Sequential streaming, binaural cues and lateralization

Marion David, Mathieu Lavandier, Nicolas Grimault

Sequential streaming, binaural cues and lateralization

Marion David[a] and Mathieu Lavandier
Université of Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue M. Audin, F-69518 Vaulx-en-Velin Cedex, France

Nicolas Grimault
Cognition Auditive et Psychoacoustique, Centre de Recherche en Neurosciences de Lyon, Université Lyon 1, UMR CNRS 5292, Avenue Tony Garnier, 69366 Lyon Cedex 07, France

(Received 16 September 2014; revised 23 October 2015; accepted 3 November 2015; published online 9 December 2015)

Interaural time differences (ITDs) and interaural level differences (ILDs) associated with monaural spectral differences (coloration) enable the localization of sound sources. The influence of these spatial cues as well as their relative importance on obligatory stream segregation were assessed in experiment 1. A temporal discrimination task favored by integration was used to measure obligatory stream segregation for sequences of speech-shaped noises. Binaural and monaural differences associated with different spatial positions increased discrimination thresholds, indicating that spatial cues can induce stream segregation. The results also demonstrated that ITDs and coloration were relatively more important cues compared to ILDs. Experiment 2 questioned whether sound segregation takes place at the level of acoustic cue extraction (ITD *per se*) or at the level of object formation (perceived azimuth). A difference in ITDs between stimuli was introduced either consistently or inconsistently across frequencies, leading to clearly lateralized sounds or blurred lateralization, respectively. Conditions with ITDs and clearly perceived azimuths induced significantly more segregation than the condition with ITDs but reduced lateralization. The results suggested that segregation was mainly based on a difference in lateralization, although the extraction of ITDs might have also helped segregation up to a ceiling magnitude.
 [http://dx.doi.org/10.1121/1.4936902]

[VMR] Pages: 3500–3512

I. INTRODUCTION

In a context of competing sound sources, the main idea of auditory scene analysis (ASA; Bregman, 1990) is that the auditory system enables the listener to group the sound events coming from the same source (i.e., integrated percept) and segregate them from other sound events coming from different sources (segregated percept). Many sound properties can influence stream formation, such as differences in fundamental frequency, temporal envelope, spectrum, or lateralization (see Moore and Gockel, 2002, 2012, for reviews). The main purpose of the present study was to focus on the influence of spatial differences on stream segregation.

Sounds coming from different locations in space present both monaural and binaural differences (Wightman and Kistler, 1992). First, the source spectrum produced at each ear depends on the listener and source positions. Sound is submitted to several frequency-dependent reflections during its propagation, and the sound at the ears results from the addition of the direct sound and a given combination of filtered reflections (Collin and Lavandier, 2013; Flanagan and Lummis, 1970; Larsen *et al.*, 2008). These monaural spectral differences associated with filtering by the head (and room, if any) will be referred to as *coloration* differences later in this paper. Second, sounds coming from different locations in space present binaural differences: interaural time and level differences (ITDs and ILDs, respectively; Carlile, 1996; Middlebrooks and Green, 1991).

In the presence of competing speech sounds, Cherry (1953) reported that the spatial separation of the target and masker signals was a major factor contributing to the improvement of speech recognition. This experiment was an example of voluntary stream segregation, since the listeners tried to hear out a target stimuli from a mixture of sounds (Bregman, 1990). Other studies, like the ones discussed below, using tasks in which performance was favored by segregation were markedly influenced by spatial cues, regardless of the type of stimuli used (pure tones, harmonic complex tones, broadband noises, or speech).

Hartmann and Johnson (1991) showed better performance in a melody recognition task using pure tones when the target and the interleaved masker differed in ILDs by 8 dB, or when they differed in ITDs by 500 μs. Sach and Bailey (2004) obtained similar results using a task where listeners had to identify a target rhythm interleaved with arrhythmic masking tones. The listeners reported a better identification accuracy when target and masker differed in ILDs (4 and 0 dB, respectively) or in ITDs (from 100 to 600 μs). Gockel

[a]Present address: Auditory Perception and Cognition Lab, Department of Psychology, University of Minnesota, Elliott Hall, 75 East River Parkway, Minneapolis, MN 55404, USA. Also at: Cognition Auditive et Psychoacoustique, Centre de Recherche en Neurosciences de Lyon, Universite Lyon 1, UMR CNRS 5292, Avenue Tony Garnier, 69366 Lyon Cedex 07, France. Electronic mail: david602@umn.edu

et al. (1999) measured to what extent ITDs could influence the threshold for detecting a change in F0 of a complex tone. In some conditions, the target was preceded and followed by harmonic complexes temporally adjacent to the target sound (i.e., temporal "fringes"). The results showed that the impairment induced by the fringes was reduced when their ITD was shifted away from the ITD of the target. Recently, Middlebrooks and Onsan (2012) reported voluntary stream segregation with broadband noise bursts based on binaural cues (ILDs and ITDs) as well as on monaural cues (i.e., coloration). Darwin and Hukin (1999) demonstrated that ITDs could influence sequential grouping of speech sounds. They showed that listeners tended to group a target word with a sentence more often if they shared the same ITD. Kidd *et al.* (2008) presented sequences of interleaved target and masker words and the task consisted of tracking the target words. The results indicated that the percentage of correct identification increased when a difference in apparent location induced by a difference in ITDs (from ± 150 to $\pm 700\,\mu$s) was applied to the target words.

In contrast to voluntary stream segregation, obligatory stream segregation refers to tasks where the listeners are biased towards grouping, but fail to group (Bregman, 1990). In these types of tasks, segregation impairs performance. Given the definitions proposed by van Noorden (1975), the temporal coherence boundary (TCB) represents the critical value of the considered parameter above which listeners are no longer able to hear the sequence as one coherent stream, and the fission boundary (FB) represents the critical value under which listeners are no longer able to hear the sequence as segregated streams. The thresholds of obligatory and voluntary stream segregation correspond to the TCB and FB, respectively. Since TCB requires a larger stimulus dissimilarity than FB, obligatory stream segregation is more restrictive than voluntary stream segregation.

Many previous studies failed to report an effect of binaural cues on obligatory stream segregation. For instance, Boehnke and Phillips (2005) found no significant improvement in segregation for broadband noises differing in ITDs in a gap discrimination task. This procedure required the listeners to integrate the streams in order to detect the gap changes. However, the sound sequences used lasted only 330 ms, and they might have been too short for the build-up of segregation to occur (Anstis and Saida, 1985; Roberts *et al.*, 2008). Stainsby *et al.* (2011) showed that ITD influenced the obligatory stream segregation of complex tones in a rhythmic discrimination procedure, but only for ITD values outside the physiological range (i.e., 1 ms and above). Finally, Füllgrabe and Moore (2012) replicated the experiment of Stainsby *et al.* (2011) with pure tones and ITDs below 500 μs. They found only a weak effect of ITDs on obligatory stream segregation. These results might be explained first by the fact that pure tones provide less binaural information than broadband stimuli, as they only involved one frequency; and second by the fact that ITDs can lead to an ambiguous perceived position for pure tones (Moore, 2007).

David *et al.* (2014) found that the monaural spectral differences associated with a difference in spatial position can influence obligatory stream segregation of broadband noise bursts. A subjective streaming task was run, where the listeners had to indicate at the end of the sequence whether they heard one single stream or two separate streams, in addition to an objective rhythmic discrimination task. The results from the two tasks were well correlated. The present study assessed whether adding the binaural cues (ILDs and ITDs) to the coloration cues could enhance obligatory stream segregation (experiment 1).

The relative importance of each spatial cue for stream segregation is still a matter of debate in the literature. In the study by Middlebrooks and Onsan (2012), a rhythmic discrimination task was conducted to measure the influence of spatial separation on stream segregation. The listeners were presented with sequences consisting of a target sequence, with a specific rhythm, and an interfering masker, with a complementary rhythm. They had to discriminate between two target rhythms, and to do so they had to separate the masker and the target into separate streams. Both target and masker were broadband noise bursts coming from sources that differed in azimuth and elevation (thus sounds containing ITDs, ILDs and monaural spectral differences). The results in the horizontal plane showed that segregation was mainly influenced by ITDs rather than by ILDs. In the vertical plane, results indicated that listeners could rely on monaural spectral differences induced by coloration to separate the sounds. However, performance was clearly worse in the vertical plane than in the horizontal plane, suggesting that monaural cues were of less importance than binaural cues for stream segregation. Bremen and Middlebrooks (2013) showed that ITDs were more important than ILDs—or coloration cues—for the segregation of complex tones. Conversely, Schwartz *et al.* (2012) demonstrated that ITDs did not help stream segregation in an identification task. The broadband stimuli used in their study presented natural speech similarities but with less grouping cues (i.e., no harmonicity nor onset/offset cues). The target and masker stimuli differed only in ITDs. They were presented simultaneously several times, followed by a 500-ms silence gap after which a probe stimulus was played. The listeners had to indicate whether the probe stimulus matched the target in the preceding sequence. The results showed that a difference in ITDs between target and masker did not produce an accurate segregation of the mixture. The study was intended to provide more information relevant to this debate by investigating the segregation of broadband noises based on spatial cues, adding these cues progressively to evaluate their relative influence (experiment 1).

Binaural cues allow for sound lateralization (Carlile, 1996; Middlebrooks and Green, 1991). In the horizontal plane, sound lateralization accuracy is largely based on the spatial dependence of the interaural difference cues. Thus, if binaural cues could facilitate segregation, it might be explained by a difference in binaural cues *per se*, but also by the corresponding difference in perceived azimuth. As far as we know, this question of whether the organization of sound events takes place at the level of acoustic cue extraction or later at the level of object formation has not been clearly addressed in the literature. Experiment 2 of the present study

investigated this question by manipulating separately ITD and lateralization to evaluate their relative influence on the obligatory segregation of broadband noise bursts.

II. GENERAL METHODS

A. Rhythmic discrimination paradigm

The rhythmic discrimination procedure described by Roberts *et al.* (2002) was used in the present study to investigate whether spatial cues could induce obligatory streaming. This procedure has been widely used to objectively measure obligatory streaming (Füllgrabe and Moore, 2012; Roberts *et al.*, 2008; Stainsby *et al.*, 2011; Stainsby *et al.*, 2004; Thompson *et al.*, 2011). It involved the presentation of two intervals of alternate noise bursts [A-B-A-B-···]. In the target interval, the first six AB pairs were regularly spaced by 40 ms (i.e., the B's were placed at the temporal midpoint between two consecutive A's). The B's were then progressively delayed by equal steps for the next four pairs. Thus, the seventh B was delayed by δT, and the three next B's were delayed by $2\delta T$, $3\delta T$, and $4\delta T$, respectively. Finally, the cumulative delay ΔT ($\Delta T = 4\delta T$) was kept constant for the last two pairs. In the reference interval, the silence duration between consecutive stimuli was always 40 ms, leading to a regular rhythm. The silence duration between two intervals was set to 1 s. Figure 1 illustrates the rhythmic discrimination paradigm. The listener's task was to identify the target sequence with the delayed B's among the two intervals.

According to this paradigm, the perceived rhythm of the target interval depends on whether the listener hears a single stream or two segregated streams (see Fig. 1). The delay applied to the B's is more easily detectable when a single stream is heard because successive time intervals [A-B] and [B-A] are compared and ΔT is not negligible compared to these interval durations. Conversely, the delay applied to the B's is more difficult to detect when the streams are segregated because successive [B-B] intervals are compared and these intervals are longer compared to ΔT. Thus, the rhythmic irregularities are better detected when the percept is integrated; segregation impairs task performance.

The rhythmic discrimination paradigm leads to a threshold measurement for detecting anisochrony (more details are given in Sec. II C). Low thresholds indicate that the listeners are able to fuse the streams to detect the irregular sequence. So, lower thresholds mean higher performance. Conversely, high thresholds indicate a failure to fuse the streams and to identify the irregular sequence. Thus, this paradigm gives an indirect measure of stream segregation, but Roberts *et al.* (2002) and Micheyl and Oxenham (2010) showed consistency between this measure and perceived-segregation judgments. This result has also been demonstrated by David *et al.* (2014) with the same experimental design as in the present study (i.e., type and duration of stimuli, onset-to-onset time). In other words, higher thresholds are associated with a clearer percept of segregation while lower thresholds are associated with a clearer percept of integration.

B. Global characteristics of the stimuli

Bursts of speech-shaped noise (SSN) were used to generate the stimuli A and B. SSNs were stationary noises with a spectrum similar to the long-term spectrum of speech, so approximately flat from 0 to 1000 Hz and then decreasing by 20 dB per octave (ANSI, 1989). It is worth noting that the spectrum of an excerpt of SSN depends on its duration. Indeed, as shown in Fig. 2, the spectrum of a long SSN is approximately flat, but those of short excerpts present fluctuations. These fluctuations increase when the stimuli are shortened. So, in order to limit this spectral variability, the stimuli had to be longer than what is traditionally used in obligatory streaming studies (around 60 ms, Roberts *et al.*, 2002).

The stimuli also had to be short enough to produce rapid sequences, otherwise no obligatory stream segregation could be observed. Indeed, van Noorden (1975) reported that once the tone repetition time (TRT) exceeded about 150 ms, the probability of reporting obligatory stream segregation became negligible. However, van Noorden used 40-ms

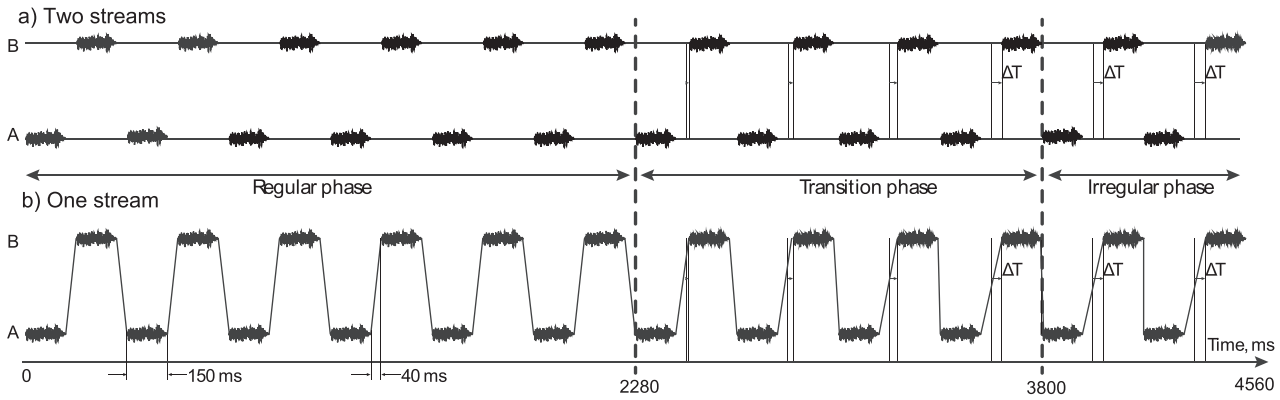

FIG. 1. Schematic representation of irregular sequences presented for the rhythmic discrimination task. The sequences consisted of 12 pairs of alternate noise bursts. In the irregular sequence, the B bursts were initially positioned at the exact temporal midpoint between two successive A bursts (regular phase), then they were progressively delayed in the transition phase. In the irregular phase the cumulative delay applied to the B bursts was kept constant. In the regular sequence (not plotted), the B bursts remained at the temporal midpoint between the A bursts for the entire sequence. The irregular sequence could lead to two different percepts depending on the segregation state of the streams A and B [segregated in the top panel (a) and integrated in the bottom panel (b)].

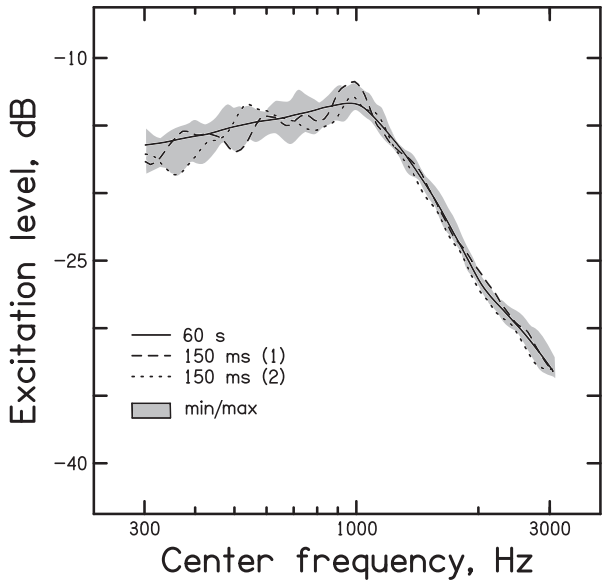

FIG. 2. Excitation patterns of a long speech-shaped noise (SSN) of 60 s and of two different SSN excerpts of 150 ms extracted from this long SSN. The patterns of the short-duration SSNs depend on the particular time epoch where the excerpt was extracted. The gray-colored zone represents the maximum and minimum excitation levels of 10 different short SSNs (upper and lower limits, respectively).

stimuli separated by 110-ms of silence. Bregman *et al.* (2000) showed that for a given TRT, longer tones—shorter inter-tone intervals—produced greater segregation. That might explain why van Noorden (1975) did not report obligatory stream segregation above a 150-ms TRT. Micheyl and Oxenham (2010) investigated to what extent segregation of pure tones can be influenced by the frequency difference between the tones (ΔF), the length (N) and the rate of the sequence. The time interval between two consecutive tones was the same as the tone duration (T). Since T was set to 50 or 100 ms, the measure tested two TRT (i.e., onset-to-onset), 100 and 200 ms. In an objective task where segregation impaired performance (experiment 2), the thresholds were significantly higher (so more segregation) in the fast condition compared to the slow condition. However, the results in the slow condition were significantly influenced by N and ΔF as in the fast condition. This result showed that obligatory stream segregation can be observed with a TRT as long as 200 ms. A reasonable trade-off was reached in the present study by generating 150-ms stimuli, which led to a TRT of 190 ms since the longer inter-stimulus interval was 40 ms. These stimuli allowed for obligatory stream segregation to be investigated, and at the same time presented only small spectral variability.

Because the spectrum of a short excerpt of SSN depends on which segment of a long SSN is chosen (see Fig. 2), ten samples of SSN were used in order to average out the spectral peculiarities associated with the choice of a particular sample (as in David *et al.*, 2014). In Fig. 2, the gray-colored zone represents the maximum and minimum excitation levels obtained with these ten short SSNs. The ten "frozen" samples were used to synthesize each A-B pair. The same ten samples had to be used in all conditions to prevent the potential confounding effect of spectral difference associated with the choice of a random SSN.

C. Experimental design

Thresholds for detecting anisochrony were estimated with a two-interval, two alternative forced-choice method. The delay applied to the B bursts was adapted according to a three-down, one-up rule and varied on a logarithmic scale. This method determines the 79.4% of correct responses on the psychometric curve (Levitt, 1971). It is worth noting that higher thresholds suggest a greater tendency for the streams to be heard as segregated. The maximum delay was 40 ms, which corresponded to the maximum delay without overlap between two consecutive stimuli. The initial value of the delay was 28.28 ms. A measurement reached saturation when ten consecutive incorrect answers were provided with a ΔT of 40 ms. Saturated measurements were assigned a threshold of 40 ms (Roberts *et al.*, 2002). If more than 50% of the measurements reached saturation for a given listener, his/her data were discarded from the analysis (Devergie *et al.*, 2011). This occurred only for one participant, in experiment 2.

Since a set of ten pairs of stimuli was used, each pair had to have the same probability of being presented to the listeners. So the method of threshold estimation differed slightly from the one used by Roberts *et al.* (2002). Each run (i.e., each adaptive staircase) was divided into successive blocks of ten trials. For each trial, one SSN was drawn among the set of available SSNs (ten to start with), without replacement, to synthesize A and B. For the next trial, a different SSN was drawn from the remaining samples without replacement, and so on for the next trials until the ten samples were used. Note that the runs consisted of blocks of ten trials rather than using random draws to ensure that the same set of samples of SSNs were used in each condition.

When the listeners gave three correct answers, the delay applied to the B bursts decreased by a factor of 1.414, and it increased by the same factor when listeners gave one wrong answer. If at least two reversals were obtained at the end of a block of ten trials, the step factor was reduced to 1.189; otherwise, it was kept constant for another block until at least two reversals were obtained. Once the step factor was reduced, the number of reversals was reset, and the procedure continued until an even number of reversals greater than or equal to four was obtained at the end of a block. Finally, thresholds were estimated using the geometric mean of the reversals from the entire set of blocks which used the smallest step factor. As the number of reversals which could be obtained in one block varied from zero to four, thresholds were calculated at the end of each run with four, six, eight or ten reversals. The present procedure was at least as accurate as the procedure of Roberts *et al.* (2002) where thresholds were estimated using four reversals.

The experimental design was similar in experiments 1 and 2, only the tested stimuli differed. Both experiments were run in a double-walled sound-attenuated booth. The stimuli were sampled at 44.1 kHz and presented through Sennheiser HD650 headphones using a LynxTwo-B

soundcard. The mean sound level across the two ears was set to 70 dB sound pressure level (SPL) for each condition of the two experiments.

Listeners used a computer keyboard and mouse to enter their answers on a graphical interface visible on a screen placed outside the booth. One run lasted approximately 10 min and five runs were completed for each condition in each experiment. The experiments were divided in five 1-h sessions and all the conditions were tested once in each session. In order to get familiar with the task, participants did 20 trials before each session, where ΔT took pseudo-random values between 0 and 40 ms. For this familiarization session, diotic SSN samples of 150 ms were presented, and visual feedback was given by displaying a green or a red square after a correct or a wrong answer, respectively. After one or two familiarization sessions, the listeners verbally reported that they clearly understood the task. No feedback was given during the test session to prevent any possibility for the listeners to understand the three-down, one-up rule and thus give unreliable responses.

III. EXPERIMENT 1: REALISTIC ITDS AND ILDS

A. Rationale

The aim of experiment 1 was to investigate whether obligatory stream segregation could be influenced by spatial cues (i.e., coloration cues, ILD and ITD). These cues, associated with real positions in an anechoic room, were introduced progressively in the different conditions, to assess to what extent they were useful in the streaming process. Four conditions were tested. In the first condition, stimuli A and B were identical and diotic (reference condition). In the second condition, A and B were also diotic, but monaural spectral differences induced by head coloration were introduced (coloration condition). ILDs were introduced in the third condition (ILD condition), and ITDs were added in the fourth condition (ILD+ITD condition).

As the stimuli were synthesized based on real Head Related Impulse Responses (HRIRs), the ILDs and ITDs—when present—were preserved as a function of frequency (Feddersen *et al.*, 1957; Kuhn, 1977). The aim was to obtain images through headphones lateralized around the head. Because the non-individualized HRIRs used were measured with a KEMAR mannequin and thus did not perfectly match with the individual HRIRs of each listener, a good sound externalization outside the head was not expected. A localization task was used to quickly verify this point.

B. Stimuli synthesis

Spatial information was conveyed to stimuli A and B by convolving the bursts of SSN with HRIRs. The HRIRs were measured by Gardner and Martin (1995) in the horizontal plane, using loudspeakers mounted at a distance of 1.4 m from a KEMAR mannequin (Knowles Electronic model DB-4004) in an anechoic room. SSNs of 700 ms were convolved with the HRIRs, then a window with 12.5-ms raised-cosine on- and off-sets was applied to the middle of the convolved SSNs, leading to stationary bursts of 150-ms duration. Finally, the left-ear and right-ear channels of each stimulus were divided by the root-mean-square (rms) value averaged across the two channels. This equalization procedure preserved ILDs (when present) and led to a mean level across the ears of 70 dB SPL, as measured with an artificial ear (Larson Davis AEC101 and 842; ANSI, 1995).

In the reference condition, A and B resulted from the convolution of the SSNs with an HRIR measured at 0° azimuth. Both stimuli were identical and diotic (ILDs and ITDs of zero). For the remaining three conditions, A and B were synthesized by convolving the SSNs with HRIRs measured at +30° and −30°, respectively. In the coloration condition, only the left channel (arbitrarily chosen) of the stimuli was used and sent to both ears. This way, A and B were diotic (no interaural differences), however they were different because of the spectral differences induced at the left ear of the mannequin for a source at the two different azimuths. In the ILD condition, for each stimulus, all the fast Fourier transform (FFT) components were rotated into a null phase before computing an inverse FFT. This manipulation cancelled the ITDs—broadband as well as within-band—from the stimuli while preserving the ILDs. Finally, in the ILD+ITD condition, no processing was used on the SSNs convolved with the HRIRs, so that both ILDs and ITDs were preserved.

C. Listeners

Experiment 1 involved 14 listeners. They were students, aged between 20 and 27 year (ten females, mean age = 22 year, standard deviation [SD] = 2 year), they signed a general consent form before the experiment, and had self-reported normal hearing. They were paid an hourly wage for their participation, and came for five 1-h sessions.

D. Localization task

Eight listeners took part in a short localization task. None of them participated in any of the other experiments of the present study. They were presented with eight one-stream sequences of frozen stimuli (A and B of each condition), randomly chosen among the set of stimuli used in experiment 1. The task consisted of localizing the stimuli of the presented sequences in the azimuthal plane by drawing the perceived image of the stimuli on an illustration of a human head. Each listener completed four repetitions of this task. No precise instructions were given, so that they could point to a precise position or a spread area, inside or outside the head. The middle of the drawings was determined and the results were expressed in terms of vectors with a particular direction (pointing towards the middle of the drawing—azimuthal dimension) and length (from the middle of the head to the middle of the drawing—distance dimension). The response vectors were projected on the azimuthal and distance dimensions in order to obtain an estimation of the lateralization and externalization percepts, respectively. The results were then averaged over the four repetitions and the eight listeners.

E. Results

1. Temporal discrimination task

Figure 3 presents the geometric means across listeners of the temporal discrimination thresholds measured in experiment 1. For each condition, the mean thresholds for each listener across the five repetitions are displayed with different symbols. From left to right, the bars correspond to the reference, coloration, ILD and ILD+ITD conditions. The error bars represent the geometric standard errors across listeners. The arrow on the right indicates the direction of greater segregation. Note that two listeners had thresholds greater than the mean thresholds in all conditions. However, these particular listeners did not reach saturation in more than 50% of the trials, thus their data were included in the analysis.

The log values of these thresholds were assessed using a one-way repeated-measures analysis of variance (ANOVA), which showed that the effect of tested condition was significant [$F(3,52) = 17.11$, $p < 0.001$]. A *post hoc* analysis (*t*-test with Bonferroni corrections) indicated that the mean threshold obtained in the reference condition was significantly lower than those obtained in the three other conditions ($p < 0.0001$ in each case) and that the threshold was higher in the ILD+ITD condition compared to the coloration and ILD conditions ($p = 0.0015$ and $p = 0.0106$, respectively). There was no significant difference in thresholds between the coloration and ILD conditions.

2. Localization task

In the subjective localization task, the hypothesis was made that listeners' responses followed 1 Gaussian distribution. So, the individual results of the localization task were first projected on the azimuthal dimension, fitted with Gaussian distributions for which the mean and the SD corresponded to the middle and the spread of the drawing, and finally averaged across listeners. Figure 4 shows the mean distributions for each condition. The mean perceived positions of stimuli A and B correspond to the dotted gray and filled black lines, respectively. The mean distributions were normal in each condition according to a Kolmogorov-Smirnov test. The parameters of these mean distributions were estimated using a maximum likelihood estimation (MLE) procedure. As expected, the stimuli in the diotic conditions (reference and coloration) were perceived at 0° azimuth, and the stimuli in the ILD+ITD condition were lateralized to the left (80.5° for A) and to the right (−81.6° for B). Note that even if the distributions in the ILD condition were found statistically normal, no reliable estimation of the parameter could be evaluated. The values of the mean distributions were assessed using a one-way ANOVA. The main effect of the tested condition was significant ($p < 0.001$) and a *post hoc* analysis [least significant difference (LSD)] showed that the distributions of A and B were significantly different only in the ILD+ITD condition. This test also indicated that the distributions of the A's in the ILD and ILD+ITD conditions were significantly different, as well as for the B's in these two conditions.

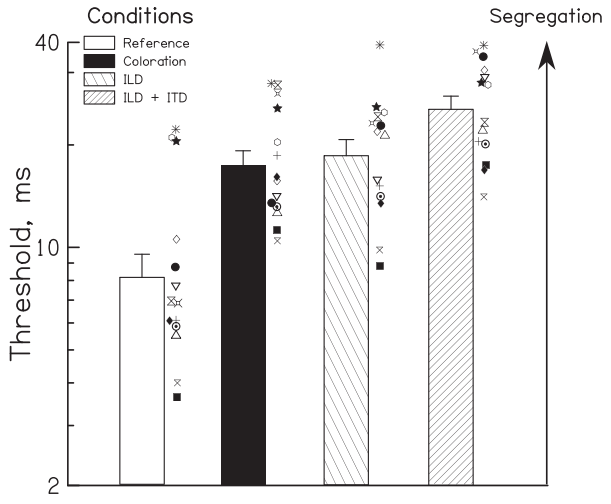

FIG. 3. Geometric mean thresholds in ms with geometric standard errors across participants for detection of the delay applied to the B bursts in experiment 1. The different symbols represent the mean data of individual listeners. From left to right, the bars correspond to the reference, coloration, ILD and ILD+ITD conditions. The arrow on the right indicates the direction of greater segregation.

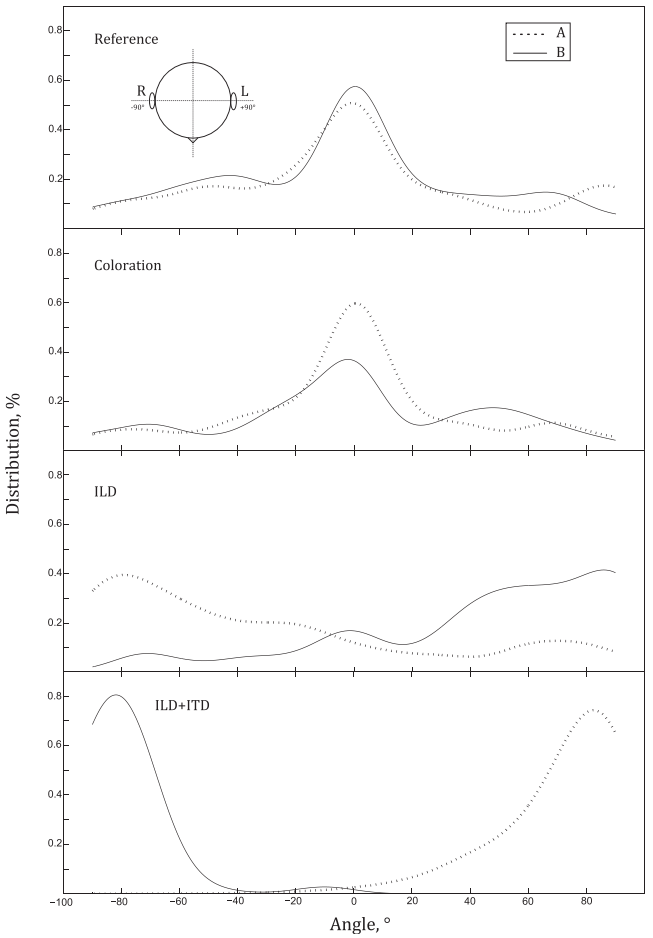

FIG. 4. Mean distributions of the perceived azimuth for the stimuli of experiment 1. The curves represent the mean distribution across eight naive listeners for the stimuli A (dotted lines) and B (solid lines) in each condition. From top to bottom, the panels correspond to the reference, coloration, ILD and ILD+ITD conditions.

The response vectors were also projected on the distance dimension in order to discriminate between intracranial and extracranial percepts (Hartmann and Wittenberg, 1996).

The results are given as a ratio (distance in cm compared to the head radius in cm). When the ratio was smaller than 1, the stimulus was perceived inside the head and when it was greater than 1, the stimulus was perceived outside the head. The results show that the stimuli were perceived inside the head in all the conditions except the ILD+ITD condition. Indeed, the ratio was equal to 0.7 in the two first conditions (standard error = 0.08), 1 in the ILD condition (standard error = 0.08) and 1.3 (standard error = 0.07) in the ILD+ITD condition.

It is worth noting that the stimuli in the ILD+ITD condition were perceived at around ±80° azimuth (see Fig. 4, ILD+ITD condition) while the stimuli were filtered with HRIRs from ±30° azimuth. This result might be explained by the use of non-individualized HRIRs that could lead to an issue with the virtual acoustic space of the stimuli.

F. Discussion

In a previous study, David *et al.* (2014) used the same SSNs and rhythmic discrimination task as in the present study, and they also ran a control experiment consisting of sequences of ABA triplets. In this subjective experiment, the listeners had to indicate, at the end of the sequence, whether they heard one single stream or two separate streams. The results of the subjective task were highly consistent with those of the objective task. Because the same procedure and type of stimuli were used in the present study, it can be assumed that the thresholds were related to perceived-segregation judgments. Thus, low thresholds reflected high performance and a greater tendency to hear the sequence as a single stream and high thresholds reflected low performance and a greater tendency to hear the sequences as two separate streams.

The mean threshold obtained in the reference condition was significantly lower compared to the coloration condition. In the latter, the stimuli presented monaural spectral differences (including a broadband level difference and within-band differences induced by coloration) associated with the difference in spatial locations. These results are consistent with those obtained by David *et al.* (2014), since they observed obligatory streaming by introducing only monaural within-band differences associated with coloration. The effects on segregation of the within-band spectral differences and of the broadband level difference could be additive, because the difference in thresholds induced by coloration was larger in the present study (9.25 ms) compared to the previous study (3.40 ms) which used the same stimuli without the broadband level difference. The present results are also in agreement with those obtained by Middlebrooks and Onsan (2012) who showed that voluntary streaming could be obtained in the median plane where head filtering induced subtle spectral differences which depend on position.

The mean threshold was significantly lower in the reference condition compared to the ILD condition. This result confirms that SSN bursts can be segregated when they differed in ILDs, in agreement with Stainsby *et al.* (2004). In addition, there was no significant difference in mean threshold between the coloration and ILD conditions. This suggests that adding the "binaural component" of ILD did not influence streaming. The monaural spectral differences induced by coloration at each ear are sufficient to explain the increase of discrimination thresholds and the corresponding improvement in segregation. Thus, listeners seemed to organize the incoming sounds based on the spectral variations across time at each ear rather than on the spectral difference across ears.

The mean threshold was significantly higher in the ILD+ITD condition compared to the ILD condition. This result suggests that ITDs significantly favored obligatory streaming between the sounds coming from competing sources which were spatially separated. The stimuli were generated using recorded HRIRs, so the ITDs were in the physiological range (±272 μs). According to this result, realistic ITDs could favor segregation whereas previous studies showed that ITD had no or only a limited influence on obligatory streaming (Füllgrabe and Moore, 2012; Stainsby *et al.*, 2011). This difference within the literature might be explained by the nature of the stimuli used: the broadband noises used in the present study might have lead to stronger binaural cues than the pure or complex tones used in previous studies.

The localization task indicated that ILDs alone produced an inaccurate percept of lateralization. There was a difference in lateralization between stimuli in the coloration and ILD conditions, but no difference in streaming. It is worth noting that these two experiments should be compared with caution because they involved different listeners. Nevertheless, one hypothesis could be that ILDs did not induce enough lateralization to enhance obligatory stream segregation. When both ILDs and ITDs were present the stimuli were clearly lateralized. Wightman and Kistler (1992) showed that ITD information is more likely to influence the judgment of source location than ILDs when low frequencies are present because ITD is relatively constant across frequencies whereas ILD is highly dependent on frequency. The present result is coherent with this finding. Adding ITDs leads to a clear percept of lateralization and enhanced obligatory stream segregation.

IV. EXPERIMENT 2: ITDS VERSUS LATERALIZATION

A. Rationale

Experiment 1 showed a significant effect of ITD on stream segregation which could be due to the interaural difference *per se* and/or to the associated lateralization. The aim of experiment 2 was to investigate separately the potential influence of these two cues. The distinction between interaural differences and the corresponding lateralizations might be useful to determine whether the sounds were separated at the level of acoustic cue extraction, or later at the level of object formation (Darwin and Hukin, 1999).

In order to test for the relative importance of ITDs and perceived azimuth, the stimuli were high- and low-pass filtered, allowing the two frequency regions to be manipulated separately, i.e., different ITDs were applied to each frequency region as done by Edmonds and Culling (2005). Note that in this experiment, artificial ITD values were used, so that stimuli were expected to be perceived inside the

head, more or less lateralized depending on the condition tested. When ITDs were consistent across the whole spectrum, it was assumed that perceived azimuths would be clearly identified. This lateralization percept would be blurred when ITDs were inconsistent between high- and low-frequency regions. This way, a "consistent" condition—with ITDs and associated perceived azimuths—could be compared to an "inconsistent" condition—with ITDs but reduced lateralization.

B. Stimuli synthesis

To synthesize the stimuli, 10-s SSNs were first high- and low-pass filtered using fourth-order Butterworth filters and a cutoff frequency of 550 Hz. This cutoff frequency was chosen first to have approximately the same energy at high and low frequencies, and second to be sure that ITDs were usable in the two frequency regions (Feddersen *et al.*, 1957). The high- and low-stop frequencies were 592 and 507 Hz, respectively, leading to a gap to avoid any overlap between the two regions containing a potentially different ITD (Edmonds and Culling, 2005). Temporal delays were applied across the ears in each of the two frequency regions to simulate broadband ITDs. Thus, within each frequency region, the ITDs were artificial and did not depend on frequency, but they could still allow for sound lateralization. Then, the two parts of the spectrum were concatenated and an inverse FFT was performed. The stimuli were time-windowed in the middle of the 10-s signal using 12.5-ms on- and off-cosine ramps to obtain stationary bursts of 150 ms. Finally, the two channels of each stimulus were independently equalized in rms to reach the same level of 70 dB SPL at each ear. Note that conversely to experiment 1, the stimuli were not convolved with HRIR and did not present ILDs. Instead, they presented only ongoing broadband ITDs.

C. Conditions and lateralization evaluation

Figure 5 presents the conditions tested in experiment 2. In the first condition (reference condition), the A and B bursts were identical, without ITD, with the hypothesis that they would both be perceived as coming from the middle of the head. In the second condition (272-μs consistent condition), A and B had consistent ITDs across the whole spectrum of 272 and -272 μs, respectively, which correspond to the broadband ITDs in experiment 1. A and B were supposed to be lateralized to the left and to the right, respectively. The third condition (500-μs consistent condition) was identical to the second condition except that the ITD magnitude was increased to 500 μs. In this condition, A and B were expected to be lateralized to the left and to the right, respectively, with a larger azimuth than in the second condition. Finally, in the fourth condition (500-μs inconsistent condition), the high-frequency band of stimuli A and the low-frequency band of stimuli B were presented with an ITD of 500 μs while the low-frequency band of stimuli A and the high-frequency band of stimuli B were presented with an ITD of -500 μs. The lateralization of the two stimuli was supposed to be blurred in this condition. Applying ITDs of equal magnitude but opposite signs in each frequency region kept the interaural difference at 500 μs in each region, as in the third condition. The only thing that differed from one region to the other was to which ear the signal was leading. So, if interaural differences are the main factor influencing stream segregation, one would expect the same thresholds in the 500-μs consistent and inconsistent conditions. Otherwise, if lateralization is

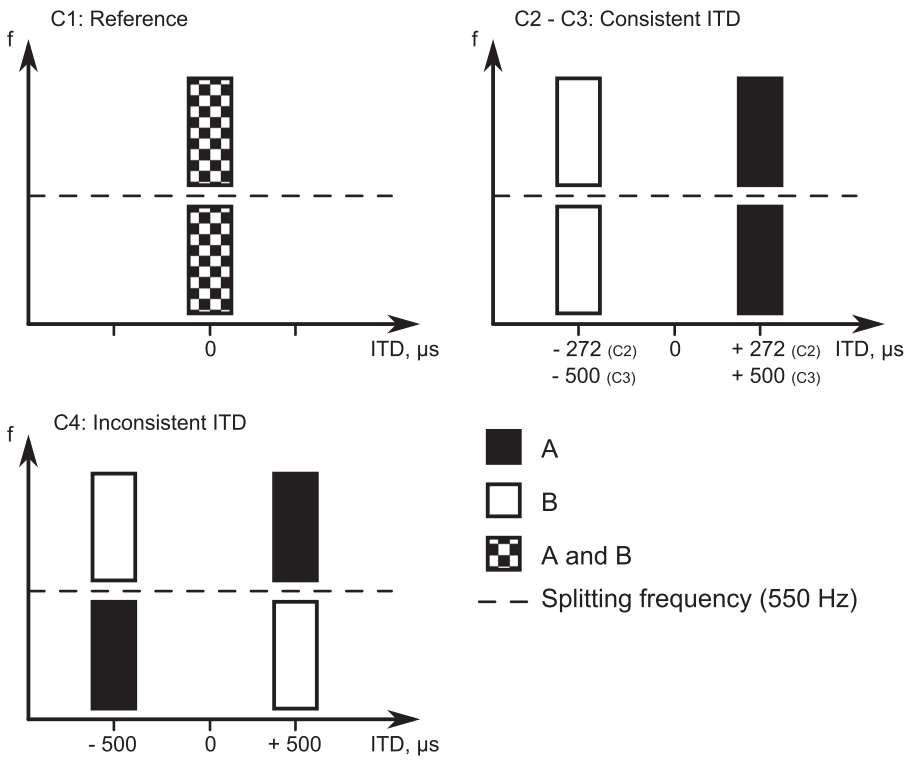

FIG. 5. Conditions tested in experiment 2. Stimuli were spectrally divided into high-and low-frequency bands, with a splitting frequency of 550 Hz. In the reference, 272 and 500 μs conditions, the ITD was consistent across frequencies, so the two bands had the same ITD. In the 500 μs inconsistent condition, the lateralization was blurred by manipulating the ITD independently in each frequency band (i.e., the high-frequency band of stimuli A and the low-frequency band of stimuli B were presented with a +500 μs ITD while the low-frequency band of stimuli A and the high-frequency band of stimuli B were presented with a -500 μs ITD).

the main factor for stream segregation, one would expect a significantly higher threshold in the 500-μs consistent condition compared to the 500-μs inconsistent condition.

Temporal discrimination thresholds were measured using the rhythmic discrimination procedure described in the General methods. After the rhythmic discrimination task (i.e., 5 sessions), the listeners did a short subjective lateralization task (about 10 min long). In this experiment, the stimuli were not convolved with HRIRs, so they were expected to be perceived inside the head. That is why a lateralization task was conducted instead of the localization task used in experiment 1. The listeners were asked to evaluate the perceived azimuth of the different stimuli used in experiment 2, ten different A's and B's for each of the four conditions presented in a random order. The sequences played during the lateralization task corresponded to a single stream of the rhythmic discrimination task (the A's or the B's). For each tested condition, 20 one-stream sequences were generated, ten used the different A's stimuli and ten used the different B's stimuli. Before making their judgments, listeners could play the sequence as many times as they wanted. They had to draw on a protractor the perceived azimuth of the corresponding sound source. The protractor was graduated from $+90°$ (left hand side) to $-90°$ (right-hand side) with 5°-steps. Listeners had to judge the bearing of all sequences twice. No specific indications were given, thus they could indicate either a precise point, a single or several area(s), varying in width, from which they perceived the sounds. None of the listeners ever indicated more than two distinct areas.

D. Listeners

Experiment 2 involved fourteen listeners, but the results of one participant had to be discarded because saturation was reached in more than 50% of the trials, as previously indicated. The 13 remaining listeners were students, aged between 20 and 30 year (six females, mean age = 26 year, SD = 3 year), and signed a general consent form before the experiment. One of them participated in the first experiment, represented with a black square in Fig. 3 and Fig. 6. All listeners had audiometric thresholds of 20 dB hearing level or less in each ear at octave frequencies between 250 and 4000 Hz. They were paid an hourly wage for their participation, and they came for five sessions lasting roughly 1 h.

E. Results

1. Temporal discrimination task

Figure 6 shows the geometric means across listeners of the temporal discrimination thresholds measured in experiment 2. The different symbols represent the mean results for each listener across the five repetitions in each condition. From left to right, the bars correspond to the reference, 272-μs consistent, 500-μs consistent, and 500-μs inconsistent conditions. The arrow on the right indicates the direction of greater segregation. Note that three listeners had thresholds greater than the mean thresholds in all conditions but the reference. However, these listeners did not reach saturation in more than 50% of the trials, thus their data were included in the analysis.

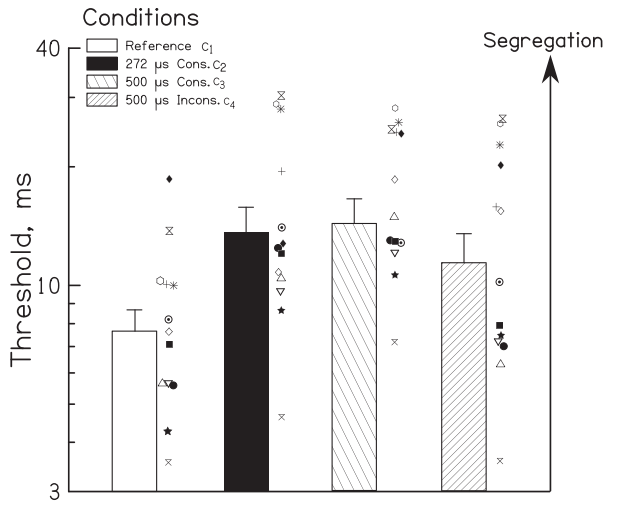

FIG. 6. Geometric mean thresholds in ms with geometric standard errors across participants for detection of the delay applied to the B bursts in experiment 2. The different symbols represent the mean data of individual listeners. From left to right, the bars correspond to the reference, 272-μs consistent, 500-μs consistent, and 500-μs inconsistent conditions. The arrow on the right indicates the direction of greater segregation.

The log values of these thresholds were assessed using a one-way repeated-measures ANOVA, which showed that the effect of tested conditions was significant [$F(3,48) = 3.528$, $p < 0.05$]. A *post hoc* analysis (t-test with Bonferroni corrections) indicated that the mean threshold obtained in the reference condition was significantly lower than in the other conditions ($p < 0.001$ in each case), and that the mean threshold obtained in the 500-μs consistent condition was significantly higher than in the 500-μs inconsistent condition ($p = 0.041$). There was no significant difference in thresholds between the 272- and 500-μs consistent conditions, nor between the 272-μs consistent and 500-μs inconsistent conditions.

2. Lateralization task

As in the localization task presented above, the hypothesis was made that listeners' responses followed one (or two) Gaussian distribution(s). So the individual results were fitted with one or two Gaussians for which the mean(s) and the SDs corresponded to the mean(s) and the spread(s) of the drawings. When a listener perceived a stimulus as coming from two different positions at the same time—which happened sometimes with the stimuli of the 500-μs inconsistent condition—his/her drawings clearly presented distinct patterns, and were fitted with two Gaussian distributions. For each condition and each listener, the Gaussian distributions were averaged over the two repetitions and the ten SSN excerpts (A's and B's) in each condition, even when the responses presented two Gaussians. The individual distributions were then averaged across listeners.

Figure 7 presents the mean distributions of perceived direction as a function of the tested condition. The mean perceived azimuth of stimuli A and B correspond to the dotted

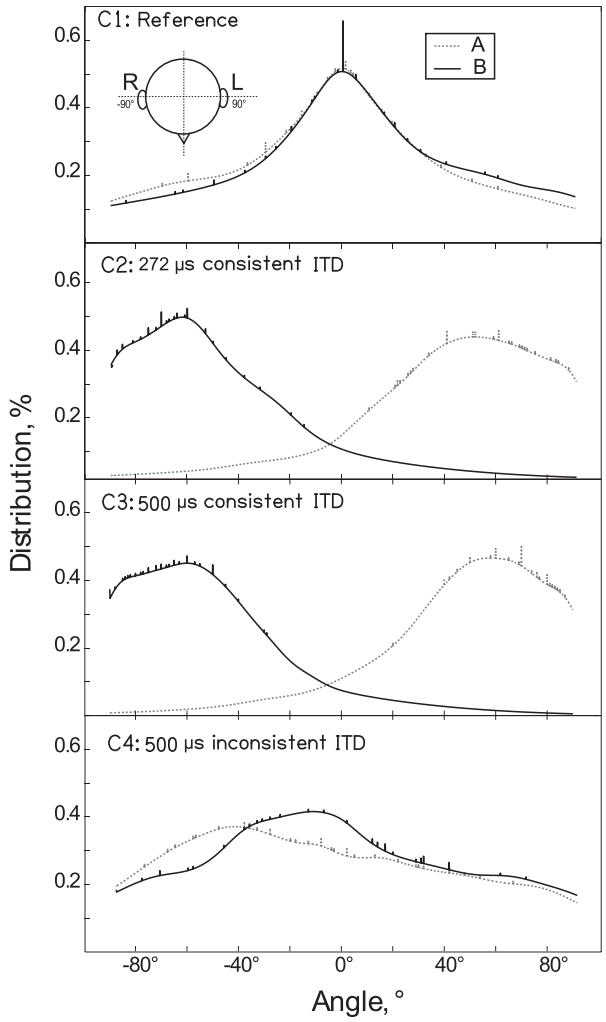

FIG. 7. Mean distributions of the perceived azimuth for the stimuli of experiment 2. The curves represent the mean distribution across thirteen listeners of experiment 2, for the stimuli A (dotted gray lines) and B (filled black lines) in each condition. From top to bottom, the panels correspond to the reference condition, the 272- and 500-μs consistent conditions and the 500-μs inconsistent condition.

gray and filled black lines, respectively. Some listeners (6 out of 13) often lateralized the stimuli at very precise positions and even pointed at only one position. Thus, the SD of their response was zero, resulting in peaks in the mean distributions.

A Kolmogorov-Smirnov test showed that the mean distributions were normal in each condition ($p < 0.001$ in each case). As in experiment 1, the parameters of these mean distributions were estimated using a MLE procedure. For the reference condition (top panel), the mean distribution of the responses was centered at 0°, and the SD was equal to 4.6 and 4.8° for A and B, respectively. For the 272- and 500-μs consistent conditions (second and third panels), the mean distributions were centered at 52.7 and 59.2° (stimuli A) and at −62.4 and −62.1° (stimuli B), respectively. SDs were equal to 9.3 and 8.2° (stimuli A) and 2.6 and 2.7° (stimuli B), for the 272- and 500-μs consistent conditions, respectively. Note that the MLE procedure seemed to underestimate the SD of stimuli B in these conditions. An explanation might be that the data were truncated at −90° and a larger azimuth might be needed to obtain a better estimation. For the 500-μs inconsistent condition (bottom panel), the mean distributions were centered around −40 and 0° for A and B, respectively, and were flattened compared to the other distributions, with SDs equal to 20 and 8.7° for A and B, respectively.

The values of these mean distributions were assessed using a one-way ANOVA, which showed that the effect of the tested conditions was significant ($p < 0.001$). A *post hoc* analysis (LSD) indicated that there was no significant difference between the mean distributions of A and B in the reference condition. The difference between the distributions of A and B was significant in all the other conditions ($p < 0.0001$ for the 272- and 500-μs consistent conditions, and $p = 0.015$ for the 500-μs inconsistent condition). There was no significant difference between the distributions of stimuli A nor between the distributions of stimuli B across the 272- and 500-μs consistent conditions.

F. Discussion

According to the experimental paradigm, low thresholds indicate high performance and a greater tendency to integrate the streams, and high thresholds indicate low performance and a greater tendency to separate the streams. The mean threshold was significantly lower in the reference condition compared to the other conditions. This finding confirms the result of experiment 1, indicating that ITD and the associated perceived azimuth can favor obligatory stream segregation.

The mean threshold was significantly higher in the 500-μs consistent condition compared to the 500-μs inconsistent condition. In this last condition, the ITDs were swapped between high and low frequencies of A and B and that led to blurred lateralization. Even though the difference in perceived azimuth might not have been completely eliminated, it was strongly reduced (see Fig. 7). Since the extent of interaural differences was constant in these two conditions in each frequency band, this result suggests that streams tended to be more segregated when the difference in lateralization was more salient.

The mean threshold was significantly lower in the reference condition compared to the 500-μs inconsistent condition. The distributions of the perceived azimuth of the stimuli were comparable in the ILD condition of experiment 1 and the 500-μs inconsistent condition of experiment 2 (see Fig. 4, panel 3 and Fig. 7, panel 4), suggesting that the percept of lateralization was comparable in these two conditions. In experiment 1, this percept was not salient enough to introduce a difference in segregation, so it should not have been an influencing factor in experiment 2. Thus, the difference in thresholds between the reference and the 500-μs inconsistent conditions in experiment 2 was probably mainly due to the difference in ITD *per se*. This result indicates that ITD itself can induce obligatory stream segregation.

The rhythmic discrimination task did not show a significant difference in threshold in the 272-μs consistent condition compared to the 500-μs consistent condition. Thus, the increase of ITD from 272 to 500 μs did not significantly

improve the segregation, i.e., the effect of ITD on segregation reached a ceiling. This result is not in accordance with previous results showing that an increase of ITD above the physiological range increased its effect on segregation (Füllgrabe and Moore, 2012; Stainsby *et al.*, 2011). Once again, this difference might be due to the nature of the stimuli used (see Sec. III E).

Besides, the lateralization task did not show any difference in perceived azimuth between the A or B in the 272-μs compared to the 500-μs consistent conditions. This difference in ITD did not lead to a difference in perceived azimuth, nor a difference in perceived segregation. This result supports the idea of a ceiling effect of ITD *per se*. It also suggests that a clear difference in perceived azimuth is required to induce obligatory stream segregation.

The lateralization task showed that A and B were perceived at significantly different azimuths in all conditions except the reference condition. This result indicated that the differences in ITD were sufficient to introduce a perceptual lateralization difference. This task also showed that the estimated azimuths in the 500-μs consistent condition were consistent with the tested ITD values—an ITD of $\pm 500\,\mu$s should correspond to an azimuth of $\pm 60°$ (Feddersen *et al.*, 1957). However, the perceived azimuths in the 272-μs consistent condition were overestimated—an ITD of $\pm 272\,\mu$s should correspond to an azimuth of $\pm 30°$. An explanation for this poor accuracy in the lateralization task could be due to the fact that a unique ITD value was applied across frequency, while real ITD is dependent on frequency because of the diffraction effects around the head (Kuhn, 1977). Furthermore, the lateralization task used in the present study was presumably less accurate than a pointing task for example, where listeners have to move a narrow band of noise to match the perceived position of the target stimulus (Bernstein and Trahiotis, 1985). However, the present lateralization task—designed for a qualitative purpose rather than a quantitative evaluation—allowed us to verify that the lateralization was actually reduced from a condition with a consistent ITD across frequency to a condition with an inconsistent ITD across frequency.

V. GENERAL DISCUSSION

Objective tasks, such as the rhythmic discrimination task used in the present study, give an indirect measure of stream segregation, controlling interpretation bias. To ensure that stream segregation has been actually observed using an objective task, results can be correlated with a subjective task (Anstis and Saida, 1985; David *et al.*, 2014; Füllgrabe and Moore, 2012; Roberts *et al.*, 2002; Stainsby *et al.*, 2011, 2004; Thompson *et al.*, 2011). The present study used the same type of stimuli (i.e., SSN convolved with HRIRs) and experimental design (i.e., same onset-to-onset time, same stimuli duration and same sequence length) as in David *et al.* (2014). In the previous study, the results obtained with the rhythmic discrimination task were highly correlated with those obtained with a subjective task. These previous results indicate that the paradigm used in the present experiments measured stream segregation.

With the experimental design used in the present study, it is conceivable that the listener can perform the task based only on the last pairs of the sequence—ignoring the beginning of the sequence—to determine whether the rhythm was regular or irregular. In this case, the time for segregation to build-up is substantially reduced. Thus, one might wonder if the obtained results reflected auditory streaming or impairment in gap detection within a single stream. Indeed, the latter can be affected by the perceived dissimilarities between adjacent sounds (Grose *et al.*, 2001). However, Oxenham (2000) showed that gap detection is affected by dissimilarities only if these dissimilarities involve large spectral differences. Experiment 1 involved only slight spectral differences and experiment 2 induced only temporal differences, so it is likely that these results are due to stream segregation rather than gap detection. Moreover, according to the study by Deike *et al.* (2012), stream segregation can occur after just a few presentations (as in a gap detection task). In their study, the listeners were presented with sequences of tones with alternating frequency values and were asked to indicate the number of streams they heard as fast as possible. The results showed that the first percept of the listener was often segregation. Thus, at least in principle, the difference in thresholds obtained in the present study can be explained by a difference in stream segregation, even if the listener made his/her judgment on the last two pairs of stimuli.

Experiment 1 showed that the auditory system benefits more from a difference in ITDs and monaural coloration cues to separate bursts of SSN than from a difference in ILDs. This result is consistent with the findings of Middlebrooks and Onsan (2012) and Bremen and Middlebrooks (2013) who showed the importance of ITDs over ILDs for stream segregation. Middlebrooks and Onsan (2012) also showed the importance of ITDs over monaural cues. In this previous study, monaural cues consisted of the spectral differences induced by different positions in the vertical plane, whereas in the present study, monaural cues also involved the broadband level variations associated with different positions in the horizontal plane. The combination of these two studies suggests that monaural broadband level cues might be important for the segregation of broadband noises.

Although Schwartz *et al.* (2012) showed that ITDs alone did not allow correct identification of their specific stimuli (i.e., broadband noises presenting speech similarities but without harmonicity and onset/offset cues), their results indicated that the listeners reported hearing two separate sources when they were asked to localize the independent sources within a mixture of sounds (i.e., target plus masker) based on ITDs. This finding showed that segregation could occur when the spatial position of the auditory objects were identified. The present results confirmed that the auditory system separates the streams once the sources are clearly perceived as coming from distinct azimuths, at the level of auditory object formation. In fact, stream segregation was reduced when the percept of lateralization was reduced (in the 500-μs inconsistent condition in experiment 2). This interpretation is in agreement with the results of Darwin and Hukin (1999). In their first experiment, they showed that listeners tend to group a target word with the sequence which

presents the same ITD. In two other experiments, they investigated the extent to which this result was due to the listeners' ability to track a common ITD (i.e., exploitation of the individual frequency components) or to track an attended position (i.e., exploitation of grouped objects). They modified the ITD of a harmonic close to the first formant in a vowel. This modification showed only a little effect on the vowel recognition, even when the vowel was presented within a sentence which had the same ITD as the main part of the vowel. Darwin and Hukin (1999) concluded that when the listeners attend to a particular source, they track the particular location of the auditory object instead of tracking the frequency components that share the same ITD. Furthermore, the importance of a difference in lateralization over a difference in ITD could explain the weak effect of ITDs found by Füllgrabe and Moore (2012) and Stainsby *et al.* (2011). Pure and complex tones provide less accurate lateralization than broadband noises (Sandel *et al.*, 1955).

In experiment 1, introducing ILDs did not result in a segregation enhancement (i.e., ILD condition compared to coloration condition). This result might be due to the fact that the lateralization induced by ILDs was not salient enough to produce obligatory stream segregation. Indeed, the localization task indicated that ILD alone produced less lateralization than ILD+ITD, which is in agreement with the findings of Wightman and Kistler (1992). In a set of lateralization experiments, their listeners had to evaluate the perceived position of conflicting sounds (i.e., ITD gave a cue towards one direction while ILD gave a cue towards another direction). Their results showed that the perceived position was determined based principally on ITD as long as the stimuli contained low frequencies. Thus, for the present study using broadband speech-shaped noises, ITD was the dominant cue for lateralization compared to ILD. This result supports the hypothesis that the auditory system relies on a clear difference in perceived positions to segregate the streams.

In experiment 2, the lateralization percept was substantially reduced in the 500-μs inconsistent condition, albeit not completely suppressed. So it is unlikely that the observed segregation (compared to the reference condition) was only due to the difference in perceived azimuth. Indeed, in this condition, the lateralization was comparable to the ILD condition in experiment 1, which was not salient enough to induce stream segregation. The ITDs available in this condition might have been used to segregate the streams. This suggests that segregation might also occur at the level of cue extraction. But this influence of ITD *per se* seems limited by a ceiling effect since increasing ITDs from 272 to 500 μs did not increase stream segregation.

The present study showed that a difference in lateralization can favor segregation and thus enables one to follow a stream across time, which could be seen as a cue reducing informational masking. This result is in agreement with those of Kidd *et al.* (1994). In their experiment, listeners had to track target words simultaneously presented with two other talkers. Their results showed that listeners could easily attend to the target speech, and thus ignore the masker voices, when they knew the particular location of the target.

Edmonds and Culling (2005) investigated the influence of ITD on spatial unmasking. They measured speech reception thresholds (SRTs) for target speech presented with a concurrent speech masker, and assessed whether the mechanisms underlying spatial unmasking rely on a difference in ITD within each frequency channel or a difference in ITD consistent across frequencies. ITD is a useful cue for sound localization as long as it is consistent across frequencies. Their results showed that listeners could rely on ITD in each frequency band to reach high performance. Spatial unmasking was not impaired by inconsistent ITDs across frequency as long as target and interferer differed in ITD. Thus, ITD differences could be exploited within each frequency band, even if this led to unclear lateralization, to segregate target and masker. Their results seemed in opposition with the results of the present study which suggested that consistent ITD across frequency is needed to favor segregation of auditory streams. However, the stimuli used by Edmonds and Culling (2005) contained strong streaming cues other than lateralization, such as differences in pitch, timbre and level, so that their study mainly concerned energetic masking rather than informational masking. It appears that differences in perceived lateralization might not be a relevant cue in this case where differences in ITD are crucial. Differences in perceived lateralization would be a segregation cue when informational masking is the overriding factor, like in the present study or the one of Kidd *et al.* (1994).

In a multi-talker environment, if it can be assumed that the spatial configurations of speakers and listeners remain sufficiently constant over a given period of time, the consistency of the location cues associated with the different sources positions could be used by the auditory system. Thus, the consistency of the spatial differences could be relevant for the segregation of competing voices. The present study does not allow us to conclude on this particular point, as frozen stationary noises were used. In fact, contrary to frozen noises, speech sounds present spectro-temporal variability. In this respect, a first follow-up of this study would be to assess the influence of binaural cues using unfrozen stimuli with spectro-temporal variability to get a step closer to real-world situations.

VI. SUMMARY AND CONCLUSIONS

Experiment 1 showed that listeners had a greater tendency to segregate sequences of speech-shaped noises when the stimuli presented spatial cues (coloration, ILD and ITD). The results also indicated that the monaural spectral level variations across time were more important for stream segregation than the interaural level differences.

Experiment 2 investigated whether the influence of ITD was due to the interaural difference *per se* and/or to the corresponding differences in perceived lateralization. The results indicate that sequences were more segregated when the percept of lateralization was salient rather than blurred. Thus, the difference in lateralization associated with ITDs had an important influence on obligatory stream segregation. The results also showed that ITDs helped to segregate

sounds up to a ceiling ITD value, above which segregation was not further improved.

ACKNOWLEDGMENTS

The authors would like to thank the listeners who took part in the experiments, Andrew Oxenham and Eugene Brandewie for their helpful comments. This work was supported by an institutional grant from the LabEX CeLyA (ANR-10-LABX-0060/ANR-11-IDEX-0007) operated by the French National Research Agency.

ANSI S3.6 (**1989**). *American National Standard Specification for Audiometers* (American National Standards Institute, New York).
ANSI S3.7 (**1995**). *Methods for Coupler Calibration of Earphones* (American National Standards Institute, New York).
Anstis, S. M., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol.: Human Percept. Perf. **11**, 257–271.
Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of low-frequency, complex wave-forms: The use of envelope based temporal disparities," J. Acoust. Soc. Am. **77**, 1868–1880.
Boehnke, S. E., and Phillips, D. P. (**2005**). "The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences," Percept. Psychophys. **67**, 1088–1101.
Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), Chap. 1, pp. 1–45.
Bregman, A. S., Ahad, P. A., Crum, P. A., and O'Reilly, J. (**2000**). "Effects of time intervals and tone durations on auditory stream segregation," Percept Psychophys. **62**, 626–636.
Bremen, P., and Middlebrooks, J. C. (**2013**). "Weighting of spatial and spectro-temporal cues for auditory scene analysis by human listeners," PLoS ONE **8**, e59815.
Carlile, S. (**1996**). *The Physical and Psychophysical Basis of Sound Localization* (R. G. Landes Company and Springer-Verlag, Berlin Heidelberg), Chap. 2, pp. 27–78.
Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.
Collin, B., and Lavandier, M. (**2013**). "Binaural speech intelligibility in rooms with variations in spatial location and modulation depth of noise interferers," J. Acoust. Soc. Am. **134**, 1146–1159.
Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol.: Human Percept. Perf. **20**, 617–629.
David, M., Lavandier, M., and Grimault, N. (**2014**). "Room and head coloration can induce obligatory stream segregation," J. Acoust. Soc. Am. **136**, 5–8.
Deike, S., Heil, P., Bockmann-Barthel, M., and Brechmann, A. (**2012**). "The build-up of auditory stream segregation: A different perspective," Front. Psychol. **3**, 461.
Devergie, A., Grimault, N., Gaudrain, E., Healy, E. W., and Berthommier, F. (**2011**). "The effect of lip-reading on primary stream segregation," J. Acoust. Soc. Am. **130**, 283–291.
Edmonds, B. A., and Culling, J. F. (**2005**). "The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.
Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (**1957**). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988–991.
Flanagan, J. L., and Lummis, R. C. (**1970**). "Signal processing to reduce multipath distortion in small rooms," J. Acoust. Soc. Am. **47**, 1475–1481.
Füllgrabe, C., and Moore, B. C. J. (**2012**). "Objective and subjective measures of pure-tone stream segregation based on interaural time differences," Hear. Res. **291**, 24–33.
Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.
Gockel, H., Carlyon, R. P., and Micheyl, C. (**1999**). "Context dependence of fundamental-frequency discrimination: Lateralized temporal fringes," J. Acoust. Soc. Am. **106**, 3553–3563.
Grose, J. H., Hall, J. W. I., Buss, E., and Hatch, D. (**2001**). "Gap detection for similar and dissimilar gap markers," J. Acoust. Soc. Am. **109**, 1587–1595.
Hartmann, W. M., and Johnson, D. (**1991**). "Stream segregation and peripheral channeling," Music Percept. **9**, 155–183.
Hartmann, W. M., and Wittenberg, A. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am. **99**, 3678.
Kidd, G. J., Best, V., and Mason, C. R. (**2008**). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.
Kidd, G. J., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, S. H. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.
Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.
Larsen, E., Iyer, N., Lansing, C. R., and Feng, A. S. (**2008**). "On the minimum audible difference in direct-to-reverberant energy ratio," J. Acoust. Soc. Am. **124**, 450–461.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Micheyl, C., and Oxenham, A. J. (**2010**). "Objective and subjective psychophysical measures of auditory stream integration and segregation," J. Assoc. Res. Otolaryngol. **11**, 709–724.
Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Ann. Rev. Psychol. **42**, 135–159.
Middlebrooks, J. C., and Onsan, Z. A. (**2012**). "Stream segregation with high spatial acuity," J. Acoust. Soc. Am. **132**, 3896–3911.
Moore, B. C. J. (**2007**). *An Introduction to the Psychology of Hearing*, 5th ed. (Elsevier Academic Press, London, UK), Chap. 7, 236–237.
Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust. Acust. **88**, 320–332.
Moore, B. C. J., and Gockel, H. E. (**2012**). "Properties of auditory stream formation," Philos. Trans. R. Soc. B. **367**, 919–931.
Oxenham, A. J. (**2000**). "Influence of spatial and temporal coding on auditory gap detection," J. Acoust. Soc. Am. **107**, 2215–2223.
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.: Human Percept. Perf. **34**, 992–1006.
Sach, A. J., and Bailey, P. J. (**2004**). "Some characteristics of auditory spatial attention revealed using rhythmic masking release," Percept. Psychophys. **66**, 1379–1387.
Sandel, T. T., Teas, D. C., Feddersen, W. E., and Jeffress, L. A. (**1955**). "Localization of sound from single paired sources," J. Acoust. Soc. Am. **27**, 842–852.
Schwartz, A., McDermott, J. H., and Shinn-Cunningham, B. (**2012**). "Spatial cues alone produce inaccurate sound segregation: The effect of interaural time differences," J. Acoust. Soc. Am. **132**, 357–368.
Stainsby, T. H., Füllgrabe, C., Flanagan, H. J., Waldman, S. K., and Moore, B. C. J. (**2011**). "Sequential streaming due to manipulation of interaural time differences," J. Acoust. Soc. Am. **130**, 904–917.
Stainsby, T. H., Moore, B. C. J., Medland, P. J., and Glasberg, B. R. (**2004**). "Sequential streaming and effective level differences due to phase-spectrum manipulations," J. Acoust. Soc. Am. **115**, 1665–1673.
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.: Human Percept. Perform. **37**, 1253–1262.
van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, University of Technology, Eindhoven, the Netherlands.
Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.