Stream biasing by different induction sequences: Evaluating stream capture as an account of the segregation-promoting effects of constant-frequency inducers

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

Articles you may be interested in

Does the semantic content or syntactic regularity of masker speech affect speech-on-speech recognition?
The Journal of the Acoustical Society of America **144**, 3289 (2018); 10.1121/1.5081679

Short-term implicit voice-learning leads to a Familiar Talker Advantage: The role of encoding specificity
The Journal of the Acoustical Society of America **144**, EL497 (2018); 10.1121/1.5081469

Binaural ambiguity amplifies visual bias in sound source localization
The Journal of the Acoustical Society of America **144**, 3118 (2018); 10.1121/1.5079568

Development and validation of a spectro-temporal processing test for cochlear-implant listeners
The Journal of the Acoustical Society of America **144**, 2983 (2018); 10.1121/1.5079636

A mathematical model of bowel sound generation
The Journal of the Acoustical Society of America **144**, EL485 (2018); 10.1121/1.5080528

Immediate phonetic convergence in a cue-distractor paradigm
The Journal of the Acoustical Society of America **144**, EL528 (2018); 10.1121/1.5082984

Stream biasing by different induction sequences: Evaluating stream capture as an account of the segregation-promoting effects of constant-frequency inducers[a]

Saima L. Rajasingam, Robert J. Summers, and Brian Roberts[b]

Psychology, School of Life and Health Sciences, Aston University, Birmingham B4 7ET, United Kingdom

Stream segregation for a test sequence comprising high-frequency (H) and low-frequency (L) pure tones, presented in a galloping rhythm, is much greater when preceded by a constant-frequency induction sequence matching one subset than by an inducer configured like the test sequence; this difference persists for several seconds. It has been proposed that constant-frequency inducers promote stream segregation by capturing the matching subset of test-sequence tones into an on-going, pre-established stream. This explanation was evaluated using 2-s induction sequences followed by longer test sequences (12–20 s). Listeners reported the number of streams heard throughout the test sequence. Experiment 1 used LHL– sequences and one or other subset of inducer tones was attenuated (0–24 dB in 6-dB steps, and ∞). Greater attenuation usually caused a progressive increase in segregation, towards that following the constant-frequency inducer. Experiment 2 used HLH– sequences and the L inducer tones were raised or lowered in frequency relative to their test-sequence counterparts ($\Delta f_I = 0$, 0.5, 1.0, or $1.5 \times \Delta f_T$). Either change greatly increased segregation. These results are concordant with the notion of attention switching to new sounds but contradict the stream-capture hypothesis, unless a "proto-object" corresponding to the continuing subset is assumed to form during the induction sequence.
https://doi.org/10.1121/1.5082300

[AKCL]

I. INTRODUCTION

Auditory stream segregation refers to the phenomenon in which a sequence of sounds is perceived as comprising more than one auditory stream, each corresponding to a distinct acoustic source in the environment (Bregman and Campbell, 1971). This phenomenon—a key aspect of auditory scene analysis (Bregman, 1990)—has been researched extensively using sequences of sounds with a wide range of properties, but most often using sequences of pure tones alternating rapidly between low (L) and high (H) frequencies (e.g., Miller and Heise, 1950; Bregman and Campbell, 1971). These sequences can be heard either as one stream of sounds moving back and forth in pitch (integrated) or as two independent and monotonous streams of different pitch (segregated); the organization heard is a bistable, characterized by spontaneous switches between the two percepts (e.g., Pressnitzer and Hupé, 2006). For these stimuli, either a larger frequency separation or a faster rate of presentation increases the likelihood of perceiving two streams (e.g., Bregman and Campbell, 1971; van Noorden, 1975). There is also a dynamic aspect of stream segregation—the likelihood of hearing an alternating-frequency (AF) sequence of unchanging frequency separation and rate as two streams

builds up over time (van Noorden, 1975; Bregman, 1978; Anstis and Saida, 1985). The time course of this build-up is fairly slow; the initial (faster) phase occurs during the first ~10 s of the stimulus but the tendency for stream segregation may continue to increase gradually for at least 1 min (Anstis and Saida, 1985).

Factors influencing subsequent stream segregation are often referred to as stream biasing effects (e.g., Beauvois and Meddis, 1997; Snyder *et al.*, 2008). A convenient arrangement for exploring how the perceptual organization of later sounds is influenced by earlier sounds involves a stimulus configuration in which a standardized AF test sequence is preceded without break by an induction sequence whose properties are manipulated across conditions (e.g., Rogers and Bregman, 1993). The effect of a given induction sequence on the perception of the subsequent test sequence can then be assessed by comparing it with two control cases—one in which the properties of the induction sequence match exactly those of the test sequence and one in which the induction sequence is replaced by silence or continuous wideband noise. Studies using this or related approaches have shown that there is typically a near-immediate loss of build-up, referred to as resetting, following a sudden change of sufficient magnitude in the acoustic properties of the AF sequence—e.g., a change in ear of presentation, center frequency, or lateralization (e.g., Anstis and Saida, 1985; Rogers and Bregman, 1998).

The inducer-test configuration has also been used to study another segregation-promoting effect, one that occurs

---

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

without frequency alternation in the earlier sounds (Rogers and Bregman, 1993; Beauvois and Meddis, 1997; Roberts *et al.*, 2008; Haywood and Roberts, 2010, 2013). These studies have shown that using a constant-frequency (CF) induction sequence composed of tones matching one or other subset in the test sequence has a strong segregation-promoting effect. Indeed, the extent of stream segregation for an AF test sequence presented in a galloping rhythm (e.g., LHL–LHL–···) is much greater when preceded by a short CF induction sequence (2.0 s) matching one subset of the test-sequence tones (e.g., L–L–L–L–···) than by an induction sequence of the same duration configured like the test sequence, and this difference persists for several seconds after the test sequence begins (Haywood and Roberts, 2013).

The relationship between the segregation-promoting effects of a matched-AF inducer and those of a matched-CF inducer remains unclear, most notably because (with one exception) there are considerable differences in their temporal characteristics. As noted above, stream segregation during an unchanging AF sequence builds up over many seconds. In contrast, the number of tones in a matched-CF induction sequence can be reduced from 10 (2.0 s) to 3 (0.6 s) without any diminution of subsequent stream segregation and even one inducer tone can be sufficient to promote some segregation (Haywood and Roberts, 2013). Another difference is that, compared with the magnitude of the changes usually required for substantial resetting of build-up to occur in an AF tone sequence, making the final tone of an otherwise matched-CF induction sequence a "deviant" on some dimension (e.g., in frequency, in duration, or by replacement with silence) is usually sufficient to cause substantial resetting (Haywood and Roberts, 2010). The exception is that the decay of segregation promotion during a silent interval is near-complete for most listeners in ∼4 s for AF and for CF inducers (Bregman, 1978; Beauvois and Meddis, 1997).

The rapid onset and strong promotion of stream segregation induced by matched-CF stimuli has usually been explained in terms of the capture of a subset of test-sequence tones into an on-going stream, already formed from the unvarying inducer tones (e.g., Rogers and Bregman, 1993; Haywood and Roberts, 2013); a similar effect attributed to stream capture had previously been observed in an objective task using a different but related stimulus configuration (Bregman and Rudnicky, 1975). The experiments reported here tested this account and explored further the relationship between the stream biasing effects of matched-AF and matched-CF inducers by creating AF stimuli with properties intermediate between them—hybrid-AF stimuli—in which one subset of inducer tones always precisely matched the acoustic properties of its counterpart in the test sequence but the other subset did not. We used a stimulus arrangement in which a short induction sequence was followed without break by a long test sequence. Compared with the more typical use of short test sequences (e.g., Rogers and Bregman, 1993, 1998; Haywood and Roberts, 2010), the advantage of this approach is that it allows observation not only of the initial effect of the induction sequence on streaming but also of its time course and persistence during the test sequence.

Listeners attended to the entire stimulus, but responded only during the test sequence.

Each test sequence comprised two subsets of pure tones, A and B, presented in a repeating triplet pattern (i.e., ABA–ABA–···). Following the method of Haywood and Roberts (2013, experiment 3), the extent of stream segregation was assessed throughout the test sequence—listeners continuously monitored the test sequence and reported when they heard it as one stream and when as two streams. Subjective measures, based on introspection, are widely used for research in auditory scene analysis and provide an efficient and direct measure of the streaming experienced by listeners, rather than one that must be inferred from changes in accuracy of performance (for a review, see Bregman, 2015). Note that results obtained in streaming studies using subjective and objective measures are usually concordant (e.g., Roberts *et al.*, 2002; Farkas *et al.*, 2016), and both measures have their advantages, but there are some circumstances in which their outcomes can differ (e.g., Billig and Carlyon, 2016).

The properties of the accompanying induction sequences were manipulated in various ways and their effects on subsequent stream segregation were measured in two experiments. Both experiments included conditions involving a standard AF inducer and a matched CF inducer created by deleting one or other subset of tones. Other conditions were created by manipulating the relative level (experiment 1) or frequency (experiment 2) of one subset of tones in the standard AF induction sequence, leaving the other unchanged. In Sec. IV, we evaluate the stream-capture hypothesis in the context of the results obtained for these other conditions and conclude that the notion of stream capture—at least as currently conceived—must either be modified or rejected. We also consider attention switching to new sounds as a possible alternative or additional explanation for the results obtained.

II. EXPERIMENT 1

This experiment examined the effect of level differences between the A and B subsets in the induction sequence—here represented by L and H tones, respectively—on the perception of the test sequence. One or other subset was attenuated to different extents and the other remained identical to its counterpart in the test sequence. The case where neither subset was attenuated corresponded to the standard alternating-frequency (AF) induction sequence; where one or other subset was attenuated completely, these cases corresponded to the matched constant-frequency (CF) induction sequences. Partial attenuation of one or other subset created the intermediate cases—hybrid induction sequences involving frequency alternation but for which the tones of the attenuated subset did not fully match their counterparts in the test sequence.

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 defined in relation to each listener's profile of responses in the reference condition (see Sec. II A 3). Twelve listeners (3 males) successfully completed the experiment (mean age = 23.4 years, range = 19.9–28.3). This research was approved by the Aston University Ethics Committee.

2. Stimuli and conditions

The test sequence used was 20 s long and comprised 50 LHL– cycles. Each tone was 100-ms long (including 10-ms raised cosine onset/offset ramps). The silence at the end of each triplet was also 100-ms long and so the duration per cycle was 400 ms. This rate of presentation is known to facilitate stream segregation based on frequency separation (e.g., Bregman and Campbell, 1971; van Noorden, 1975). The frequency of the L subset of tones was kept constant at 1 kHz (reference frequency) and the frequency of the H subset was set according to the desired high-low (HL) frequency difference for the test sequence (Δf_T), which was 4, 6, or 8 semitones (ST). Hence, the frequency of the H subset was 1260, 1414, or 1587 Hz, respectively. This range of frequency separations was used to protect against ceiling and floor effects,

and to provide information on any interactions that might occur between frequency separation and induction condition. All tones in the test sequence were presented at 73 dB sound pressure level (SPL); tones in the induction sequence were presented at this reference level except where indicated.

Ten induction conditions were used; a schematic illustrating them is shown in Fig. 1, for which the panel numbers correspond to condition numbers. The induction sequences differed in the extent of attenuation (if any) applied to one of the subsets of tones; note that the HL frequency difference for the induction sequence (Δf_I) was always identical to Δf_T in this experiment. In the silent-induction condition (panel 1), the test sequence was preceded by 2 s of silence; this condition provided a measure of test-sequence streaming in the absence of any opportunity for prior build-up. In the standard AF-induction condition (panel 2), the induction sequence was 2 s long and consisted of 5 LHL– triplets; these triplets were identical to those comprising the test sequence and so the transition at the induction/test boundary was seamless. This condition provided a measure of the segregation-promoting effect of an unaltered AF induction sequence; it has been shown previously that build-up for an attended sequence occurs at the same rate whether or not listeners can respond (Haywood and Roberts, 2013). For the other eight conditions, one or other subset of inducer tones was attenuated by 6, 12, 24 dB, or completely (∞; i.e., replaced by silence) relative to the reference level (H tones = left-hand
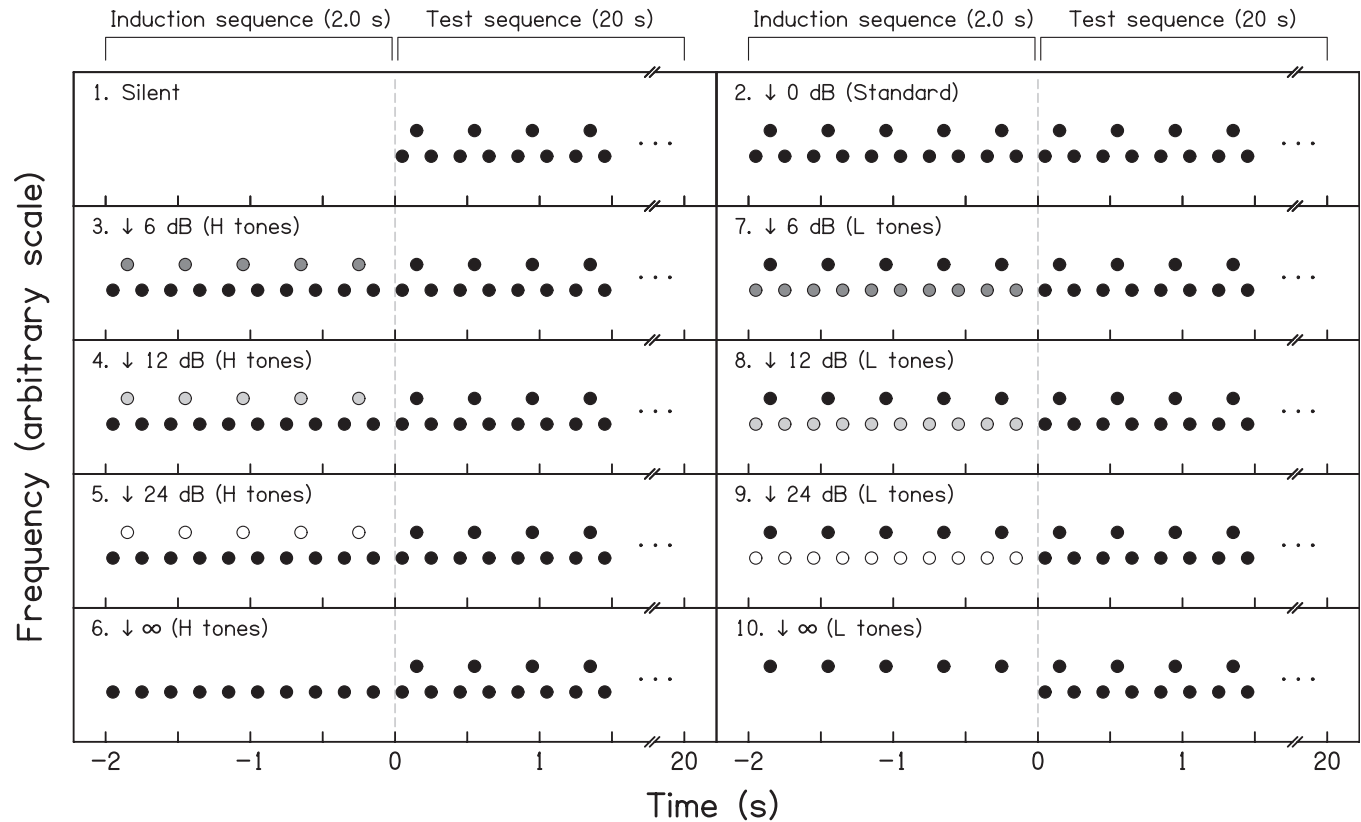

FIG. 1. Stimuli for experiment 1—illustration of the induction conditions used (1–10). Each panel displays a different induction sequence paired with a subsequent LHL– test sequence; note that the test sequence continued for 20 s. Circles represent pure tones and progressively greater attenuation of the manipulated subset of tones (A or B) is indicated by the transition in shading from filled to unfilled. Induction sequences involving attenuation of the H tones (B subset) and L tones (A subset) are shown in the left- and right-hand panels, respectively. Conditions involving the complete attenuation of one or other subset (∞) correspond to constant-frequency induction sequences.

J. Acoust. Soc. Am. **144** (6), December 2018

Rajasingam et al.   3411

panels, 3–6; L tones = right-hand panels, 7–10). Given that the repetition rate of the L tones (A subset) was twice that of the H tones (B subset), the two CF induction conditions created by complete attenuation of one or other subset differed in tone density by a factor of 2 (cf. panels 6 and 10).

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 PA5; Alachua, FL), and a headphone buffer (TDT HB7). Output levels were calibrated using a sound-level meter (Brüel and Kjaer, type 2209; Nærum, Denmark) coupled to the earphones by an artificial ear (type 4153). Diotic presentation was used throughout this study.

3. Procedure

Listeners completed the experiment in a single-walled sound-attenuating chamber (Industrial Acoustics 401A; Winchester, UK) housed within a quiet room. They were free to take breaks between trials whenever they wished. After reading the instructions, listeners completed one training block of trials identical to those used in the main experiment (see below); a second training block was offered but rarely required. During the training and the main experiment, stimuli were presented in a new quasi-random order in each block for each listener. Completing all stages of the procedure (screening, training, and main experiment) typically took ∼3.5 hours, 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.

On each trial, a single combination of an induction sequence and a test sequence was presented once. Each trial was initiated 1 s after the listener pressed "enter" on the computer keyboard. Listeners were instructed to monitor the stimulus continuously throughout, but not to respond during the induction sequence. At the start of the test sequence, the on-screen message changed from "please wait" to "please respond" and listeners were asked to indicate as soon as possible whether they were hearing integration (one stream) or segregation (two streams) by pressing either the "A" or "L" keys, respectively. Thereafter, listeners were asked to press the appropriate key every time their perception of the test sequence changed. They were asked to avoid listening actively for either integration or segregation, but simply to report which of the two percepts they heard at that moment; on occasions when the percept was ambiguous, listeners were asked to report the more dominant (cf. Haywood and Roberts, 2013). 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 could decay before the onset of the next trial; earlier studies have shown near-complete decay of build-up for a silent interval of 4 s (e.g., Bregman, 1978; Beauvois and Meddis, 1997).

Each combination of induction condition (10 levels) and Δf_T (3 levels) was presented ten times in the main experiment, once in each block, giving 300 trials. Using three different Δf_T values also provided a means of defining 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 (e.g., van Noorden, 1975; Anstis and Saida, 1985). Therefore, any listener whose data did not show a systematic effect of Δf_T on judgments of stream segregation in the AF conditions (silent inducer and standard inducer) was excluded from the study and replaced; this happened only for one listener.

4. Data analysis

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 during which the listener reported the test sequence as segregated was calculated from the timings of individual key presses. This percentage 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 limited number of trials meeting this criterion for the 0–1 s time bin (∼15%–20%; cf. Haywood and Roberts, 2013), responses made during that time bin 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. For each listener, the data for a given time bin were averaged across trial blocks separately for each combination of induction condition and frequency separation. Each mean was calculated only from the trials for which that time bin met the acceptance criterion described above. On occasions when one of these means was missing (12 cases, corresponding to ∼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 the twelve listeners to yield, for each combination of induction condition and frequency separation, 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.

All statistical analyses were computed using SPSS (SPSS statistics version 21, IBM Corp.). 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 (η^2_p). Two-tailed pairwise comparisons were conducted using the restricted least-significant-difference test (Snedecor and Cochran, 1967; Keppel and Wickens, 2004). The analysis involved three factors—frequency separation between tone subsets A and B in the test sequence (Δf_T), induction condition (C), and time interval (T). Two versions of each ANOVA were computed—a primary version excluding the silent-induction condition and a supplementary version including it. This condition was excluded

3412   J. Acoust. Soc. Am. **144** (6), December 2018
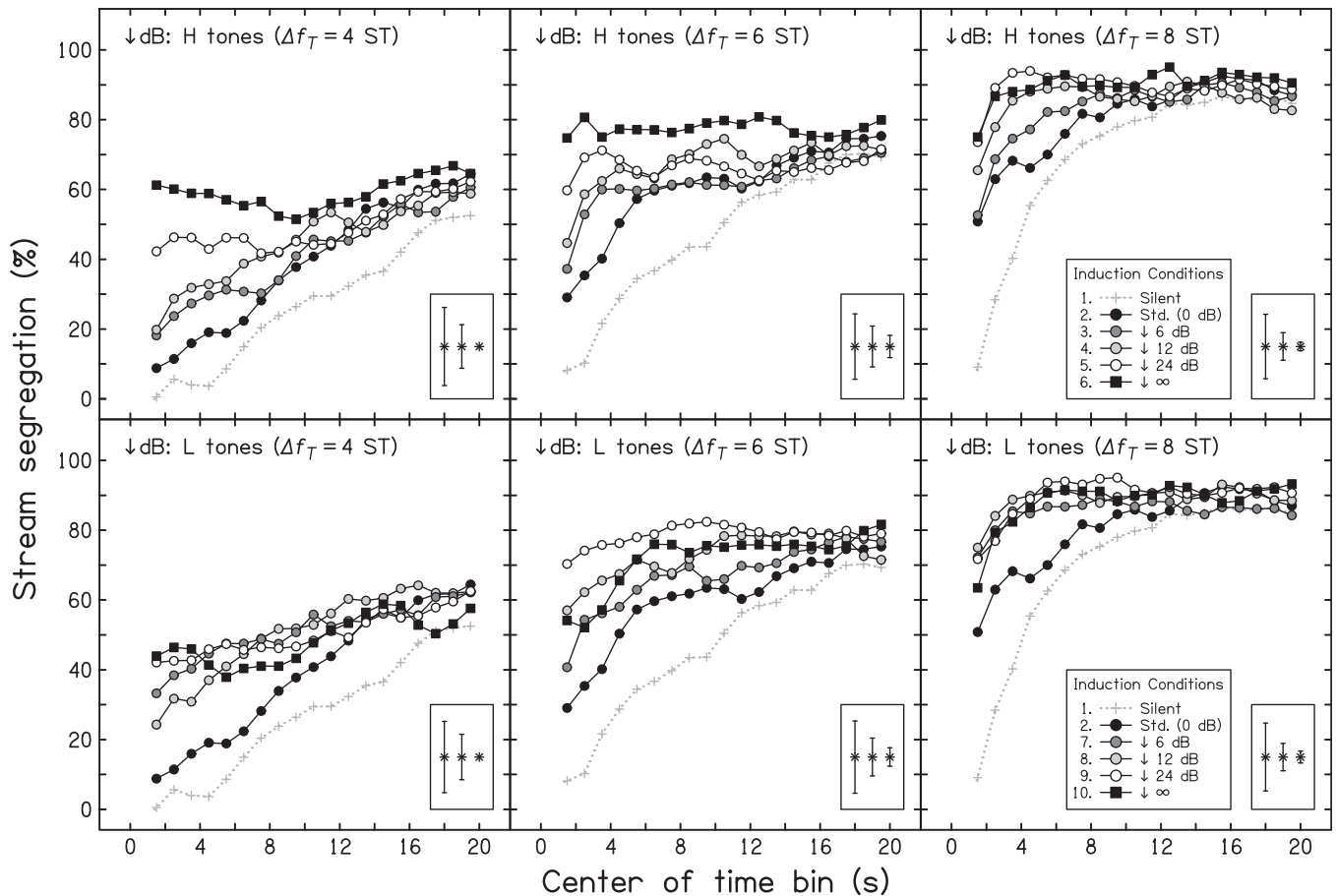
Rajasingam *et al.*

FIG. 2. Results for experiment 1—effects of induction condition (1–10, see insets in right-hand panels) and frequency separation in the test sequence on the extent and time course of reported stream segregation (n = 12). Responses for each separate trial are divided into 1-s time bins, for which the results are averaged across repetitions of the same type, and finally across listeners. Data for the first time bin (0–1 s) are excluded owing to the limited number of responses made during this interval (see main text). Note that the time indicated on the abscissa corresponds to the center of the corresponding time bin. Results for each frequency separation are displayed in separate panels from left to right; the results for attenuation of the H tones (B subset) and the L tones (A subset) are shown in the upper and lower panels, respectively. For clarity, the mean values 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).

from the primary version because it is the only case for which no induction sequence was presented before the test sequence and so it is, in effect, equivalent to the standard AF-induction condition delayed by 2 s (cf. Haywood and Roberts, 2013, experiment 3). Only the primary version of each ANOVA is presented here; the supplementary version was computed simply to allow pairwise comparisons within the condition factor between the results for the silent-induction case and for the various induction sequences used.

B. Results and discussion

The results averaged across all listeners are shown in Fig. 2. Previous research has shown that the fast phase of the build-up of stream segregation takes place over the first 10–12 s of a repeating sequence (e.g., Anstis and Saida, 1985; Haywood and Roberts, 2013), and it is during this part of the test sequence that the greatest differences between the induction conditions can be seen. Therefore, a three-factor repeated-measures ANOVA was conducted on the first 10 s of response data available for analysis (frequency separation × induction condition × time interval: *time bins 1–2 s to*

10–11 s, inclusive); the statistical outcomes are presented in Table I. This analysis showed significant main effects of frequency separation, induction condition, and time interval ($p < 0.001$ in all cases). Clearly, all three factors influenced stream segregation during the first half of the test sequence—segregation was greater for larger frequency separations (means including C1: 4 ST = 40.8%, 6 ST = 66.0%, and 8 ST = 84.4%), varied substantially across conditions (means for C1–C10: 34.2%, 49.5%, 55.8%, 61.8%, 67.0%, 74.0%, 62.3%, 67.2%, 70.6%, and 65.3%, respectively), and tended to change over time (usually increased).

The origin of the main effect of condition (C2–C10) was explored using pairwise comparisons. Stream segregation was greater for both CF inducers than for the standard AF inducer (C6 vs C2, $p = 0.001$; C10 vs C2, $p = 0.027$). Greater attenuation of the H tones (B subset) led to a clear and progressive increase in segregation of the subsequent test sequence, reaching a maximum for the infinite-attenuation condition (i.e., the matched-CF inducer). The increase in stream segregation was significant for attenuations of ≥ 12 dB (C2 vs C4–C6, $p = 0.009 – p = 0.001$); note that an attenuation of 6 dB (C3) also had a significant effect

TABLE I. Results for experiment 1—summary of the three-way repeated-measures ANOVA for time bins 1–2 s to 10–11 s, excluding the silent-induction condition. All significant terms are shown in bold.

Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf_T)	**(2, 22)**	**57.871**	**<0.001**	**0.840**
Induction condition (C)	**(8, 88)**	**6.832**	**<0.001**	**0.383**
Time interval (T)	**(9, 99)**	**6.571**	**<0.001**	**0.374**
$\Delta f_T \times$ C	(16, 176)	1.607	0.071	0.127
$\Delta f_T \times$ T	(18, 198)	1.597	0.064	0.127
C $\times$ T	**(72, 792)**	**1.960**	**<0.001**	**0.151**
$\Delta f_T \times$ C $\times$ T	**(144, 1584)**	**1.426**	**0.001**	**0.115**

if the time bins included in the comparison were restricted to the first five (p = 0.037). A broadly similar pattern was observed when the L tones (A subset) were attenuated; the increase in stream segregation was significant for attenuations of ≥ 6 dB (C2 vs C7–C10, p = 0.027 – p = 0.002). However, there is a suggestion in these data that the infinite-attenuation condition was less effective at promoting segregation than the 24-dB case, particularly for the intermediate frequency separation ($\Delta f_T = 6$ ST). Most likely, this partial reversal of the effect of applying increasing attenuation to the A subset in the infinite-attenuation case (C10) was due to the two-thirds reduction in the number of onsets present relative to the AF condition, which slows the inducer rhythm substantially. The equivalent case for the B subset (C6) reduced the number of onsets present by only one third. Longer tone onset-to-onset times and lower tone density for the earlier sounds are both factors known to decrease their segregation-promoting effects (e.g., van Noorden, 1975; Rogers and Bregman, 1993). Compared with the silent-induction case, every other condition promoted stream segregation (p < 0.001 in all cases).

As well as the significant main effects of all three factors, one of the two-way interactions (induction condition ×time interval, p < 0.001) and the three-way interaction (p = 0.001) were also significant. The origin of these interactions is evident in Fig. 2. First, the pattern of change in stream segregation during the first half of the test sequence was strongly dependent on induction condition—the segregation-promoting effect of CF inducers was most evident early on, such that the differences between the effect of the standard AF inducer and those of the other inducers typically decreased over several seconds. In particular, the characteristic rising profile for the AF induction condition tended to flatten as the attenuation of one or other subset of tones increased. Indeed, Haywood and Roberts (2013) found that the initial segregation-promoting effect of a CF inducer could be so great that the mean reported segregation actually declined over the first several seconds of the test sequence for the largest Δf_T that they tested (9 ST). Second, the relationship between condition and time interval was influenced by frequency separation—typically, the effects of induction condition over time were most differentiated when $\Delta f_T = 6$ ST and there was evidence of ceiling effects influencing the results when $\Delta f_T = 8$ ST.

A similar analysis was conducted on the response data for the final 9 s of the test sequence (frequency separation × induction condition × time interval: time bins 11–12 s to 19–20 s, inclusive); the statistical outcomes are presented in Table II. The ANOVA showed significant main effects of frequency separation (means including C1: 4 ST = 56.5%, 6 ST = 73.1%, and 8 ST = 89.3%), induction condition (means for C1–C10: 63.7%, 71.5%, 68.8%, 70.6%, 70.1%, 77.2%, 73.8%, 74.8%, 75.3%, and 74.1%, respectively), and time interval (p ≤ 0.001 in all cases). These outcomes indicate that (1) larger frequency separations increased the extent of reported segregation, even after the period of most substantial change in the tendency to hear two streams was over; (2) although smaller, effects of induction condition on stream segregation persisted into the latter half of the sequence—in particular, mean stream segregation remained greatest following the L-tones-only induction sequence (C6); (3) although more slowly, reported stream segregation continued on average to rise in the latter portion of the test sequence. Two of the two-way interactions were also significant—induction condition × time interval (p = 0.002) and Δf_T × time interval (p < 0.001). The former has the same origin as its counterpart for the first half of the sequence; the latter mainly arises because the overall tendency for stream segregation to continue increasing during the second half of the sequence is greater for smaller frequency separations.

In summary, the most important finding of this experiment is that—unless the consequent reduction in tone density is too great—attenuating one subset of tones in a matched-AF induction sequence leads to a smooth and progressive increase in its segregation-promoting effect towards that following a matched-CF induction sequence. The consequence of changing the frequency, rather than the relative level, of one subset of inducer tones remains to be established. According to the theory of indispensable attributes (Kubovy, 1981; Kubovy and van Valkenburg, 2001; van Valkenburg and Kubovy, 2003), visual objects are formed in space-time but auditory objects are formed in frequency-time, and so frequency and frequency differences play a special role in auditory perceptual organization. Therefore, one might predict even stronger effects on subsequent stream segregation when the frequency of one subset of the inducer tones is changed.

III. EXPERIMENT 2

This experiment examined the effect of differences between the induction and test sequence in Δf_I and Δf_T

TABLE II. Results for experiment 1—summary of the three-way repeated-measures ANOVA for time bins 11–12 s to 19–20 s, excluding the silent-induction condition. All significant terms are shown in bold.

Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf_T)	**(2, 22)**	**45.310**	**<0.001**	**0.805**
Induction condition (C)	**(8, 88)**	**3.554**	**0.001**	**0.244**
Time interval (T)	**(8, 88)**	**5.253**	**<0.001**	**0.323**
$\Delta f_T \times$ C	(16, 176)	1.081	0.376	0.089
$\Delta f_T \times$ T	**(16, 176)**	**4.343**	**<0.001**	**0.283**
C $\times$ T	**(64, 704)**	**1.614**	**0.002**	**0.128**
$\Delta f_T \times$ C $\times$ T	(128, 1408)	0.767	0.973	0.065

3414   J. Acoust. Soc. Am. **144** (6), December 2018

Rajasingam *et al.*

values for the A and B subsets—here represented by H and L tones, respectively—on the perception of the test sequence. Hill *et al.* (2012) found no difference in streaming reports for HLH– vs LHL– sequences and so the choice of stimulus arrangement used here was arbitrary. One subset (H tones) always remained identical to its counterpart in the test sequence; the other (L tones) was adjusted in frequency relative to its test-sequence counterpart or each tone was replaced with silence. The condition for which $\Delta f_I = \Delta f_T$ corresponded to the standard AF induction sequence; the condition for which the L subset was replaced by silence corresponded to the matched CF induction sequence. The intermediate cases—the hybrid-AF conditions—used induction sequences involving a greater or lesser extent of frequency alternation relative to the test sequence. Note that decreasing or increasing Δf_I by changing the frequency of only one subset of tones inevitably introduces a change in the center frequency of the stimulus at the induction/test boundary; in contrast, previous studies investigating the effect on streaming of altering the Δf between earlier and later sounds in AF sequences have usually done so by raising the frequency of one tone subset and lowering the frequency of the other, without changing the center frequency of the stimulus.

A. Method

Except where described, the same method was used as for experiment 1. Twelve listeners (2 males, mean age = 25.1 years, range = 19.8–29.4) took part and successfully completed the experiment; no listeners were excluded and replaced. Two of the listeners also took part in experiment 1. The results of experiment 1 indicated that differences between conditions were most apparent during the first 10–12 s of the test sequence (cf. Haywood and Roberts, 2013, experiment 3), and so there was considerable scope to shorten it from 20 s without significant loss of analytical power. This allowed all stages of the procedure to be completed in a single session, which typically took ~1.5 h. The test sequence used was 12 s long, comprising 30 HLH– cycles. In this experiment, the H tones were set to 1 kHz (reference frequency) and Δf_T was set to 4, 6, or 8 ST by lowering the frequency of the L tones to 794, 707, or 630 Hz, respectively. All tones in the test and induction sequences were presented at 70 dB SPL.

There were six induction conditions in this experiment; a schematic illustrating them is shown in Fig. 3. As for experiment 1, these conditions included the standard AF-induction (panel 5) and silent-induction (panel 1) cases; the experiment also included one of the possible CF-induction cases (high subset only; panel 2). For the other three conditions, Δf_I was manipulated by raising or lowering the frequency of the L subset of inducer tones relative to its test-sequence counterpart. By this means, Δf_I was set to 0, 0.5, 1.0 (i.e., standard AF), or $1.5 \times \Delta f_T$ (panels 3–6, respectively). Note that the special case for which the frequency of the L tones was set to match that of the H tones ($\Delta f_I = 0$; panel 3) is like the high-subset-only case, but with a 50% increase in the number of tone onsets during the induction sequence. Each combination of induction condition (6
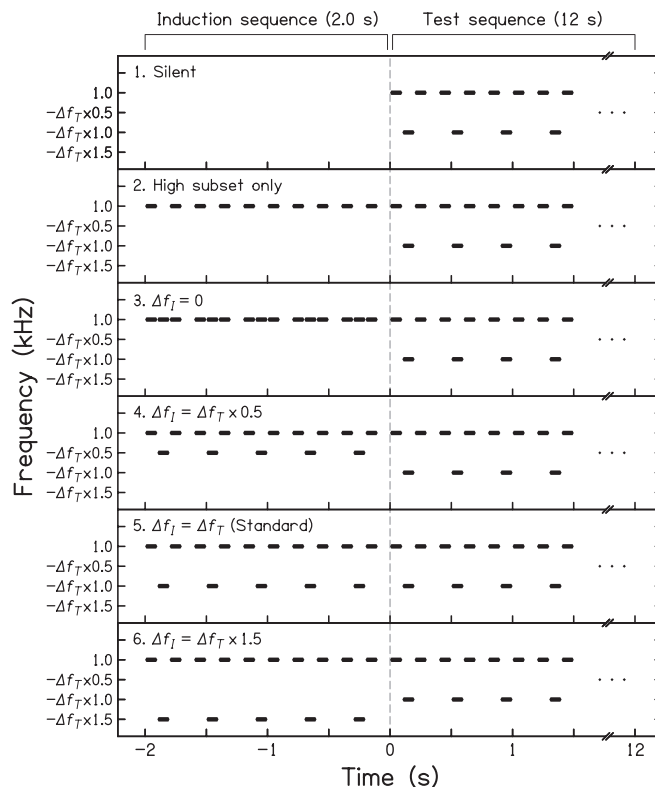

FIG. 3. Stimuli for experiment 2—illustration of the induction conditions used (1–6). Each panel displays a different induction sequence paired with a subsequent HLH– test sequence; note that the test sequence continued for 12 s. Solid lines represent pure tones whose frequency was set relative to that of the H subset (reference = 1 kHz). Note that some conditions involved raising or lowering the frequency of the L tones in the induction sequence relative to their counterparts in the test sequence; others used induction sequences containing only H tones.

levels) and Δf_T (3 levels) was presented ten times in the main experiment, once in each block, giving 180 trials.

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 (66 cases, corresponding to ~2.7% of the data and all occurring within the first few time bins), mean imputation was used to replace the missing value. Once again, the results were analyzed using three-factor repeated-measures ANOVA and the silent-induction condition was excluded from the primary version of the analysis. Owing to the shorter test sequence used here, all time bins (i.e., 1–2 s to 11–12 s) were included within the same analysis.

B. Results and discussion

The results averaged across all listeners are shown in Fig. 4 and the statistical outcomes are presented in Table III. The analysis showed significant main effects of frequency separation, induction condition, and time interval ($p \leq 0.001$ in all cases). As for experiment 1, all three factors influenced stream segregation—segregation was greater for larger frequency separations (means including C1: 4 ST = 52.5%, 6 ST = 72.2%, and 8 ST = 82.6%), varied substantially across conditions (means for C1–C6: 37.8%, 73.7%, 78.0%, 76.3%, 48.9%, and 68.7%, respectively), and tended to change over

J. Acoust. Soc. Am. **144** (6), December 2018
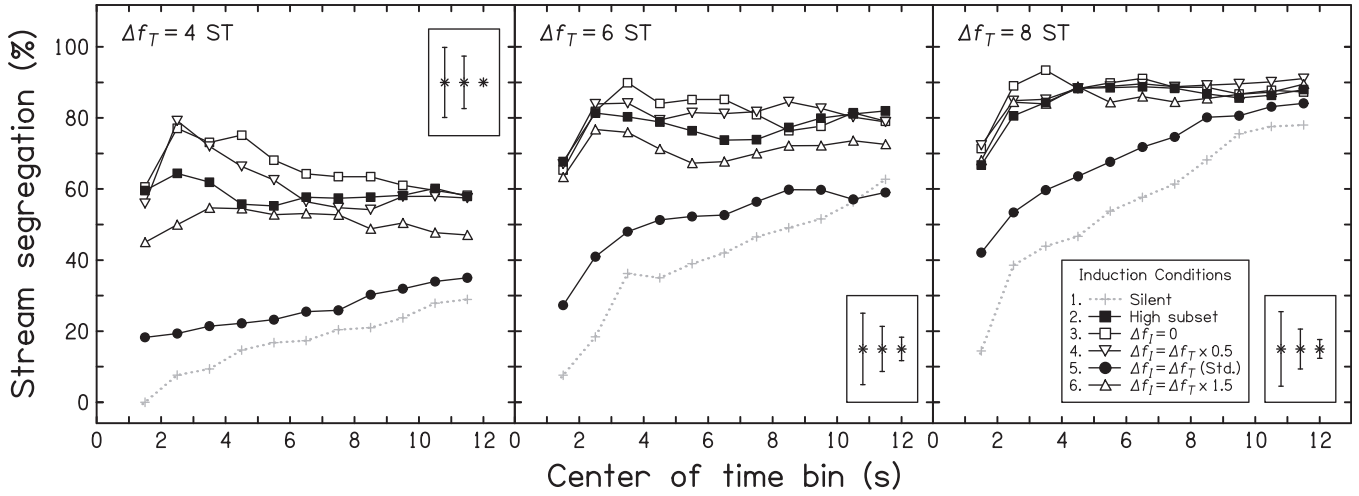
Rajasingam *et al.*   3415

FIG. 4. Results for experiment 2—effects of induction condition (1–6, see inset in right-hand panel) and frequency separation in the test sequence on the extent and time course of reported stream segregation (n = 12). Responses for each separate trial are divided into 1-s time bins, for which the results are averaged across repetitions of the same type, and finally across listeners. Data for the first time bin (0–1 s) are excluded owing to the limited number of responses made during this interval (see main text). Note that the time indicated on the abscissa corresponds to the center of the corresponding time bin. Results for each frequency separation are displayed in separate panels from left to right. For clarity, the mean values 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).

time. Compared with the silent-induction case, every other condition promoted significant stream segregation (p < 0.001 in all cases). The origin of the main effect of condition (C2–C6) was explored using pairwise comparisons. Relative to the standard AF inducer (C5: $\Delta f_I = \Delta f_T$), segregation was significantly greater for the H-subset-only condition (C2; p < 0.001) and for all other induction sequences tested (C3, C4, and C6: $\Delta f_I = 0$, 0.5, and $1.5 \times \Delta f_T$, respectively; p < 0.001 in all cases). This outcome indicates that a stimulus arrangement in which there is an exact match in frequency for one subset of inducer and test tones—but a mismatch in frequency for the other subset—is strongly segregation-promoting, biasing listeners towards a two-stream percept.

Two other aspects of the main effect of induction condition also merit comment. First, the highest nominal mean occurred for C3 ($\Delta f_I = 0$) rather than for C2 (H subset only). Although this difference is not significant when all time bins are included in the comparison (p = 0.165), note that it becomes significant when the time bins included are restricted to the first five (p = 0.020). This difference cannot be explained in terms of the greater number of onsets (50%

more; i.e., 3 per ABA– cycle) and higher tone density for the induction sequence used in C3 relative to that used in C2. This is because it has already been established that doubling the number of onsets (and the associated tone density) in a CF induction sequence relative to an exact match with the corresponding subset in the AF test sequence (i.e., from 2 to 4 per ABA– cycle) causes a small but significant *decrease* in stream segregation (Rogers and Bregman, 1993). Note that there is a growing body of evidence that predictability and rhythm are factors that can influence auditory perceptual organization (e.g., Jones *et al.*, 1981; Snyder and Weintraub, 2011; Bendixen *et al.*, 2013), and so it is possible that the explanation lies in the rhythmic difference between the induction sequences used in C2 (isochronous) and C3 (3 beats and 1 pause per ABA– cycle). However, it should be acknowledged that the difference between C2 and C3 is not evident when the reports begin (1–2 s time bin), but seems to emerge ∼2–4 s after the start of the test sequence. It is not clear how this delay might arise in the context of an explanation based on rhythmic differences between induction sequences.

Second, setting $\Delta f_I < \Delta f_T$ (including where $\Delta f_I = 0$), such that the frequency separation for the sequence increased at the induction/test boundary, promotes more segregation than setting $\Delta f_I > \Delta f_T$, such that the frequency separation decreased (C3 vs C6, p = 0.001; C4 vs C6, p = 0.033). This secondary outcome most probably reflects a contrast effect, in which an increase in Δf at the induction/test boundary biases judgments of the test sequence towards more segregated percepts and vice versa. Snyder *et al.* (2008) reported an across-trial effect of this kind in a streaming task using long AF sequences (10.8 s)—less streaming was reported for a given Δf in the current trial with increasing Δf in the previous trial, despite the silent interval between them (≥1.44 s). This effect persisted over several seconds and was similar to the duration of auditory sensory memory (Cowan, 1984).

TABLE III. Results for experiment 2—summary of the three-way repeated-measures ANOVA for all time bins (1–2 s to 11–12 s), excluding the silent-induction condition. All significant terms are shown in bold.

Factor	df	F	p	η^2_p
Frequency separation in test sequence (Δf_T)	**(2, 22)**	**53.624**	**<0.001**	**0.830**
Induction condition (C)	**(4, 44)**	**26.704**	**<0.001**	**0.708**
Time interval (T)	**(10, 110)**	**3.429**	**0.001**	**0.238**
$\Delta f_T \times$ C	(8, 88)	1.920	0.067	0.149
$\Delta f_T \times$ T	**(20, 220)**	**4.672**	**<0.001**	**0.298**
C × T	**(40, 440)**	**2.912**	**<0.001**	**0.209**
$\Delta f_T \times$ C × T	(80, 880)	1.142	0.193	0.094

Similarly, in the current experiment, for which there was no break between the induction and test sequences, the contrast effect observed is mostly sustained throughout the duration of the test sequence. Snyder *et al.* (2009) found that the effect of prior context not only extended over gaps of several seconds but could also be separated into stimulus-related (whether prior Δf was larger or smaller than current Δf) and perception-related components (whether the prior stimulus was perceived as one or two streams). In the context of the current study, note that the contrast effect associated with prior Δf is additive with the primary segregation-promoting effect of the mismatch in frequency (in either direction) between the L subset of inducer tones and its test-sequence counterpart. Hence, as can be seen in Fig. 4, the results for the conditions where $\Delta f_I = 0.5 \times \Delta f_T$ (C4) and $\Delta f_I = 1.5 \times \Delta f_T$ (C6) do not bracket those for the notionally intermediate standard AF condition, for which $\Delta f_I = \Delta f_T$ (C5), but rather bracket those for the high-subset-only condition (C2).

In addition to the significant main effects of all three factors, two of the two-way interactions were also significant—induction condition $\times$ time interval ($p < 0.001$) and $\Delta f_T \times$ time interval ($p < 0.001$). First, as for experiment 1, the pattern of change in stream segregation across the test sequence was strongly dependent on induction condition—the segregation-promoting effect of induction sequences in which one subset of tones did not match its counterpart in the test sequence was most apparent early on, such that the differences between the effects of the standard AF inducer and the other inducers typically declined over several seconds. Second, when averaged across conditions, larger values of Δf_T were associated with greater rises in reported stream segregation over the course of the test sequence.

IV. GENERAL DISCUSSION

It has long been known that the build-up of stream segregation that occurs during an unchanging AF tone sequence is typically reduced or lost altogether following a sudden change of sufficient size in both subsets of tones, leading to more integrated percepts (e.g., Anstis and Saida, 1985; Rogers and Bregman, 1998). The experiments reported here have shown, to our knowledge for the first time, that induction sequences for which one subset of tones precisely matches its counterpart in the test sequence, but the other does not, have the opposite effect—they cause more segregated percepts. This outcome illustrates a change in behavior towards that induced by listening to matched-CF sequences. There is no straightforward way to equate the magnitude of changes on different physical dimensions, such as the differences in level or frequency used here, but it merits comment that changes of 6 or 12 dB are quite substantial and yet the corresponding induction sequences caused considerably less promotion of stream segregation than the matched-CF inducer in experiment 1. In contrast, all changes in Δf_I used in experiment 2 led to subsequent levels of segregation much closer to those for the matched-CF inducer than to those for the standard-AF inducer. Although not conclusive, this outcome is consistent with the notion of the critical

importance of frequency differences in auditory perceptual organization proposed in the theory of indispensable attributes (Kubovy, 1981). A secondary effect was also apparent in experiment 2; the enhanced stream segregation observed following the abrupt frequency change for one subset of tones at the induction/test boundary was modulated by the size of the HL frequency difference for the induction sequence relative to that for the test sequence. This outcome may have the same origin as the context effect of prior Δf previously reported by Snyder and his colleagues (Snyder *et al.*, 2008, 2009; see also Snyder and Weintraub, 2011; Weintraub *et al.*, 2014).

Different cognitive accounts have previously been proposed for the stream-biasing effects caused by matched-AF and matched-CF induction sequences that can explain qualitatively differences in their strength and temporal characteristics. Bregman (1978) proposed that the default assumption of the auditory system is that a sequence of tones arises from one source and that the relatively slow build-up of stream segregation during an unchanging AF sequence reflects a conservative process of evidence accumulation in favor of a two-stream interpretation. In contrast, as noted earlier, the rapid-onset and strong promotion of segregation induced by matched-CF stimuli has usually been explained in terms of stream capture (Bregman and Rudnicky, 1975; Rogers and Bregman, 1993; Haywood and Roberts, 2013). The results of the current study—particularly those for experiment 2—represent a major challenge for this account of how stream biasing is caused by matched-CF inducers because, with the exception of the case where $\Delta f_I = 0$, the induction sequences with intermediate properties used here involved frequency alternation. Given the slow time constant for build-up in AF sequences, this must have reduced considerably—and in some cases completely prevented—the formation of a pre-established stream from the matched subset of inducer tones during the short interval available (2 s) before the test sequence began. To illustrate this point, consider the results for experiment 2 when $\Delta f_T = 4$ ST and $\Delta f_I = 0.5 \times \Delta f_T$ (C4), for which the corresponding $\Delta f_I = 2$ ST. Given that the results for the silent-induction case (C1) when $\Delta f_T = 4$ ST indicate a mean extent of stream segregation below 10% about 2 s into the test sequence, it seems likely that build-up during the 2-s induction sequence would be negligible when $\Delta f_I = 2$ ST. Nonetheless, overall stream segregation was close to (actually greater than) that for the corresponding matched-CF case (C2) and substantially higher than for the corresponding standard-AF condition (C5). Without the establishment of a monotonous stream composed of the matched-CF subset of tones during the induction sequence, on what basis could its counterpart tones in the test sequence be captured?

This argument against a capture account assumes the necessity of overtly experiencing the perception of segregated monotonous streams corresponding to the two subsets of inducer tones, but it cannot be ruled out that an internal representation of the two-stream interpretation exists without reaching conscious awareness. For example, a model of streaming by Mill *et al.* (2013) proposes a framework in which "proto-objects"—a set of candidate perceptual objects

J. Acoust. Soc. Am. **144** (6), December 2018

Rajasingam *et al.* 3417

consisting of predictable patterns (e.g., ABA–, A–A–, –B–)—are discovered for a sound sequence and evidence for them is accumulated over time, based on how well they can account for the sensory input. Different combinations of these proto-objects can together form a perceptual organization, and alternative perceptual organizations compete with one another to reach conscious awareness (i.e., the reported state, in this case "one stream" or "two streams"). Two assumptions are required to account for the segregation-promoting effects of the hybrid-AF inducers used in the experiments reported here. First, a proto-object must be capable of capturing subsequent sounds into the corresponding organization in the same way that has been supposed for an overtly experienced stream (cf. Bregman and Rudnicky, 1975; Rogers and Bregman, 1993). Second, proto-objects associated with the two-stream (as well as the one-stream) interpretation must be discovered rapidly (i.e., within the 2-s duration of the inducers). Consider, for example, how the stimuli used in experiment 2 for C4 ($\Delta f_I = 0.5 \times \Delta f_T$) and C6 ($\Delta f_I = 1.5 \times \Delta f_T$) might be represented if these assumptions were met. When the test sequence begins, the ABA– and –B– proto-objects discovered during the induction sequence are no longer supported, but the A–A– proto-object is consistent with the new scene and the new –B– proto-object is soon discovered. Note that, although this "proto-object capture" account of our results is plausible in principle, determining how best to evaluate it experimentally may prove challenging.

Thus far, neural models of auditory stream segregation have focused primarily on accounting for behavioral results obtained using unchanging AF sequences (for an exception, see Rankin et al., 2017). An early proposal was that the build-up of stream segregation in an AF sequence may be due to the adaptation of hypothetical frequency-jump detectors (van Noorden, 1975; Anstis and Saida, 1985) but, as noted by Rogers and Bregman (1993), the concept of frequency-jump detectors cannot account for the strong segregation-promoting effects of CF inducers because frequency-jump detectors would not respond—and so would not adapt—during that type of induction sequence. Fishman et al. (2001) performed the first direct investigation into the neural basis of streaming by recording multi-unit activity from primary auditory cortex (A1) in awake macaques during presentations of AF sequences of pure tones. The A-tone frequency was set at, or close to, the best frequency of the recording site and the B-tone frequency was varied. Consistent with behavioral reports of a more segregated percept, the neural response of A1 units to the B tones was attenuated at faster tone repetition rates and larger frequency separations. Subsequently, Micheyl et al. (2005) found that the suppression of B-tone responses increased throughout a 10-s sequence—similar to the time course observed behaviorally for the main phase of the build-up of stream segregation—indicating a progressive narrowing of frequency tuning for A1 units stimulated at best frequency. Neither of these studies included conditions in which one or other subset of tones changed abruptly.

Similar findings have since been reported for physiological studies of build-up in a variety of species and at different levels along the auditory pathway ranging from cochlear nucleus to auditory cortex (e.g., Pressnitzer et al., 2008; Bee et al., 2010). The physiological mechanism suggested to mediate the multi-second adaptation seen in response to unchanging AF sequences is long-term synaptic depression (Pressnitzer et al., 2008). In principle, this adaptation need not necessarily require stimulation away from a unit's best frequency—narrowing of frequency tuning might occur during CF as well as AF sequences. However, a more substantial modification of this neural model of streaming would be needed to account for the strong and rapid-onset segregation-promoting effect observed in human listeners for matched-CF induction sequences (Rogers and Bregman, 1993; Beauvois and Meddis, 1997; Roberts et al., 2008; Haywood and Roberts, 2013) and for inducers in which only one of the two tone subsets matched its counterpart in the test sequence (as used here). The role of attention in streaming tasks offers a means of bridging this gap.

There are many examples of the ways in which attention can influence the perceptual organization of tone sequences. For example, it has long been known that listening set—trying to hold a sequence together as a single stream or trying to attend to one or other subset of tones—influences both the overall likelihood of stream segregation and the effects of manipulating tone repetition rate and frequency separation (van Noorden, 1975). Although it is difficult to rule out the possibility that perceptual reports are influenced by response bias associated with the demand characteristics of the task, a recent study using stimulus-locked magnetoencephalographic activity in auditory cortex as a measure of whether listeners were experiencing one or two streams has provided evidence that the effect of intention on stream segregation is at least partly a low-level perceptual effect (Billig et al., 2018). There are also other contexts in which attention is known to influence stream segregation. For example, Thompson et al. (2011) have shown that the detection of a delay imposed on the B tone of a single ABA– triplet 12.5 s into a long sequence can be improved if build-up is disrupted by preventing listeners from attending to the sequence during the first 10 s, by requiring them to perform a task on competing stimuli presented in the other ear. Also, Kondo et al. (2012) showed that changes in lateralization cues in an AF sequence can cause resetting of build-up even if they arise from self-induced head motions, suggesting that stream segregation is directly influenced by a listener's active sensing of their environment, such as orienting the head towards relevant acoustic stimuli.

The segregation-promoting effect of matched-CF inducers, and of the hybrid-AF inducers used here (i.e., one matched tone subset and one mismatched), can be considered as another example of the attentional modulation of streaming. What both these types of induction sequence share is the continuity in acoustic properties of one set of tones and the sudden transition at the induction/test boundary—from silence to a new set of tones in the former case, or in the latter case a salient change in the properties of the other set either in level (experiment 1) or frequency (experiment 2). Thompson et al. (2011) suggested that matched-CF inducers may be segregation-promoting not because of

stream capture but because the attention of listeners is biased towards the novel tones in the test sequence. This argument can be extended to the hybrid-AF induction sequences used here—the sudden change in the properties of one subset of tones when the test sequence begins causes the new sounds to grab attention, leading to a fast-acting bias for stream segregation. It is also worth noting that an attention-switching account of the segregation-promoting effects of hybrid-AF inducers does not require an assumption that an internal representation of a two-stream organization (perceived or not) has formed by the end of the induction sequence.

Although the experiments reported here were not designed to test the attention-switching hypothesis (Thompson *et al.*, 2011), the observed outcomes are clearly compatible with it if we assume that the extent of attention switching is governed by the salience of the change. For experiment 1, the smooth and progressive rise in stream segregation found for greater attenuation of one or other subset of inducer tones—in the absence of any frequency change—can be interpreted in terms of a progressive rise in the salience of the sudden increase in level for the mismatched subset of tones. For experiment 2, the minimum frequency change at the induction/test boundary was 2 ST (for the case where $\Delta f_T = 4$ ST), and so it seems probable that all the sudden changes in pitch would have been highly salient, leading to strong promotion of stream segregation in all conditions where there was a change in frequency for one subset of tones. Note that, in principle, attention switching and proto-object capture may jointly contribute to perceptual organization—both the salience of the new sounds (switching) and the continuity of the old sounds (capture) may increase the likelihood of stream segregation following CF and hybrid-AF inducers. In terms of the neural model of streaming outlined above, the effects of selective attention on the responses of frequency-tuned units in the central auditory pathway—which have been found as early as in the cochlear nucleus—may arise from fast-acting efferent control of these units via descending projections of the medial olivo-cochlear efferent system (cf. Pressnitzer *et al.*, 2008).

In conclusion, the experiments reported here contradict the notion that the stream biasing associated with a matched constant-frequency induction sequence arises because the constituent tones capture their counterparts in the alternating-frequency test sequence into the on-going experience of a pre-established auditory stream. This is because the strong and fast-acting promotion of segregation associated with matched-CF inducers also occurs for hybrid-AF inducers, for which the tones of one subset match their counterparts in the test sequence but the others do not. For the short induction sequences used here, the presence of frequency alternation should greatly reduce—and, in some cases, eliminate—the possibility of experiencing a segregated monotonous stream capable in principle of capturing its test-sequence counterparts. As noted above, a modified version of the stream-capture hypothesis based on the role of unconscious proto-objects in perceptual organization cannot be ruled out at this point. However, it can only provide a plausible account of the results for the hybrid-AF conditions if it is assumed that a proto-object corresponding to the continuing subset of tones emerges during the short induction sequence, despite (except for the $\Delta f_I = 0$ case) the presence of frequency alternation throughout. Alternatively, or in addition, the results for matched-CF and hybrid-AF inducers are both compatible with the idea that the onset of the test sequence biases the attention of listeners towards the novel tones. The findings reported here help further to refine our understanding of the dynamics of auditory stream segregation.

Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–271.

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 **196**, 543–557.

Bendixen, A., Böhm, T. M., Szalárdy, O., Mill, R., Denham, S. L., and Winkler, I. (**2013**). "Different roles of similarity and predictability in auditory stream segregation," Learn. Percept. **5**, 37–54.

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.

Billig, A. J., Davis, M. H., and Carlyon, R. P. (**2018**). "Neural decoding of bistable sounds reveals an effect of intention on perceptual organization," J. Neurosci. **38**, 2844–2853.

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. (**2015**). "Progress in understanding auditory scene analysis," Music Percept. **33**, 12–19.

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., and Rudnicky, A. I. (**1975**). "Auditory segregation: Stream or streams?," J. Exp. Psychol. Hum. Percept. Perform. **1**, 263–267.

Cowan, N. (**1984**). "On short and long auditory stores," Psychol. Bull. **96**, 341–370.

Farkas, D., Denham, S. L., Bendixen, A., and Winkler, I. (**2016**). "Assessing the validity of subjective reports in the auditory streaming paradigm," J. Acoust. Soc. Am. **139**, 1762–1772.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (**2001**). "Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey," Hear. Res. **151**, 167–187.

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]" (Belmont, MA).

J. Acoust. Soc. Am. **144** (6), December 2018

Rajasingam *et al.* 3419

Hill, K. T., Bishop, C. W., and Miller, L. M. (**2012**). "Auditory grouping mechanisms reflect a sound's relative position in a sequence," Front. Hum. Neurosci. **6**, 158.

Jones, M. R., Kidd, G., and Wetzel, R. (**1981**). "Evidence for rhythmic attention," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1059–1073.

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

Kondo, H. M., Pressnitzer, D., Toshima, I., and Kashino, M. (**2012**). "Effects of self-motion on auditory scene analysis," Proc. Nat. Acad. Sci. **109**, 6775–6780.

Kubovy, M. (**1981**). "Concurrent-pitch segregation and the theory of indispensable attributes," in *Perceptual Organization*, edited by M. Kubovy and J. R. Pomerantz (Erlbaum, Hillsdale, NJ), pp. 55–98.

Kubovy, M., and van Valkenburg, D. (**2001**). "Auditory and visual objects," Cognition **80**, 97–126.

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.

Mill, R. W., Bőhm, T. M., Bendixen, A., Winkler, I., and Denham, S. L. (**2013**). "Modelling the emergence and dynamics of perceptual organisation in auditory streaming," PLoS Comput. Biol. **9**, e1002925.

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

Pressnitzer, D., and Hupé, J. M. (**2006**). "Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization," Current Biol. **16**, 1351–1357.

Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (**2008**). "Perceptual organization of sound begins in the auditory periphery," Current Biol. **18**, 1124–1128.

Rajasingam, S. L. (**2016**). "The dynamics of auditory stream segregation for tone sequences with gradually and abruptly varying stimulus properties," Doctoral thesis, Aston University, Birmingham, UK.

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 modeling," 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.

Snedecor, G. W., and Cochran, W. G. (**1967**). *Statistical Methods*, 6th ed. (Iowa University Press, Ames, IA).

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.

Snyder, J. S., and Weintraub, D. M. (**2011**). "Pattern specificity in the effect of prior Δf on auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **37**, 1649–1656.

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

van Valkenburg, D., and Kubovy, M. (**2003**). "In defense of the theory of indispensable attributes," Cognition **87**, 225–233.

Weintraub, D. M., Metzger, B. A., and Snyder, J. S. (**2014**). "Effects of attention to and awareness of preceding context tones on auditory streaming," J. Exp. Psychol. Hum. Percept. Perform. **40**, 685–701.