Perceiving the numerosity of rapidly occurring auditory events in metrical and nonmetrical contexts

BRUNO H. REPP
Haskins Laboratories, New Haven, Connecticut

Experiment 1 determined the fastest tempo at which participants could tap in synchrony with every nth tone ($n = 2$ to 9) in an isochronous sequence. Tapping was difficult with every 5th or 7th tone but easy with every 2nd, 4th, or 8th tone, suggesting that evenly divisible groups of n tones are automatically subdivided into equal groups of 2 or 3—a form of auditory subitizing that generates metrical hierarchies commonly found in Western music. Experiments 2 and 3 sought evidence of subitizing and subdivision in timed explicit enumeration of short, rapidly presented tone sequences ($n = 2$ to 10). Enumeration accuracy decreased monotonically with n. Response time increased monotonically up to $n = 5$ or 6, but less between 2 and 3 than between 3 and 4. Thus, a single group of 2 or 3 tones perhaps can be subitized, but subdivision of larger groups into subgroups of 2 or 3 tones seems to be specific to a repetitive, metrical context.

Rate Limits of Accurate Numerosity Perception

The perception of numerosity is a topic of great scientific interest and has generated a considerable amount of research, especially in recent years (e.g., Barth, Kanwisher, & Spelke, 2003; Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004; Gallistel & Gelman, 1992, 2000; Lipton & Spelke, 2003; Whalen, Gallistel, & Gelman, 1999). One important issue being discussed in this literature is the extent to which numerical cognition in adult humans relies on nonverbal magnitude representations of number such as those used by infants and nonhuman animals. There is considerable evidence in favor of a dual number system in adult humans, with the discrete verbal counting system being overlaid upon and mapped onto a phylogenetically older analog system (Gallistel & Gelman, 1992, 2000). When people are prevented from using the verbal system, they rely on the nonverbal one, and verbal counting may always be accompanied by nonverbal representations. Most of the relevant research has been concerned with the perception of visual objects presented simultaneously, or sometimes sequentially. Relatively few studies have dealt with perceiving the number of sequential auditory events, although this is important in music performance.

For adults, counting successive objects or events is a fairly trivial task, unless time limits or other constraints are imposed. Therefore, research on numerosity perception usually concerns situations in which outright counting is difficult. In visual studies, simultaneous displays of multiple objects are presented briefly to prevent sequential counting. When presentation is sequential but in a fixed position, a high presentation rate is used to prevent partici-

pants from simply keeping count. The maximal event rate at which accurate enumeration is still possible is of theoretical interest: Does it depend on the maximal rate at which number words can be uttered or imagined? Are there differences between auditory and visual presentation? Such differences seem unlikely, if speech production is the limiting factor.

In a pair of classic studies, Taubman (1950a, 1950b) presented sequences of 1–10 short tones or light flashes at various constant interonset intervals (IOIs), and participants had to report the number of events. In the auditory study, performance was found to be perfect at an IOI of 125 msec. Errors, mainly underestimations, began to emerge at 100 msec and increased as the IOI was made even shorter. In the visual study, performance was perfect at an IOI of 500 msec. Errors, again primarily underestimations, began to emerge at 333 msec (the next shortest IOI used) and increased further at even shorter IOIs. These results reveal a striking difference in favor of the auditory modality. Massaro (1976) gives the fastest rate of subvocal (i.e., inner) speech as 6 syllables per second (IOI = 167 msec), which is in accord with empirical data by Landauer (1962) for reciting or imagining the English number words for 1–10. Taubman's results thus suggest that auditory events can be counted accurately at a faster rate than number words can be imagined, whereas visual events require a much slower rate. This rules out a common limit imposed by the rate of inner speech. Rather, the rate limits of accurate enumeration seem to represent modality-specific (and perhaps stimulus-specific) temporal limits of perceptual processing.

The rate limit for enumerating auditory events, with which the present study is concerned, has been investi-

B. H. Repp, repp@haskins.yale.edu

gated in several additional studies. Some researchers noted that Taubman (1950a) had blocked the stimulus sequences by IOI, which made it possible to respond on the basis of total sequence duration rather than on the basis of number of events, although that strategy was specifically discouraged in Taubman's instructions. However, Lechelt (1975) used a randomized design and found nearly perfect performance for 2–9 tones at an IOI of 125 msec (the shortest IOI in his experiment), which agrees with Taubman's results (in whose study 125 msec was the longest IOI). Garner (1951), too, concluded from a control experiment that randomization of IOIs made little difference. Other authors, however, have found that experimental design does matter, especially at very fast presentation rates (John, 1972; Robinson, 1992). The training of participants also plays a role. Both Taubman (1950a) and Lechelt (1975) used highly trained participants. Studies with less trained participants have found less accurate performance. Garner (1951) presented sequences of 1–20 tones at various IOIs in a blocked design. Even for single-digit counts, errors still occurred at IOIs of 167 msec, and near perfect performance was reached only at the next larger IOI, at 250 msec. Massaro (1976) presented sequences of 5 to 8 tones at various IOIs in a randomized design and obtained about 85% correct responses at an IOI of 125 msec and 95% at 155 msec. (See also Massaro, 1977.) These results suggest that the rate limit for enumerating auditory stimuli accurately lies somewhere between 100 msec and 200 msec, depending on participant training and experimental design.

Some studies have used only very fast presentation rates because they focused on the degree of underestimation when accurate enumeration is impossible, not on the rate limit as such. Cheatham and White (1954) hypothesized that a central process with a fixed rate of 9–11/sec determines the perceived numerosity and thus the degree of underestimation of extremely fast event sequences. Interestingly, they obtained a similar estimate of subjective rate for auditory and visual stimuli (Cheatham & White, 1952; Forsyth & Chapanis, 1958; White, 1963; White & Cheatham, 1959; White, Cheatham, & Armington, 1953), and Cheatham and White (1954) thought that this amodal limit was related to the alpha rhythm of brain waves. However, John (1972) expressed reservations about the constancy of subjective rate in view of the effect of experimental design factors, and both he and others (Pollack, 1968; Robinson, 1992) have pointed out that there is an inevitable confound between numerosity perception and perception of sequence duration, especially at fast rates at which numerosity can only be estimated. Perhaps as a consequence of this realization, there has been little recent research on auditory numerosity perception.

Rate Limits of Synchronization

A possibly related issue concerns the rate limit of synchronizing a rhythmic action (finger taps) with events in a sequence. Here one theoretical question of interest is whether this rate limit is biomechanical or motoric in nature (in which case it should be similar for auditory and visual stimuli) or whether it is a modality-specific (and perhaps stimulus-specific) sensorimotor or perceptual limit. Inspired by early observations of Dunlap (1910) and Bartlett and Bartlett (1959), Repp (2003) attempted to determine these rate limits (which he called *synchronization thresholds*) for both tones and light flashes. To bypass the biomechanical rate limit of finger tapping, which, at 5–7/sec, happens to be similar to that of subvocal speech (see Todor & Kyprie, 1980; Truman & Hammond, 1990), Repp asked participants to tap with every 4th tone in isochronous auditory sequences whose IOI decreased from 170 msec to 80 msec. Musically trained participants were able to maintain synchrony up to an IOI of 123 msec, on average. With visual sequences, however, much slower rates were required. Even though the task was as simple as making a tap with every light flash, synchronization broke down at an IOI of 459 msec, on average.

These different synchronization thresholds for auditory and visual stimuli are remarkably similar to the rate limits for enumerating auditory and visual events obtained by Taubman (1950a, 1950b), which suggests that enumeration and synchronization are subject to the same perceptual rate limits. It is unlikely that modality-specific temporal limits of enumeration limit synchronization, but the reverse seems more plausible: Enumeration may require synchronization of internal actions (such as beats or subvocal counts) to external events, and visual sequences are more difficult to synchronize with than are auditory sequences, because the visual system has poorer temporal resolution and is less strongly coupled to the action system (Fraisse, 1948; Repp & Penel, 2004). The common rate limit of enumeration and synchronization may thus represent a modality-specific temporal constraint on sensorimotor coordination, or on the perceptual individuation of sequential events.

One reason why accurate enumeration of auditory events can be achieved at rates faster than subvocal speech may be that it is not necessary to generate all number words in sequence; rather, it may be possible to count by twos, threes, or fours, just as it is possible to tap with every 2nd, 3rd, or 4th tone in a sequence. Indeed, it may not be necessary to generate any number words at all when counting implicitly and repetitively, as in tapping with every 2nd, 3rd, or 4th tone in a sequence. Instead, a form of *subitizing* (Kaufman, Lord, Reese, & Volkmann, 1949; Mandler & Shebo, 1982) may be involved. Subitizing is discussed in more detail later.

The Present Study

The present study was concerned specifically with the auditory modality and consisted of three experiments. Experiment 1 addressed rate limits of synchronization in tapping tasks involving repetitive grouping of events in long isochronous sequences, where the group length, or "count," n is fixed and prescribed. By contrast, Experiments 2 and 3 reinvestigated the accuracy of explicit reports of the number of events in short rapid sequences, where n is variable but is within a known range. The focus of Experiment 1 was on the hypothesis that different counts n may have different rate limits, as explained below. Experiments 2 and 3 were conducted to confirm

that the results of Experiment 1 are specific to repetitive implicit counting, and to further investigate the possibility of auditory subitizing for small n. Experiments 2 and 3 will be introduced in more detail after Experiment 1 has been reported.

Experiment 1 made use of an adaptive staircase procedure, employed successfully in recent synchronization experiments (Repp, 2005, 2007), to determine the synchronization thresholds for tapping with every nth tone in an isochronous monotone sequence, with n ranging from 2 to 9. By definition, these tasks required repeated grouping (implicit counting) of n events. The questions of interest were: Are these tasks all subject to the same sequence rate limit, or does the synchronization threshold increase with n? If the threshold increases, is the increase monotonic? There was one reason to predict that there would be an increase, and another reason to suspect that the increase would be nonmonotonic.

First, a well-established result in the literature on visual numerosity perception is that small numbers of simultaneously presented objects can be enumerated almost instantly, without any serial counting. This is known as *subitizing* (Kaufman et al., 1949; Mandler & Shebo, 1982). Kaufman et al. thought subitizing extended up to 6 objects, but Mandler and Shebo divided that range into a true subitizing range (1–3) and a range within which objects can be "kept in consciousness" and counted (4–6 or 4–7). The number of objects in larger arrays, presented briefly, can only be estimated. The main evidence for subitizing is a response-time (RT) function with a shallow slope (<100 msec per item) up to counts of 3 but a much steeper slope up to 7 or so, indicating subvocal counting, after which an asymptote is reached because larger numbers can only be estimated. Subitizing has been attributed to perceptual grouping (Beckwith & Restle, 1966; Warren, 1897) and recognition of familiar spatial patterns formed by objects (Mandler & Shebo, 1982; Wender & Rothkegel, 2000).

Subitizing in the auditory modality has been much less investigated and discussed, and apparently no study has been done to examine RTs for auditory numerosity judgments. Fraisse and Fraisse (1937) presented tone sequences of various lengths, and prevented counting by means of a secondary speech production task. Participants had to reproduce the sequences by tapping. The authors concluded that 2–5 successive sounds could be apprehended immediately without counting, although accuracy was far from perfect. Taubman (1950a) and Garner (1951) reported continuously increasing error rates in numerosity estimates as a function of the number of auditory events in a rapid sequence. Garner's participants made some errors with counts of 3, and even with counts of 2. However, Taubman's highly practiced participants hardly made any errors for counts of 1 to 3, regardless of rate, and few with counts of 4. Thus, it is reasonable to expect that it would be easier to implicitly count small numbers (up to 4) repeatedly than to count larger numbers repeatedly. Counting up to 2, 3, or 4 is extremely common in music because these are common metrical subdivisions, whereas counting up

to larger numbers is rarely required. Therefore, it was predicted that musically trained participants would find the synchronization tasks for $n = 2$–4 easier (i.e., exhibiting a higher rate or lower IOI limit) than those for $n = 5$–9.

Second, if small numbers of successive events are much easier to apprehend than large numbers, strategies of subdivision and hierarchical counting suggest themselves when dealing with larger numbers. Sequences of 6, 8, or 9 events can be subdivided into groups of equal size. Such even subdivision corresponds to the commonly used time signatures (meters) in Western music, which are notated as fractions: 3/4 (= 2 + 2 + 2) or 6/8 (= 3 + 3) for $n = 6$; 4/4 (= 2 + 2 + 2 + 2) or, rarely, 2/2 (= 4 + 4) for $n = 8$; and 9/8 (= 3 + 3 + 3) for $n = 9$. Sequences of 5 or 7 events, however, cannot be subdivided into equal groups. If they are subdivided, they yield unequal groups, such as 2 + 3 or 3 + 2 for $n = 5$, and 3 + 2 + 2 (or 3 + 4), 2 + 2 + 3 (or 4 + 3), or 2 + 3 + 2 for $n = 7$ (London, 2004; Repp, London, & Keller, 2005). Subdivision into unequal groups, or implicit counting of larger groups without subdivision, may well make greater demands on cognitive resources than subdivision into equal groups. Therefore, tapping with every 5th or 7th tone in a sequence may be more difficult than tapping with every 6th, 8th, or 9th tone. In other words, it was predicted that the synchronization threshold would increase with n in a decidedly nonmonotonic fashion.

Experiment 1 was first conducted with a group of musically trained individuals, but later a group of participants with little or no musical training (called "nonmusicians" henceforth) was added. The comparison of musicians and nonmusicians addressed two issues: the generality of automatic subdivision and the role of subvocal counting. First, if automatic subdivision of large groups is a consequence of musical training, nonmusicians should not subdivide and show more nearly monotonic increases in their synchronization threshold with n. Second, because nonmusicians are more likely than musicians to have high synchronization thresholds (in terms of IOI), the sequence IOIs might become long enough for subvocal counting of the tones as they occur; this could reduce, or even eliminate, differences among the tasks with different n. Thus, there were two hypotheses predicting smaller differences among tasks for nonmusicians than for musicians.

EXPERIMENT 1

Method

Participants. The musician group consisted of 7 paid volunteers (4 women, 3 men) and the author. All were regular participants in synchronization experiments. They included 4 classically trained professional musicians (2 violinists, 1 violist, 1 clarinetist), one amateur classical pianist (the author), 2 amateur percussionists who played nonclassical styles (one of whom was also a pianist), and one individual who had had 5 years of piano instruction but no longer played.[1] Ages ranged from 18 to 35, except for the author, who was 58 at the time.

The nonmusician group initially included 11 individuals who had responded to a campus advertisement that asked for participants with little or no musical training; they were paid for their participation. Four of them were subsequently excluded, which left 7 partici-

pants (6 women and 1 man, age range 18–32) who yielded usable data.[2] Answers to a brief questionnaire about their musical experience revealed that 3 had had no musical training at all, whereas the other 4 had had up to four years of instruction, but had stopped playing their instruments at least six years previously. One participant said she was a dancer.

Materials and equipment. The auditory sequence consisted of identical tones produced by a Roland RD-250s digital piano under control of a program written in MAX 4.0.9. The software ran on an iMac G4 computer connected to the digital piano via a MOTU Fastlane-USB MIDI translator. The tones had a high pitch (E_7, MIDI pitch 100, 2637 Hz), sharp onsets, and a nominal duration of 20 msec. (There was some residual decay following the nominal offset.) At the beginning of the sequence, as described below, some of the high tones were accompanied by low tones (E_3, MIDI pitch 52, 165 Hz), having a nominal duration of 50 msec. All tones were produced at the same nominal intensity (MIDI key velocity).

Participants sat in front of a computer monitor on which the current trial number was displayed, listened to the sequences over Sennheiser HD540 II headphones at a comfortable loudness level, and tapped with their preferred hand (the right hand for all but one) on a Roland SPD-6 percussion pad held on the lap. Most participants rested the wrist and other fingers of the right hand on the surface of the pad and tapped by moving the index finger only; some, however, also moved the wrist and elbow. The impact of the finger on the rubber pad was audible as a thud, in proportion to the tapping force.

Procedure. The experiment comprised two sessions, which were typically one week apart. There were eight tasks: to tap with every *n*th tone, with *n* ranging from 2 to 9. During the first session, the tasks were done in order of ascending *n*, and in the reverse order during the second session.

Each task comprised a variable number of trials called a *run*. A run typically lasted between 5 and 10 min. At the beginning of a run, the IOI duration of the tone sequence was set to a value at which the task could be managed. For musicians, the initial IOI was always 200 msec. Nonmusicians could choose from the values 200, 250, and 300 msec. They were encouraged to try a short IOI first but to restart the run with a longer IOI if they could not manage the first trial after repeated attempts.

Participants initiated a trial by pressing the space bar of the computer keyboard. After a delay of 2 sec, the sequence started playing. The first high tone and every *n*th high tone thereafter (until tapping started) were accompanied by a low tone that indicated the spacing of the target tones for synchronization. Participants were free to start tapping whenever they felt ready. The low tone disappeared as soon as tapping started. The computer kept track of the number of taps made and of their asynchronies relative to the target tones.

For a trial to be considered successful, each of *N* successive taps had to be within ±IOI/2 of a target tone. *N* varied inversely with *n* so as not to make trials excessively long. It ranged from 40 (*n* = 2) to 12 (*n* = 9), according to the formula $N = 40 − 4 * (n − 2)$. After *N* successful taps had been made, the sequence stopped, a positive message appeared on the computer screen, and the IOI for the next trial was reduced by Δ, with $\Delta = 10$ msec initially. If any tap did not meet the accuracy criterion of ±IOI/2, the sequence stopped immediately, a negative message appeared on the computer screen, the IOI for the next trial was increased by Δ, and Δ for subsequent trials was reduced by 2 msec. As soon as Δ reached zero (i.e., after five unsuccessful trials), a chime sounded to indicate the end of the run. The final IOI duration was considered the estimate of the synchronization threshold. The asynchronies between taps and target tones were also recorded and saved.

Results

Synchronization thresholds. The mean synchronization thresholds of musicians and nonmusicians in the eight tasks are shown in Figure 1. The data were subjected to a 2 × 2 × 8 mixed-model ANOVA with the
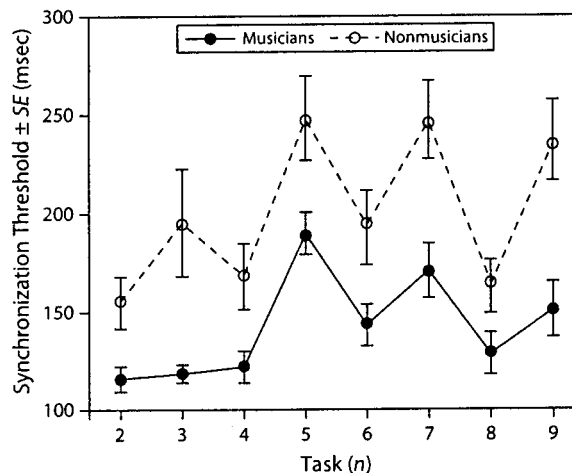

Figure 1. Mean synchronization thresholds of musicians and nonmusicians in the eight tapping tasks of Experiment 1, with standard error bars.

variables of group (musicians vs. nonmusicians), session (first vs. second), and task (*n* = 2 to 9).[3] As expected, synchronization thresholds were substantially lower for musicians than for nonmusicians [$F(1,13) = 14.93, p < .002$]. In addition, the synchronization threshold varied significantly across the eight tasks [$F(7,91) = 17.40, p < .001$].[4] As predicted, the "5" and "7" tasks were the most difficult, whereas the "2," "4," and "8" tasks were the easiest. Contrary to predictions, the differences among tasks were more pronounced for nonmusicians than they were for musicians; nonmusicians also seemed to have relatively greater difficulty with "3," "7," and "9" than did musicians. However, the group × task interaction was not significant [$F(7,91) = 1.58, p < .22$]. There were no significant effects involving session, indicating neither improvement with practice nor effects of the order of the tasks.

Separate ANOVAs on the data of the two participant groups showed the main effect of task to be significant for both musicians [$F(7,49) = 13.11, p < .001$] and nonmusicians [$F(7,42) = 7.61, p < .005$]. To determine which tasks differed significantly within each participant group, the Newman–Keuls procedure for repeated measures (Winer, 1971) was used with a criterion of $p < .05$. For musicians, "5" was significantly more difficult than all other conditions except "7," and "7" was significantly more difficult than "2," "3," "4," and "8." For nonmusicians, "7" was significantly more difficult than "2" and "8," and "5" was significantly more difficult than "2."

A less conservative test, an ANOVA with the "5" and "7" tasks omitted, still yielded a significant main effect of task, for both musicians [$F(5,35) = 5.30, p < .02$] and nonmusicians [$F(5,30) = 5.32, p < .02$]. This supports the impression from Figure 1 that "6" and "9" (as well as perhaps "3" for nonmusicians) were more difficult than "2," "4," and "8." When only the "2," "3," "4," and "8" tasks were retained in the ANOVA, the main effect of task was no longer significant for either group of participants.

Mean asynchronies and standard deviations. Analyses of asynchronies were restricted to successful trials.[5] The means and standard deviations (*SD*) of the asynchronies were calculated for each trial and then averaged across all trials in a run.

Neither musicians nor nonmusicians showed the negative mean asynchrony that is typically observed in simple 1:1 synchronization tasks. This can be attributed to the presence of explicit subdivisions (Repp, 2003), as well as to the fact that negative asynchronies tend to decrease with IOI duration (e.g., Mates, Radil, Müller, & Pöppel, 1994; Peters, 1989). On average, the asynchronies were close to zero, but there were considerable individual differences, with some participants tapping early and others late. An ANOVA on these data did not reveal any significant differences between groups of participants, tasks, or sessions.[6]

Figure 2 shows the mean within-trial *SD*s. Clearly, the musicians tapped with much less variability than the nonmusicians [$F(1,13) = 12.77, p < .003$]. In addition, however, there were significant main effects of task [$F(7,91) = 4.95, p < .02$] and of session [$F(1,13) = 4.71, p < .05$], as well as a group $\times$ session interaction [$F(1,13) = 10.56, p < .006$]. A separate ANOVA on the musicians' data showed a significant main effect only of task [$F(7,49) = 8.62, p < .007$]. It is evident that the asynchronies for the "2," "3," and "4" tasks were less variable than were those of the other tasks. There was no tendency, however, for variability to be higher in the successful trials of the "5" and "7" tasks than in those of the "6," "8," and "9" tasks, or to increase monotonically in parallel to the increase of the intertap interval with *n*. A separate ANOVA on the nonmusicians' data did not reveal any significant difference among tasks, although the low variability in the "2" task might be noted. Only the main effect of session was significant [$F(1,6) = 9.21, p < .03$], indicating some improvement between sessions.
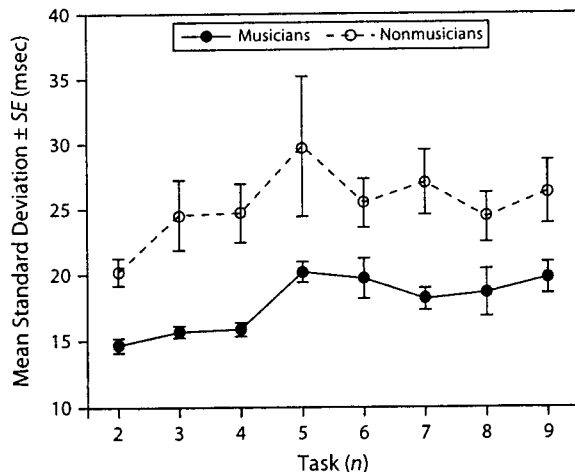

Figure 2. Mean within-trial standard deviations of asynchronies for musicians and nonmusicians in the eight tapping tasks of Experiment 1, with standard error bars.

One potential problem with this analysis is that the trial runs for the easier tasks contained wider ranges of IOIs (i.e., extending to shorter IOIs) than the difficult tasks. If *SD*s changed systematically with IOI (and intertap interval) duration, this would contaminate comparisons among tasks. To assess whether such changes occurred, the *SD* was linearly regressed on IOI within each individual run that had at least five trials and covered a range of at least 20 msec. Because runs with few trials or a narrow range of IOIs still yielded untrustworthy results, both the number of trials and the IOI range were used as joint multiplicative weights in computing weighted averages of the slopes of the regression lines for each participant across tasks. They were -0.03 for musicians [$t(7) = -2.66, p < .04$], and -0.05 for nonmusicians [$t(6) = -2.33, p < .06$]. Thus, *SD*s exhibited a slight tendency to increase as IOI decreased. Therefore, the finding of significantly smaller *SD*s in the "2," "3," and "4" tasks than in the other tasks for musicians (Figure 2) can be safely interpreted as indicating the low difficulty level of these tasks.

Discussion

Musicians' synchronization thresholds for the "2," "3," and "4" tasks were in the vicinity of 120 msec, similar to the thresholds obtained previously in a "4" task (Repp, 2003) and in a "3" task (Repp, 2007) with different groups of musically trained participants.[7] Not surprisingly, nonmusicians had thresholds that were considerably higher, on average. These synchronization thresholds either reflect a limit of perceptual temporal resolution beyond which it is no longer possible to track individual events, or a sensorimotor limit (e.g., of feedback processing or error correction) that prevents the accurate placement of taps in such a fast sequence. (For a more detailed discussion, see Repp, 2006.) While the exact nature of the limit remains to be clarified, it is clear that it varies with task demands, as well as with musical training.

The main finding of Experiment 1 is that the synchronization thresholds differed significantly among the eight tasks. They were highest in the "5" and "7" tasks, intermediate in the "6" and "9" tasks, and lowest in the "2," "3," "4," and "8" tasks (though some nonmusicians had difficulty with "3"). The only plausible explanation of these nonmonotonic changes is that they reflect subdivision of the groups of *n* events defined by the taps into smaller groups. Formation of subgroups is equivalent to creation of an additional metrical level below the main beat (defined by the taps).

In the "5" and "7" tasks, participants must have either refrained from subdivision or have subdivided the large groups into unequal subgroups (2 + 3 or 3 + 2 for "5"; 2 + 2 + 3, 2 + 3 + 2, or 3 + 2 + 2 for "7"). Indeed, musicians' mean synchronization thresholds in those conditions were similar to those obtained in an earlier study, in which musicians had to tap unequal subdivisions explicitly while synchronizing with a fast isochronous sequence (Repp et al., 2005). The lower thresholds in the "6," "8," and "9" tasks suggest that participants subdivided those large groups into equal subgroups of 2, 3, or 4, and that such even subdivision facilitated synchronization com-

pared with uneven (or no) subdivision. The "2," "3," and "4" tasks were easiest because they did not require subdivision (although binary subdivision into 2 × 2 may have occurred in the "4" task). The finding that the "8" task was no more difficult than the "2" and "4" tasks suggests that binary subdivision (2 × 4, 4 × 2, or even 2 × 2 × 2) is a particularly natural process. The intermediate difficulty of the "6" and "9" tasks suggests that ternary subdivision (2 × 3 or 3 × 2 for "6"; 3 × 3 for "9") requires more processing resources than binary subdivision. For some nonmusicians, even the "3" task presented difficulties. This is consistent with other evidence in the literature suggesting that ternary subdivision is more demanding than binary (Bergeson & Trehub, 2006; Repp, 2003; but see also Drake, 1997).

It was not determined to what extent participants were aware of their subdivision strategies and what particular strategies they used when several were available (e.g., 2 × 3 or 3 × 2 for "6"), although in hindsight that might perhaps have been a good thing to do. The instructions did not mention subdivision, although for musicians it is of course a very natural thing to do. Yet the author as a participant can testify that he was not really aware of subdividing the large groups and certainly did not pursue a deliberate subdivision strategy. For nonmusicians, such a spontaneous strategy would also seem less likely. Therefore, binary and ternary subdivision probably occurred quite automatically.

Support for that claim comes from a closely related involuntary perceptual phenomenon that has been known for a long time: *subjective grouping* or *rhythmization* (Bolton, 1894; MacDougall, 1903). Bolton presented uniform isochronous sequences at various rates to listeners who reported their subjective experience. They commonly reported hearing the auditory events in groups of twos, threes, or fours, depending on the sequence rate. The periodic recurrence of these groups implies imposition of a subjective (most likely, group-initial) beat onto the sequence. The sequence events thus functioned as subdivisions of the subjective beat, constituting the lower level of a simple two-level metrical hierarchy. In the present tasks, the participants' taps defined a main beat that constrained subjective grouping. The subjective grouping thus took place within the confines of an explicit beat that demarcated larger groups.

Temporal grouping of successive auditory events into twos, threes, and perhaps fours is analogous to the spatial grouping of small numbers of simultaneous objects in space (i.e., subitizing) observed in visual enumeration tasks, which likewise occurs automatically and does not depend on subvocal counting (Logie & Baddeley, 1987). Temporal grouping may reflect an auditory short-term memory with a capacity of up to four items, as postulated by Cowan (2001), who cites subitizing research in support of his theory. The automatic subdivision of larger groups may be a direct consequence of the engagement of this capacity-limited memory (see also Klahr, 1973). The results of the present experiment could thus be regarded as indirect evidence for an auditory subitizing process.

The occurrence of automatic subdivision or subjective grouping can be made even more plausible by considering internal periodicities (oscillators or timekeepers) that might be entrained by a stimulus sequence (Large & Jones, 1999). An isochronous sequence will entrain not only an internal oscillator whose period corresponds to the IOI but also others with periods of 2*IOI or 3*IOI, because events recur periodically at these intervals. Those multiple periodicities will differ in amplitude or relative salience, according to a broad sensorimotor resonance function (peaking at 500–700 msec) that determines preferred beat frequencies in music (Parncutt, 1994; Todd, Lee, & O'Boyle, 2001; van Noorden & Moelants, 1999). At sequence rates approaching the synchronization threshold, the IOI oscillator becomes weak and eventually disappears. This means that oscillators with a period of 2 * IOI or 3 * IOI become stronger than the IOI oscillator and automatically assume control of the task. Which of them takes over is determined by the tapping period (n * IOI), which reinforces one or the other, according to whether n is evenly divisible by 2 or 3.

Why does even subdivision lead to a lower synchronization threshold, and why is triple subdivision less effective than duple? One likely reason is the fact that variability increases with interval duration (Peters, 1989; Wing & Kristofferson, 1973). Thus the present results can be explained at least in part by the size of the smallest oscillator period, once a period equal to the IOI can no longer be maintained internally: 2 * IOI for "2," "4," and "8" (assuming 4 × 2 or 2 × 2 × 2 subdivision); 3 * IOI for "3," "6" (assuming 2 × 3), and "9"; and n * IOI for "5" and "7." Multiple mutually coupled oscillators provide dynamic support for the internal representation of a hierarchical metrical structure (Large, 2000, 2001; Large & Kolen, 1994; Large & Palmer, 2002). By encouraging subdivision of the tapping beat, the present tasks thus can be said to have led to the spontaneous emergence of metrical structure.

Contrary to expectations, nonmusicians did not show smaller differences among the tasks than did the musicians. This implies not only that musical training is not a prerequisite for the observed differences, but also that automatic subdivision occurs even when the sequences are slow enough to permit verbal counting. A glance at Figure 1 shows that nonmusicians' synchronization thresholds were often well above 170 msec, the estimated maximal rate of single-digit counting (Landauer, 1962; Massaro, 1976). Indeed, at least one nonmusician participant was observed counting along with the sequence tones. However, this did not seem to reduce the differences in the relative difficulty of the various tasks.

Even though musical training affected only overall performance, musicians and nonmusicians shared exposure to Western music, and it is possible that that experience is reflected in the results. In Western musical styles, duple meters (4/4, 2/4) are more common than triple meters (3/4, 6/8, 9/8), and uneven meters (5/8, 7/8) are rare. In future research, it would be interesting to test participants from cultures such as Bulgaria or Macedonia, in which uneven meters are common (Hannon & Trehub, 2005).[8]

EXPERIMENT 2

The results of Experiment 1 raise the question of whether auditory subitizing—henceforth considered synonymous with automatic subdivision and subjective grouping into twos or threes, and perhaps fours—occurs only when auditory grouping occurs repeatedly within an emergent metrical rhythmic framework, or whether it occurs also in simple enumeration. In other words, is enumerating 5 or 7 tones, presented just once, more difficult than enumerating 6, 8, or 9 tones? And is enumerating 2, 3, or 4 tones equally easy?

The auditory enumeration studies of Taubman (1950a) and Garner (1951), which present error data in considerable detail, do not suggest any special difficulty of counting 5 or 7 tones. In general, the proportion of errors increased monotonically with n, the number of tones presented. Errors were naturally scarce with 2–4 events, but this cannot be considered strong evidence for subitizing. It seems that no auditory enumeration study in the literature has measured RT. By contrast, response time has been the primary measure in studies of visual numerosity judgment (Mandler & Shebo, 1982), and it has provided the main empirical evidence for subitizing, namely a more gradual increase in RT for small numbers than for larger numbers.

The purpose of Experiment 2, therefore, was to analyze both error rates and RT for numerosity judgments elicited by rapid sequences of tones. The two main questions were (1) whether it would take participants longer to determine that 5 or 7 tones have occurred than to decide that 6, 8, or 9 tones have occurred; and (2), whether RTs for small numbers of tones are nearly constant, but then increase steeply for larger numbers of tones. Positive answers to these questions would indicate the operation of a subitizing mechanism in simple auditory enumeration. Negative answers would suggest that auditory subitizing occurs only in the context of longer, metrical sequences.

The participants in Experiment 2 were the musicians from Experiment 1. This made it possible to compare enumeration performance with synchronization thresholds. One question of interest was whether participants who have low synchronization thresholds are also proficient at perceiving the number of rapidly presented tones. Such a correlation between synchronization performance and enumeration accuracy would suggest that the two tasks engage similar perceptual and cognitive processes. A second question is whether the two tasks are subject to similar temporal limits. If so, enumeration of tones should be accurate as long as synchronization is accurate. Earlier auditory studies (reviewed above) suggest that, depending on the design of the study and on the participants' experience, enumeration may reach perfection anywhere between 125 and 200 msec of IOI, which is not inconsistent with the synchronization thresholds observed in Experiment 1. However, if a short sequence does not elicit automatic subdivision, enumeration may be more severely time-limited than synchronization.

Method

Participants. The participants were the same 8 musicians as in Experiment 1.

Materials and equipment. The stimulus sequences consisted of the same high-pitched tones as in Experiment 1 but were much shorter, comprising 2 to 10 events. The IOI durations ranged from 80 to 170 msec in steps of 10 msec. Thus there were 90 different sequences (trials), which were presented in 10 different randomized orders (blocks). The equipment was the same as in Experiment 1, except that the percussion pad for tapping was no longer used.

Procedure. Participants came for two sessions, typically one week apart, and listened to the same 10 blocks of 90 trials in each session.[9] At the beginning of Session 1, a few sample trials were presented, and it was pointed out that the number of tones (n) ranged from 2 to 10. After listening to each sequence, participants typed the number of events they thought it contained on the numeric pad of the extended computer keyboard, using the "0" key for "10." The next sequence started 2 sec after the response. Participants were instructed to respond quickly and not to deliberate for a long time or to count events in auditory short-term memory. RT was measured from the onset of the last tone in each trial. At the end of each block, participants saved the data in a file and selected the next block. Each session lasted about 75 min.

Analysis. The data were analyzed in terms of three dependent variables: percent correct (PC), deviation from the correct count (DC), and RT. First, the extent to which PC and RT changed between sessions and across blocks within each session, especially within Session 1, was examined. Then the data were pooled across blocks and sessions and each of the dependent variables was subjected to an ANOVA with the independent variables of n (9 levels) and IOI (10 levels). However, because for PC and DC there were data cells without variation (due to a ceiling effect), the statistics program (SYSTAT) refused to test the two-way interaction. Instead, two one-way ANOVAs were conducted on these data after averaging them across all levels of either n or IOI.

RTs varied greatly, and participants apparently could not always resist the temptation to mull over a difficult count. To filter out extremely long RTs, the 90 RTs of each block were subjected to a recursive procedure that eliminated all values that fell beyond three SDs from the mean. Subsequently, the median RT was computed across the 10 blocks within a session for each of the 90 combinations of n and IOI, regardless of the correctness of the response. These medians were then averaged across the two sessions.

Results

Practice effects. A 2 (sessions) × 10 (blocks) repeated measures ANOVA on the PC scores yielded no significant main effects but a nearly significant interaction [$F(9,54) = 2.90, p < .06$]. This was due to somewhat lower scores (by about 5%) in the first block of Session 1. From the second block onward, however, scores were fairly constant, and it was not deemed necessary to exclude the data of the first block. A similar ANOVA on the median RTs yielded no significant effects. Thus, performance was rather stable throughout the experiment, and no improvement seemed to occur after Block 1.

Individual differences. The grand mean PC score was 70%, with individual scores ranging from 47% to 86%. The correlation of the individual PC scores with the individual mean synchronization thresholds in Experiment 1 was −.50 ($p > .1$). The grand mean RT was 1,078 msec, with individual mean RTs ranging from 846 to 1,336 msec. The correlation of these values with the individual mean synchronization thresholds was .50. Neither correlation

reached significance because of the small number of participants, but they do suggest a possible relationship: Those with low synchronization thresholds tended to be faster and more accurate in the simple enumeration task. The correlation between PC and RT was $-.43$.

Percent correct. The PC data are shown in Figure 3 as a function of n with IOI as the parameter. The heavy line without symbols represents the mean. The main effect of n was pronounced $[F(8,56) = 25.72, p < .001]$, with a clear discontinuity between 5 and 6: Up to 5 tones were generally enumerated accurately, but 6 or more were not. Within each of these ranges, accuracy decreased gradually as n increased. There was no evidence of a dip in PC for 5 and/or 7 tones. Enumeration of 2–5 tones was not strongly affected by IOI because of a ceiling effect. Surprisingly, 6–8 tones seemed to be affected more by IOI than 9 and 10 tones, mainly due to the shortest IOIs, at which accuracy for 9 and 10 tones was relatively improved. There was a significant main effect of IOI $[F(9,63) = 15.82, p < .001]$, but it was rather gentle: PC increased from about 55% correct at the shortest IOI (80 msec) to only about 75% correct at the longest IOI (170 msec). Inspection of individual data revealed that three participants showed fairly steep increases in PC as IOI increased, four showed shallower increases, and one showed no change at all.

Deviation from correct count. The DC data are shown in Figure 4. Deviations were almost always in the direction of underestimation, which is in agreement with previous studies. DC was dependent on n $[F(8,56) = 9.63, p < .02]$: Numbers of 2–5 were not underestimated, but 6–10 were increasingly underestimated. IOI had no significant main effect on the extent of underestimation $[F(9,63) = 0.65]$.

RTs. The RT data are shown in Figure 5. A strong dependence of RT on n is evident $[F(8,56) = 11.81, p < .001]$: RT increased steeply as the number of tones in-

Figure 4. Results of Experiment 2: Deviation from correct count as a function of n, with IOI as the parameter.

creased from 2 to 6 and then remained fairly constant. (Only at the shortest IOI did a decrease in RT occur, when the number of tones exceeded 7.) A regression line fitted to the mean RTs for 2–6 tones had a slope of 120 msec. It is evident, however, that RTs increased much more between 3 and 4 and between 5 and 6 than between 2 and 3 and between 4 and 5. A separate ANOVA confirmed that the RT increase between 2 and 3 was significantly smaller than that between 3 and 4 $[F(1,7) = 7.69, p < .03]$. RTs tended to decrease gradually as IOI increased, except for 2–3 tones; however, the main effect of IOI $[F(9,63) = 2.12, p < .18]$ and the IOI $\times$ n interaction $[F(72,504) = 1.39, p < .27]$ were not significant.

There were substantial individual differences in RT as a function of n for the larger numbers: Four participants showed functions that more or less resembled the average function in Figure 5; two showed functions that peaked at 7 and then decreased; and two showed functions that increased steadily up to 9 and only then took a slight downturn. One participant showed a twofold increase in RT between 2–5 and 6–10; all others showed more gradual functions.

Sequence duration. Despite the randomized presentation of all combinations of n and IOI, there was an inevitable confound between n and sequence duration: Only sequences with small n could have very short total durations (from the onset of the first tone to the onset of the last tone, minimum $= 80$ msec), and only sequences with large n could have very long durations (maximum $= 1,530$ msec). Thus, sequence duration could have served as a cue to the number of tones in a sequence. Figure 6 shows PC and RT as a function of sequence duration, with n as the parameter. Figure 6A shows that, although PC decreased as sequence duration increased, at each level of n there were substantial effects of IOI duration orthogonal to the variation in sequence duration. Therefore, sequence
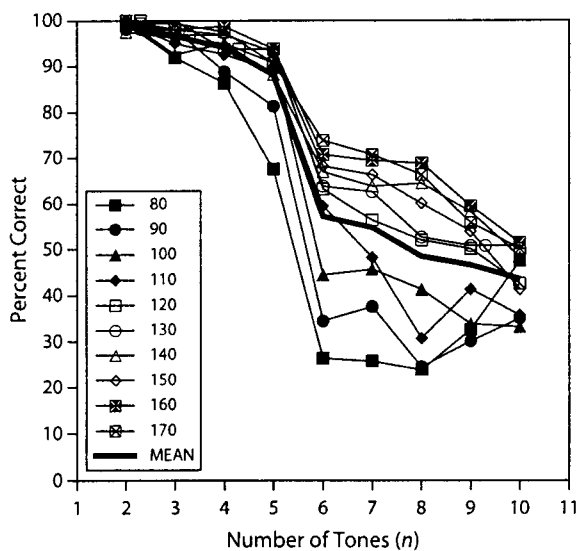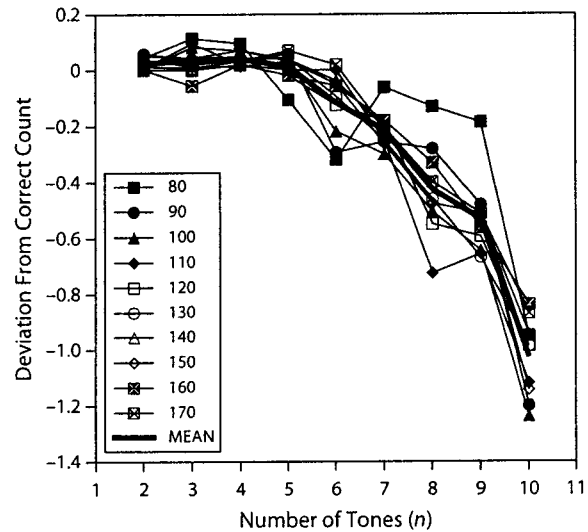

Figure 3. Results of Experiment 2: Percent correct as a function of n, with interonset interval (IOI) as the parameter.
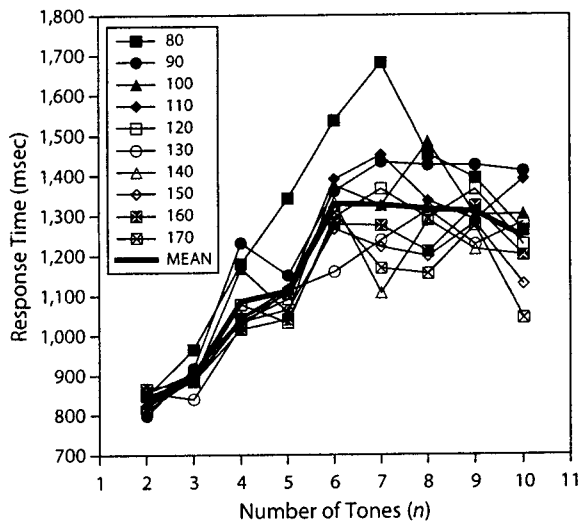
Figure 5. Results of Experiment 2: Response time as a function of *n*, with IOI as the parameter.

duration probably contributed little to enumeration accuracy. Figure 6B gives no indication that very short sequence duration accelerated responses to 2 or 3 tones, unless the flat RT function for 2 tones is considered to constitute support for that hypothesis. For *n* greater than 2, RT tended to decrease as sequence duration (and IOI) increased. The fact that this trend was no more pronounced for the largest numbers (9, 10) than for other numbers suggests that very long sequence duration did not affect RT.

Discussion

With regard to measures of enumeration accuracy, PC and DC, the present data are reasonably consistent with the previous studies by Taubman (1950a) and Garner (1951). However, the increase in PC as a function of IOI was less steep than expected, perhaps because of the completely randomized design. Even though the longest IOIs were close to those at which verbal counting of successive events is possible, enumeration performance was still far from perfect. Also, an abrupt decrease in PC between 5 and 6 tones was observed here, but not in previous studies. The results do replicate earlier findings that larger numbers of events are underestimated, and that the absolute amount of underestimation increases with *n*.

One question addressed by Experiment 2 was whether simple enumeration is subject to the same temporal limits as is the implicit tracking of different numbers of events during sensorimotor synchronization (Experiment 1). Taubman's (1950a) study in conjunction with Repp's (2003) data suggested that this might be the case, but the present data indicate that, for musicians, enumeration is more difficult than synchronization. For example, in Experiment 1, all musicians could synchronize accurately with every 8th tone in a sequence having an IOI of 170 msec (and most could do this at much shorter IOIs), whereas their enumeration of 8 tones was far from perfect at an IOI of 170 msec in Experiment 2

(see Figure 3). There are several possible explanations of this discrepancy. For example, there was uncertainty about the number of tones in Experiment 2 but not in Experiment 1, and enumeration may depend to a greater extent on subvocal counting than does synchronization. The most likely reason, however, is that automatic subdivision into smaller groups does not occur automatically when a group of tones is presented only once; in other words, auditory subitizing seems to be specific to a situation that can give rise to metrical structure by means of entrainment of internal periodicities, which requires the cyclic repetition of groups of *n* tones (defined by the taps). It is possible that deliberate grouping strategies would aid enumeration, but such strategies do not seem to have been adopted spontaneously by the present participants.
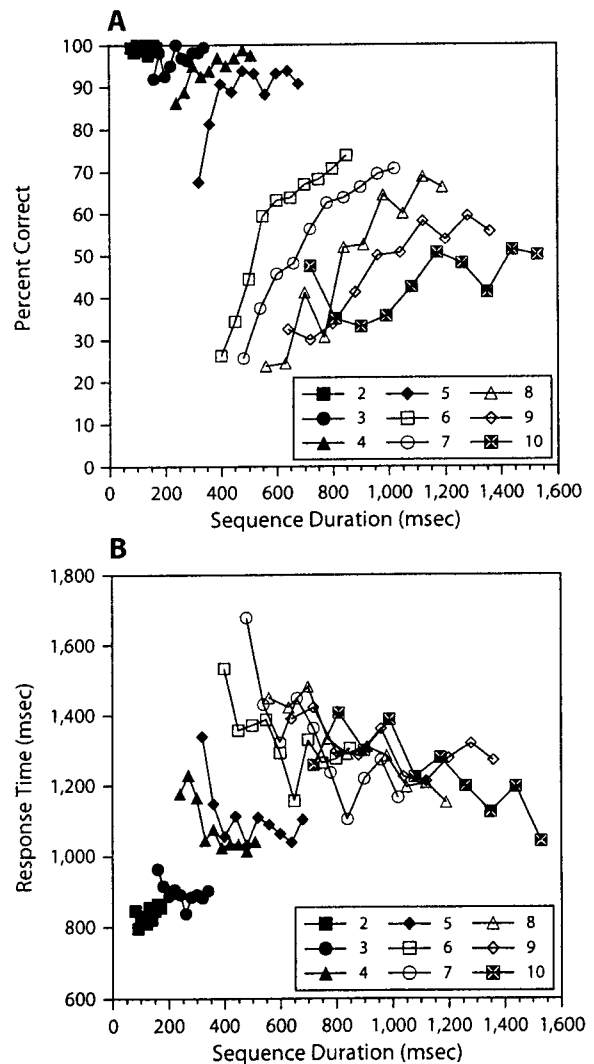

Figure 6. Results of Experiment 2: (A) Percent correct as a function of sequence duration, with *n* as the parameter; (B) Response time as a function of sequence duration, with *n* as the parameter.

The data of Experiment 2 reveal no relative disadvantage in enumerating 5 or 7 tones, rather than 6, 8, and 9 tones. This reinforces the conclusion that division into subgroups did not take place. Furthermore, the RT data do not show the telltale signs of subitizing: There was a fairly steep (albeit somewhat irregular) increase in RT between 2 and 6 tones, which seems to argue against subitizing. However, the increase in mean RT from 2 to 3 tones was rather small (62 msec), and significantly smaller than that between 3 and 4 tones.[10] This could be considered evidence that groups of 2 or 3 tones can be apprehended almost instantly. However, there was an even smaller RT increase from 4 to 5 tones (26 msec), which cannot be interpreted in that way.

There were two clear discontinuities in the data. A sudden decrease in enumeration accuracy occurred between numbers 5 and 6, suggesting that 2–5 tones were easier to enumerate than larger numbers, which probably could only be estimated. Also, RT ceased to increase beyond 6 tones, probably for the same reason. Thus, while there is little evidence for subitizing, a range of small "enumerable" numbers of events can be distinguished from a range of larger, merely "estimable" numbers.

The increase in RT within the small number set is of considerable theoretical interest, because it suggests that a serial process of some sort took place after a short sequence had ended. The slope of the linear increase was 120 msec, too fast for serial verbal (subvocal) counting. It may be that a fast "replay" of the auditory sequence from short-term memory is necessary to determine or confirm the number of elements (even though such a strategy was specifically discouraged in the instructions). Several authors have made observations that encourage such an explanation. For example, Lechelt (1975) found that "auditory trains 'resounded,' enabling subjects to recite the train immediately after its cessation" (p. 106). In an experiment that required comparing auditory and visual numerosity, Barth et al. (2003) observed that "when the tone sequence came first, [participants] often reported 'playing it back' and their response times reflected this strategy" (p. 212). Trick (2005), referring to temporal enumeration (probably with successive visual stimuli in mind, as used by Logie & Baddeley, 1987), states that "either there would have to be a detailed memorial record of the sequence of events that could be 'replayed,' or the units would have to be . . . assigned number names immediately" (p. 680).

The slope of the RT function for small numbers of events did not depend on IOI. Thus, if there was replay from auditory memory, it did not occur at the rate of presentation but at a fixed rate of about 120 msec per event. Interestingly, that rate is similar to the mean synchronization threshold for small n (Experiment 1; see also Repp, 2003, 2007). In addition, it is not much slower than the rate of the central process that, according to Cheatham and White (1954), determines the perceived number of rapidly presented auditory or visual events (see our introduction). Thus, there may be a connection between the maximal rate of serial retrieval from short-term memory and the maximal rate of entrainable internal periodicities.

For a review of various other temporal phenomena in the 8–10 Hz range, see Repp (2006).

Experiment 2 also addressed the question of whether there is a relationship between the synchronization threshold and enumeration accuracy. The data suggest a modest correlation, but this will have to be confirmed in future research with a larger sample of participants.

EXPERIMENT 3

Although little specific evidence was found in Experiment 2 that sequence duration functioned as a cue to numerosity, the possibility remains that it affected RTs to small numbers of events, which are of special interest here. Experiment 3 used an alternative design in which not sequence duration but IOI duration (i.e., rate of presentation) was confounded with n. IOI duration is more likely to act as a cue when it is short than when it is long, because at long IOIs there is little uncertainty about numerosity. This design thus has the advantage that any undesirable cueing effects of the confounding variable are more likely to occur for large numbers (presented at fast rates) than for small numbers (presented at slow rates). In that way, Experiment 3 reexamined the possibility of subitizing for small n.

Because it was clear from Experiment 2 and earlier studies that there is nothing special about enumerating 5 or 7 events, a smaller range of n was used than in Experiment 2, thereby reducing the length of the experiment. The participants were nonmusicians. The change in design prevented a direct comparison with the musicians' results in Experiment 2, but it seemed likely that musical training would matter less in an enumeration task than in a synchronization task.[11]

Method

Participants. The participants were the same 7 nonmusicians as in Experiment 1.

Materials and equipment. The number of tones in a sequence, n, ranged from 2 to 6. Each of these sequences was presented at five durations, ranging from 240 to 480 msec (measured from the onset of the first tone to the onset of the last tone) in steps of 60 msec. IOIs thus ranged from a minimum of 48 msec ($n = 6$) to a maximum of 240 msec ($n = 2$). The 25 sequences were presented 20 times in 10 randomized blocks of 50 trials each, which were preceded by an additional practice block of 50 trials. The equipment was the same as previously.

Procedure and analysis. These were the same as in Experiment 2. The experiment was conducted in a single session lasting about 45 min.

Results

Practice effects. One-way repeated measures ANOVAs on the PC scores and median RTs for the 10 blocks showed no significant differences. Two participants showed notably slower RTs in the first block; otherwise, performance was stable throughout the session and showed no evidence of improvement.

Individual differences. The grand mean PC score was 74%, with individual scores ranging from 64% to 84%. The correlation of the individual PC scores with the individual

mean synchronization thresholds in Experiment 1 was $-.40$. The grand mean RT was 868 msec, with individual mean RTs ranging from 661 to 1,299 msec. The correlation of these values with the individual mean synchronization thresholds was $-.20$. These values are far from significance, and only the one for PC scores is in the expected direction. The correlation between PC and RT was .04.

Percent correct. The percent correct scores are shown as a function of n in Figure 7. There was a rapid fall-off in PC as the number of tones increased [$F(4,24) = 47.81$, $p < .001$]. Clearly, there was little benefit for $n = 6$ from the fact that it represented the end of the range; rather, its enumeration suffered because high presentation rates (short IOIs) were associated with it, despite their potential cue value. Sequence duration, the parameter in Figure 7, had only a mild effect on PC overall [$F(4,24) = 5.53$, $p < .05$], but its effect depended strongly on the number of tones [$F(16,96) = 6.23, p < .01$]. Sequences of 2 or 3 tones showed a ceiling effect; sequences of 4 or 5 tones were poorly identified at sequence durations below 360 msec; and sequences of 6 tones were poorly identified at all sequence durations, but with relatively best performance at the shortest sequence durations. This reversal suggests that a very rapid rate of presentation did serve as a cue to the largest number of tones (6).

Deviation from correct count. These data are shown in Figure 8, and they confirm once again that underestimation was rampant, most strongly for $n = 6$. The main effect of number was significant [$F(4,24) = 37.22, p < .001$], but the main effect of sequence duration and the interaction were not.

RTs. The RT data are shown in Figure 9. RTs increased up to $n = 4$ or 5 [$F(4,24) = 7.84, p < .02$], and then decreased for $n = 6$ when the sequence duration was short. This too suggests that the rapid presentation rate at those durations served as a cue to the largest possible number of events. RT decreased as sequence duration increased
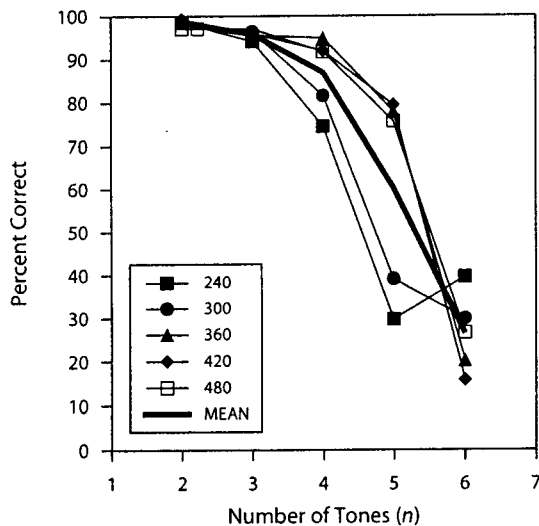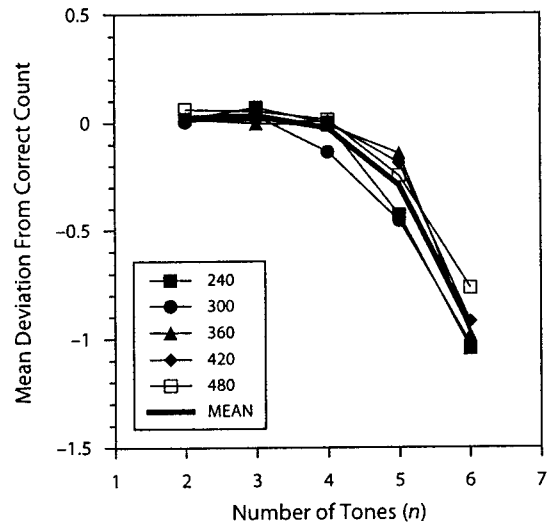
Figure 8. Results of Experiment 3: Deviation from correct count as a function of n, with IOI as the parameter.

[$F(4,24) = 12.21, p < .002$], and the interaction with n was not significant. As in Experiment 2, the increase in RT between 3 and 4 was larger than between 2 and 3 and between 4 and 5. This difference was most pronounced at the shorter sequence durations (360 msec or less), whereas the RT function for the two longest sequence durations was more nearly linear between 2 and 5. A separate ANOVA was conducted to test whether the RT difference between 2 and 3 was significantly smaller than that between 3 and 4. The difference only approached significance [$F(1,6) = 4.85, p < .07$].

IOI duration. Finally, Figure 10 shows PC and RT as a function of IOI duration (rate of presentation), which was confounded with n in this experiment. Figure 10A shows that PC fell off steeply as IOI went below 120 msec, and that PC at these short IOIs also depended strongly on n. However, the data do not reflect a strong contribution of sequence duration separate from the effect of IOI. Figure 10B presents a similar picture for RT, which seemed to depend primarily on IOI. One curious feature of these data is the substantially longer RT for 2 than for 3 tones at an IOI of 240 msec. It is unclear what caused this difference. The leveling off of the RTs at long IOIs for $n = 2$ can be explained by the fact that participants needed to wait for a time equal to the duration of the IOI before they could be sure that the sequence had ended. When the IOI was short, that waiting time accounted for only a small proportion of the RT, but when the IOI became as long as 420 or 480 msec, the waiting time accounted for up to 75% of the RT. The mean RT for $n = 2$ with an IOI of 480 msec was only 165 msec longer than the waiting time; this represents a very fast response to the absence of a third tone.

Discussion

Compared to that of Experiment 2, the design of Experiment 3 eliminated confounding effects of sequence

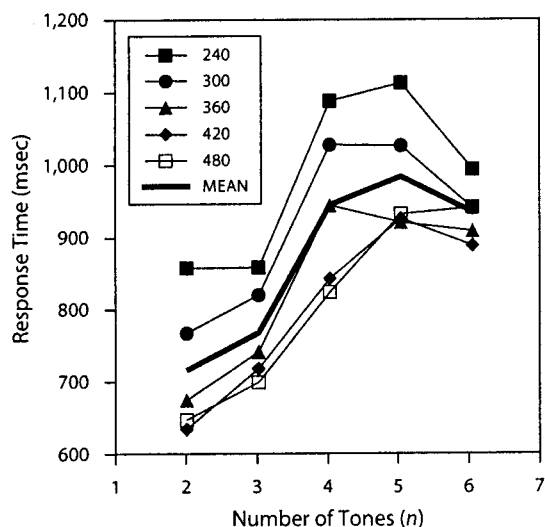Figure 7. Results of Experiment 3: Percent correct as a function of n, with IOI as the parameter.

Figure 9. Results of Experiment 3: Response time as a function of *n*, with interonset interval as the parameter.

duration and instead confounded IOI duration (rate of presentation) with *n*. Whereas sequence duration may have had some cue value for small numbers in Experiment 2 (although there was no specific evidence suggesting such an effect), the cue value of IOI duration in the present experiment clearly applied to large numbers (mainly *n* = 6), but hardly to small ones. Therefore, the data for small numbers can be considered uncontaminated. Nevertheless, they replicated the pattern found in Experiment 2: a small increase in RT from 2 to 3, followed by a larger increase from 3 to 4. This is suggestive of subitizing for 2 and 3, consistent with the role of these small groupings in metrical structures. There is no evidence, however, that longer sequences of tones, such as 4 or 6, were ever subdivided into smaller groups in the enumeration task. Subdivision seems to require a metrical context.

GENERAL DISCUSSION

The present study compared synchronized 1:*n* tapping with explicit enumeration of tone sequences presented at rapid rates. Although the tapping task seems to require the repeated implicit enumeration of *n* events, the results suggest that performance in these two tasks is governed by different processes.

In the synchronization task, the cyclic repetition of a group of *n* tones, defined initially by an explicit beat and later by the participant's taps, appears to lead automatically to subdivision of longer groups into shorter groups of 2 or 3 events. This can be seen as the consequence of the entrainment of multiple internal periodicities (with periods of IOI, 2 * IOI or 3 * IOI, and *n* * IOI), resulting in a hierarchical representation of metrical structure (Large, 2001; Large & Jones, 1999; Large & Kolen, 1994; Large & Palmer, 2002). Tracking the numerosity of *n* is both unnecessary and impossible at the fast rates employed in this study. In fact, the internal oscillator with period IOI

(the basic pulse defined by the event rate) is likely to be weak or nonexistent at these rates, and a slower oscillator with period 2 * IOI or 3 * IOI is likely to be in control. Although it is still necessary to keep track of the number of groups within a cycle of *n* events, that can be done easily because of the slower rate at which these groups occur, and because their number is again only 2 or 3. The temporal grouping of auditory events into twos or threes, which seems to occur automatically, can be regarded as a form of subitizing, analogous to the spatial grouping of two or three visual events (Mandler & Shebo, 1982).

Subdivision helps synchronization performance when *n* can be subdivided evenly, compared to when *n* can only be
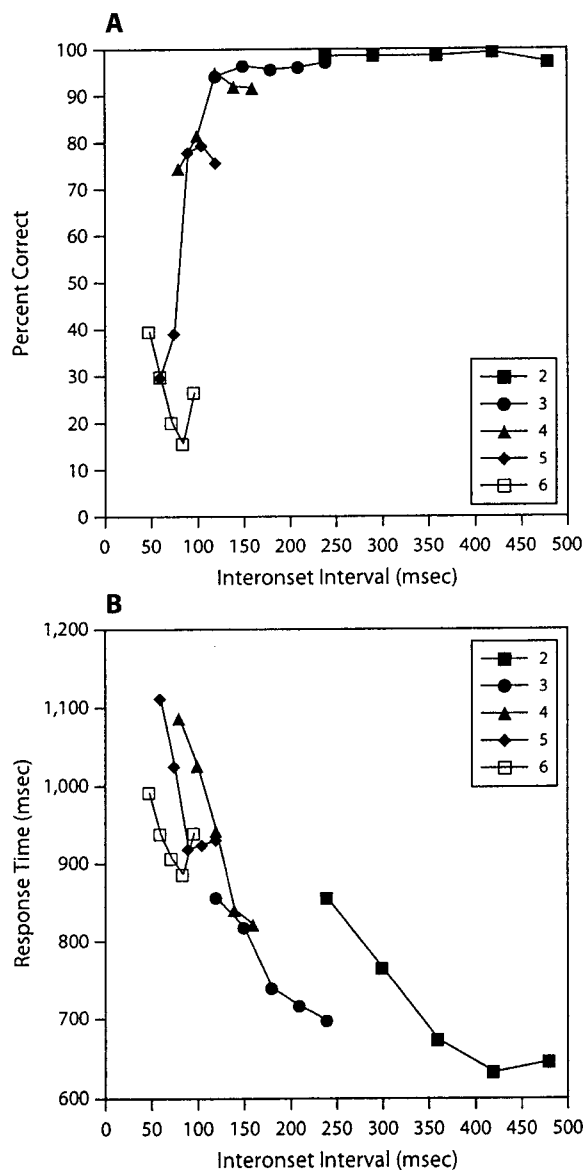

Figure 10. Results of Experiment 3: (A) Percent correct as a function of IOI duration, with *n* as the parameter; (B) Response time as a function of IOI duration, with *n* as the parameter.

subdivided unevenly. The main finding of Experiment 1 is that, for 1:n synchronization to be successful, groups of 5 or 7 tones require a slower rate of presentation than evenly divisible groups of tones. For musicians, that slower rate was just about at the temporal limit of verbal counting, though for nonmusicians it was considerably slower. It remains unclear whether participants relied on subvocal counting for $n = 5$ or 7, or whether they created unequal subgroups of events in their minds. Although unequal subdivision is difficult to conceptualize in terms of oscillators, it is certainly a humanly possible strategy, especially for musicians and particularly for those who have been exposed to music employing uneven meters (Hannon & Trehub, 2005). It can be conceptualized as alternating between two different internal timekeepers or phase-reset oscillators (McAuley & Jones, 2003). Naturally, such alternation will affect timing accuracy and variability, unless the individual has had a lot of practice with it.

Interestingly, musicians and nonmusicians showed similar patterns of results in Experiment 1, even though the latter group required slower sequence rates at which it should have been possible to count the sequence tones individually. Thus, metrical subdivision (for even meters, at least) seems to occur even when counting is in principle possible. Some nonmusicians had difficulty with grouping of events by threes, which suggests that even this simple strategy requires some practice.

The results of the explicit enumeration task (Experiments 2 and 3) present a very different picture. When a short sequence of n tones is presented only once, automatic subdivision into smaller groups does not seem to occur, as indicated by the fact that 5 or 7 tones are not more difficult to enumerate than 4, 6, or 8 tones. This was already clear from earlier studies in the literature (Garner, 1951; Taubman, 1950a). The novel contribution of the present experiments was the measurement of RT for auditory enumeration. In the literature on visual numerosity perception, a shallow increase in RT for small numbers constitutes the prime evidence for subitizing. A steeper increase for larger numbers is taken to indicate subvocal counting based on a spatial memory representation of an initial static display. By contrast, an auditory sequence unfolds in time. If the rate of presentation is slow enough for subvocal counting, no further processing is needed at the end of the sequence, and RT should reflect only the time needed to execute the response. If the rate is too fast for counting, however, other strategies come into play. The present results suggest that there may be two or three such strategies, dependent on n, as follows.

First, although the data are not very strong on this point, they do suggest that sequences of 2 or 3 tones can be apprehended immediately—that is, subitized. Although RT was somewhat longer for 3 than for 2 tones, that increase can be attributed to a difference in complexity, analogous to a similar increase between 2 and 3 events in visuospatial subitizing. Thus, subitizing of 2 or 3 events may be possible even in the absence of a metrical framework, and this ability may in fact form the basis for the automatic subdivision occurring in metrical contexts. It corresponds

to what Mandler and Shebo (1982) considered the true subitizing range in vision.

Second, RT increased more steeply between 3 and 4 tones and continued to increase up to 5 or 6 tones, at least in Experiment 2. This corresponds to the range within which simultaneously shown visual objects can be "kept in consciousness" and counted (Mandler & Shebo, 1982). It is also possible to view the present data as suggesting a linear increase in RT from 2 to 6 tones. The slope of that linear increase was less than would be required for subvocal counting, but it was similar to the subjective rate of fast sequences inferred by Cheatham and White (1952, 1954) from their research. One possibility is that rapid sequences of up to 5 or 6 tones are stored in a short-term memory and subsequently read out and matched against a pattern template, rather than counted sequentially. The increase in RT need not reflect a serial readout process; it could instead reflect the increasing complexity of the auditory template.

Third, RT ceased to increase beyond 6 tones, indicating that those larger numbers were merely estimated. Here participants presumably relied on the nonverbal system of magnitude representation that forms the ancient basis of the number concept (Gallistel & Gelman, 1992, 2000; Whalen et al., 1999). That system took into account both the event rate and the duration of the sequence. The fact that errors were almost exclusively underestimations suggests that the event rate (and perhaps duration, too) was underestimated.

For musicians, the rate limits of synchronization were lower (in terms of IOI) than those of perfect enumeration, which may reflect expertise with rhythms and synchronization tasks. For nonmusicians, the rate limits of the two tasks may be more similar, although the limit for enumeration was not determined precisely in the present study. Whereas synchronization accuracy increased steeply with IOI duration, enumeration accuracy increased only very gradually. There was a suggestion that synchronization and enumeration performance are correlated, but this requires confirmation with a larger sample of individuals.

In summary, the present research has revealed that two superficially similar tasks can give rise to very different processes of perceptual organization, with hierarchical metrical structuring being unique to the synchronization task, due to the cyclic repetition it involves. The motor response is probably immaterial, and metrical subdivision would almost certainly occur also if the large recurring groups were marked by special auditory events such as added beats or accents (Palmer & Krumhansl, 1990), or even without any markings, as in subjective rhythmization (Bolton, 1894). In those cases, an internal periodicity takes the place of the motor response in synchronization. What links the synchronization and enumeration tasks, if anything, is the possibility of subitizing groups of 2 or 3 events, which form the basic constituents of metrical structure.

AUTHOR NOTE

REFERENCES

BARTH, H., KANWISHER, N., & SPELKE, E. (2003). The construction of large number representations in adults. Cognition, 86, 201-221.

BARTLETT, N. R., & BARTLETT, S. C. (1959). Synchronization of a motor response with an anticipated sensory event. Psychological Review, 66, 203-218.

BECKWITH, M., & RESTLE, F. (1966). Process of enumeration. Psychological Review, 73, 437-444.

BERGESON, T. R., & TREHUB, S. E. (2006). Infants' perception of rhythmic patterns. Music Perception, 23, 345-360.

BOLTON, T. L. (1894). Rhythm. American Journal of Psychology, 6, 145-238.

CHEATHAM, P. G., & WHITE, C. T. (1952). Temporal numerosity: I. Perceived number as a function of flash number and rate. Journal of Experimental Psychology, 44, 447-451.

CHEATHAM, P. G., & WHITE, C. T. (1954). Temporal numerosity: III. Auditory perception of number. Journal of Experimental Psychology, 47, 425-428.

COWAN, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral & Brain Sciences, 24, 87-185.

DEHAENE, S. (1997). The number sense: How the mind creates mathematics. New York: Oxford University Press.

DRAKE, C. (1997). Motor and perceptually preferred synchronization by children and adults: Binary and ternary ratios. Polish Quarterly of Developmental Psychology, 3, 43-61.

DUNLAP, K. (1910). Reactions to rhythmic stimuli, with attempt to synchronize. Psychological Review, 17, 399-416.

FEIGENSON, L., DEHAENE, S., & SPELKE, E. (2004). Core systems of number. Trends in Cognitive Sciences, 8, 307-314.

FORSYTH, D. M., & CHAPANIS, A. (1958). Counting repeated light flashes as a function of their number, their rate of presentation, and retinal location stimulated. Journal of Experimental Psychology, 56, 385-391.

FRAISSE, P. (1950). Rythmes auditifs et rythmes visuels [Visual and auditory rhythms]. L'Année Psychologique, 49, 21-41.

FRAISSE, P., & FRAISSE, R. (1938). Études sur la mémoire immédiate I—L'appréhension des sons [Studies of immediate memory I—The apprehension of sounds]. L'Année Psychologique, 38, 48-85.

GALLISTEL, C. R., & GELMAN, R. (1992). Preverbal and verbal counting and computation. Cognition, 44, 43-74.

GALLISTEL, C. R., & GELMAN, R. (2000). Non-verbal numerical cognition: From reals to integers. Trends in Cognitive Sciences, 4, 59-65.

GARNER, W. R. (1951). The accuracy of counting repeated short tones. Journal of Experimental Psychology, 41, 310-316.

HANNON, E. E., & TREHUB, S. E. (2005). Metrical categories in infancy and adulthood. Psychological Science, 16, 48-55.

JOHN, I. D. (1972). Some variables affecting judgments of auditory temporal numerosity. Australian Journal of Psychology, 24, 347-352.

KAUFMAN, E. L., LORD, M. W., REESE, T. W., & VOLKMANN, J. (1949). The discrimination of visual number. American Journal of Psychology, 62, 498-525.

KLAHR, D. (1973). A production system for counting, subitizing, and adding. In W. G. Chase (Ed.), Visual information processing (pp. 527-546). New York: Academic Press.

LANDAUER, T. K. (1962). Rate of implicit speech. Perceptual & Motor Skills, 15, 646.

LARGE, E. W. (2000). On synchronizing movements to music. Human Movement Science, 19, 527-566.

LARGE, E. W. (2001). Periodicity, pattern formation, and metric structure. Journal of New Music Research, 30, 173-185.

LARGE, E. W., & JONES, M. R. (1999). The dynamics of attending: How we track time-varying events. Psychological Review, 106, 119-159.

LARGE, E. W., & KOLEN, J. F. (1994). Resonance and the perception of musical meter. Connection Science, 6, 177-208.

LARGE, E. W., & PALMER, C. (2002). Perceiving temporal regularity in music. Cognitive Science, 26, 1-37.

LECHELT, E. C. (1975). Temporal numerosity discrimination: Intermodal comparisons revisited. British Journal of Psychology, 66, 101-108.

LIPTON, J. S., & SPELKE, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. Psychological Science, 14, 396-401.

LOGIE, R. H., & BADDELEY, A. D. (1987). Cognitive processes in counting. Journal of Experimental Psychology: Learning, Memory, & Cognition, 13, 310-326.

LONDON, J. (2004). Hearing in time: Psychological aspects of musical meter. Oxford: Oxford University Press.

MACDOUGALL, R. (1903). The structure of simple rhythm forms. Psychological Review Monograph Supplements, 4, 309-416.

MANDLER, G., & SHEBO, B. J. (1982). Subitizing: An analysis of its component processes. Journal of Experimental Psychology: General, 111, 1-22.

MASSARO, D. W. (1976). Perceiving and counting sounds. Journal of Experimental Psychology: Human Perception & Performance, 2, 337-346.

MASSARO, D. W. (1977). Perceptual grouping in audition. Perception, 6, 541-553.

MATES, J., RADIL, T., MÜLLER, U., & PÖPPEL, E. (1994). Temporal integration in sensorimotor synchronization. Journal of Cognitive Neuroscience, 6, 332-340.

MCAULEY, J. D., & JONES, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. Journal of Experimental Psychology: Human Perception & Performance, 29, 1102-1125.

PALMER, C., & KRUMHANSL, C. L. (1990). Mental representations for musical meter. Journal of Experimental Psychology: Human Perception & Performance, 16, 728-741.

PARNCUTT, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. Music Perception, 11, 409-464.

PETERS, M. (1989). The relationship between variability of intertap intervals and interval duration. Psychological Research, 51, 38-42.

POLLACK, I. (1968). Number or duration discrimination for auditory pulse trains. Journal of Experimental Psychology, 78, 660-665.

REPP, B. H. (2003). Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. Journal of Motor Behavior, 35, 355-370.

REPP, B. H. (2005). Rate limits of on-beat and off-beat tapping with simple auditory rhythms: 2. The roles of different kinds of accent. Music Perception, 23, 167-189.

REPP, B. H. (2006). Rate limits of sensorimotor synchronization. Advances in Cognitive Psychology, 2, 163-181.

REPP, B. H. (2007). Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. Cognition, 102, 434-454.

REPP, B. H., LONDON, J., & KELLER, P. E. (2005). Production and synchronization of uneven rhythms at fast tempi. Music Perception, 23, 61-78.

REPP, B. H., & PENEL, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. Psychological Research, 68, 252-270.

ROBINSON, G. H. (1992). Single-parameter power law psychophysics of auditory numerosity and the psychological moment hypothesis. Perception & Psychophysics, 51, 363-378.

TAUBMAN, R. E. (1950a). Studies in judged number: I. The judgment of auditory number. Journal of General Psychology, 43, 167-194.

TAUBMAN, R. E. (1950b). Studies in judged number: II. The judgment of visual number. Journal of General Psychology, 43, 195-219.

TODD, N. P. MCAGUS, LEE, C. S., & O'BOYLE, D. J. (2002). A sensorimotor theory of temporal tracking and beat induction. Psychological Research, 66, 26-39.

TODOR, J. I., & KYPRIE, P. M. (1980). Hand differences in the rate and variability of rapid tapping. Journal of Motor Behavior, 12, 57-62.

TRICK, L. M. (2005). The role of working memory in spatial enumeration: Patterns of selective interference in subitizing and counting. Psychonomic Bulletin & Review, 12, 675-681.

TRUMAN, G., & HAMMOND, G. R. (1990). Temporal regularity of tapping by the left and right hands in timed and untimed finger tapping. Journal of Motor Behavior, 22, 521-535.

VAN NOORDEN, L., & MOELANTS, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, **28**, 43-66.

WARREN, H. C. (1897). Studies from the Princeton Psychological Laboratory, VI. The reaction time of counting. *Psychological Review*, **6**, 569-591.

WENDER, K. F., & ROTHKEGEL, R. (2000). Subitizing and its subprocesses. *Psychological Research*, **64**, 81-92.

WHALEN, J., GALLISTEL, C. R., & GELMAN, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, **10**, 130-137.

WHITE, C. T. (1963). Temporal numerosity and the psychological unit of duration. *Psychological Monographs: General & Applied*, **77**(Whole No. 575).

WHITE, C. T., & CHEATHAM, P. G. (1959). Temporal numerosity: IV. Comparison of the major senses. *Journal of Experimental Psychology*, **58**, 441-444.

WHITE, C. T., CHEATHAM, P. G., & ARMINGTON, J. C. (1953). Temporal numerosity: II. Evidence for central factors influencing perceived number. *Journal of Experimental Psychology*, **46**, 283-287.

WINER, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

WING, A. M., & KRISTOFFERSON, A. B. (1973). The timing of interresponse intervals. *Perception & Psychophysics*, **13**, 455-460.

NOTES

1. This participant was included in the musician group because she had a good sense of rhythm and was part of the regular crew of "master tappers" that performed weekly in the author's laboratory.

2. Two participants (both without any musical training) were not able to manage even the easiest tasks at the slowest tempo and were excused at that point. A third participant did not do the tasks in the required order and skipped tasks he had difficulty with ($n = 5, 7$). The fourth participant did not return for the second session; however, her results in the first session were similar to the mean results reported here.

3. Among musicians, there were five runs in which all trials were unsuccessful, and among nonmusicians there were four such runs. In the statistical analyses, these thresholds were entered as the final IOI of the run, although the real thresholds were presumably higher. Thus, the mean thresholds of the most difficult tasks were somewhat underestimated.

4. The p levels of all F values whose degrees of freedom in the numerator exceeded 1 were subjected to the Greenhouse–Geisser correction.

5. Because of a programming error, the last asynchrony of each trial for musicians was not saved correctly and therefore could not be included. The problem was corrected before the nonmusicians were run.

6. For the ANOVA, missing data points (for runs containing only unsuccessful trials) were replaced by reasonable estimates based on the participant's other data.

7. One amateur drummer, however, achieved thresholds of nearly 80 msec in the "8" task, the lowest individual synchronization thresholds observed in the author's research to date.

8. So far, the author has been able to test two Bulgarian students (both classically trained musicians) and one Serbian student (who had lived in Bulgaria and had only a few years of musical training), each in a single session. The Serbian participant gave results similar to those of the present nonmusician group. The two Bulgarian musicians, however, performed exceptionally well in all conditions and showed synchronization thresholds for "5" and "7" as low as those for "6," "8," and "9," but higher than those for "2," "3," and "4." Unless these results represent a ceiling effect due to these participants' exceptional rhythmic acuity, they do suggest that Bulgarians find it easier to master the "5" and "7" tasks because of their early exposure to uneven meters.

9. The author, who served as the first participant, completed only a single session containing 12 blocks and considered the first two blocks as practice. His data were omitted from analyses that included session as a variable.

10. It could be argued that the large increase in RT between 3 and 4 tones was due to the arrangement of the response keys: 2 and 3 were in the front row, whereas 4–6 were in the middle row of the 3 × 3 numeric keypad. However, if the arrangement of keys had had an effect on RT, there should have been a similar increase in RT between 6 and 7, which was not the case. Instead, there was a large increase between 5 and 6, the keys for which were in the same row. For evidence against an effect of key arrangement on RT, see also the "copy digits" data in Figure 1 of Logie and Baddeley (1987).

11. The author (who is musically trained) ran himself as a pilot participant in Experiment 3 and obtained a pattern of results very similar to the one reported below, although his RTs were faster.