2023 07:35:38

FIG. 3. Results for experiment 2—effects of condition (reference = 1–5, experimental = 6–11; see insets in left panels) and  $\Delta f$  on the extent and time course of reported stream segregation ( $n = 12$ ). Results for the reference conditions (no transition) are shown in the top panels; results for the conditions involving transitions between stimulus types PT and FD, PT and NB, or SA and FA are displayed in the upper middle, lower middle, and bottom panels, respectively. Otherwise as for Fig. 1.

interaction was mainly a result of changes in the size of the difference scores for different stimulus pairs, particularly the much smaller effect of transition direction for the FA and SA pair. The effects of individual transitions over the

selected 8.0-s time interval were explored further using the pairwise comparisons summarized in Table VI (parts a and b). Figure 4 displays the mean changes in segregation score and corresponding inter-subject standard errors across

TABLE IV. Results for experiment 2—effects of stimulus type (no-transition reference cases). Summary of the three-way repeated-measures ANOVA for the segregation scores from all time bins (1–2 s to 17–18 s) across conditions C1–C5. Factor  $\Delta f$  has 3 levels (2, 4, and 6 ST) and stimulus type has 5 levels—PT, SA, FA, NB, and FD. All significant terms are shown in bold.

Factor	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
<b>Frequency separation in sequence (<math>\Delta f</math>)</b>	<b>(1.342, 14.766)</b>	<b>27.678</b>	<b>&lt;0.001</b>	<b>0.716</b>
Stimulus type (S)	(1.415, 15.564)	0.690	0.468	0.059
<b>Time interval (T)</b>	<b>(1.803, 19.838)</b>	<b>31.903</b>	<b>&lt;0.001</b>	<b>0.744</b>
$\Delta f \times S$	<b>(3.580, 39.377)</b>	<b>3.580</b>	<b>0.017</b>	<b>0.246</b>
$\Delta f \times T$	<b>(4.880, 53.681)</b>	<b>11.619</b>	<b>&lt;0.001</b>	<b>0.514</b>
<b>S <math>\times</math> T</b>	<b>(6.741, 74.155)</b>	<b>3.310</b>	<b>0.004</b>	<b>0.231</b>
$\Delta f \times S \times T$	(7.690, 84.588)	1.727	0.107	0.136

$\Delta f$  for each stimulus pair in a separate panel; the top axis again summarizes the mean difference scores between the two transition directions.

Other than the SA-to-FA case when  $\Delta f$  was 2 ST, all transitions predicted to be more integration-promoting (part a) led to substantial and significant resetting (range:  $-15.3\%$  pts to  $-33.1\%$  pts,  $p = 0.008$ — $p < 0.001$ ). A more complex pattern emerged for the transitions predicted to be less integration-promoting (part b). For the 2-ST case, there was clear and significant overshoot for the PT-to-FD transition ( $+22.7\%$  pts,  $p = 0.002$ ); for the other stimulus pairs the overshoot was smaller and approached but did not quite reach significance (PT to NB:  $+18.4\%$  pts,  $p = 0.057$ ; FA to SA:  $+13.2\%$  pts,  $p = 0.054$ ). This outcome is unsurprising given the variability in the data. For the larger  $\Delta f$ s, most cases were associated with only small and non-significant effects, but there were two exceptions. First, the PT-to-NB transition led to significant overshoot for the 4-ST case. Second, the FA-to-SA transition led to significant resetting for the 6-ST case. More generally, transitions between FA and SA stimuli were associated with considerably smaller mean difference scores than for the other stimulus pairs. This

TABLE V. Results for experiment 2—effects of abrupt changes in stimulus properties. Summary of the three-way repeated-measures ANOVA for scores corresponding to the difference in stream segregation between transition and reference cases over a single 8.0-s time interval. Factor  $\Delta f$  has 3 levels (2, 4, and 6 ST), stimulus pair has 3 levels (stimulus pairs: FD & PT, NB & PT, and FA & SA), and direction of change has 2 levels (forward and reverse within-pair order, corresponding to transitions predicted to be more and less integration-promoting, respectively). All significant terms are shown in bold.

Factor	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
<b>Frequency separation in sequence (<math>\Delta f</math>)</b>	<b>(1.715, 18.863)</b>	<b>16.200</b>	<b>&lt;0.001</b>	<b>0.596</b>
Stimulus pair (P)	(1.811, 19.925)	0.460	0.619	0.040
<b>Direction of change (D)</b>	<b>(1.000, 11.000)</b>	<b>123.718</b>	<b>&lt;0.001</b>	<b>0.918</b>
$\Delta f \times P$	(2.504, 27.541)	1.874	0.165	0.146
$\Delta f \times D$	(1.455, 16.006)	1.818	0.198	0.142
<b>P <math>\times</math> D</b>	<b>(1.781, 19.587)</b>	<b>7.316</b>	<b>0.005</b>	<b>0.399</b>
$\Delta f \times P \times D$	(2.587, 28.457)	0.302	0.795	0.027

outcome is consistent with the idea that the perceived dissimilarity between FA and SA stimuli was smaller than for the other stimulus pairs, which may account for the clear resetting observed for the 6-ST case (see also Sec. IV). When collapsed across  $\Delta f$  and stimulus pair, the transitions predicted to be more integration-promoting on average led to 26.7% pts less segregation than their counterparts. As for experiment 1, the overall pattern of response profiles suggests a summation of the effects of three factors, one based on transition direction, one on  $\Delta f$ , and one that biases subsequent streaming judgments towards segregation.

The results for the FD-to-PT and PT-to-FD transitions across  $\Delta f$  were similar to those for their exact counterparts in experiment 1. The results for the NB-to-PT and PT-to-NB transitions had broadly similar profiles across  $\Delta f$  to their counterparts involving FD stimuli, suggesting that the underlying asymmetry in the effect of changes between fluctuating and steady stimuli, and vice versa, is like that for changes between modulated and unmodulated tones. Of course, it should be acknowledged that the amplitude fluctuations of a noise band become slower and more salient as its bandwidth narrows. There will inevitably be some differences in peripheral channeling cues between corresponding pure tones and narrow-band noises with a 2-ST passband, and there may also be small differences in loudness when matched for rms power. However, there is again no obvious mechanism by which these differences might account for the strong directional effects observed for these transitions.

The finding here of directional effects, albeit smaller ones, of transitions between pure tones with different amplitude envelopes (damped vs ramped) further extends the range of stimulus types known to produce these asymmetric effects on stream segregation. If the basis for the observed effects of direction did arise from differences in perceived duration, these will have been small compared with the differences in physical duration between targets and distractors used by Cusack and Carlyon (2003), which might account for why the difference scores seen here were relatively small. Another factor meriting comment is that the perceptual center of a sound (P center; Morton *et al.*, 1976) is influenced by its amplitude envelope, such that the P centers of damped sounds (for which the energy peak occurs early) are before those of ramped sounds (e.g., Howell, 1984). Hence, a sudden change from damped to ramped sounds affects the rhythm of a physically isochronous sequence by slightly lengthening the perceived beat duration across the transition (and vice versa for ramped to damped). However, this effect is quite small and does not seem to offer any obvious explanation for the asymmetry seen at smaller  $\Delta f$ s for the direction of the transition.

#### IV. GENERAL DISCUSSION

The experiments reported here have established that sudden coherent changes in a range of acoustic properties without substantial peripheral-channeling cues can lead to asymmetric effects on subsequent streaming. One direction of change may lead to major loss of built-up segregation (resetting) whereas the other leads to little or no resetting,

TABLE VI. Results for experiment 2—effects of abrupt changes in stimulus properties. Parts (a) and (b) show a summary of pairwise comparisons (t tests) exploring the change in stream segregation following more and less integration-promoting transitions, respectively, relative to their reference cases, for each transition type and  $\Delta f$ . All significant cases are shown in bold.

Part (a): Results for more integration-promoting transitions (differences re reference cases)			
Transition type	$\Delta f = 2$ ST (mean, $p$ )	$\Delta f = 4$ ST (mean, $p$ )	$\Delta f = 6$ ST (mean, $p$ )
FD to PT	<b>-15.3% pts, 0.008</b>	<b>-20.2% pts, 0.004</b>	<b>-26.5% pts, 0.002</b>
NB to PT	<b>-24.3% pts, &lt;0.001</b>	<b>-33.1% pts, &lt;0.001</b>	<b>-29.3% pts, &lt;0.001</b>
SA to FA	-2.8% pts, 0.393	<b>-19.9% pts, &lt;0.001</b>	<b>-19.3% pts, 0.007</b>
Part (b): Results for less integration-promoting transitions (differences re reference cases)			
Transition type	$\Delta f = 2$ ST (mean, $p$ )	$\Delta f = 4$ ST (mean, $p$ )	$\Delta f = 6$ ST (mean, $p$ )
PT to FD	<b>+22.7% pts, 0.002</b>	+7.4% pts, 0.372	-3.1% pts, 0.276
PT to NB	+18.4% pts, 0.057	<b>+10.8% pts, 0.028</b>	+3.0% pts, 0.453
FA to SA	+13.2% pts, 0.054	-7.5% pts, 0.066	<b>-14.8% pts, 0.002</b>

and in some cases may *promote* rather than reset stream segregation (overshoot). Transitions from unmodulated to modulated tones (see also Rajasingam *et al.*, 2021) or to noise bands, or from slower to faster rates of AM were less integration-promoting than the reverse changes. Transitions from fast-attack (damped) to slow-attack (ramped) amplitude envelopes also tended to be less integration-promoting than vice versa, but the effect of direction was smaller at all  $\Delta f$ s (and negligible at 6 ST). Overall, the effect of the less integration-promoting direction was usually manifest as overshoot for smaller  $\Delta f$ s and as reduced resetting for larger  $\Delta f$ s. Overshoot is a striking phenomenon in this context—it is easy to see why a sudden change in one or other subset of sounds might promote subsequent streaming (e.g., Rajasingam *et al.*, 2018), but it is not immediately obvious why a sudden change in both subsets might also do so. More generally, the asymmetries observed are hard to explain using classical accounts of streaming, because the extent of perceived dissimilarity between the two stimulus types in a pair does not depend on the direction of the transition.

Instead, we propose an account based on the different effects of switching attention towards or away from sounds possessing a salient additional feature (or with greater activation on that perceptual dimension).

Overall, the response profiles obtained here after abrupt transitions suggest a summation of the effects of three broadly independent factors—one based on the direction of the transition (one strongly integration-promoting, the other having little effect), one based on  $\Delta f$  (for smaller values, there is little prior build-up and so little resetting is possible), and one causing a subsequent bias towards segregation lasting several seconds (analogous to the effect of preceding an alternating-frequency sequence with a constant-frequency sequence corresponding to one or other subset). In this framework, overshoot is observed only when the scope for loss of segregation is minimized (i.e., the less integration-promoting transition direction combined with a small  $\Delta f$ ), such that the third factor dominates. The effects of all these factors for the transition types tested are clearly visible in Figs. 2 and 4. The outcomes for the FA and SA pair can be accommodated within this framework by assuming a smaller difference in salience between the constituents than for the other pairs. This would have led to reduced asymmetry—i.e., FA-to-SA transitions would have retained more integration-promoting properties—such that, following a transition, the increased scope for loss of segregation for sequences with larger  $\Delta f$ s soon became the dominant factor.

Comparison of the results of the current study ( $\Delta f = 2-6$  ST) with those of Rajasingam *et al.* (2021;  $\Delta f = 4-8$  ST) suggests that the effect of  $\Delta f$  is influenced by context—overshoot requires a smaller  $\Delta f$  to be manifest when the overall mean  $\Delta f$  for the stimulus set is lower (or the highest and lowest values are lower). Context effects arising from differences in stimulus range are often found in perception research. Other kinds of context effects involving  $\Delta f$  have been reported for sequences of alternating-frequency tones—less streaming was perceived for a given  $\Delta f$  in the current trial with increasing  $\Delta f$  in the previous trial (e.g., Snyder *et al.*, 2008; Snyder *et al.*, 2009).

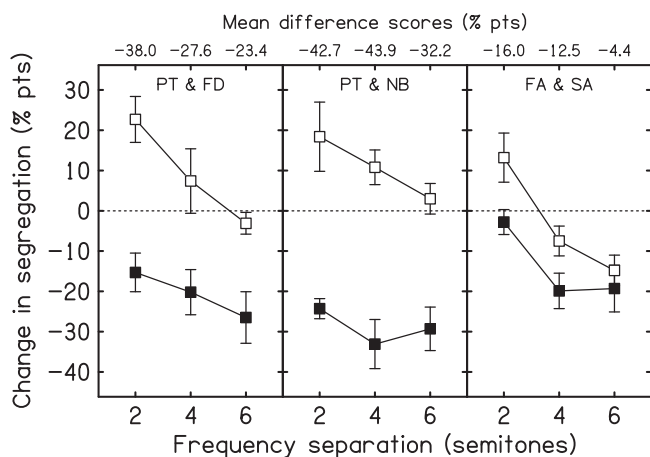


FIG. 4. Results for experiment 2—summary of the effects of abrupt changes in stimulus properties over the selected 8.0-s interval. Changes in segregation (% pts) are shown for the experimental conditions relative to their corresponding reference cases. Results for each stimulus pair are displayed in separate panels from left to right. Otherwise as for Fig. 2.

As far as we are aware, there is only one other finding in the streaming literature akin to the overshoot that can occur after a sudden change in the properties of an alternating-frequency sequence. Rankin *et al.* (2017) used pure-tone sequences composed of six LHL– triplets into which a perturbation could be introduced—either an additional tone with a frequency 2 ST above that of the H tones was inserted in the usual gap between the third and fourth triplets (distractor case) or the H tone of the third triplet was replaced by a tone 2-ST higher (deviant case). Listeners were asked to report at the end of each trial whether the final triplet was heard as one stream or two. Compared with the standard case (no perturbation), the distractor and deviant cases both promoted stream segregation. Nonetheless, there are important differences in stimulus configuration and task between the two studies. Modifying our stimulus configuration to include transients rather than transitions (e.g., a single deviant triplet starting at 6 s) and comparing the effects of changes affecting both subsets (deviant triplet) with those affecting only one subset (deviant L or H tone) should help illuminate the relationship between our results and those of Rankin *et al.* (2017). More generally, the results reported by Rankin *et al.* (2017) may correspond to the effects of the third factor in the framework described above. One possible account of why some of the transitions in the current study may have produced a similar segregation-promoting effect is considered later in this discussion.

Our understanding of the neurophysiological bases for the build-up and resetting of stream segregation has not changed greatly since the study of Micheyl *et al.* (2005). By this account, build-up is assumed to arise from slow adaptation leading to a progressive narrowing of the receptive fields of frequency-tuned neurons in primary auditory cortex (A1). This leads to the emergence of two distinct subpopulations—one responding to the L subset and the other to the H subset—and hence to the perception of two streams. The slow adaptation is presumably the result of long-term synaptic depression (e.g., Pressnitzer *et al.*, 2008); resetting triggered by sudden change is assumed to reflect fast recovery from this adaptation. This account can be adapted to explain the directional effects of  $\pm 12$ -dB transitions seen by Rogers and Bregman (1998) and Rajasingam *et al.* (2021). Specifically, a rising transition (softer-to-louder) raises neural activity, partly offsetting the build-up of adaptation; this widens the A1 receptive fields again leading to greater overlap between populations tuned to the L and H frequencies, favoring a one-stream interpretation until further adaptation erodes it again. A falling transition (louder-to-softer) cannot raise neural activity, but it may not necessarily lower greatly the cortical response. This is because rising transients usually affect central neural responses more than falling ones, leading in this instance to little or no effect on cortical receptive field bandwidths and hence on streaming. However, this kind of account does not seem able to explain the overshoot observed for some kinds of transition at smaller  $\Delta f$ s. Transitions involving the loss or gain of a new stimulus feature (e.g., the presence or absence of modulation

or fluctuation) presumably involve the activation of other populations of neurons coding those features, and hence more complex physiological models will be needed to explain the asymmetric effects on streaming observed for those properties.

It has long been known that focused attention can influence auditory streaming, but the extent to which the build-up of segregation takes place automatically or requires focused attention (e.g., Carlyon *et al.*, 2001) remains to be answered fully. Recent research suggests that at least part of this build-up occurs automatically because subsequent recovery from the reduced segregation caused by performing a competing auditory task during the first part of an alternating sequence occurs more rapidly than the initial build-up when that sequence is attended throughout (Billig and Carlyon, 2016). It is well established, however, that sudden switches in attentional focus affect subsequent streaming (Cusack *et al.*, 2004; Thompson *et al.*, 2011; Rajasingam *et al.*, 2018). In the context of the current study, a possible candidate for the general bias towards segregation after a transition is that the sudden change always began with the L subset of sounds. This may have triggered a switch in attention to that subset. Indeed, there is evidence that preceding an ABA– sequence with only a single tone, matching one or other subset, can promote subsequent segregation (Haywood and Roberts, 2013). In this conception, the directional effects on streaming found for transitions between, e.g., unmodulated and modulated tones arise from an attentional asymmetry (Rajasingam *et al.*, 2021).

As noted earlier, the auditory search literature (e.g., Asemi *et al.*, 2003; Cusack and Carlyon, 2003) indicates that some sounds have additional features that are salient and elicit exogenous orienting (e.g., modulated sounds) whereas others do not (e.g., unmodulated sounds). By this account, a switch in attention away from sounds with primary attention-grabbing properties (loss of a salient feature or reduced activation on that dimension) leads to resetting but a switch towards them (gain or greater activation of a salient feature) does not. Where there is less difference in salience between the stimulus types, the asymmetry may soon be eclipsed by the effects of other factors (e.g., as we have proposed for the FA and SA pair at larger  $\Delta f$ s). By this account, the broader question of what constitutes a relevant additional feature remains to be answered—e.g., why is a percussive onset apparently not such a feature? Further research with the task and stimulus configuration used here might be extended from amplitude- to frequency-modulated (FM) tones. FM is highly salient and strongly associated with asymmetry in auditory search tasks (e.g., Cusack and Carlyon, 2003). On that basis, transitions from steady pure tones to FM tones would be predicted to cause strong overshoot for smaller  $\Delta f$ s.

In conclusion, for sequences of sounds in which there is a correlated transition 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. Whether subsequent stream segregation is increased or decreased, compared with the no-transition reference case, appears to depend on the balance between the factors promoting resetting (e.g., transitions from modulated to unmodulated sounds favor integration) and those favoring a general bias towards stream segregation (probably through switches in attention). Further research is needed to uncover the basis for these directional asymmetries in the underlying neurophysiology and cognitive architecture of auditory stream segregation. This approach should help elucidate further the functional significance of stream segregation in the changing auditory scenes encountered in daily life.

### ACKNOWLEDGMENTS

This research was supported by Aston University's Visiting Scholars' Scheme, which part-funded a research visit by B.R. to N.R.H. at the Hearing Hub, Macquarie University, Sydney, Australia, in November and December 2019. Our thanks go to James Hill for his assistance with data collection and analysis for a pilot version of experiment 1 to explore the effects of sudden transitions in the rate of AM on subsequent stream segregation. We are also grateful to James Rankin for drawing our attention to the effects of introducing single perturbations into a sequence of pure-tone triplets, as reported in his 2017 article, and to Brian Moore and Stephen McAdams for their comments on an earlier version of this manuscript. A preliminary poster presentation on the experiments presented here was given at the Ear and Hear Meeting (Southampton, United Kingdom, September 2022). This article is dedicated to the memory of Al Bregman (1936–2023), whose research and seminal book on auditory scene analysis has inspired our work and continues to inspire others in this field.

<sup>1</sup>Note that a pair of unresolved components also creates frequency modulation at the output of the cochlear filters, which depends on the relative amplitude of each component (Hartmann, 1998).

<sup>2</sup>Note that the edges of successive time bin edges were adjacent but strictly non-overlapping – i.e., the labels 0-1 s, 1-2 s, etc., describe the cases 0 s to <1 s, 1 s to <2 s, etc.

<sup>3</sup>See <https://doi.org/10.17036/researchdata.aston.ac.uk.00000601> (Last viewed July 10, 2023).

<sup>4</sup>See supplementary material at <https://doi.org/10.1121/10.0020172> for alternative versions of Figs. 1 and 3.

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 Ac.* **89**, 346–354.

Beauvois, M. W., and Meddis, R. (1997). "Time decay of auditory stream biasing," *Percept. Psychophys.* **59**, 81–86.

Billig, A. J., and Carlyon, R. P. (2016). "Automaticity and primacy of auditory streaming: Concurrent subjective and objective measures," *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 339–353.

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 Press, Cambridge, MA).

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.

Bregman, A. S., Liao, C., and Levitan, R. (1990). "Auditory grouping based on fundamental frequency and formant peak frequency," *Can. J. Psychol.* **44**, 400–413.

Bridges, D., Pitiot, A., MacAskill, M. R., and Peirce, J. W. (2020). "The timing mega-study: Comparing a range of experiment generators, both lab-based and online," *PeerJ* **8**, e9414.

Carlyon, R. P. (2000). "Detecting coherent and incoherent frequency modulation," *Hear. Res.* **140**, 173–188.

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.

Cusack, R., and Roberts, B. (2000). "Effects of differences in timbre on sequential grouping," *Percept. Psychophys.* **62**, 1112–1120.

Dannenbring, G. L., and Bregman, A. S. (1976). "Stream segregation and the illusion of overlap," *J. Exp. Psychol. Hum. Percept. Perform.* **2**, 544–555.

Dolležal, L.-V., Beutelmann, R., and Klump, G. M. (2012). "Stream segregation in the perception of sinusoidally amplitude-modulated tones," *PLoS ONE* **7**(9), e43615.

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

Gockel, H., Moore, B. C. J., and Patterson, R. D. (2002). "Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression," *J. Acoust. Soc. Am.* **111**, 2759–2770.

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.

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.

Hellman, R. P. (1972). "Asymmetry of masking between noise and tone," *Percept. Psychophys.* **11**, 241–246.

Howell, P. (1984). "An acoustic determinant of perceived and produced anisochrony," in *Proceedings of the Tenth International Congress of Phonetic Sciences*, edited by A. Cohen and M. P. R. van den Broecke (De Gruyter Mouton, Berlin), pp. 429–433.

Iverson, P. (1995). "Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes," *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 751–763.

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.

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. (2002). "Factors influencing sequential stream segregation," *Acta Acust. united Ac.* **88**, 320–333.

Moore, B. C. J., and Gockel, H. (2012). "Properties of auditory stream formation," *Philos. Trans. R. Soc. B* **367**, 919–931.



- Moore, B. C. J., Vickers, D. A., Baer, T., and Launer, S. (1999). "Factors affecting the loudness of modulated sounds," *J. Acoust. Soc. Am.* **105**, 2757–2772.
- Morton, J., Marcus, S. M., and Frankish, C. (1976). "Perceptual centers (P-centers)," *Psychol. Rev.* **83**, 405–408.
- Nie, Y., and Nelson, P. B. (2015). "Auditory stream segregation using amplitude modulated bandpass noise," *Front. Psychol.* **6**, 1151.
- Pearce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J. (2019). "PsychoPy2: Experiments in behavior made easy," *Behav. Res.* **51**, 195–203.
- 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.
- 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.
- Rajasingam, S. L., Summers, R. J., and Roberts, B. (2021). "The dynamics of auditory stream segregation: Effects of sudden changes in frequency, level, or modulation," *J. Acoust. Soc. Am.* **149**, 3769–3784.
- Rankin, J., Osborn Popp, P. J., and Rinzel, J. (2017). "Stimulus pauses and perturbations differentially delay or promote the segregation of auditory objects: Psychoacoustics and modelling," *Front. Neurosci.* **11**, 198.
- 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.
- Singh, P. G., and Bregman, A. S. (1997). "The influence of different timbre attributes on the perceptual segregation of complex-tone sequences," *J. Acoust. Soc. Am.* **102**, 1943–1952.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods*, 6th ed. (Iowa University Press, 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.
- Stecker, G. C., and Hafter, E. R. (2000). "An effect of temporal asymmetry on loudness," *J. Acoust. Soc. Am.* **107**, 3358–3368.
- 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, The Netherlands.
- Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (1999). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," *J. Acoust. Soc. Am.* **106**, 938–945.
- von Békésy, G. (1963). "Hearing theories and complex sounds," *J. Acoust. Soc. Am.* **35**, 588–601.
- Wang, J., Qin, L., Chimoto, S., Tazunoki, S., and Sato, Y. (2014). "Response characteristics of primary auditory cortex neurons underlying perceptual asymmetry of ramped and damped sounds," *Neurosci.* **256**, 309–321.
- Wilson, A. S., Hall, J. W., and Grose, J. H. (1990). "Detection of frequency modulation (FM) in the presence of a second FM tone," *J. Acoust. Soc. Am.* **88**, 1333–1338.
- Yost, W. A., and Sheft, S. (1989). "Across-critical-band processing of amplitude-modulated tones," *J. Acoust. Soc. Am.* **85**, 848–857.