Rate Limits of Sensorimotor Synchronization

Bruno H. Repp

Haskins Laboratories, New Haven, CT

ABSTRACT

Empirical evidence for upper and lower rate limits of sensorimotor synchronization (typically, finger tapping with an auditory or visual event sequence) is reviewed. If biomechanical constraints are avoided, the upper rate limit can be as high as 8–10 Hz (sequence event inter-onset intervals of 100–125 ms) with auditory stimuli, but has been found to be less than 2.5 Hz (> 400 ms) with simple visual stimuli (flashes of light). The upper rate limit for auditory stimuli varies with task difficulty and musical experience; that for visual stimuli requires further investigation. The lower rate limit, according to one definition, tends to be at about 0.56 Hz (1800 ms), regardless of modality. Attentional, perceptual, and sensorimotor explanations of these limits are considered. Rate limits of sensorimotor synchronization place important constraints on musical ensemble performance and other forms of rhythmic coordination.

INTRODUCTION

Sensorimotor synchronization (SMS) is the coordination of a rhythmic action with a rhythmic event sequence, the prime example being musical ensemble performance. In the laboratory, SMS is studied most often in the form of finger tapping with a simple auditory or visual pacing sequence, but there are also investigations of more complex rhythmic tapping or continuous limb movement as well as experiments using complex pacing sequences such as music. For recent reviews, see Aschersleben (2002) and Repp (2005c, in press–a). The present review focuses on one specific aspect of SMS — its temporal range.

SMS can be carried out successfully only within a certain range of event rates, and the limits of that range place important constraints on music performance and other forms of rhythmic coordination. I am not concerned here with the biomechanical rate limit imposed by the maximal rate of finger tapping, which typically is between 5 and 7 Hz, corresponding to intertap intervals of 150–200 ms (Keele & Hawkins, 1982; Keele, Pokorny, Corcos, & Ivry, 1985; Peters, 1980, 1985; Repp, 2005a; Todor & Kyprie, 1980; Truman & Hammond, 1990). Rather, the rate limits to be dis-

cussed here are attentional, perceptual, or perhaps sensorimotor in nature: They concern the rate of the pacing events, not that of the paced actions.

I will discuss upper rate limits first, and then lower rate limits. This terminology can be confusing because a high rate (measured in Hz or beats per second) corresponds to short event inter-onset intervals (IOIs), whereas a low rate corresponds to long IOIs. Thus the upper rate limit is the lower IOI limit, and the lower rate limit is the upper IOI limit. I will generally refer to IOI duration rather than rate in Hz.

UPPER RATE (LOWER IOI) LIMITS OF SMS

Empirical results

The upper rate limit of SMS is the fastest rate of a pacing sequence at which an action can still be coordinated with specific sequence events. For an isochronous auditory sequence, that rate typically exceeds the maximal fin-

Correspondence concerning this article should be addressed to Bruno H. Repp, Haskins Laboratories, 300 George Street, New Haven, CT 06511-6624, e-mail: repp@haskins.yale.edu

ger-tapping rate, and therefore the task cannot be 1:1 synchronization if biomechanical limits are to be avoided. Rather, participants must make fewer taps than there are sequence events, such as tapping with every other event (1:2) or with every fourth event (1:4). With visual stimuli this is not a concern, as we will see. Moreover, a more or less arbitrary criterion must be set for what constitutes successful SMS performance. The length of the pacing sequence clearly plays a role because the probability of failure increases with sequence length.

Early observations: Different rate limits for auditory and visual stimuli

Some of the earliest relevant observations were made by Dunlap (1910), who noted that IOIs of 250 ms and even 333 ms were too short for successful synchronization of finger taps with visual sequences (flashes of light). By contrast, auditory sequences (clicks) with similar IOIs did not present any difficulty. These observations pointed toward an intriguing modality difference.

Almost half a century elapsed before Bartlett and Bartlett (1959) carried out another study on this topic. Their participants were required to make a single tap, such that it coincided with any single event in an isochronous click or flash sequence, presented at various rates. In that task, failure to synchronize was reflected in a random distribution of tap times relative to the nearest event onset time (i.e., of relative phases) across a series of trials. In one experiment, where the IOIs used were 125, 250, and 500 ms, all of three participants evidently failed to synchronize at 125 ms in the auditory condition, but they failed at both 125 and 250 ms in the visual condition. In a subsequent experiment, two highly skilled participants (a professional musician and a Morse telegraphist) attempted the same tasks at IOIs of 167, 200, 250, 333, and 500 ms. None of the auditory conditions presented any difficulty to them. In the visual conditions, however, one participant performed above chance level only at 333 and 500 ms, the other one only at 500 ms. These results suggest that the rate limit for auditory stimuli was roughly between IOIs of 125 and 167 ms, whereas the rate limit for visual stimuli was somewhere between 300 and 500 ms. Bartlett and Bartlett concluded that "some sensory effect fixes the fastest rate at which events can be followed" (p. 217).

The "synchronization thresholds" (STs) for auditory and visual stimuli

Little pertinent research was conducted for almost another half century, until I decided to re-investigate the upper rate limits of SMS with auditory and visual sequences (Repp, 2003b). To determine these limits more precisely, I used isochronous sequences with ten different IOI durations, spaced closely within the relevant range for each modality. For auditory sequences, which consisted of identical high-pitched piano tones, IOIs ranged from 80 to 170 ms in steps of 10 ms. For visual sequences, generated by a flashing light, IOIs ranged from 320 to 680 ms in steps of 40 ms. Each set of sequences was presented 10 times in random order. In the auditory condition, the task was to tap with every fourth tone in the sequence (1:4 synchronization). Each sequence was preceded by six tones that indicated the spacing of the target tones (i.e., their IOIs were four times as long). In the visual condition, the task was simple 1:1 synchronization. In each condition, 55 taps were made in each trial. Failures to synchronize were generally quite obvious and consisted of a progressive phase drift of the taps relative to the target events; in other words, participants tapped at a different rate from that of the target sequence, usually without noticing it. (Mere phase slips, with subsequent stable synchronization, were rare and were not treated as errors.) The participants, with one exception, had various amounts of musical training, but only some had experience in SMS tasks.

The mean percentages of successful (i.e., no-drift) trials are shown as a function of IOI duration in Figure 1 (data points with error bars). The IOI at which synchronization was successful in 50% of the trials was estimated by linear interpolation and was considered the *synchronization threshold* (ST) in each condition. The mean auditory ST was 123 ms, with individual STs ranging from 97 to more than 170 ms (i.e., outside the range of IOIs used). The single musically untrained participant in the group had the highest ST, whereas the more experienced musicians in the group all had low STs. In stark contrast, the mean visual ST was 459 ms, with individual STs ranging from 413 to 540 ms. The visual ST did not seem to depend on musical experience, and there was no significant correlation between the auditory and visual STs.

These results are consistent with the early observations of Dunlap (1910) and Bartlett and Bartlett (1959). Moreover, there is an interesting coincidence between the auditory and visual STs and the maximal rates at which the number of auditory or visual (flashing light) events in a short sequence can be reported accurately (Taubman, 1950a, 1950b). This suggests that SMS and enumeration are subject to the same modality-dependent rate limits, perhaps because enumeration requires synchronization of an internal rhythm with the event sequence. (The auditory rate limit is too fast for verbal counting.)
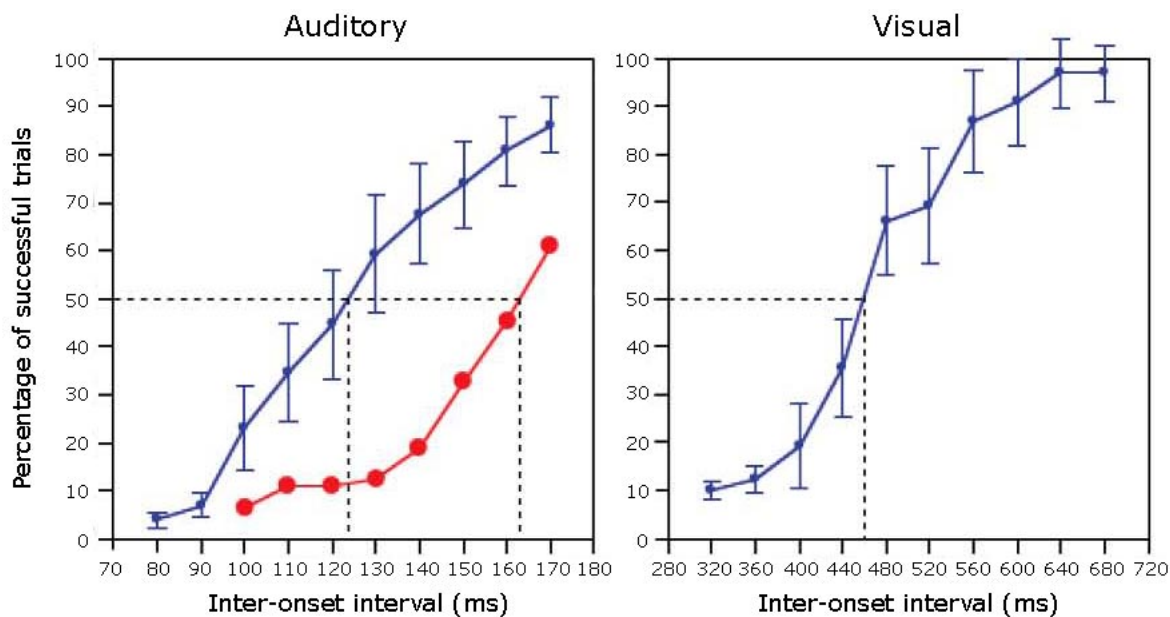
Figure 1.

Percentage of successful synchronization trials as a function of sequence inter-onset interval duration in auditory (1:4) and visual (1:1) synchronization conditions (dots with standard error bars; Repp, 2003b), and for tapping of uneven rhythms in time with an isochronous auditory sequence (circles; Repp et al., 2005). Dotted lines indicate the 50% synchronization thresholds on the x-axis.

A caveat about the visual ST

So far, there has been no further research on the visual ST. It seems likely that the visual ST will be found to be dependent on the nature of the visual sequence events. A stationary flashing light may be a particularly challenging stimulus for SMS. Although several studies (cited in Repp, 2003b) have encountered similarly low rate limits with an oscillating visual stimulus, Roerdink, Peper, and Beek (2005) recently reported a somewhat higher critical frequency of 2.59 Hz (= 386 ms IOI between inflection points) in a task requiring in-phase manual tracking of a continuously oscillating light source. Obviously, that critical IOI is still far above the auditory ST. However, other kinds of visual display may prove to be easier to synchronize with. For example, informal observations have suggested to me that a lower critical IOI duration might be obtained when the sequence elements are color changes in a stationary computer display. Displays that include various forms of spatial displacement or biological motion (e.g., a conductor's movements) may further facilitate SMS. Research on these and related issues is now under way in several laboratories. Nevertheless, it seems unlikely to me that the visual ST will ever be found to approach the auditory ST.

The auditory ST depends on rhythm complexity

Repp, London, and Keller (2005) determined the STs for "uneven" (i.e., non-isochronous) rhythms that had to be tapped in synchrony with rapid isochronous auditory sequences. There were eight rhythms, with the following nominal interval ratios: 2:3, 3:2, 2:2:3, 2:3:2, 3:2:2, 2:3:3, 3:2:3, and 3:3:2. Thus, for example, when tapping the 2:3 rhythm (which represents a musical meter of 5/8), participants tapped with the first and third tones in each cycle of five pacing tones. Otherwise, the procedure was similar to that of my earlier study (Repp, 2003b), except that for each rhythm the sequences were presented just once in order of decreasing IOI, with IOIs ranging from 170 to 100 ms in steps of 10 ms. Each isochronous sequence was preceded by seven cycles of an exact rhythm template and continued for the equivalent of 20 rhythm cycles.

The results, averaged across rhythms and musically trained participants, are included in Figure 1 (data points without error bars). It is clear that SMS performance with uneven rhythms was much poorer than with simple 1:4 tapping (Repp, 2003b). The mean ST for uneven rhythms was 163 ms, with individual STs ranging from 143 ms (a professional percussionist) to more than 170 ms (the participant with the least musical training). There were also differences among the eight rhythms, which shall not concern us here. What these results demonstrate is that an increase in rhythm complexity raises the ST, perhaps because action planning diverts attentional resources from the synchronization task or because the inherent impreci-
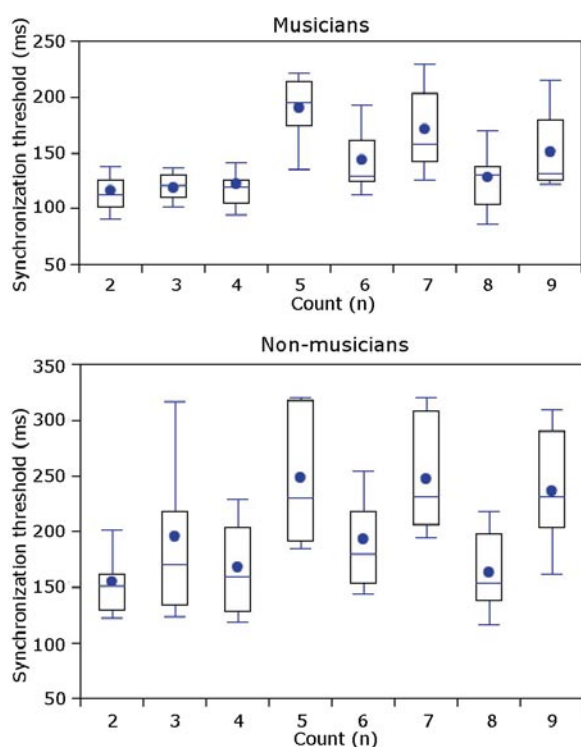
Figure 2.

Synchronization thresholds for 1:n tapping tasks in musicians and nonmusicians (Repp, 2005d). Each box shows the mean (circle), median (horizontal line), quartiles (box limits), and 10th and 90th percentiles (whiskers).

sion of uneven rhythm production (see Repp et al., 2005) causes greater instability of synchronization.

The auditory ST also depends on the tap-to-event ratio

Stimulated in part by the approximate coincidence between the mean auditory ST in Repp (2003b) and Taubman's (1950a) auditory enumeration results, I recently investigated whether the ST depends on the number of events that intervene between the target tones in an isochronous sequence – in other words, whether it depends on the tap-to-event ratio (Repp, 2005d). The experiment included eight 1:n tapping tasks, with n ranging from 2 (i.e., tapping with every other tone) to 9 (tapping with every 9th tone). The target tones in each trial were marked initially by a low accompanying tone, but that tone disappeared as soon as tapping started.

The procedure differed from the previous experiments in that a simple adaptive staircase algorithm was used to estimate the ST. Each task comprised a variable number (a "run") of trials. A successful trial consisted of 40 taps that all were within ±IOI/2 of their respective target tones. In the first trial of a run,

the sequence IOI was set at a value at which the task could be performed without difficulty. After each successful trial, the IOI was decreased by x, with x = 10 ms initially. If a tap did not meet the accuracy criterion, the sequence stopped immediately, the IOI for the next trial was increased by x, and x was decreased by 2 ms.[1] A run ended when x reached zero (i.e., after five unsuccessful trials). The final IOI was the estimate of the ST.

In this experiment (Repp, 2005d) I tested both highly trained musicians and nonmusicians who at best had had a few years of music instruction in childhood. (Some nonmusicians were not up to the task and had to be excluded.) The results are shown in Figure 2. Not surprisingly, the musicians had much lower mean STs than the nonmusicians. For musicians, the mean STs for 1:2, 1:3, 1:4, and 1:8 tapping did not differ significantly and were similar to the mean ST obtained earlier for 1:4 tapping (Repp, 2003b).[2] However, the STs for 1:6 and 1:9 tapping were higher, and those for 1:5 and 1:7 tapping were much higher. Non-musicians showed a similar pattern, except that some of them had relatively more difficulty with 1:3 and 1:9 tapping than the musicians. Given that all these tasks require keeping track of ("counting") the number of tones intervening between target tones, the results suggest that people cannot count repeatedly to 5 or 7 (prime numbers) as quickly as they can count to 6 or 9 (divisible by 3), and that they are best at counting to 4 and 8 (divisible by 2). People have no particular difficulty with counting to 5 or 7 when that number of tones is presented just once at a rapid rate (Garner, 1951; Repp, 2005d; Taubman, 1950a). Thus the difficulty with prime numbers is specific to repeated (rhythmic) counting, which generates a metrical structure. Interestingly, it is not dependent on the inability to count verbally, because in non-musicians it occurs at rates that are slow enough to permit serial verbal (subvocal) counting of the tones.[3]

A straightforward explanation of these findings is provided by the phenomenon of *subjective rhythmicization*, which has been known for quite a long time (Bolton, 1894; MacDougall, 1903). At moderately fast rates of presentation, an isochronous sequence of identical sounds tends to organize itself spontaneously into perceived groups of two or four sounds, less often of three – that is, it becomes a simple metrical structure with regular beats (usually perceived as group-initial) and subdivisions (elementary pulses). In my 1:n synchronization tasks, such spontaneous subjec-

tive grouping presumably took place within the constraints of the tapped beat, resulting in a three-level metrical hierarchy consisting of beats, subdivisions (groupings), and elementary pulses. Within-beat counts of 4 and 8 can be subdivided into groups of two or four elements (in the case of 8), and counts of 6 and 9 can be subdivided into groups of three (or perhaps two in the case of 6), though apparently with somewhat greater difficulty. Counts of 5 or 7, however, cannot be subdivided evenly, which means that these beats are either not subdivided at all or are subdivided into groups of two and three (like the uneven rhythms in the study of Repp et al., 2005); in either case, the task is more demanding and thus the ST is raised. These differences correspond to the relative prevalence of even meters (e.g., 2/4, 4/4, 3/4, 6/8) and relative rarity of uneven meters (5/8, 7/8) in Western music. Uneven meters are not uncommon, however, in the folk music of southeastern Europe, and one question worth investigating is whether extensive experience with uneven rhythms will alleviate the difficulty with 1:5 and 1:7 synchronization (cf. Hannon & Trehub, 2005).

It is also worth noting that Bolton (1894) already arrived at the conclusion that an IOI of 115 ms constitutes the upper rate limit of subjective rhythmicization, which corresponds quite well to the mean ST for 1:2, 1:3, 1:4, or 1:8 tapping in musicians. In other words, when uniform events occur at a rate too fast to be tracked individually, they also can no longer be grouped, and consequently they cannot carry a beat either. However, this does not apply to sequences of non-uniform events in which recurrent differences (e.g., intensity or pitch accents) may form a higher-order periodicity that can function as a beat, as is often the case in music.

The ST for off-beat tapping is higher than that for on-beat tapping

It is well known that tapping between the tones of an isochronous sequence (offbeat, anti-phase, or syncopated tapping) is more difficult than tapping in synchrony with the tones (e.g., Fraisse & Ehrlich, 1955; Keller & Repp, 2004; Pressing, 1998; Volman & Geuze, 2000). However, the ST for off-beat tapping was not determined precisely in these previous studies. A preliminary experiment, using a methodology similar to that in Repp (2003b), suggested that off-beat tapping becomes difficult for musically trained participants around an IOI of 350 ms (Repp, 2005a). Subsequently I used the adaptive staircase procedure described earlier to determine the off-beat ST more precisely (Repp, 2005b: Exp. 1). In order
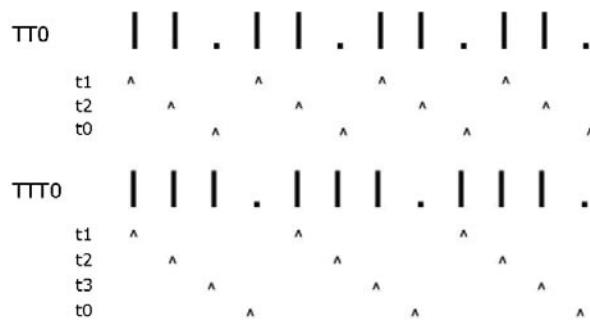

Figure 3.

Two types of rhythmic sequence (T or | = tone, 0 or . = no tone), and different on-beat (t1, t2, t3) and off-beat (t0) tapping tasks (t or ^ = tap), as used in Repp (2005b).

to compare on-beat and off-beat tapping, I had to use the same criterion for successful synchronization in both tasks. In the 1:n on-beat tapping experiment described above (Repp, 2005d), the accuracy criterion was ±IOI/2. That criterion will not do for off-beat tapping because a deviation of ±IOI/2 from the target point (the IOI midpoint) amounts to on-beat tapping. (The same deviation in on-beat tapping amounts to off-beat tapping, of course, but one tends to disregard this when only on-beat tapping is of interest.) Therefore a criterion of ±IOI/4 was used for both tasks. With that more stringent criterion, a group of musically trained participants produced mean STs of 182 and 350 ms for on-beat and off-beat tapping, respectively. The finding that the ST for off-beat tapping is about twice as high as that for on-beat tapping is consistent with the fact that off-beat taps bisect the IOIs, leading to a joint event rate of alternating tones and taps that is twice as high as that of the coincident tones and taps in on-beat tapping. Thus, the two STs may reflect essentially the same rate limit. Moreover, they were positively correlated ($r = .62$, $p < .05$).

In the same experiment (Repp, 2005b: Exp. 1), I also determined the STs for generalized forms of on-beat and off-beat tapping. Here the auditory sequences were no longer isochronous but consisted of groups of two or three tones separated by IOIs that were twice as long as the within-group IOIs. These sequences (referred to as TT0 and TTT0, respectively) are illustrated schematically in Figure 3. The on-beat tapping tasks required tapping with the first, second, or third tone (if present) in each group, whereas the off-beat tapping tasks required tapping in the middle of the between-group IOI, filling in the missing event as it were. The mean ST for off-beat tapping in the TTT0 sequence was at a between-group IOI duration of
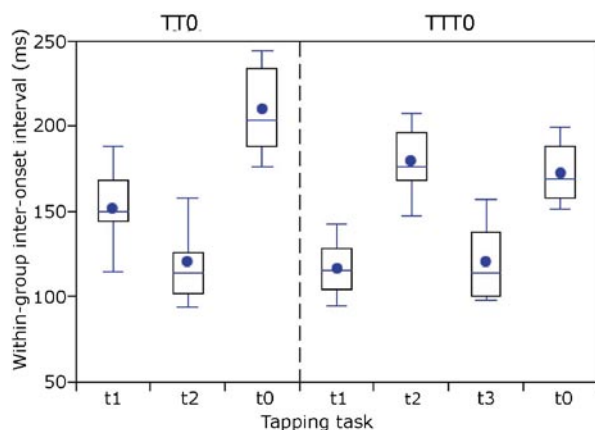
Figure 4.

Synchronization thresholds for generalized on-beat and off-beat tapping with the rhythmic sequences shown in Figure 3. Data from Repp (2005b: Exp. 1).

344 ms, similar to the ST for off-beat tapping between single tones (350 ms). However, the mean ST for off-beat tapping in the TT0 sequence was significantly higher (420 ms). This suggests again that temporal coordination is more difficult within a triple meter than within a duple or quadruple meter. With one exception, these generalized off-beat tapping tasks were more difficult than the generalized on-beat tapping tasks, which are discussed in more detail in the next section. This finding cannot easily be explained on the basis of the joint event rate of taps and tones, although it is still true that the mean interval duration was less in off-beat than in on-beat tapping. However, the relative difficulty of a tapping task may also depend on the relative salience of the tapping target (tone vs. silence) in a rhythmic pattern.

The ST is affected by rhythmic grouping accent

In classic experiments, Povel and colleagues (Povel & Okkerman, 1981; Povel & Essens, 1985) showed that some sounds occurring in temporal groups are perceptually more prominent than others, even though they are physically identical: In groups of two tones, the second tone is perceived as accented (unless the between-group IOI approaches the within-group IOI), and in longer groups the initial and final tones are perceived as accented. These *rhythmic grouping accents* are important for beat induction: A beat is felt when the accents are evenly spaced, but not (or to a lesser extent) when they occur at irregular intervals. In the rhythmic sequences of my study (Repp, 2005b: Exp. 1), the grouping accents were always evenly spaced, but the different on-beat tapping tasks forced participants to tap either on accented or

on unaccented tones (see Figure 3). Tapping on unaccented tones was predicted to be more difficult than tapping on accented tones.

This prediction was borne out by the data, as can be seen in Figure 4. The STs are expressed here in terms of the within-group IOI duration. The accuracy criterion was ±50% of the within-group IOI (= ±25% of the between-group IOI). It is evident that the STs for tapping with tones that carried a rhythmic grouping accent (t2 in the TT0 sequence, t1 and t3 in the TTT0 sequence) were low, as predicted, and similar to the ST for simple 1:n (e.g., 1:2, 1:4) on-beat tapping in isochronous sequences (cf. Figs. 1 & 2). However, the STs for tapping with unaccented tones were significantly higher, and the ST for tapping with the second tone of the TTT0 sequence (t2) was as high as that for offbeat tapping (t0), which suggests that the middle tone of the group did little to facilitate SMS, compared to tapping in the middle of a silent IOI.

A subsequent experiment (Repp, 2005b: Exp. 2) demonstrated that, not surprisingly, the ST is also affected by physical accents (increased intensity and raised pitch) on tones in a sequence: The ST was lower for tapping with a physically accented tone than with one that was not so accented, and this effect was independent of that of rhythmic grouping accent, at least in the TTT0 sequence. Surprisingly, the expected effect of grouping accent in the TT0 sequence vanished in that experiment (see also Repp, 2005a). Some individuals appear to find it easier to tap on the first than on the second tone in a group of two.

The ST is also affected by metrical accent — at least under certain conditions

Rhythmic grouping accent seems like a subjective phenomenon, but it is clearly a consequence of the temporal separation of events, which affects the relative distinctiveness and strength of the auditory representations of the events in a group (see, e.g., Todd, O'Boyle, & Lee, 1999). Indeed, the same factors that create rhythmic grouping accent may also be responsible for the rate limit in isochronous sequences, which can be seen as being caused by a reduced distinctiveness of events due to temporal crowding, as I will discuss later. By contrast, metrical accent – the feeling of a regular beat – is genuinely subjective. To be sure, a beat is usually induced by physical properties of a rhythm, such as physical accents and temporal structure, but the perceived beat can be dissociated intentionally from these stimulus properties (as long as they are not overwhelmingly powerful), and this

identifies it as an internal process that is under cognitive control.

When tapping with unaccented tones in TT0 and TTT0 sequences, a person can adopt one of two strategies: (1) to keep hearing the beat on the accented tones and tap off that beat (i.e., on an explicit subdivision of the beat), or (2) to force oneself to hear the beat on the unaccented tones and tap on that relocated beat. I did not inquire about the strategies adopted by the participants in my study, but the second strategy seemed more natural to me when I performed the tapping tasks myself.

In my next experiment, I attempted to manipulate the location of the beat in a top-down fashion (Repp, 2005b: Exp. 3). Still using the TT0 and TTT0 sequences, I instructed the participants to *hear* the beat on one of the tones in each group while either tapping on that tone or on another tone, for all possible combinations of beat location and tapping target. To facilitate the subjective manipulation of the beat, I tried to neutralize the effect of rhythmic grouping accent by increasing the relative intensity of the rhythmically unaccented tones. My prediction was that tapping on any given tone in a group should be easier (i.e., result in a lower ST) when that tone is metrically accented (coincides with the self-imposed beat) than when it is not.

That experiment was not a great success. Most participants, even though they were highly trained musicians, had considerable difficulty maintaining the subjective beat on rhythmically unaccented tones, despite the intensity boost. Sometimes participants were entirely unable to hear the beat in the required location, or – more often – they found that the beat shifted spontaneously to a different location as the sequence rate increased during a run of trials. That location was usually the one of the target tones they were tapping with. The only participant who not only had relatively little difficulty with the beat manipulation but also showed the expected effect of metrical accent on the ST was I myself. At least that convinced me that metrical accent *can* affect SMS, but it was not a compelling demonstration for a wider audience. It seems that the subjective beat is difficult to manipulate reliably in non-isochronous sequences.

Therefore, I turned next to isochronous melodies (Repp, in press–b: Exp. 1), with considerably greater success. A sequence of 12 tones varying in pitch was repeated cyclically, and the location of the beat was cued by musical notation, as shown in Figure 5. Three melodies were notated in 3/8 meter, which implies four groups of three tones each (albeit without any temporal structure) and a beat on the first element of each

Figure 5.

Notation used to prompt different metrical interpretations of the same cyclically repeated pitch sequence (Repp, in press–b: Exp. 1), and the different tapping tasks (^ = tap).

group. Each melody started on a different tone (C, D, or E) of the 12-tone pitch sequence. Thereby, the phase of the subjective beat was shifted. Participants were shown the musical notation and were instructed to hear the notated melody while tapping on the first, second, or third tone of each group (a 1:3 tapping task). Thus there were nine conditions – three metrical interpretations and three tapping targets – in three of which (C1, D1, E1) participants tapped on the beat, and in the other six of which they did not. STs were predicted to be lower in the former conditions than in the latter.

The results are shown in Figure 6. Each compartment of the figure shows three conditions in which participants tapped on the very same tones of the continuously repeated melodic sequence. Letters (C, D, E) indicate the three metrical interpretations (different phases of the beat), whereas numbers indicate the target tone locations within the groups implied by each metrical interpretation. In each case, target location 1 coincides with the self-imposed beat. STs were significantly lower in those conditions, and once again they were in the vicinity of 125 ms. The other six STs were not significantly different from each other. This experiment thus provided a convincing demonstration that metrical interpretation – a purely subjec-
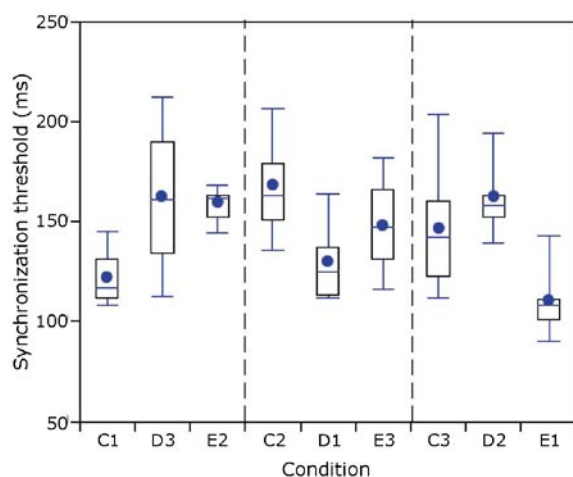
Figure 6.

Synchronization thresholds for 1:3 tapping with a continuous melody (Repp, in press–b: Exp. 1). Different metrical interpretations are indicated by letters (C, D, E), whereas tapping target locations within each metrical interpretation are indicated by numbers (1, 2, 3). Location 1 is on the beat.

tive construct – can affect SMS. Both the absence of temporal structure and the presence of pitch structure presumably facilitated the manipulation of the subjective beat, although difficulties (such as spontaneous phase shifts of the beat) were not eliminated entirely. Any such difficulties, however, worked in favor of the null hypothesis by reducing differences among metrical interpretation conditions.[4]

In a second experiment with the same pitch sequence, both the period and the phase of the beat were manipulated (Repp, in press–b: Exp. 2). The sequence was notated in either 12/8 meter, which implies four groups of three tones (like 3/8 meter), or in 3/2 meter, which implies three groups of four tones, as indicated by the beaming of eighth notes. The sequence always started on C. Participants made either three or four taps per 12-tone cycle, and made the first tap either on C or on D. Significantly lower STs were obtained in those two (out of eight) conditions in which all tapping targets coincided with the prescribed group-initial beat.

Summary

The experiments reviewed here show consistently that the mean ST for auditory 1:2, 1:3, 1:4, or 1:8 on-beat synchronization tasks, carried out by musically trained participants, is in the vicinity of 125 ms. They further demonstrate that the ST is higher in tasks that make greater demands on attention and rhythmic skill, such as tapping a non-isochronous rhythm, tapping at prime-number (>3) tap-to-tone ratios, tapping off the

main beat, or tapping on a weak element of a rhythmic group. Thus the auditory ST is highly sensitive to task demands. Clearly, it also depends strongly on participants' rhythmic ability and indeed measures one facet of that ability, namely the ability to synchronize.

Much less is currently known about the visual ST, except that it is much higher than the auditory one. The slow sequence rates required for visual SMS, at least with flashing lights, preclude analytic experiments in which rhythm or meter are varied in a meaningful way. Indeed, it seems that a sequence of flashing lights, at least, cannot carry a metrical structure (Patel et al., 2005). The visual ST also seems to be unrelated to rhythmic ability. More research on the visual ST is needed.

Theoretical explanations

What crossing the ST feels like

The phenomenology of on-beat (e.g., 1:4) tapping with an isochronous auditory sequence at or beyond the ST is that the sequence tones still sound distinct and successive, but that it is difficult to tell whether one's taps are in synchrony with the target tones. Unlike off-beat tapping, where the taps typically shift spontaneously into an on-beat mode around the ST (see, e.g., Repp, 2005a) – a shift that musically trained participants are acutely aware of and struggle to prevent – the crossing of the ST for on-beat tapping is typically smooth: Tapping just gets out of phase, often without participants' awareness. The variability of inter-tap intervals, which increases near the ST for off-beat tapping as a sign of instability (Repp, 2005a), is unaffected by the ST for on-beat tapping and in fact continues to decrease beyond the ST as the tapping rate increases (Repp, 2003b; Repp et al., 2005). The experience of crossing the visual ST is similar.

In the following sections, I will consider three possible explanations for the auditory ST: attentional, perceptual, and sensorimotor. One thing to keep in mind is that the mean on-beat ST of about 125 ms is for musically trained individuals; others are likely to have higher STs (see Figure 2). Therefore, temporal phenomena with a range of up to 200 ms or even 250 ms may be relevant to explanations of the ST.

The ST as a limit of attention

One promising hypothesis is that the ST reflects the limits of a dynamic attentional process. Large and Jones (1999) have proposed a model of temporal modulation of attentional energy that is formally quite similar to models of SMS (e.g., Mates, 1994; Repp & Keller, 2004). The main difference is that,

instead of a motor response, an attentional pulse is entrained by a periodic stimulus sequence. The attentional pulse is envisioned as a waxing and waning of attentional energy over time, with the energy peak being centered on the point of temporal expectancy induced by the sequence. The width of the attentional pulse is adjustable in the model and represents the degree of attentional focus. Such a process is subject to two temporal limitations: the minimal width of the attentional pulse, and the maximal rate at which attentional pulses can be emitted. Moreover, the minimal width constrains the maximal rate: If the minimal width is x ms, the energy peaks will merge into each other if the interval between energy peaks is less than 2x ms. Nevertheless, partially merged energy peaks may still support the tracking of events, until some minimal modulation depth of the energy flux is reached. It seems reasonable to assume that synchronization with a rapid sequence of events requires that each individual event coincide with a peak of attentional energy, even if not every event is a target for synchronization. If so, and if the ST reflects an attentional limit, then an ST of about 125 ms suggests that the maximal rate of attentional energy modulation is about 8 Hz.[5]

This estimate is not inconsistent with other estimates of attentional speed in the literature. Two forms of attentional shift have been distinguished in research on covert visual spatial attention (i.e., without eye movements): endogenous (cognitively controlled) and exogenous (prompted by events in the environment) (Jonides, 1981; Posner, 1980). A closely related distinction is that between sustained and transient attention (Nakayama & Mackeben, 1989). The attentional pulses that are being entrained by an auditory sequence (Large & Jones, 1999) clearly are exogenous (albeit anticipatory rather than reactive) and transient in nature. (It is unclear whether they can be maintained endogenously after a sequence ends.) Nakayama and Mackeben (1989) found that a burst of transient visual attention following a single exogenous cue reached its peak between 70 and 150 ms and then decayed. The temporal modulation of auditory attention could be subject to a similar limit if it depends on the serial emission of transient energy pulses. There could be a supramodal temporal process of attentional control that governs both visual-spatial and auditory-temporal attention.

Wright and Fitzgerald (2004) investigated the shape of a single auditory attentional window in an auditory signal detection task. Temporal expectations were created by varying the probability of different times of occurrence of the signal. Although the shape of the region of enhanced detection was not constant, on average it seemed to extend from about 150 ms before the signal onset to 100 ms afterwards, with peak attention anticipating the occurrence of the signal.

Another possibly related phenomenon is the so-called attentional blink. Most often demonstrated in vision, it involves a rapidly presented series of spatially superimposed stimuli (usually with IOIs of about 100 ms) containing one or two targets that are to be detected. Detection of the second target tends to be impaired if it occurs within about 500 ms of the first target (Raymond, Shapiro, & Arnell, 1992). A similar effect has also been obtained with rapid series of auditory or cross-modal stimuli (Arnell & Jolicoeur, 1998; Tremblay, Vachon, & Jones, 2005). The magnitude of the effect is rate-dependent: Arnell and Jolicoeur (1998: Exp. 4) used four IOI durations ranging from 105 to 150 ms and found an auditory attentional blink only with IOIs of 105 and 120 ms. This limit, which resembles the mean ST for simple in-phase tapping (~125 ms), may represent a failure of attentional pulses to entrain to the individual stimuli in a rapid sequence, which may be a precondition for the attentional blink.

The visual attentional blink also decreased as IOI increased but was still significant at the longest IOI (150 ms). Nevertheless, this finding can hardly be reconciled with the visual ST for flashing lights. From the literature surveyed, it seems that the temporal limits of attentional modulation are not radically different in audition and vision. Thus, IOIs of 400 ms, say, should not place any serious strain on attentional processes; yet SMS with a flashing or oscillating light is difficult to achieve at that rate. Therefore, other kinds of explanations need to be considered for the visual ST that may apply to the auditory modality as well.

The ST as a perceptual integration phenomenon

Even though there is increasing evidence that attention can influence early stages of perceptual processing (e.g., Carrasco, Ling, & Read, 2004; Carrasco, Williams, & Yeshrun, 2002; Correa, Lupíañez, & Tudela, 2005), the ST could in principle reflect limitations of perceptual processing that are independent of attention. More specifically, it may reflect a temporal integration window within which successive events cannot be treated as independent, even though their multiplicity can be perceived.

Hirsh (1974) argued that IOIs ranging from 20 to 100 ms define a range within which sequences of audi-

tory events form figures or patterns, whereas beyond that range "a series of clearly discernible separate events" is perceived (Divenyi, 2004, p. 230). Michon (1964) observed a rapid increase in the tempo discrimination threshold for auditory sequences when the rate exceeded about 10 Hz. Research on auditory sequences containing elements differing in pitch has shown that segregation into separate streams occurs inevitably when the IOIs are shorter than about 100 ms, whereas at longer IOIs the listener's intention to hear one or two streams co-determines the percept (Bregman 1990; Van Noorden, 1975). It can be inferred that, in the absence of pitch differences, the successive events are tightly integrated when the IOIs approach 100 ms.

There are possible neurophysiological correlates of these observations. Brosch and Schreiner (2000) tested sequence-sensitive neurons in the auditory cortex of cats, which showed a maximum response to the second of two tones of different frequency when the IOI was 100 ms. This enhancement was absent, however, when the two tones had the same frequency. Loveless and Hari (1993) presented sequences of two identical noise bursts to humans and found an enhancement of the neuromagnetic N100m response when the offset-to-onset interval was 100-200 ms. Although they found that the effect was not related to IOI duration, which makes it seem less relevant to the ST, they discuss their findings specifically in terms of perceptual grouping.[6] Specifically, they speculate that the observed enhancement of the N100m response reflects the turnaround time of a thalamo-cortical loop, and they refer to another study (Tiihonen et al., 1991) as suggesting that the auditory cortex has an intrinsic rhythm of 8–10 Hz in the absence of sensory input. Of course, the fact that the first major evoked neuroelectric or neuromagnetic response, the N100 or N100m, occurs about 100 ms after stimulus onset may in itself be relevant. Carver et al. (2002) found a linear decrease in the N100m amplitude as the rate of a sequence was increased, and a total disappearance of the N100m at IOIs between 98 and 162 ms. It seems quite possible that grouping of sequence elements and loss of element individuality correspond to the disappearance of that important brain response. Studies relating individual differences in the ST to temporal neurophysiological measures are called for.

Of particular interest are electrophysiological studies of auditory grouping using the so-called mismatch negativity, a brain response that is considered to reveal preattentional sensory integration processes (Kanoh et al., 2004; Shinozaki et al., 2003; Tervaniemi et al., 1994; Yabe et al., 1997, 1998, 2001a, 2001b). This research has provided evidence that two or more successive stimuli are treated as a single unit by the auditory system when their IOI is less than 150 ms or less than 200 ms, depending on the study. Furthermore, investigations of the evoked neural gamma-band (20–60 Hz) response to auditory stimuli (Snyder & Large, 2005; Tallon-Baudry & Bertrand, 1999) have indicated that the response lasts about 2.5 cycles before it dies away, which corresponds to about 100 ms. The gamma-band response is believed to reflect temporal binding of object properties that are present simultaneously (typically in vision), but conceivably gamma-band response overlap could also lead to the binding of successive stimuli to each other.

Additional behavioral findings could be cited that converge on the same temporal range as the auditory ST. SMS experiments in which an auditory target sequence is interleaved with an auditory distractor sequence suggest that target and distractor tones are perceptually integrated within a time window of 120–150 ms duration (Repp, 2003a, 2004). Similar findings have been obtained with an auditory target sequence and a visual distractor sequence (Repp & Penel, 2004). A slightly narrower time window for audiovisual integration has been demonstrated in a task that required spatial location judgments (Lewald, Ehrenstein, & Guski, 2001; Slutsky & Recanzone, 2001). Cross-modal temporalm integration serves to bind multimodal information pertaining to a single stimulus, whereas unimodal temporal integration binds successive stimuli into a group.

The phenomenon of cutaneous saltation or "rabbit" – the illusory perception of earlier stimuli, presented in one location, as moving in the direction of later-occurring stimuli, presented at another location – occurs when the IOIs between tactile stimulations are less than 200 ms (Geldard & Sherrick, 1972; Eimer, Forster, & Vibell, 2005). An analogous auditory localization phenomenon demonstrated by Hari (1995) requires IOIs of less than 150 ms. Hari noted that these IOIs "correspond well to typical sensory integration times" (pp. 29–30). Another perceptual curiosity, the "time shrinking" phenomenon, requires that the first two of three successive tones have an IOI of less than 200 ms, and that the difference between the two IOIs be less than 120 ms (Nakajima et al., 2004). The resulting perceptual shrinking of the second IOI may be yet another grouping phenomenon, and indeed Gestalt principles have been shown to play a role in

the perception of IOI duration in sequences of four tones (Sasaki et al., 2002).

There are probably many additional temporal phenomena in the 100–200 ms range that could be cited and that may or may not be related to the auditory ST. One general problem with such perceptual accounts is that the ranges of the various perceptual or neurophysiological phenomena tend to exhibit smaller individual differences than the auditory ST does, and in particular they are less likely to depend on musical training, in contrast to the auditory ST. These are important empirical questions, however, which need to be investigated in more detail: What other behavioral or neurophysiological phenomena does the auditory ST correlate with? Does musical training affect (specifically, does it narrow) the range of auditory temporal integration?

None of the perceptual findings reviewed in this section provide potential explanations for, or correlates of, the high visual ST obtained with a flashing light. Being much less familiar with the visual than with the auditory literature, I may well have overlooked some relevant research, but I suspect that the answer lies in the sensorimotor realm.

The ST as a sensorimotor limit

It is possible that the ST is neither an attentional nor a purely perceptual phenomenon but that it crucially involves action and sensorimotor integration. In other words, the rate limit represented by the ST may specific to SMS.

SMS requires phase error correction or (in dynamic systems terminology) entrainment. The auditory ST is the sequence rate at which entrainment (to a simple fraction of the sequence rate) is no longer possible, and this could mean that the perceptual information on which phase error correction is based becomes unreliable. That information is commonly assumed to be provided by the asynchronies between sequence events and taps, although an alternative view (Hary & Moore, 1985, 1987; Repp, 2005c) is that the event and tap onset times provide alternating or competing reference points for phase resetting. Accordingly, the ST may reflect the time required either to implement phase correction on the basis of perceived asynchronies or to reset the phase based on temporal reference points. Considering that the inter-tap interval at the ST, and hence the interval between perceived asynchronies, can be quite long (e.g., in 1:8 tapping) and thus should leave ample time for asynchrony-based phase correction, the phase resetting model has greater explanatory value. According to that model, each sequence event (regardless of whether or not it coincides with a tap) serves as a temporal reference, particularly the most recent event. If the event rate gets high, there is not enough time for a phase reset in response to the most recent event, which may favor phase resetting based on the taps alone. This is equivalent to saying that entrainment ceases and tapping becomes an autonomous periodic activity.

Alas, the foregoing is more a description or definition of the ST than it is an explanation. The nagging question remains why there is a temporal limit to phase resetting or entrainment, and why the limit is what it is. There is still no clear answer to these questions. The limit may be one of sensorimotor integration, perhaps the cycle time of a sensorimotor neural loop in the brain. Consider that in SMS each tap is a *timed reaction* to a preceding sequence event, timed so as to coincide with the expected arrival time of the next event. For the musically trained individuals I have been primarily investigating, the auditory ST is shorter than the shortest reaction times to unpredictable auditory stimuli (see, e.g., Jaśkowski et al., 1990). However, it may well correspond to the shortest reaction times to *predictable* auditory events (cf. Fraisse, 1966). Perhaps, then, events lose their effectiveness as temporal references when they occur too rapidly to be reacted to individually.

Viewing the ST as a sensorimotor limit offers some hope of being able to account for large individual differences, and even for the large difference between the auditory and visual STs. It is plausible that individual differences in motor skill and in musical training will be reflected in the speed of reacting to predictable auditory events. The fastest reaction times to light flashes, even though they are typically slower than those to auditory stimuli (Jaśkowski et al., 1990), are much shorter than the visual ST. However, it has been noted by Fraisse (1948) and others (e.g., Patel et al., 2005; Repp & Penel, 2004) that auditory rhythms seem to have a stronger connection to the motor system than visual rhythms do. In dynamic systems terms, that can be seen as a difference in inherent sensorimotor coupling strength. If coupling strength decreases as the sequence rate increases, a point of decoupling will be reached, and that point will be reached sooner in vision than in audition. From that perspective then, the ST is a measure of the degree to which a rhythmic stimulus sequence in a given modality engages and entrains the motor system: the stronger the entrainment, the lower the ST. The deeper reasons for the difference in sensorimotor coupling strength for auditory and

visual stimuli still remain to be discovered. It may be predicted, however, that different kinds of visual sequences will lower the ST to the extent that they engage the action system.

Summary

Thus there are three types of explanation for the upper rate limit of SMS: attentional, perceptual, and sensorimotor. They are not mutually exclusive and may make combined contributions to the rate limits. There are a variety of behavioral and neurophysiological temporal phenomena in the 100–250 ms range that may or may not reflect the same kind of limitation as the ST. Correlational and factor-analytic studies will be required to sort this out. At the moment, a sensorimotor explanation for both auditory and visual STs seems most plausible to me.

THE LOWER RATE (UPPER IOI) LIMIT OF SMS

Empirical findings

The lower rate limit of SMS (there seems to be only one) can be dealt with more briefly than the upper rate limits because the research findings and their explanation are relatively straightforward. Also, I have done much less research on the lower than on the upper rate limit and therefore have little to add to what others already have said in print. The lower rate (upper IOI) limit is less well defined than the upper rate (lower IOI) limit. If the same accuracy criterion were applied as in determining the ST (viz., that each tap must be closer to its target tone than to any other tone), there would not be any limit at all because phase drift does not occur with slow sequences. Instead, the variability of the asynchronies and inter-tap intervals increases with IOI duration, and at some point there emerges a tendency to react to (rather than anticipate) sequence tones.

Woodrow (1932) estimated what he called the "vanishing point of the capacity for synchronization" with auditory sequences by considering the IOI (about 3.4 s) at which the mean absolute asynchrony approaches the mean reaction time to single auditory stimuli. However, as long as some asynchronies are negative (i.e., anticipations), the comparison with reaction times seems misguided to me. MacDorman (1962) similarly ignored the sign of asynchronies when he examined their variability as a function of IOI. Both researchers concluded, however, that synchronization first becomes difficult around IOIs of 1.8 s. This interval corresponds to what has long been regarded as the upper limit for rhythm perception beyond which

successive tones are perceived as individual, unrelated events (e.g., Bolton, 1894; Fraisse, 1982; MacDougall, 1903). Fraisse (1966) also noted a marked increase in the variability of asynchronies between IOIs of 1500 and 3000 ms.

Three more recent studies have examined the lower rate limit in some detail. Mates et al. (1994) investigated the distribution of asynchronies across a wide range of widely spaced IOIs (300 to 4800 ms) and found that up to 1800 ms the distribution was unimodal, approximately normal, and had a negative mean, reflecting the well-known anticipation tendency in 1:1 SMS. At IOIs of 2400 and 3600 ms, the distribution became bimodal, due to the increased occurrence of positive asynchronies with a mean of about 150 ms, which evidently represented reactions to the sequence tones. At the longest IOI of 4800 ms, reactions predominated, although there were large individual differences in their proportion.

Miyake, Onishi, and Pöppel (2004) replicated these results and showed furthermore that the proportion of reactions at long IOIs increased when participants' attention was diverted by a word memory task. In other words, attentional resources were required to anticipate tones at long intervals. At the longest IOI, 6000 ms, three of six participants showed almost only reactions, but the other three still had about 60% anticipations. Depending on the individual participant, reactions began to emerge as early as 1200 ms or as late as 3600 ms.

Taking a dynamic systems approach, Engström, Kelso, and Holroyd (1996) investigated contact-free finger flexion (rather than tapping) in synchrony with a visual metronome at a number of rates (specified in Hz). They also included a syncopation condition in which participants flexed their finger between light flashes, and a reaction condition in which participants were told to react to the flashes. Their SMS results were quite similar to those of Mates et al. (1994) and Miyake et al. (2004), although reactions began to emerge somewhat sooner, at a rate of 0.75 Hz (1333 ms). Reactions predominated at the slowest rates of 0.25 Hz (4000 ms) and 0.125 Hz (8000 ms). These results suggest that the lower rate limit is similar for auditory and visual sequences. In the condition where participants were told to react, there were of course only reactions at the slow rates, but interestingly anticipations began to emerge at about 0.75 Hz (1333 ms) and constituted an appreciable minority of responses at the fastest rate used, 1.375 Hz (727 ms).[7] Syncopation did not show any bimodal distribution of "asynchronies" relative to the IOI midpoints, the obvious reason being that there were no events to react to at the IOI midpoints.

Together with an undergraduate student, Rebecca Doggett, I recently conducted a small study on the lower rate limit of SMS (Repp & Doggett, 2005). The main purpose of the experiment was to determine whether the lower rate limit of musically trained participants is similar to that of the participants in previous studies, who were not described as musically trained (and were in fact not musically trained in the study of Miyake et al., 2004, as confirmed by Miyake in personal communication). A second aim was to test whether SMS at slow rates is facilitated when sequence tones are physically connected (*legato*) and form an ascending and descending musical scale, which we thought might increase their perceptual coherence. We further included a syncopation condition, to replicate the findings of Engström et al. (1996) with auditory stimuli and with tapping on a surface. Our results were similar to those of the previous studies, and the musical scale did not facilitate SMS. In particular, a comparison with the results of Miyake et al. (2004) for a shared range of IOIs showed little difference. Figure 7 shows the percentage of reactive responses in the two studies as a function of IOI duration. As far as these data go (and our range of IOIs was much narrower than the ranges in previous studies), they suggest that musical training does not affect the lower rate limit of SMS. In the study by Engström et al. (1996), reactive responses were more common and emerged earlier, but this could reflect differences in method (such as contact-free tapping).

Theoretical explanations

What the lower rate limit feels like

There is a distinct phenomenology associated with the lower rate limit of SMS. At IOIs up to 1500 ms or so, SMS seems to proceed effortlessly and automatically, but the task begins to feel laborious as the IOI approaches 1800 ms. Beyond that duration, the task becomes essentially one of interval estimation: Each tap must be placed consciously at the remembered duration of the previous IOI, and it seems that any error must be compensated for deliberately. By contrast, error correction occurs automatically during SMS at shorter IOIs, except in the case of very large errors.

Related perceptual phenomena

Bolton (1894) observed that, on average, subjective rhythmicization (grouping) of isochronous sequences did not occur beyond IOIs of about 1600 ms, although there were considerable individual differences. MacDougall (1903) located the limit between 1500 and 2000 ms. Fraisse (1982), in reviewing the
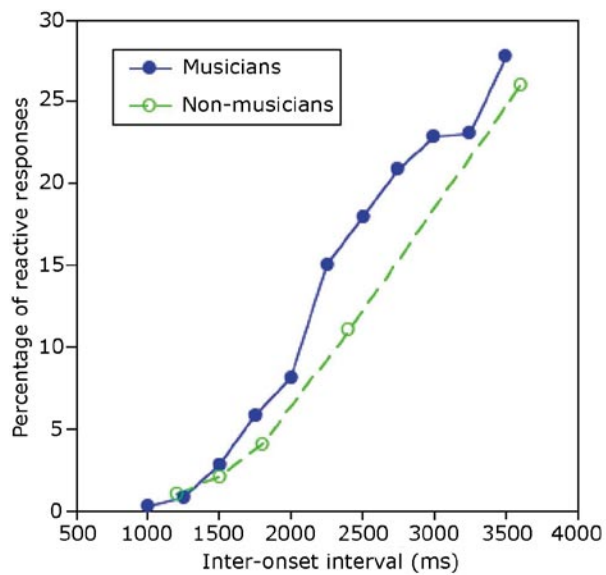
Figure 7.

Percentage of reactive responses as a function of sequence inter-onset interval duration for musicians (Repp & Doggett, 2005) and non-musicians (estimated from Figure 3, condition "N", of Miyake et al., 2004).

evidence, opted for 1800 ms, which is in agreement with the SMS results reviewed above. Szelag et al. (1996) asked participants to deliberately create a subjective rhythm by mentally accenting equidistant events in an isochronous auditory sequence whose IOIs were varied from 200 to 1000 ms. The maximal interval between subjective accents (the length of a subjective group) depended on the IOI and ranged from 1 to 3 s. At a moderate tempo (IOI = 500 ms), it was close to 2 s (i.e., a group of four events). Many other authors have pointed out that events separated by more than about 2 s are perceived as unconnected, unrelated, and individual. By contrast, at shorter intervals the events form "units which form an almost static synthesis of change" (Fraisse, 1964, p. 97).

Warren et al. (1991) asked listeners to identify eight familiar melodies that were instantiated as isochronous sequences and were repeated cyclically with constant IOIs ranging from 40 to 3600 ms. At very slow rates, participants were unable to identify the melodies. The limit was in the vicinity of 2 s for four of the melodies, but closer to 1 s for the other four. The differences may have to do with the pitch structure and relative familiarity of the melodies, but the maximal values are in the same range as the lower rate limit of SMS.

The subjective present

The lower rate limit is unlikely to constitute a sensorimotor limit in view of the slow tempo, which leaves ample time for error correction. Likewise, a specifically

attentional limit in this time range seems implausible, as attention can be maintained over much longer time spans than 2 s. The most plausible explanation is a perceptual and/or memory-based one. Fraisse (1984) distinguished between perception and estimation of duration, the former occurring within 2–3 s and being of "a quantity whose beginning has not yet been stored in memory" (p. 10). Pöppel (1997) has long argued for a low-frequency neural mechanism that "binds successive events of up to 3 s into perceptual units" (p. 58; see also Pöppel, 2004; Wittmann & Pöppel, 2000). This time period has also been termed the "subjective present", a notion that goes back to the beginnings of scientific psychology (James, 1890; Michon, 1978).

Miyake et al. (2004) specifically link the lower rate limit of SMS to the construct of working memory, particularly the "phonological loop" hypothesized by Baddeley (1986) to account for rehearsal of verbal materials. Using interference paradigms and correlational analyses, Saito (1994, 2001; Saito & Ishio, 1998) and Grube (1996, 1998) have provided evidence that the phonological loop, whose temporal span is estimated to be not more than 2 s, is involved not only in verbal memory tasks but also in the reproduction of auditory rhythms. This makes it seem likely that the lower rate limit of SMS reflects a temporal limit of auditory working memory. For a tap to anticipate the occurrence of the next tone in a sequence, the previous tone must still be "present", as it were. If it has already entered a more permanent memory, a process of deliberate time estimation takes over.

The concept of the subjective present can account for a similar lower rate limit with visual stimuli because it is essentially amodal. Explanations that rely specifically on limits of auditory working memory may apply if it is the case that rhythmic visual stimuli are automatically translated into an internal auditory code (Guttman, Gilroy, & Blake, 2005).

RELEVANCE OF THE RATE LIMITS OF SMS TO MUSIC PERFORMANCE

Both rate limits of SMS place important constraints on music performance. However, these constraints may not be very obvious to musicians because music is generally designed to stay within the rate limits. That is, usually the event rates in ensemble performance are such that the musicians are able to coordinate their actions with each other. Musicians also can rely on much additional information that was not present in the simple laboratory experiments reviewed here. Such infor-

mation includes higher-level periodic structure induced by pitch or intensity accents or by articulation (such as bowing), which makes it possible to synchronize even when the event rate is high. Moreover, visual cues are important in ensemble performance, and larger ensembles are often coordinated by a conductor's gestures. Visual cues are especially important when the event rate is very slow, or after long rests, where the continuity of the music is broken. Such local disruptions of continuity (> 2 s) occur frequently in music, and they amount to starting a new "trial" of SMS. Importantly, visual cues in playing music always involve continuous movement, not discrete flashes as in the visual SMS experiments reviewed above. The accuracy of SMS with visible movement, such as a conductor's gestures, is an important topic for further investigation. Interesting preliminary work along these lines has been done by Luck (2002; Luck & Toiviainen, 2005).

It must also be kept in mind that the upper rate limit of SMS may be reduced considerably for complex rhythms or other difficult forms of coordination, such as polyrhythms. Many musical compositions, particularly those from the 20th century, incorporate complex metrical frameworks that make temporal coordination within an ensemble difficult. The upper rate limit of SMS may often be approached in such works, and special performance expertise may be required to overcome it.

In connection with the upper rate limit of SMS for tapping with a simple metronome, I have noted that tones separated by less than about 100 ms are difficult to perceive and track as individual events; they may form a single continuous pattern instead (cf. Hirsh, 1974). This seems relevant to music structure and perception at a local level: Tones that are separated by less than 100 ms no longer function as rhythmic elements but become attached to adjacent tones as ornaments, or they form arpeggi and glissandi. Fraisse (1964) observed that the fastest melody notes in music tend to occur about 150 ms apart, and this was also approximately the rate at which the participants in the study of Warren et al. (1991) were no longer able to identify cyclically repeated familiar melodies when they were sped up, rather than slowed down. Friberg and Sundström (2002) measured the duration of the short second interval in jazz drummer's performance of the "swing rhythm" at a wide range of tempi and found that it bottomed out at 90–100 ms. However, a biomechanical limit could have played a role here, given that the rhythms were played unimanually.

On a keyboard, there is no biomechanical limit for successive key presses with different fingers or hands. Timmers et al. (2002) measured the duration of grace

notes in excerpts from piano performances and found durations of roughly 50–100 ms in two contexts, and of 100–150 ms in a third context where the grace note arguably had more of a melodic and rhythmic function. (See also Windsor et al., 2001.) Some years ago, I measured the timing of arpeggi in performances of a very slow Romantic piano piece and observed IOIs both below and above 100 ms, depending on the musical context and the individual performer (Repp, 1997). The slower arpeggi probably sounded more like a fast melody than like a broken chord. Asynchronies between tones that are intended to be played simultaneously are almost always well below 100 ms (Rasch, 1979; Repp, 1996).

These scattered observations suggest that the rate limits of SMS are not only directly relevant to the range of event densities employed in music but also are indirectly related to various local timing properties of performed music.

Acknowledgments

Notes

[1] The criterion that all taps had to meet the accuracy criterion may seem severe, but it yielded a mean ST for 1:4 tapping very similar to that obtained previously (Repp, 2003b). Typically, a deviant tap represents incipient phase drift. Of course, when the IOI gets very short and the proportional criterion gets very tight, deviant taps may occur simply as the consequence of motor variability, so there is an inherent lower limit to the ST. However, for musically trained individuals, whose mean asynchronies are near zero and whose standard deviations of asynchronies are generally less than 20 ms, that lower limit is probably lower than the observed ST (e.g., at IOI = 120 ms a deviant tap is still more than 3 standard deviations away from the mean).

[2] The lowest individual ST I have observed to date was achieved by an amateur percussionist in 1:8 tapping (83 ms, average of two separate estimates).

[3] Clearly, the duration of the inter-tap intervals (ITIs), which increases with n, has little to do with the variation in the ST as a function of n. The increase in variability that is generally observed as ITI increases in 1:1 SMS or in self-paced tapping (see, e.g., Madison, 2001) is effectively neutralized when the ITIs are physically subdivided by sequence tones (cf. Repp, 2003b). Also, although non-musicians usually tap ahead of the sequence tones in 1:1 SMS (see, e.g., Aschersleben, 2002), their mean asynchronies were close to zero in 1:n tapping, as were those of the musicians. This shows that asynchronies and their variability are controlled by the rate of pacing events, not by the rate of produced actions.

[4] Interestingly, the advantage for tapping on a self-imposed beat disappeared in a variant of the task in which, instead of tapping in synchrony with computer-controlled target tones in the melodic sequence, participants controlled those tones with their taps, thus filling gaps in the computer-controlled sequence (Repp, in press–b: Exp. 1). In that case, when participants tapped on the beat, the tones carrying the beat were self-controlled and hence variable in their timing; therefore, they either served as unstable temporal references for synchronization and/or participants used the stable off-beat tones as references. In off--beat tapping, however, the tones carrying the beat remained stable and could still serve as references. All tasks had similar mean STs, comparable to those of the off-beat tapping tasks in the original version of the experiment.

[5] Alternatively, it might be hypothesized that only the target events need to coincide with an energy peak. In that case, the ST would be reached when events adjacent to the target tone fall within the attentional focus. That alternative hypothesis can be rejected, however, on the basis of the finding, reviewed above, that 1:2, 1:3, 1:4, and 1:8 tapping have similar STs (Repp, 2005d), which would imply that the width of the attentional pulse is independent of pulse rate. This seems unrealistic because it is contrary to the increase in temporal uncertainty with interval duration (Weber's law). Even if people could emit equally narrow attentional pulses in 1:2 and 1:8 tapping, it would remain unclear how people can keep track of the number of events between targets or subdivide the beat. Therefore, it seems more plausible to assume that each event must receive an attentional pulse.

[6] In that connection, I should mention that I used tones without distinct offsets in all my auditory ST experiments. They were high-pitched synthetic piano tones whose amplitude decayed in a roughly exponential fashion within about 100 ms. For participants with very low STs (approaching 100 ms), the sequence tones may have become perceptually less distinct for psy-

choacoustic reasons (but see Footnote 2). It is presently unknown whether the ST is sensitive to the offset characteristics or duration of sequence tones.

[7] Klemmer (1967) showed long ago that intended reactions become anticipations at even faster sequence rates, sooner with lights than with tones.

References

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition, 48*, 66–79.

Arnell, K. M., & Jolicoeur, P. (1998). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 630–648.

Baddeley, A. (1986). *Working memory*. New York: Oxford University Press.

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

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology, 6*, 145–238.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Brosch, M., & Schreiner, C. E. (2000). Sequence sensitivity of neurons in cat primary auditory cortex. *Cerebral Cortex, 10*, 1155-1167.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7*, 308–313.

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision, 2*, 467–479.

Carver, F. W., Fuchs, A., Jantzen, K. J., & Kelso, J. A. S. (2002). Spatiotemporal analysis of the neuromagnetic response to rhythmic auditory stimulation: rate dependence and transient to steady-state transition. *Clinical Neurophysiology, 113*, 1921-1931.

Correa, A., Lupiáñez, J., & Tudela, P. (2005). Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychonomic Bulletin & Review, 12*, 328–334.

Divenyi, P. L. (2004). The times of Ira Hirsh: Multiple ranges of auditory temporal perception. *Seminars in Hearing, 25*, 229–239.

Dunlap, K. (1910). Reactions to rhythmic stimuli, with attempt to synchronize. *Psychological Review, 17*, 399–416.

Eimer, M., Forster, B., & Vibell, J. (2005). Cutaneous saltation within and across arms: A new measure of the saltation illusion in somatosensation. *Perception & Psychophysics, 67*, 458–468.

Engström, D. A., Kelso, J. A. S., & Holroyd, T. (1996). Reaction–anticipation transitions in human perception–action patterns. *Human Movement Science, 15*, 809–832.

Fraisse, P. (1948). Rythmes auditifs et rythmes visuels. [Visual and auditory rhythms.] *L'Année Psychologique, 49*, 21–41.

Fraisse, P. (1964). *The psychology of time* (transl. J. Leith). London: Eyre & Spottiswoode. [French original published in 1963.]

Fraisse, P. (1966). L'anticipation de stimulus rythmiques: Vitesse d'établissement et précision de la synchronisation. [Anticipation of rhythmic stimuli: Speed of establishment and precision of synchronization.] *L'Année Psychologique, 66*, 15–36.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *Psychology of music* (pp. 149–180). Orlando, FL: Academic.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology, 35*, 1-36.

Fraisse, P., & Ehrlich, S. (1955). Note sur la possibilité de syncoper en fonction du tempo d'une cadence [Note on the possibility of syncopation as a function of sequence tempo]. *L'Année Psychologique, 55*, 61–65.

Friberg, A., & Sundström, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythm pattern. *Music Perception, 19*, 333–349.

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

Geldard, F., & Sherrick, C. (1972). The cutaneous "rabbit": a perceptual illusion. *Science, 178*, 178–179.

Grube, D. (1996). Verarbeitung akustisch dargebotener Zeitintervalle im Sekundenbereich: Eine Leistung der phonologischen Schleife des Arbeitsgedächtnisses? [Processing of acoustically presented temporal intervals in the second range: An achievement of the phonological loop of working memory?] *Zeitschrift für Experimentelle Psychologie, 43*, 527–546.

Grube, D. (1998). Die Kapazität des phonetischen Speichers des Arbeitsgedächtnisses als 'auditive Präsenzzeit' und ihr Einfluß auf die Reproduktion von Zeitmustern. [The capacity of the phonetic store of working memory as the 'auditory present' and its influence on the reproduction of temporal patterns.] In U. Kotkamp & W. Krause (Eds.), *Intelligente Informationsverarbeitung* (pp. 223–231). Deutscher Universitätsverlag.

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science, 16*, 228–235.

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

Hari, R. (1995). Illusory directional hearing in humans. *Neuroscience Letters, 189*, 29–30.

Hary, D., & Moore, G. P. (1985). Temporal tracking and synchronization strategies. *Human Neurobiology, 4*, 73–77.

Hary, D., & Moore, G. P. (1987). Synchronizing human movement with an external clock source. *Biological Cybernetics, 56*, 305–311.

Hirsh, I. J. (1974). Temporal order and auditory perception. In H. R. Moskowitz, B. Scharf, & J. C. Stevens (Eds.), *Sensation and measurement* (pp. 251–258). Dordrecht, The Netherlands: Reidel.

James, W. (1890). *The principles of psychology*. New York: Henry Holt.

Jaśkowski, P. Jaroszyk, F., & Hojan-Jezierska, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research, 52*, 35–38.

Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.

Kanoh, S., Futami, R., & Hoshimiya, N. (2004). Sequential grouping of tone sequence as reflected by the mismatch negativity. *Biological Cybernetics, 91*, 388–395.

Keele, S. W., & Hawkins, H. L. (1982). Explorations of individual differences relevant to high level skill. *Journal of Motor Behavior, 14*, 3–23.

Keele, S. W., Pokorny, R. A., Corcos, D. M., & Ivry, R. (1985). Do perception and motor production share common timing mechanisms: a correlational analysis. *Acta Psychologica, 60*, 173–191.

Keller, P. E., & Repp, B. H. (2004). When two limbs are weaker than one: Sensorimotor syncopation with alternating hands. *Quarterly Journal of Experimental Psychology, 57A*, 1085–1101.

Klemmer, E. T. (1967). Sequences of responses to signals encoded in time only. *Acta Psychologica, 27*, 197–203.

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

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioural Brain Research, 121*, 69–79.

Loveless, N. E., & Hari, R. (1993). Auditory evoked fields covary with perceptual grouping. *Biological Psychology, 35*, 1–15.

Luck, G. (2002). Conductors' gestures: Perception of, and synchronization with, visual beats. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the Seventh International Conference on Music Perception and Cognition* (p. 638). Adelaide, Australia: Causal Productions (CD-ROM).

Luck, G., & Toiviainen, P. (2005). *Expert ensemble musicians' synchronization with expert conductors' gestures*. Paper presented at the Tenth Rhythm Perception and Performance Workshop, Bilzen, Belgium, 2–6 July.

MacDorman, C. E. (1962). Synchronization with auditory models of varying complexity. *Perceptual and Motor Skills, 15*, 595–602.

MacDougall, R. (1903). The structure of simple rhythm forms. *Psychological Review Monograph Supplements, 4*, 309–416.

Madison, G. (2001). Variability in isochronous tapping: Higher order dependencies as a function of intertap interval. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 411–422.

Mates, J. (1994). A model of synchronization of motor acts to a stimulus sequence. I. Timing and error corrections. *Biological Cybernetics, 70*, 463–473.

Mates, J., Radil, T., Müller, U., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience, 6*, 332–340.

Michon, J. A. (1964). Studies on subjective duration: I. Differential sensitivity in the perception of repeated temporal intervals. *Acta Psychologica, 22*, 441–450.

Michon, J. A, (1978). The making of the present: a tutorial review. In J. Requin (Ed.), *Attention and performance VII* (pp. 89–111). Hillsdale, NJ: Erlbaum.

Miyake, Y., Onishi, Y., & Pöppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta Neurobiologiae Experimentalis, 64*, 415–426.

Nakajima, Y., ten Hoopen, G., Sasaki, T., Yamamoto, K., Kadota, M., Simons, M., & Suetomi, D. (2004). Time-shrinking: the process of unilateral temporal assimilation. *Perception, 33*, 1061–1079.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29*, 1631–1647.

Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research, 163*, 226–238.

Peters, M. (1980). Why the preferred hand taps more quickly than the non-preferred hand: Three experiments on handedness. *Canadian Journal of Psychology, 34*, 62–71.

Peters, M. (1985). Constraints in the performance of bimanual tasks and their expression in unskilled and skilled subjects. *The Quarterly Journal of Experimental Psychology, 37A*, 171–196.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences, 1*, 56–61.

Pöppel, E. (2004). Lost in time: a historical frame, elementary processing units and the 3-second window. *Acta Neurobiologiae Experimentalis, 64*, 295–301.

Posner, M. I. (1980). Orienting of attention. Quarterly *Journal of Experimental Psychology, 32*, 3–25.

Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*, 411–440.

Povel, D.-J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics, 30*, 565–572.

Pressing, J. (1998). Error correction processes in temporal pattern production. *Journal of Mathematical Psychology, 42*, 63–101.

Rasch, R. (1979). Synchronization in performed ensemble music. *Acustica, 40*, 121–131.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 849–860.

Repp, B. H. (1996c). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America, 100*, 3917-3932.

Repp, B. H. (1997c). Some observations on pianists' timing of arpeggiated chords. *Psychology of Music, 25*, 133-148.

Repp, B. H. (2003a). Phase attraction in sensorimotor synchronization with auditory sequences: Effects of single and periodic distractors on synchronization accuracy. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 290–309.

Repp, B. H. (2003b). 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. (2004). On the nature of phase attraction in sensorimotor synchronization with interleaved auditory sequences. *Human Movement Science, 23*, 389–413.

Repp, B. H. (2005a). Rate limits of on-beat and off-beat tapping with simple auditory rhythms: 1. Qualitative observations. *Music Perception, 22*, 479–496.

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

Repp, B. H. (2005c). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*, 969–992.

Repp, B. H. (2005d). *Counting auditory events at rapid rates: Emergence of metrical structure from automatic subdivision of repeated counts*. Manuscript submitted for publication.

Repp, B. H. (in press–a). Musical synchronization. In E. Altenmüller, J. Kesselring, & M. Wiesendanger (Eds.), *Music, motor control, and the brain*. Oxford, UK: Oxford University Press.

Repp, B. H. (in press–b). Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. *Cognition*.

Repp, B. H., & Doggett, R. (2005). *Rate limits of on-beat and off-beat tapping with simple auditory rhythms:3. The lower rate limit in musicians*. Manuscript submitted for publication.

Repp, B. H., & Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *Quarterly Journal of Experimental Psychology, 57A*, 499–521

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.

Roerdink, M., Peper, C. E., & Beek, P. J. (2005). Effects of correct and transformed visual feedback on rhythmic visuo-motor tracking: Tracking performance and visual search behavior. *Human Movement Science, 24*, 379–402.

Saito, S. (1994). What effect can rhythmic finger tapping have on the phonological similarity effect? *Memory & Cognition, 22*, 181–187.

Saito, S. (2001). The phonological loop and memory for rhythms: an individual difference approach. *Memory, 9*, 313–322.

Saito, S., & Ishio, A. (1998). Rhythmic information in working memory: effects of concurrent articulation on reproduction of rhythms. *Japanese Psychological Research, 40*, 10–18.

Sasaki, T., Suetomi, D., Nakajima, Y., & ten Hoopen, G. (2002). Time-shrinking, its propagation, and Gestalt

principles. *Perception & Psychophysics, 64*, 919–931.

Shinozaki, N., Yabe, H., Sato, Y., Hiruma, T., Sutoh, T., Matsuoka, T., & Kaneko, S. (2003). Spectrotemporal window of integration of auditory information in the human brain. *Cognitive Brain Research, 17*, 563–571.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport, 12*, 7–10.

Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research, 24*, 117–126.

Szelag, E., von Steinbüchel, N., Reiser, M., de Langen, E. G., & Pöppel, E. (1996). Temporal constraints in processing of nonverbal rhythmic patterns. *Acta Neurobiologiae Experimentalis, 56*, 215–225.

Tallon-Baudry, C., & Betrand, O. (1999). Oscillatory gamma activity in humans and its role in object recognition. *Trends in Cognitive Sciences, 3*, 151–162.

Taubman, R. E. (1950a). Studied in judged number: I. The judgment of auditory number. *Journal of General Psychology, 43*, 167–194.

Taubman, R. E. (1950b). Studied in judged number: II. The judgment of visual number. *Journal of General Psychology, 43*, 195–219.

Tervaniemi, M., Saarinen, J., Paavilainen, P. Danilova, N., & Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology, 38*, 157–167.

Tiihonen, J., Hari, R., Kajola, M., Karhu, J., Ahlfors, S., & Tissari, S. (1991). Megnetoencephalographic 10-Hz rhythm from the human auditory cortex. *Neuroscience Letters, 129*, 303–305.

Timmers, R., Ashley, R., Desain, P., Honing, H., & Windsor, W. L. (2002). Timing of ornaments in the theme from Beethoven's Paisiello Variations: Empirical data and a model. *Music Perception, 20*, 3–33.

Todor, J. I., & Kyprie, P. M. (1980). Hand differences in the rate and variability of rapid tapping. *Journal of Motor Behavior, 12*, 57–62.

Todd, N. P. McA., O'Boyle, D. J., & Lee, C. S. (1999). A sensory-motor theory of rhythm, time perception and beat induction. *Journal of New Music Research, 28*, 5–28.

Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Perception & Psychophysics, 67*, 195–208.

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. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology.

Volman, M. J. M., & Geuze, R. H. (2000). Temporal stability of rhythmic tapping "on" and "off the beat": A developmental study. *Psychological Research, 63*, 62–69.

Warren, R. M., Gardner, D. A., Brubaker, B. S., & Bashford, J. A., Jr. (1991). Melodic and nonmelodic sequences of tones: Effects of duration on perception. *Music Perception, 8*, 277–290.

Windsor, W. L., Desain, P., Aarts, R., Heijink, H., & Timmers, R. (2001). The timing of grace notes in skilled musical performance at different tempi: a case study. *Psychology of Music, 29*, 149–169.

Wittmann, M., & Pöppel, E. (2000). Temporal mechanisms of the brain as fundamentals of communication – with special reference to music perception and performance. *Musicae Scientiae, Special Issue 1999–2000*, 13–28.

Woodrow, H. (1932). The effect of rate of sequence upon the accuracy of synchronization. *Journal of Experimental Psychology, 15*, 357–379.

Wright, B. A., & Fitzgerald, M. R. (2004). The time course of attention in a simple auditory detection task. *Perception & Psychophysics, 68*, 508–516.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport, 8*, 1971–1974.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology, 35*, 615–619.

Yabe, H., Koyama, S., Kakigi, R., Gunji, A., Tervaniemi, M., Sato, Y., & Kaneko, S. (2001a). Automatic discriminative sensitivity inside temporal window of sensory memory as a function of time. *Cognitive Brain Research, 12*, 3948.

Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T., & Kaneko, S. (2001b). Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. *Brain Research, 897*, 222–227.