

The auditory continuity phenomenon: Role of temporal sequence structure

Carolyn Drake, and Stephen McAdams

Citation: *The Journal of the Acoustical Society of America* **106**, 3529 (1999);

View online: <https://doi.org/10.1121/1.428206>

View Table of Contents: <http://asa.scitation.org/toc/jas/106/6>

Published by the *Acoustical Society of America*

The auditory continuity phenomenon: Role of temporal sequence structure^{a)}

Carolyn Drake^{b)}

Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, EPHE, 28 rue Serpente, F-75006, Paris, France

Stephen McAdams

Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, EPHE, 28 rue Serpente, F-75006, Paris, France and Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), 1 place Igor-Stravinsky, F-75004, Paris, France

(Received 19 February 1999; accepted for publication 4 August 1999)

The auditory continuity phenomenon occurs when listening to a series of alternating high- and low-level tones: instead of perceiving this intermittence, listeners often report hearing a continuous tone upon which is superimposed a series of intermittent tones. The temporal limits to the perception of this phenomenon are investigated, as well as effects of task instructions on it. A loudness-matching paradigm developed previously [McAdams *et al.*, *J. Acoust. Soc. Am.* **130**, 1580–1591 (1998)] provided both an objective indication of this phenomenon and an indication of its strength. In the studies reported here, the phenomenon was observed reliably when the low-level tones were at least half the duration of the high-level tones: the greater the duration ratio between low-level and high-level tones, the stronger the phenomenon. Duty-cycle duration did not affect the strength of this phenomenon. It proved to be particularly robust as its strength was unaffected by task instructions, level of expertise, and repetition. A model is proposed in which adjusted levels depend on the relative levels of the high- and low-level tones weighted by relative duration and attentional focusing. © 1999 Acoustical Society of America. [S0001-4966(99)00412-9]

PACS numbers: 43.66.Cb, 43.66.Mk, 43.66 Lj [SPB]

INTRODUCTION

The phenomenon of auditory continuity or auditory induction may occur when listening to a series of alternating high- and low-level tones: instead of perceiving this alternation, listeners often report hearing a continuous low-level tone upon which is superimposed a series of intermittent tones (Thurlow and Elfner, 1959; Warren, Obusek, and Ackroff, 1972). Traditional loudness models (Zwicker, 1977) have difficulty accounting for this continuity phenomenon (henceforth referred to simply as “the phenomenon”). However, Warren (1982) proposed an explanation in terms of a subtractive process derived from his auditory induction theory which proposes that a portion of the internal representation of the louder sound is subtracted or reallocated to serve as a substrate for synthesis of the apparently continuous fainter sound. Bregman (1990) has also proposed a theoretical framework which can interpret these findings: the auditory system organizes the auditory scene before computing perceptual attributes of events. Thus, the phenomenon occurs because of a bottom-up perceptual mechanism which he has called the “old-plus-new” listening strategy: the “old” information is composed of the low-level tones and

an equivalent part of the high-level tones, and the “new” information is composed of the remaining part of the high-level tones. The auditory system considers that the old information does not stop, but continues throughout the time period occupied by the new information. An interpolation is performed between the properties of the low-level sounds occurring before and after the interrupting sound (Ciocca and Bregman, 1987). Thus, the high-level tones are in fact separated into two parts: one attributed to the continuous tone and one attributed to the intermittent tones.

The old-plus-new strategy allows two predictions concerning the perception of the phenomenon. First, it should only be perceived if the old events (the low-level tones) are well established. Thus, the longer the old tones are in relation to the new tones, the stronger should be the phenomenon. Second, it should be very robust because it results from a relatively low-level process, occurring before loudness attribution. Thus, it should be: (1) observed in all subjects, (2) unaffected by task constraints and instructions, and (3) unaffected by training and practice.

Previous studies have examined some limiting factors to the perception of this phenomenon: it does not occur if the high- and low-level tones are separated from each other by short pauses and nonabrupt transitions, or are presented to separate ears. It only occurs if the two sets of tones differ by certain level differences and are in the same spectral range (Bregman and Rousseau, 1991; Houtgast, 1972; McAdams, Botte, and Drake, 1998; Thurlow, 1957; van Noorden, 1977; Verschuure, Rodenburg, and Maas, 1976; Warren, 1984; Warren *et al.*, 1994; Warren, 1999). To our knowledge, few

^{a)}Preliminary work leading to this study was reported at the 4th French Congress of Acoustics, Marseille (C. Drake and S. McAdams, “The Continuity illusion: Role of temporal sequence structure,” *Actes du He Congrès Français d’Acoustique*, Marseille, 1997, Vol. 1, pp. 519–522).

^{b)}Address correspondence to C. Drake, Laboratoire de Psychologie Expérimentale, 28 rue Serpente, 75006, Paris, France. Electronic mail: drake@idf.ext.jussieu.fr

studies have systematically studied the role of temporal factors in the continuity phenomenon. McAdams *et al.* (1994, 1998), for example, found that the phenomenon was observed under all temporal conditions examined: the high-level tones were equal to or shorter in duration than the low-level tones (high/low duration ratios of 100/100, 200/200, 100/300, and 100/700 ms). In the present study, we examine two temporal factors that may play a determining role in whether or not the continuity phenomenon is observed: the relative duration of high- and low-level tones and the total duty-cycle duration (duration of high-level tone plus duration of low-level tone). We therefore examine instances where the high-level tones are both longer *and* shorter than the low-level tones and where the duty cycle ranges from short (150 ms) to long (2100 ms). We also examine the robustness of the phenomenon by comparing the role of instructions given to the subjects (instructions explicitly concerning the continuity phenomenon and instructions emphasizing the rhythmic structure without reference to the phenomenon). In addition, we also investigated the role of experience in this sort of task (by comparing experts and novices), and eventual improvements over repeated measures.

We establish whether or not listeners perceived the phenomenon by using a paradigm developed previously (McAdams *et al.*, 1998). A similar paradigm was developed independently by Warren *et al.* (1994), except that they used a modified method-of-limits procedure rather than an adjustment paradigm. In our paradigm, we measured the perceived loudness of various components of the stimulus: listeners were asked to adjust the loudness of (a) a continuous sound to equal that of the perceived continuous tone in the reference sequence, and (b) the loudness of an intermittent sequence to equal that of the perceived intermittent sequence in the reference sequence. Two distinct adjustment patterns were observed depending on the percept. When listeners perceived the phenomenon, the level of the continuous tone was adjusted very near the level of the low tones, but the level of the intermittent tone was systematically adjusted well below that of the high tones, as predicted by the old-plus-new hypothesis. When listeners did not perceive the phenomenon, the level of both components of the reference sequence was adjusted near the physical values. In the present study, we use this method to identify the temporal conditions under which listeners perceive the continuity phenomenon.

I. EXPERIMENT 1: RELATIVE DURATION OF HIGH- AND LOW-LEVEL TONES

Previous studies have shown that the phenomenon is perceived when the low-level tones (old events alone) have the same duration as or are longer than the high-level tones (old and new events together). We predict that the phenomenon will disappear in the opposite configuration when the low-level tones are shorter than the high-level tones, because the old events are not temporally well established. We therefore measure listeners' loudness matches to parts of the reference sequence in order to establish the conditions under which they perceive the phenomenon. We identify whether or not listeners perceive the phenomenon by measuring the precision with which they match the levels: previous work

has established two distinct level-matching patterns when subjects perceive and do not perceive the phenomenon. Two groups of subjects were examined (novices and experts) to test the role of expertise, and six repeated measures were made to test the role of practice.

A. Method

1. Subjects

All eight subjects reported normal hearing. They were all psychology students at the Université René Descartes. Subjects were divided into two equal-sized groups: the "novices" had not previously participated in psychoacoustic experiments, and the "experts" had previously participated in auditory experiments (at least 10 h, but not on the continuity phenomenon) and were musically experienced (had been playing an instrument for at least 5 years).

2. Apparatus

Sinusoidal signals were synthesized at a sampling rate of 20 kHz with 16-bit resolution on an Oros DSP card controlled by a Compaq computer. In the case of alternating-level signals, high and low bursts were processed in different channels. The signals were then filtered with a Kemo VBF/24 bandpass filter with cutoff frequencies of 930 and 1070 Hz and -48 dB/oct slopes. The filtering of the signals served for antialiasing. The filtered signals were then routed through Charybis D programmable attenuators with 0.25-dB resolution that were controlled by the computer. The final signals were mixed and sent to both earpieces of a TDH-49 headset. Experimental sessions took place in an IAC single-walled sound-isolation chamber. Subjects adjusted levels for comparison stimuli with a single-turn potentiometer and signaled their satisfaction with the adjustment by pressing a button on the response box. At this point the computer presented the next trial. Levels at each earpiece were verified using a Bruel & Kjaer 4153 artificial ear and 2230 sound-pressure meter.

3. Stimuli

Sequences were composed of 1-kHz pure tones with overlapping 5-ms linear onset and offset ramps. Reference sequences with alternating levels were composed of eight low-level tones interleaved with seven high-level tones. The duration of the short tones was 50 or 100 ms, and the duration of the long tones was 300, 700, or 1500 ms. All combinations of short and long tones were tested, giving 12 duration conditions illustrated in Fig. 1. Six of these duration conditions corresponded to the "classic" configuration where the high-level tones were shorter than the low-level tones (in which continuity is usually reported), and the other six duration conditions corresponded to an "inverted" configuration where the high-level tones were longer than the low-level tones. The total duration of the sequences depended on the duration combinations, varying between 2.5 and 11.2 s. The sound level of the low-level part of the reference stimulus was varied randomly from trial to trial within the set $60 \pm \{1,3,5\}$ dB. In all analyses and plots, lev-

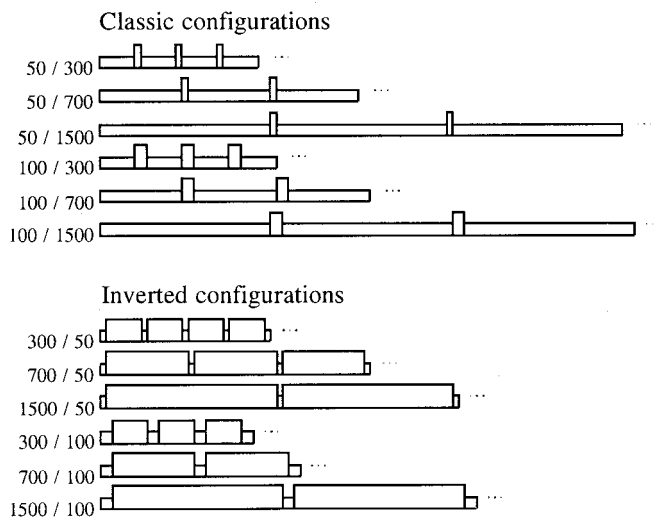


FIG. 1. Experiments 1 and 3: Stimuli used in the classic and inverted configurations (high- and low-level tone durations in ms).

els are presented relative to the mean of the roving range (60 dB). The high-level tones were 6 dB higher than the low-level tones.

There were two types of comparison sequences: (1) a continuous tone of the same total duration as the reference sequence, or (2) an intermittent sequence with tones having the same duration as the high-level tones of the reference sequence and being separated by pauses of the same duration as the low-level tones of that sequence. They were presented at an initial level of $60 \pm \{7,8,9,10\}$ dB to avoid response bias. The reference and comparison sequences were separated by 800 ms, and the next reference sequence followed 1500 ms later.

4. Procedure

Half the subjects (two experts and two novices) completed experiment 1 the first day, and then experiment 3 about a week later. The other half of the subjects completed the two experiments in the opposite order. The first experimental session was preceded by a *familiarization phase* in which all the stimuli were presented to the subjects. They were questioned as to what they heard in order to verify whether or not auditory continuity was perceived for each duration condition. The familiarization phase was followed by a *test phase* composed of 72 trials (6 duration conditions, 2 stimulus configurations, 6 repetitions). Each trial consisted of the repeated alternation between the reference and comparison sequences. During this alternation, the level of the comparison sequence could be adjusted with a single-turn potentiometer. Subjects could listen to the alternation as many times as necessary to make a satisfactory match, at which point they signaled the computer to record the level of the comparison sequence by pressing a button. The listener aligned the turnpot to a fixed reference point at the beginning of each trial. Subjects were instructed to (a) match the loudness of the intermittent comparison sequence to that of what they perceived as intermittent in the reference sequence, or (b) match the loudness of the continuous comparison sequence to that of what they perceived as continuous in the

reference sequence. The stimuli were presented in a random order. Six measures were made in each experimental condition. The dependent measure was the adjustment "error" (in dB) between the adjusted level of the comparison sequence (intermittent level = L_I ; continuous level = L_C) and that of the targeted part of the reference sequence (low level = L_L ; high level = L_H) (error = $L_L - L_C$ or $L_H - L_I$). The experiment lasted about 2 h per subject, with a pause halfway through the session.

B. Results

1. Subjective descriptions

In the familiarization phase at the beginning of the first experimental session, subjects were asked to describe the 12 sequences. No reference was made to the continuity phenomenon. In the classic configurations, 94% of the responses indicated the phenomenon, whereas only 42% indicated the phenomenon in the inverted configurations. The percentage reports of the phenomenon were more divergent across the two classes of stimuli for the expert group than for the novice group (classic: novices = 88% and experts = 100%, inverted: novices = 63% and experts = 21%), and were independent of the temporal combination.

2. Classic configuration

Figure 2 shows the mean adjusted levels for the classic and inverted duration conditions. For the *continuous* adjustable tones, matches were extremely accurate (mean adjustment = 59.9 dB; mean adjustment error = $59.9 - 60.0 = -0.1$ dB). A single-sample *t*-test revealed that the mean adjustment error did not differ significantly from zero. The matched level did not vary with the duration conditions: a mixed analysis of variance (ANOVA) on the mean adjustment errors by group (novices/experts), duration of high-level tones (50 or 100 ms), duration of low-level tones (300, 700, or 1500 ms), and repetition (6) revealed no significant differences.

However, for the *intermittent* adjustable sequence, matches were quite different from the physical levels (mean adjustment = 60.8 dB; mean adjustment error = $60.8 - 66.0 = -5.2$ dB), indicating that subjects systematically *underestimated* the level of the intermittent tones. A single-sample *t*-test revealed that the mean adjustment error differed significantly from zero [$t(7) = 5.6$; $p < 0.001$]. A mixed ANOVA on the mean adjustment errors by group (novices/experts), duration of high-level tones (50 or 100 ms), duration of low-level tones (300, 700, or 1500 ms), and repetition (6) revealed no significant differences between the two groups of subjects or over the six repetitions. The duration of the high-level tones did not influence the matches but the duration of the low-level tones did [$F(2,12) = 9.17$, $p < 0.01$]; Fig. 2 shows that the longer the low-level tones, the greater the adjustment error (difference from L_H).

Thus, in these classic conditions, adjustment errors corresponded to the patterns expected if subjects were experiencing the continuity phenomenon. The phenomenon was greater when the low-level tones were longer.

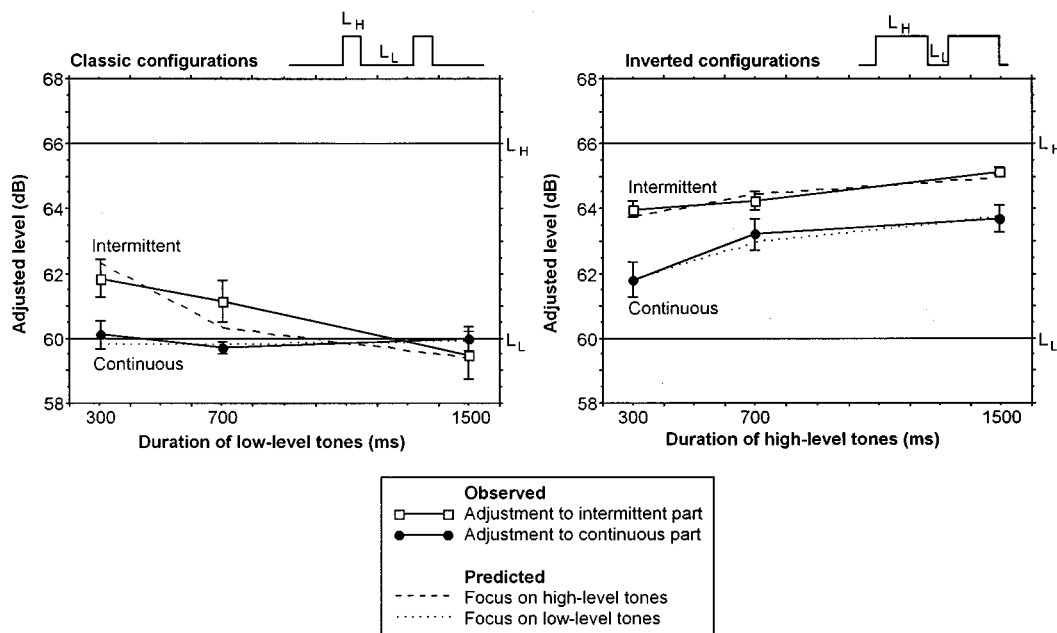


FIG. 2. Experiment 1: Solid-line curves represent mean levels adjusted to match the loudness of the continuous and intermittent comparison sequences for the six duration conditions. The data are averaged over groups, repetitions, and the short-tone durations (50 and 100 ms). Note that the abscissa corresponds to the durations of the longer tones (low-level for the classic configurations and high-level for the inverted configurations). Horizontal lines indicate physical levels of low- and high- level tones in the reference sequences. Error bars (where visible) indicate ± 1 standard error. Dashed and dotted curves represent predictions based on subject means weighted for duration with attentional focus on high-level (dashed) and low-level (dotted) parts of the reference sequence (see Sec. V).

3. Inverted configuration

For the *continuous* adjustable tones, systematic adjustment errors were observed (mean adjustment = 62.9 dB; mean adjustment error = 62.9 - 60.0 = +2.9 dB), indicating that contrary to the classic conditions, subjects systematically *overestimated* the level of the continuous tones. A single-sample *t*-test revealed that the mean adjustment error differed significantly from zero [$t(7) = 7.4$, $p < 0.001$]. A mixed ANOVA on the mean adjustment errors by group (novices/experts), duration of high-level tones (300, 700, or 1500 ms), duration of low-level tones (50 or 100 ms), and repetition (6) revealed only a significant effect of high-level tone duration [$F(2,12) = 10.24$, $p < 0.01$]: the longer the high-level tone, the higher the level match of the continuous tones.

For the *intermittent* adjustable sequence, systematic matches were also observed (mean adjustment = 64.5 dB; mean adjustment error = 64.5 - 66.0 = -1.5 dB), indicating that subjects systematically *underestimated* the level of the intermittent tones. A single-sample *t*-test revealed that the mean adjustment error differed significantly from zero [$t(7) = 8.7$, $p < 0.001$]. A mixed ANOVA on the mean adjustment errors by group (novices/experts), duration of low-level tones (50 or 100 ms), duration of high-level tones (300, 700, or 1500 ms), and repetition (6) revealed no significant differences between the two groups of subjects or over the six repetitions. As for the continuous adjustable tones, the duration of the low-level tones did not influence the matches but the duration of the high-level tones did [$F(2,12) = 22.96$, $p < 0.001$]: the longer the high-level tone, the higher the adjusted level of the intermittent tones.

Thus, in these inverted conditions, adjustment errors did

not correspond to the patterns expected if subjects were experiencing the continuity phenomenon. Instead, the matches suggest that subjects were adjusting to an intermediate level, the estimation of which was affected by the duration of the high-level tones.

C. Discussion

We have identified some temporal constraints on the continuity phenomenon: it is only observed if the old information temporally dominates the new information, that is, the low-level tones must be equal to or longer than the high-level tones. The strength of the phenomenon varied with the temporal sequence structure. Greater underestimations of the level of the intermittent sequence were observed when the low-level tones were longer, suggesting a stronger phenomenon. This could be due either to the ratio between the durations of the high- and low-level tones, or to the increased duration of the duty cycle. Experiment 2 will separate these two factors. However, when the phenomenon was not perceived, listeners adjusted both the high- and low-level tones towards an intermediate level which increased with longer durations. The adjustment behaviors are thus both quantitatively and qualitatively different for the two classes of stimuli (those that induce the phenomenon and those that do not).

In order to investigate the relation between subjective reports of the illusion and adjustment errors, two adjustment data sets were created: for each of the 12 temporal conditions, subjects were classified according to whether they reported the phenomenon or not. Very similar patterns of over- and underestimation were observed in the two data sets, once

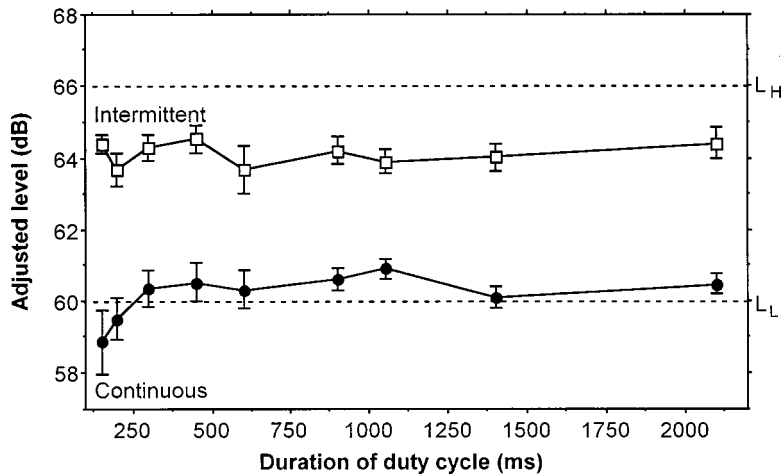


FIG. 3. Experiment 2: Mean levels adjusted for the continuous and intermittent comparison sequences adjusted to the classic, equal, and inverted configurations, where the duty-cycle duration was varied systematically. The data are averaged over high/low duration ratios and repetitions. Error bars indicate ± 1 standard error.

again emphasizing the fundamental nature of this phenomenon.

For the classic conditions, the phenomenon was surprisingly robust because it was not influenced by listeners' experience with listening to sequences (novice/expert), nor by repeated exposure to the phenomenon (repetitions). However, the instructions in this experiment may have facilitated the phenomenon because subjects were asked to adjust a continuous tone, thus emphasizing the phenomenon. In fact, this may have led to confusion for the listeners in the inverted conditions where they reported hearing the continuity phenomenon only 42% of the time. Instructions which do not orient the listener toward the phenomenon will be used in experiment 3.

II. EXPERIMENT 2: DUTY-CYCLE DURATION OR DURATION RATIO?

In experiment 1, we examined two extreme temporal conditions: in the classic configuration the high-level tones were very short (50 or 100 ms) and in the inverted configuration the low-level tones were very short (50 or 100 ms). We observed two distinct percepts (phenomenon or no phenomenon). But what happens between these two extremes? Does the phenomenon become progressively less strong as the temporal conditions change or is there an abrupt passage from the presence to the absence of the phenomenon? We tested these two possibilities by examining a range of sequences which varies systematically along two continua: the total duty-cycle duration and the duration ratio between high- and low-level tones.

A. Method

Eight new novice subjects with normal hearing participated in the experiment. They were all psychology students at the Université René Descartes. The same apparatus and procedure was used as in experiment 1. Fifteen sequences were constructed which varied in the duty-cycle duration and ratio between the duration of high- and low-level tones. Three duration ratios were examined: "classic," where the high-level tones were half the duration of the low-level tones (ratio 1:2), "equal," where the high- and low-level tones were the same duration (ratio 1:1), and "inverted," where

the high-level tones were twice as long as the low-level tones (ratio 2:1). For each duration ratio, the duty cycle varied in total duration (ratios 1:2 and 2:1—50/100, 100/200, 150/300, 300/600, 350/700, 700/1400 ms, ratio 1:1—100/100, 300/300, 700/700 ms). All other experimental details were identical to experiment 1.

B. Results

1. Subjective descriptions

Subjects were asked to describe each temporal combination. Contrary to expectations (the phenomenon should only be perceived if old temporally dominates the new, i.e., only in classic and equal conditions), all subjects described perceiving the phenomenon in all 15 duration conditions, with the exception of two subjects in the 50/100-ms condition who were unable to describe the sequence.

2. Duty-cycle duration

Figure 3 shows the mean adjusted levels averaged across the three duration ratios for each of the duty-cycle durations. In accordance with the subjective descriptions, adjusted levels for all 15 duration conditions followed the pattern associated with the perceived phenomenon (continuous correctly matched, intermittent underestimated). In all cases, for the continuous comparison sequences, a single-sample *t*-test revealed that the mean adjustment errors (difference from L_L) did not differ significantly from zero, and for the intermittent comparison sequence, a single-sample *t*-test revealed that the mean adjustment errors (difference from L_H) differed significantly from zero [$t(7) = 6.7$, $p < 0.001$]. Noteworthy is the fact that the size of the phenomenon (degree of underestimation) is constant across the different duration conditions, that is, it is unaffected by the duty-cycle duration and duration ratios examined here. A repeated-measures ANOVA on the mean adjustment errors for the continuous and intermittent comparison sequences by condition (15) revealed no significant difference between the conditions. Whereas only a limited range of ratios was examined, limiting conclusions for the moment, the duty-cycle durations covered the wide range of durations which could feasibly induce the phenomenon. Duty-cycle duration did not therefore appear to influence the perception of the phenomenon.

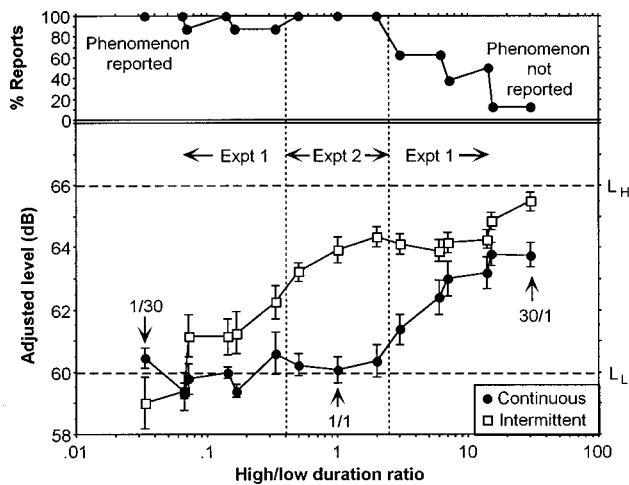


FIG. 4. Experiments 1 and 2: Percentage reports of the continuity phenomenon (upper panel) and mean adjusted levels for the continuous and intermittent comparison sequences (lower panel) as a function of the ratio between the durations of high- and low-level tones. The data are averaged over three duty-cycle durations, groups, and repetitions. Error bars indicate ± 1 standard error.

3. Duration ratios

In order to compare a wider range of duration ratios, Fig. 4 presents the mean adjusted levels and subjective descriptions for the 12 ratios used in experiment 1 and the three ratios used in experiment 2. The ratios ranged here from 1:30 (classic—50/1500 ms) to 30:1 (inverted—1500/50 ms), the wide range of duration ratios which could feasibly induce the phenomenon. The two predicted matching patterns were observed which corresponded with the subjective descriptions: that for the continuity phenomenon on the left of the figure, and that without the phenomenon on the right of the figure. Three points are noteworthy. First, as predicted, the phenomenon was observed for all ratios in which the old was longer than the new, but also for the inverted configuration where the new was twice as long as the old (2:1). Above this ratio, the phenomenon was less frequently observed. Second, when the phenomenon was perceived, the smaller the high/low duration ratio, the greater the underestimation of the level of the intermittent sequence. Underestimation was greatest for the highest ratios (-6 dB for 1:30 and 1:14), decreasing progressively to -2 dB for the 1:2, 1:1, and 2:1 ratios. Separate repeated-measures ANOVAs for the continuous and intermittent sequences on the mean adjustment errors for the seven ratios yielding reports of the phenomenon revealed a significant main effect of ratio for the intermittent comparison sequences [$F(1,6) = 14.6, p < 0.001$] but not the continuous sequences. Third, when the phenomenon was not perceived, the adjustments appeared to correspond to a global mean level: the adjustments increased gradually as the relative duration of the high tones increased, thus corresponding to an increase in mean level. A repeated-measures ANOVA on the mean adjustment errors for the four largest high/low duration ratios yielding fewer reports of the phenomenon revealed significant increases for both the continuous and intermittent comparison sequences.

C. Discussion

The results from experiments 1 and 2 indicate that the continuity phenomenon only occurred when the duration of the high-level tones was not too long in relation to the low-level tones: they must be in a ratio of at most 2:1. In these conditions, the smaller the ratio, the stronger the phenomenon. The duty-cycle duration does not influence the strength of the phenomenon. There was not an abrupt transition from the perception of the phenomenon to the absence of the phenomenon, but rather a gradual change from one state to the other, with a gradual decrease in the strength of the phenomenon as indicated by the decreases in the underestimation of the level of the intermittent sequence.

III. EXPERIMENT 3: INSTRUCTIONS EMPHASIZING THE RHYTHMIC STRUCTURE

In experiments 1 and 2, the instructions could be considered as facilitating the phenomenon because subjects were asked to adjust the intensity of a continuous comparison tone, thus emphasizing the phenomenon. Here, we test its robustness by comparing the previous results with those obtained with instructions which draw listeners' attention to the rhythmic structure rather than to an eventual continuity: listeners were required to adjust two intermittent comparison sequences, each corresponding to the duration of either the high- or low-level tones in the reference sequence. The rhythmic task might be considered to require listeners to focus on either the high- or the low-level tones independently in the reference sequence in order to match levels with the comparison sequence. However, if the phenomenon is compelling and unavoidable, then adjusting the comparison sequence with the low-level rhythm should correspond to the same pattern of results as in experiment 1, otherwise one might expect listeners to correctly match the physical levels in the reference sequence.

A. Method

The same eight subjects (four novices and four experts) as in experiment 1 participated in this experiment. Half the subjects (two experts and two novices) completed experiment 1 the first day, and then experiment 3 about a week later. The other half of the subjects did the two experiments in the opposite order. The same apparatus was used as in experiments 1 and 2. The reference sequences were identical to those in experiment 1. The comparison sequences were intermittent tones with the same onset and offset durations as either (a) the high-level tones or (b) the low-level tones in the reference sequences. They thus emphasized the rhythmic structure rather than the continuity phenomenon. All other details were as described previously. Subjects were instructed to adjust the level of the intermittent sequence to match that of the tones with the same rhythm in the reference sequence.

B. Results and discussion

1. Subjective descriptions

As for experiment 1, subjects were asked to describe the 12 sequences at the beginning of the first experimental ses-

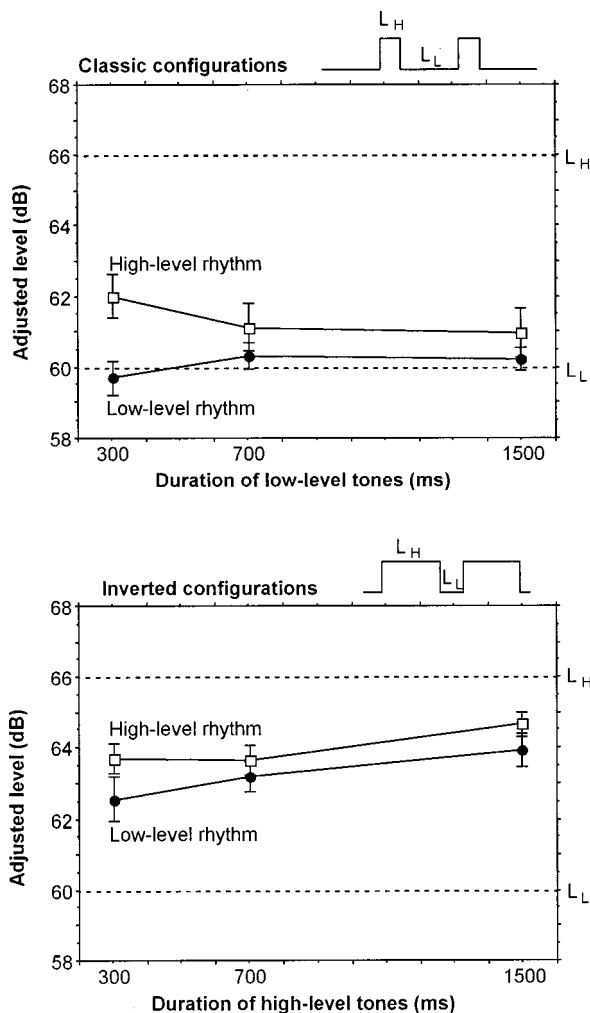


FIG. 5. Experiment 3: Mean adjusted levels for the intermittent comparison sequences for the 12 duration conditions averaged over groups, repetitions, and short-tone durations. Error bars indicate ± 1 standard error.

sion. No mention was made to the continuity phenomenon. In the classic configuration, 90% of the responses indicated the phenomenon, whereas only 45% indicated the phenomenon in the inverted configuration. The percentage reports of the phenomenon were more divergent across the two classes of stimuli for the expert group than for the novice group and were independent of the duration conditions.

2. Classic configuration

For the rhythm corresponding to the *low-level* (long-duration) tones, matches were extremely accurate (mean adjustment = 59.9 dB, mean adjustment error = 59.9 - 60.0 = 0.1 dB) (see Fig. 5). A single-sample *t*-test revealed that the mean adjustment error did not differ significantly from zero. A mixed ANOVA on the mean adjustment errors by group (novices/experts), duration of high-level tones (50 or 100 ms), duration of low-level tones (300, 700, or 1500 ms), and repetition (6) revealed no significant differences.

However, for the rhythm corresponding to the *high-level* (short-duration) tones, adjustment were far from accurate (mean adjustment = 61.4 dB; mean adjustment error = 61.4 - 66.0 = -4.6 dB), indicating that subjects systematically *underestimated* the level of these tones. A mixed ANOVA

on the mean adjustment errors by group (novices/experts), duration of high-level tones (50 or 100 ms), duration of low-level tones (300, 700, or 1500 ms), and repetition (6) revealed no significant differences between the two groups of subjects or over the six repetitions. The duration of the high-level tones did not influence the matches but the duration of the low-level tones did [$F(2,12) = 3.17$, $p < 0.04$]: Fig. 5 shows that the longer the low-level tone, the greater the adjustment error.

Thus, as for experiment 1, adjustment errors corresponded to the patterns expected if subjects were experiencing the continuity phenomenon. The phenomenon was greater when the low-level tones were longer.

3. Inverted configuration

For the rhythm corresponding to the *low-level* (short-duration) tones, systematic adjustment errors were observed (mean adjustment = 63.2 dB; mean adjustment error = 63.2 - 60.0 = +3.2 dB), indicating that subject systematically *overestimated* their level. A single-sample *t*-test revealed that the mean adjustment error differed significantly from zero [$t(7) = 8.4$, $p < 0.001$]. A mixed ANOVA on the mean adjustment errors by group (novices/experts), duration of high-level tones (300, 700, or 1500 ms), duration of low-level tones (50 or 100 ms), and repetition (6) revealed no significant differences, although the same tendency described above was observed here: the longer the high-level tone, the higher the adjustment of the low-level tones.

For the rhythm corresponding to the *high-level* (long-duration) tones, systematic adjustments were also observed (mean adjustment = 66 - 2 = 64 dB, mean adjustment error = -2.0 dB), indicating that subjects systematically *underestimated* their level. A single-sample *t*-test revealed that the mean adjustment error differed significantly from zero [$t(7) = 9.2$, $p < 0.001$]. An ANOVA on the mean adjustment errors by group (novices/experts), duration of low-level tones (50 or 100 ms), duration of high-level tones (300, 700, or 1500 ms), and repetition (6) revealed no significant differences between the two groups of subjects or over the six repetitions. The duration of the low-level tones did not influence the adjustments but the duration of the high-level tones did [$F(2,12) = 5.35$, $p < 0.02$]: the longer the high-level tone, the higher the adjustment of those tones.

Thus, as for experiment 1, in these inverted conditions, adjustment errors did not correspond to the patterns expected if subjects were experiencing the continuity phenomenon. Instead, the matches suggest that subjects were adjusting to a duration-weighted mean level: the adjusted levels increased as the relative duration of the high-level tones increased.

4. Comparison of experiments 1 and 3

The results obtained here are almost identical to those obtained in experiment 1. To test for an order effect reflecting the possibility that instructions emphasizing the phenomenon may influence its perception in later conditions, the results of experiments 1 and 3 were compared in a mixed ANOVA on the mean adjustment errors by group (novices/experts), duration of high-level tones (50 or 100 ms), dura-

tion of low-level tones (300, 700, or 1500 ms), repetition (6), and order (experiment 1/3; 3/1). Particularly worthy of note here is that the order of presentation of the two experiments and all of its interactions are not statistically significant, ruling out the possibility that the similarity in the data is due to contamination from the instructions orienting listeners toward the continuity phenomenon in experiment 1. The similarity of results from experiments 1 and 3 also occurs despite a different matching stimulus for the low-level tone, a continuous tone in experiment 1, and an intermittent tone in experiment 3. This suggests that the overestimation of the fainter sound's loudness was not a consequence of the mismatch between standard and comparison stimuli (continuous/intermittent or intermittent/continuous). Thus, the phenomenon is surprisingly robust because it occurs even when instructions should not favor perception of the phenomenon and should indeed create a bias toward perceptually decomposing the sequence.

IV. GENERAL DISCUSSION

According to the Auditory Scene Analysis framework proposed by Bregman (1990), the continuity phenomenon results from a primitive organizational process by which the auditory system is predisposed to expect existing sounds to continue throughout new events if no evidence exists to indicate that they have stopped. In this way, listeners organize the sequence into two streams. In the present case, the continuous tone is one stream (the old information) and the series of intermittent tones another (the new information). We have provided evidence of the presence of this phenomenon by asking listeners to adjust the perceived level of each stream separately. When the sequence is organized into two streams, the continuity phenomenon is perceived: the level of the low-level tones is correctly adjusted but there is a systematic underestimation of the level of the high-level tones. This underestimation suggests that the perceived level of the intermittent part (the new information) is derived from the high level by subtracting out the part assigned to the old stream.

Evidence in favor of the primitive nature of this phenomenon has been provided in three ways. First, the strength of the phenomenon is unaffected by the task: extremely similar underestimations of the level of the intermittent tones were obtained in both experiment 1 when the instructions emphasized the phenomenon and in experiment 3 when the instructions emphasized a different organizational principle, namely the interleaved rhythmic structure. The phenomenon occurs even when the listeners were asked to think about the stimuli in a different way. Second, it is thought that musical expertise often results in an enhanced ability to listen selectively to particular aspects of a sequence, and so we could have expected that musicians would be more resistant to the phenomenon: their adjustments might have been more veridical. However, this was not the case: no differences were observed between the musician and nonmusician listeners. Finally, repeated listening to the stimuli in our sequences could have led to more veridical adjustments. This was not the case either, as no improvements were observed over rep-

etitions. The phenomenon therefore appears to be particularly robust.

The main aim of this series of experiments was to establish whether the temporal structure of the old and new information influences the continuity phenomenon. Previous studies have tended to use only conditions in which the time period occupied by the old information alone is equal to or longer than that occupied by the combination of the old and new information. Is this just a random sampling error, or have researchers implicitly used the duration conditions under which the phenomenon is most clearly demonstrated? The latter is probably the case because we have demonstrated that the phenomenon only occurs under particular duration conditions: the old sounds must be relatively well established in relation to the new sounds. In experiments 1 and 3 we have shown that the phenomenon only occurs when the low-level tones (the old information) are almost equal to or longer than the high-level tones. In both these experiments we used extremely long (300, 700, and 1500 ms) and short (50 and 100 ms) tones. In experiment 2, we used the whole range of durations in order to examine the importance of the duty-cycle duration and duration ratios between the old and new information. Whereas the duty-cycle duration did not affect the degree of under- and overestimation, the relative duration is revealed to be the determining factor of whether or not the phenomenon is observed. More precisely, the high/low duration ratio must be at most 2:1. Warren (1994) states that continuity is reported when intermittent tones do not exceed 300 ms. Our results are coherent with this statement (the phenomenon is not observed in the inverted conditions where the intermittent/loud tones were 700 and 1500 ms). Contrary to previous assumptions (Houtgast, 1972; Warren *et al.*, 1972), and in agreement with data provided by Warren *et al.* (1994), the phenomenon is not all or nothing: its strength varies with the duration conditions, the degree of underestimation becoming greater as the high/low ratio decreases. This suggests that the more the old information dominates the new information, the stronger the tendency to perceptually organize the sequence according to this strategy.

Another issue concerns how well listeners are able to adjust loudness under various duration conditions, either when the continuity phenomenon is perceived or not. We have shown previously (McAdams *et al.*, 1998) that traditional loudness models positing precise loudness perception predict that tones should be adjusted close to the physical levels (66 dB for the long/high-level tones and 60 dB for the short/low-level tones). This is not what is observed for either the intermittent stream in the classic configurations nor for the inverted configurations. For the classic configurations, adjusted levels are those predicted by a pressure subtraction model (McAdams *et al.*, 1998) if the long/low-level tone duration is 1500 ms, but this predicted level is progressively overestimated as the duration of the tones decreases. In the inverted configurations, the intermittent/high-level tones were systematically underestimated (64.7 and 64.0 dB in experiments 1 and 3, respectively) and the continuous/low-level tones were systematically overestimated (62.9 and 63.2 dB, respectively).

Various explanations can be proposed for these systematic adjustment errors. A first possibility may be that listeners adjust towards a *mean level* (63.0 dB). Averaged over the different duration conditions and experiments, these predictions are quite close to the data for the inverted configuration (long/high-level=63.6 dB, short/low-level=63.0 dB). However, this explains neither the effect of duration condition (adjusted levels were higher for greater long/short ratios), nor the different adjustments for the high- and low-level tones. Further, the predictions are nowhere near the data for the classic configuration.

Second, a *duration-weighted mean*, taking into consideration the relative long-term contributions of the long- and short-duration levels (computed over the duration of the stimulus sequence), could explain the duration condition effect because the level of the longer tones would affect the mean more than the level of the shorter tones. However, for the inverted conditions the theoretical duration-weighted means (64.8, 65.4, and 65.7 dB for 300, 700, and 1500 ms, respectively) were considerably higher than the adjusted levels, whereas for the classic conditions, the duration-weighted means were in between the adjusted levels for the 300- and 700-ms long-tone durations (61.2 and 60.6 dB, respectively) and above the adjusted levels for the 1500-ms duration (60.3 dB). In all of these cases, calculations based on acoustic pressure units are nearly identical to results obtained when loudness units (sones) are used and then reconverted to dB. While the duration-weighted mean varies with the duration of the long-duration tones, it cannot account for the difference between adjustments for the long- and short-duration tones in either of the two stimulus configurations, since there are no free parameters in the equation related to focusing on a given subset of tones in the stimulus sequence.

A third possibility is that the relative weights accorded to the long-duration and short-duration tones may be modified by *attentional focusing* on them. In developing a model for this effect, we assume that listeners focus on high- or low-level tones (or long- or short-duration tones, according to the task instructions) and that this focusing is not perfectly selective, along a sensory dimension that we assume for simplicity's sake to be isomorphic to level, thus allowing influence from the other type of tone due to both tones stimulating largely the same population of auditory fibers. We hypothesize that the relative weighting of high- and low-level tones in their contribution to the perceived loudness would be different when attention is directed to one or the other (in an attempt to match their loudness) and that differences should also occur between classic and inverted configurations due to the different perceptual representations upon which the focusing operates. Our model for weighting the levels with respect to relative duration and attentional focus is given in Eq. (1)

$$L_{\text{adj}} = \alpha \times \frac{D_L}{D_L + D_H} \times L_L + \beta \times \frac{D_H}{D_L + D_H} \times L_H, \quad (1)$$

where L_{adj} is the adjusted level, D_L is the duration of the low-level tone, D_H is the duration of the high-level tone, L_L is the level (in dB) of the low-level tone, L_H is the level (in dB) of the high-level tone, and α and β are the weights for

low-level and high-level tones, respectively, for a given attentional focus. Since levels, durations, and mean adjusted levels (combined across 50-ms and 100-ms short-duration tones and using 75 ms as the duration) are known, we solve the simultaneous equations for the three pairs of long-duration tones (300, 700, 1500 ms) to determine α and β . For example, in the inverted configuration the three equations would be:

$$L_{\text{adj}}(300) = 0.2 \times 60\alpha + 0.8 \times 66\beta = 12.0\alpha + 52.8\beta$$

$$[D_L = 75 \text{ ms}, D_H = 300 \text{ ms}], \quad (2)$$

$$L_{\text{adj}}(700) = 0.097 \times 60\alpha + 0.903 \times 66\beta = 5.64\alpha + 59.60\beta$$

$$[D_L = 75 \text{ ms}, D_H = 700 \text{ ms}], \quad (3)$$

$$L_{\text{adj}}(1500) = 0.048 \times 60\alpha + 0.952 \times 66\beta = 2.88\alpha + 62.83\beta$$

$$[D_L = 75 \text{ ms}, D_H = 1500 \text{ ms}]. \quad (4)$$

Each equation can be rewritten in terms of α , e.g., for 300 and 700 ms

$$\alpha = \frac{L_{\text{adj}}(300) - 52.80\beta}{12.0}, \quad (5)$$

$$\alpha = \frac{L_{\text{adj}}(700) - 59.60\beta}{5.64}. \quad (6)$$

Assuming α equal in both cases, we solve for β

$$\beta = \frac{12.0 \times L_{\text{adj}}(700) - 5.64 \times L_{\text{adj}}(300)}{417.41}, \quad (7)$$

by inserting the mean adjusted levels for a given subject in Eq. (7). Doing this for the three pairs (300–700, 300–1500, 700–1500) gives three values of β which are averaged for each subject, since we assume that relative duration does not affect the attentional focus on high-level or low-level tones in either configuration. The mean β is reinserted into Eqs. (2)–(4) to obtain three values of α which are then averaged. The aim is to estimate a single set of coefficients for all temporal relations of a given stimulus configuration (classic, inverted) and focusing strategy (high-level tones, low-level tones). It is important to note that differences in the pattern of α and β across stimulus configurations and focusing strategies include differences due to the percept, i.e., no continuity in inverted and continuity (and thus the old-plus-new subtractive mechanism) in classic configurations. Plugging the mean coefficients averaged across subjects back into Eq. (1) gives the model predictions plotted in Fig. 2. A close correspondence exists between observed and predicted values when means are weighted both for duration and attentional focus. The mean values of α and β and their standard errors across subjects are plotted in Fig. 6 for the two stimulus configurations and two focusing conditions. If both coefficients have values close to 1, the duration-weighted mean accounts for the data. It is clear from this figure that both stimulus configuration and attentional focus affect the relative weights of the component tones in the sequence as they contribute to loudness.

These pairs of weighting coefficients vary in an interesting way across stimulus configurations and focusing strate-

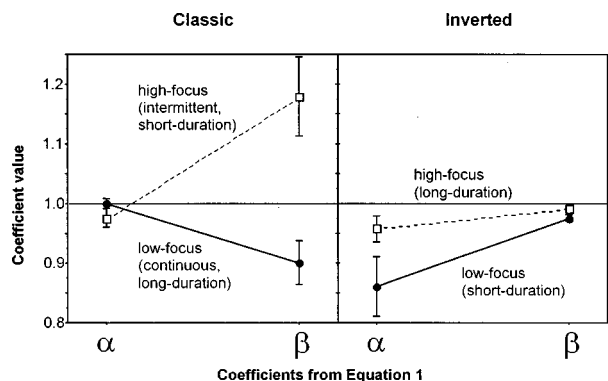


FIG. 6. Coefficients for the model that weight the levels according to relative duration and attentional focusing in loudness matches. Classic stimulus configuration in the left panel and inverted configuration in the right panel.

gies (Fig. 6). First, short- and long-duration tones do not have the same pattern of coefficients across stimulus configurations: focusing on low-level tones in the classic configuration is not the same as focusing on high-level tones in the inverted configuration and vice versa, suggesting that the differences between configurations are not due simply to a reversal of the high and low levels on the short and long durations. Second, high- and low-level tones do not have the same pattern across stimulus configurations: low-level classic is not the same as low-level inverted and high-level classic is not the same as high-level inverted, suggesting that the duration differences across configurations are critical to the perceived loudnesses. Both of these results are consistent with the differences in reports of the continuity phenomenon given by our subjects. Third, α and β are affected both by the stimulus configuration and by the focusing strategy. For the classic configuration, the weight on the low-level tones is determined entirely by the relative duration ($\alpha = 1$), whereas the weight on the high-level tones increases ($\beta > 1$) when they are the object of focus and decreases ($\beta < 1$) when they are not. A nearly inverse pattern exists for the inverted configuration: the weight on the high-level tones is largely determined by relative duration ($\beta \cong 1$), whereas the weight on the low-level tones is determined by relative duration ($\alpha \cong 1$) when they are not the object of focus and paradoxically decreases ($\alpha < 1$) when they are, indicating an overbearing influence of the high-level tones when trying to focus on the loudness of the short-duration, low-level tones in the inverted configuration.

In summary, the homophonic continuity phenomenon in the classic stimulus configuration is quite robust with respect to task instructions, listener expertise or training, and repeated exposure to the task. The phenomenon depends on the relative durations of high- and low-level tones and appears to require that the low-level portions temporally dominate the high-level portions, or at least that the ratio of their durations (high/low) be at most 2:1, in order to be perceived systematically. In stimulus conditions in which short-duration tones

alternate with long-duration tones, the effect of longer tones on loudness matches cannot be explained only by a simple mechanism involving the global level accumulated over an entire duty cycle, but also seems to involve a relative attentional weighting of the high- and low-level portions in inverted conditions and of the continuous and intermittent streams in the classic conditions. This attentional focusing is not perfectly selective and part of the unfocused portion appears to influence the perceived loudness. The nature of this relative weighting is clearly different for classic and inverted conditions: it involves an inversion of the high- and low-level weights suggesting quite different underlying sensory representations for the two stimulus configurations, consistent with subjective reports concerning the presence of the continuity phenomenon.

ACKNOWLEDGMENTS

This research was funded in part by a grant from the French Ministry of the Environment. Lucette Cysique and Sophie Savel helped run subjects and Renaud Brochard helped with DSP programming.

- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).
- Bregman, A. S., and Rousseau, L. (1991). "Auditory intensity changes can cue perception of transformation, accompaniment or replacement," *Bull. Psychonom. Soc.* **29**, 476.
- Ciocca, V., and Bregman, A. S. (1987). "Perceived continuity of gliding and steady-state tones through interrupting noise," *Percept. Psychophys.* **42**, 476–484.
- Drake, C., and McAdams, S. (1997). "The continuity illusion: Role of temporal sequence structure," *Actes du 4e Congrès Français d'Acoustique, Marseille, Vol. 1*, pp. 519–522.
- Houtgast, T. (1972). "Psychophysical evidence of lateral inhibition in hearing," *J. Acoust. Soc. Am.* **51**, 1885–1894.
- McAdams, S., Botte, M.-C., and Drake, C. (1994). "Le phénomène de la continuité auditive et la répartition de la sonie entre flux auditifs," *J. Phys. IV (C5)*, 383–386.
- McAdams, S., Botte, M.-C., and Drake, C. (1998). "Auditory continuity and loudness computation," *J. Acoust. Soc. Am.* **103**, 1580–1591.
- Thurlow, W. R. (1957). "An auditory figure-ground effect," *Am. J. Psychol.* **70**, 653–654.
- Thurlow, W. R., and Elfner, L. F. (1959). "Continuity effects with alternately sounding tones," *J. Acoust. Soc. Am.* **31**, 1337–1339.
- van Noorden, L. P. A. S. (1977). "Minimum differences of level and frequency for perceptual fission of tones sequences ABAB," *J. Acoust. Soc. Am.* **61**, 1041–1045.
- Verschuure, J., Rodenburg, M., and Maas, A. J. J. (1976). "Presentation conditions of the pulsation threshold method," *Acustica* **35**, 47–54.
- Warren, R. M. (1982). *Auditory Perception: A New Synthesis* (Pergamon, New York).
- Warren, R. M. (1984). "Perceptual restoration of obliterated sounds," *Psychol. Bull.* **96**, 371–383.
- Warren, R. M. (1999). *Auditory Perception: A New Analysis and Synthesis* (Cambridge University Press, Cambridge).
- Warren, R. M., Bashford, J. A., Healy, E. W., and Brubaker, B. S. (1994). "Auditory induction: Reciprocal changes in alternating sounds," *Percept. Psychophys.* **55**, 313–322.
- Warren, R. M., Obusek, C. J., and Ackroff, J. M. (1972). "Auditory induction: Perceptual synthesis of absent sounds," *Science* **176**, 1149–1151.
- Zwicker, E. (1977). "Procedure for calculating loudness of temporally variable sounds," *J. Acoust. Soc. Am.* **62**, 675–682.