



ELSEVIER

Human Movement Science 21 (2002) 515–532

HUMAN
MOVEMENT
SCIENCE

www.elsevier.com/locate/humov

Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions

Yanqing Chen ^{a,*}, Bruno H. Repp ^b, Aniruddh D. Patel ^a

^a *The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA*

^b *Haskins Laboratories, New Haven, CT 06511, USA*

Abstract

Spectral analysis was applied to study the variability in human rhythmic synchronization to a visual, auditory or combined auditory–visual metronome of about 2 Hz, as well as the variability in continuation tapping at the same rate with or without visual or auditory feedback. In synchronization, variability was larger in the visual condition than in the auditory and combined conditions, but only below frequencies of about 0.3 Hz. Thus, there seem to be at least two sources of variability in synchronization, one being modality-independent and limited to intervals shorter than 3 s, and the other being modality-dependent and evident as slow “drift”, especially in the visual task. In continuation tapping, variability did not depend reliably on the presence or modality of feedback. However, spectral analysis revealed a change in the temporal structure of variability around 0.08 Hz (a period of about 12 s or 24 taps), which roughly agrees with earlier findings reported in the literature.

© 2002 Elsevier Science B.V. All rights reserved.

PsycINFO classification: 2300; 2500

Keywords: Synchronization; Continuation tapping; Variability; Timing; Sensorimotor coordination; Spectral analysis; Modality differences

* Corresponding author. Tel.: +1-858-626-2069; fax: +1-858-626-2099.

E-mail address: chen@nsi.edu (Y. Chen).

1. Introduction

Variability in human timing occurs at different time scales. When tapping with a metronome, for example, slow trends of variation (*drift*) can be observed in asynchronies (defined as the timing differences between taps and stimulus onsets) and inter-tap intervals (ITIs). Drift is also observed in ITIs when participants are instructed to continue tapping at a given tempo after the metronome ceases. Such drift occurs in addition to local fluctuations in timing from tap to tap. Characterizing drift and understanding its origin has been a research interest for over a century. Early researchers (Stevens, 1886; Dunlap, 1910) noticed the ubiquity of drift in finger tapping tasks and used the terms “drifting tendency” or “wave” to describe the observed slow variations in ITIs and asynchronies. However, it is only in recent years that time series analysis techniques have been applied to such data with the aim of characterizing drift.

Gilden, Thornton, and Mallon (1995) computed the power spectrum of the ITIs in continuation tapping at various rates and showed that the low-frequency components of the ITI variations are of the $1/f$ type (see also Gilden, 2001; N. Yamada, 1995). This means that the low-frequency spectrum, when plotted in log–log coordinates, is an approximately linear function of frequency with a slope of -1 . The slower the drift, the larger its amplitude. At high frequencies, this relationship breaks down because other sources of variability with white-noise properties dominate. Chen, Ding, and Kelso (1997, 2001) similarly analyzed asynchronies in synchronization tasks and found that their power spectrum could be described as a $1/f^\alpha$ function of frequency (i.e., having a slope of $-\alpha$ in a log–power and log–frequency plot). The value of α was related to different task requirements (synchronization or syncopation) and cognitive strategies employed by the participants. $1/f^\alpha$ type noise contains variation (“waves”) at many different time scales. As the length of the time series increases, slow drift becomes increasingly prominent. Thus, spectral analysis provides an interesting mathematical description of the timing variations in rhythmic tapping, especially of the phenomenon of drift.

One possible mechanism underlying $1/f^\alpha$ noise is the aggregation of distributed neural processes acting on multiple time scales (Granger, 1966; Chen et al., 2001; Rao et al., 1997). In addition to this relatively slow variation, there are other sources of variability which are mostly responsible for local cycle to cycle variations. They include motor variability in continuation tapping, which is responsible for a negative correlation among successive ITIs (Wing, 1980; Wing & Kristofferson, 1973), and error correction in synchronization (Mates, 1994; Vorberg & Wing, 1996). The overall variability may constitute the summation of slow-scale and fast-scale processes. Madison (2001) found evidence of drift even in short trials (40 taps) of continuation tasks and attempted to distinguish these “higher order dependencies” from local dependencies arising from motor variability. However, he did not use spectral analysis methods.

One way of separating the two sources of variability is to look for a change in slope of the power spectrum function and to attribute the spectral energy on either side of this breakpoint to different processes (Gilden, 2001). Using this idea, Musha,

Katsurai, and Terauchi (1985) found a power spectrum breakpoint in continuation tapping corresponding to a time interval of 10 s and suggested that the temporal “controllability” of tapping extends within such a time span. M. Yamada (1996) (M. Yamada & Yonera, 2001) suggested that the breakpoint corresponds to 20 taps over different tapping rates, instead of a fixed time interval. In both cases, the reasons for these time spans (10 s or 20 taps) remain unclear.

The purpose of the present research was to compare the variability of rhythmic movement when pacing or feedback signals occur in different modalities, specifically audition and vision. We wanted to examine possible differences in the time scale of variability in the two conditions in synchronization tapping, and to examine if modality influences the breakpoints in the power spectrum of continuation tapping time series. Thus the experiment comprised both synchronization and continuation tasks. In the synchronization tasks, participants synchronized their finger taps with auditory, visual, or auditory–visual metronomes at a fixed tempo. In the continuation tasks, participants continued tapping at the same tempo after initially synchronizing with an auditory or visual metronome. During continuation tapping, they either did or did not receive a feedback signal (auditory or visual) from their own taps, identical to the pacing signal during synchronization. This design enabled us to determine whether the magnitude and spectral characteristics of variability are related to the modality of the perceptual input.

Several previous studies have shown that the variability of finger taps in synchronization (or closely related) tasks is greater when the pacing stimuli are visual than when they are auditory (Bartlett & Bartlett, 1959; Dunlap, 1910; Fraisse, 1948; Klemmer, 1967; Kolers & Brewster, 1985; Repp & Penel, in press). Repp and Penel (in press) also included a condition with auditory–visual pacing sequences and found that variability was similar to that in synchronization with auditory sequences, even though the participants had been instructed to synchronize with the visual sequence and to ignore the simultaneous auditory sequence. We expected to replicate these findings. The continuation tasks with sensory feedback replicated conditions in Kolers and Brewster’s (1985) study. In one experiment, these authors found that the difference in variability between auditory and visual conditions was just as large in continuation than in synchronization. In another experiment, in which only continuation data were analyzed, the modality effect was also present but decreased as the tempo decreased. Our continuation conditions without explicit sensory feedback resemble conditions in a recent study of interval subdivision by Semjen and Ivry (2001). They compared the variability of ITIs in continuation tapping following synchronization with auditory or visual pacing sequences. Variability was slightly greater when the pacing had been visual, but this held only for the shorter sub-intervals of a basic 1-s interval. Although we examined only isochronous tapping, the ITIs in our study were in the same range as the sub-intervals in Semjen and Ivry’s study, and so we wondered whether a modality effect would emerge.

We applied spectral analysis to the ITIs in all conditions, and to asynchronies in the synchronization conditions. Variance and standard deviation measures describe the variability of a time series averaged over the whole sequence, but they do not provide information about the time scale of the variation. In particular, they do

not reveal how much variance can be attributed to slow drift and how much to local fluctuations. Power spectrum analysis, on the other hand, can be viewed as a method for “decomposing” the variance of a time series into different frequency components, which provides information about the strength (power) of the variation at different time scales. By comparing the power spectra between different conditions, the time scales over which timing variations are the same or different can be identified.

2. Methods

2.1. Participants

The experiment was conducted at Haskins Laboratories. Six paid volunteers (three women, three men) and two unpaid men (author BR and a post-doctoral research associate) participated. Their ages ranged from 21 to 32, except for BR who was 56. All except the post-doc had participated in many earlier synchronization experiments. Four participants had extensive musical training (professional training in two cases), three had limited training, and one participant had none at all. All participants were right-handed.

2.2. Materials and equipment

Pacing sequences consisted of either 310 or 60 identical events occurring at fixed inter-onset intervals (IOIs) of 488 ms. Previous tapping experiments in the same lab (Haskins Labs) had used the same rate (e.g., Repp & Penel, in press). This rate has also been found to give rise to the smallest coefficient of variation of asynchronies (Bartlett & Bartlett, 1959) and is close to the preferred tapping rate for many people (see Kelso, Holt, Rubin, & Kugler, 1981). The long sequences were used in the synchronization conditions and the short ones in the continuation conditions. The sequences were instantiated as files of musical instrument digital interface (MIDI) instructions and were played by a program written in MAX running on a Macintosh Quadra 660AV computer, which also collected and saved the tapping data. Auditory sequences consisted of high-pitched tones (4168 Hz) produced on a Roland RD-250s digital piano. They were heard over Sennheiser HD540 II earphones. The tones had a sharp onset containing a strong noise component (piano key impact noise) and decayed within about 100 ms. Visual sequences consisted of discrete flashes of a green LED 3 mm in diameter. The LED was the message indicator of an Opcode MIDI Translator II box and lit up whenever a “note on” message was received from the computer. (There were no “note off” messages in the instructions.) The temporal variation of the LED luminance could not be determined precisely, but successive flashes were clearly visible as discrete events in a darkened room. The auditory and visual feedback signals during continuation tapping were identical to those making up the sequences and were triggered by a key depression without delay.

Participants tapped on a white key of a Fatar Studio 37 MIDI controller (a three-octave piano keyboard). The key moved about 10 mm and produced no sound unless

it was struck rather hard, in which case some impact noise was audible. This was true for three of the eight participants who thus received some direct auditory feedback from the taps in the purely visual conditions. (When auditory tones were provided as feedback over earphones, the direct auditory feedback was most likely masked.) The key depression was sensed during the downward movement of the key, which added up to -20 ms (depending on key velocity) to the individual asynchronies, but the magnitude of these asynchronies is not of particular interest here.

2.3. Procedure

Participants came for two sessions during each of which the seven tapping conditions were performed twice. (One participant completed only half of the second session and then became unavailable.) In each session, all conditions were performed once before they were repeated in the same order. Their order was partially fixed and partially counterbalanced across participants. Participants started with either visual or auditory synchronization, which was followed by the other of these two conditions and then by auditory–visual synchronization. The four continuation conditions followed, starting with either the visual or the auditory conditions and, within these, with either the feedback or the no-feedback condition.

The four conditions involving visual stimuli were conducted in the dark because the LED flashes were difficult to see otherwise. The room lights were extinguished, and the computer screen was dimmed and covered with a cardboard flap, which left only a faint illumination emanating from the bottom of the screen. The room lights were on during the three conditions involving only auditory stimuli. During the visual synchronization condition, participants did not wear earphones; during visual continuation conditions, the earphones were worn around the neck, so that a sound could be heard that signaled the end of the continuation tapping period.

Participants held the MIDI controller on their lap and tapped with the index finger of their preferred hand (the right hand in all cases), starting with the second sequence event and continuing beyond the end of the short sequence in the continuation conditions. The precise tapping kinematics varied across individuals; although it was recommended that the index finger remain in contact with the response key, some participants struck the key from above and thus generated some noise. For the visual and auditory synchronization conditions, participants were instructed to tap in synchrony with the flashes or tones, respectively. For auditory–visual synchronization, they were instructed to synchronize with the flashes and ignore the tones. For the continuation conditions, they were informed that at some point the sequence would stop and that the taps would be producing the flashes or tones. Participants were told to tap as steadily as possible at a constant tempo until they heard the termination signal.

Timing of the stimuli was defined by the timing of the MIDI instructions that gave rise to the stimuli. Time lags between instructions and stimuli were minimal, however, and constant in any case. Timing of a subject's tapping response was defined as the computer registered key-press. ITIs were defined as the timing differences between two adjacent taps as registered by the computer. In the synchronization

conditions, asynchronies were defined as the timing differences between the MIDI instructions for the stimulus (auditory or visual) and the associated key-press registrations. Continuous recordings from every single trial (about 350 taps in continuation and 300 taps in synchronization) form a time series for further analysis.

3. Results

3.1. Mean and variance: Synchronization conditions

Fig. 1 shows the means and standard deviations of ITIs and asynchronies in the three synchronization tasks. Fig. 1(a) shows that in the auditory and auditory–visual conditions, the mean ITIs were almost perfectly equal to the metronome period (488 ms), while in the visual synchronization task the average ITI was smaller and more variable across subjects (error bars represent standard error of the mean). This reflects the performance of three participants who had difficulty synchronizing with a visual sequence (see below). A one-way repeated-measures ANOVA, however, showed that the difference between conditions is not significant, $F(2, 14) = 1.38$, $p < 0.28$. The average standard deviation (SD) of the ITIs differed more reliably among the three tasks. Fig. 1(b) shows that the SD was significantly larger in the vi-

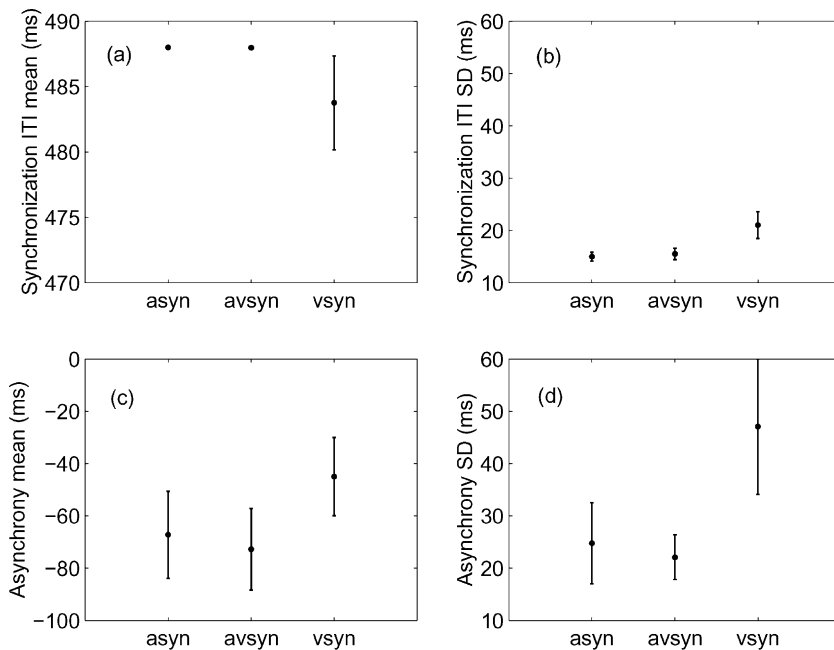


Fig. 1. Means and standard deviations of ITIs (a and b) and asynchronies (c and d) in the synchronization tasks, with standard error bars. asyn = auditory metronome; avsyn = auditory–visual metronome; vsyn = visual metronome.

sual synchronization condition than in the other two conditions, $F(2, 14) = 7.1$, $p < 0.01$. Variability in the auditory and auditory–visual synchronization conditions was similar, as predicted.

Turning from ITIs to asynchronies, we found that asynchronies could be clearly defined in only five participants in the visual synchronization condition. Three participants tapped at a slightly different tempo from the metronome, so that the phase relationship between taps and light flashes kept changing. (This also contributed to a different mean ITI in Fig. 1(a).) Asynchronies from the other two conditions (auditory and auditory–visual) were calculated from all eight participants, but ANOVAs were conducted only on the data of five participants. The means and SDs of asynchronies are shown in Fig. 1(c) and (d). Mean asynchronies in the visual condition were slightly smaller than in the other two conditions, $F(2, 8) = 9.5$, $p < 0.01$. The variability of the asynchronies (SDs) was significantly larger in the visual condition than in the auditory and auditory–visual conditions, $F(2, 8) = 23.6$, $p < 0.001$. Again, the latter two conditions did not differ in variability.

In summary, tapping to a metronome was significantly more variable when the metronome was visual than when it was auditory. When the two modalities were combined (auditory–visual metronome), variability was close to that of the auditory condition, even though the participants were explicitly told to ignore the auditory stimulus and attend to the visual one. This suggests that auditory information dominates over visual information in this task, in agreement with the findings of Repp and Penel (in press).

3.2. Mean and variance: Continuation conditions

Fig. 2 shows the means and standard deviations of the ITIs in the continuation tasks. The ITIs from the first 30 taps after the metronome stopped were analyzed separately from the remaining taps, to examine transient modality effects. The top panels (a and b) represent this initial transition period. The subsequent 260 taps reflect the “steady state” of continuation tapping and are shown in the bottom panels (c and d). Two types of ANOVA were performed, a single four-level variable (four conditions in Fig. 2 caption), and two crossed two-level variables (modality or feedback). The one-variable analysis failed to show any significant differences in mean ITIs, either in the transition period, $F(3, 21) = 2.2$, $p < 0.12$, or during the steady state, $F(3, 21) = 2.6$, $p < 0.09$. Fig. 2(b) and (d) suggest that the standard deviation of the ITIs in the auditory-feedback (afb) condition was slightly smaller than in the other three conditions. The difference was significant for the first 30 taps, $F(3, 21) = 5.0$, $p < 0.01$, but was only marginally significant for the subsequent 260 taps, $F(3, 21) = 3.0$, $p < 0.06$. The two-variable ANOVA showed the effect of feedback on variability (SD) of ITIs to be significant both for the first 30 taps, $F(1, 7) = 12.1$, $p < 0.02$, and for the subsequent 260 taps, $F(1, 7) = 6.1$, $p < 0.05$. The effect of modality (visual or auditory) was only close to significance for the first 30 taps, $F(1, 7) = 5.6$, $p < 0.06$, and was not significant for the subsequent 260 taps. Thus, auditory feedback may contribute to a smaller variance, especially during the initial transition period soon after a metronome ceases, but the difference is not

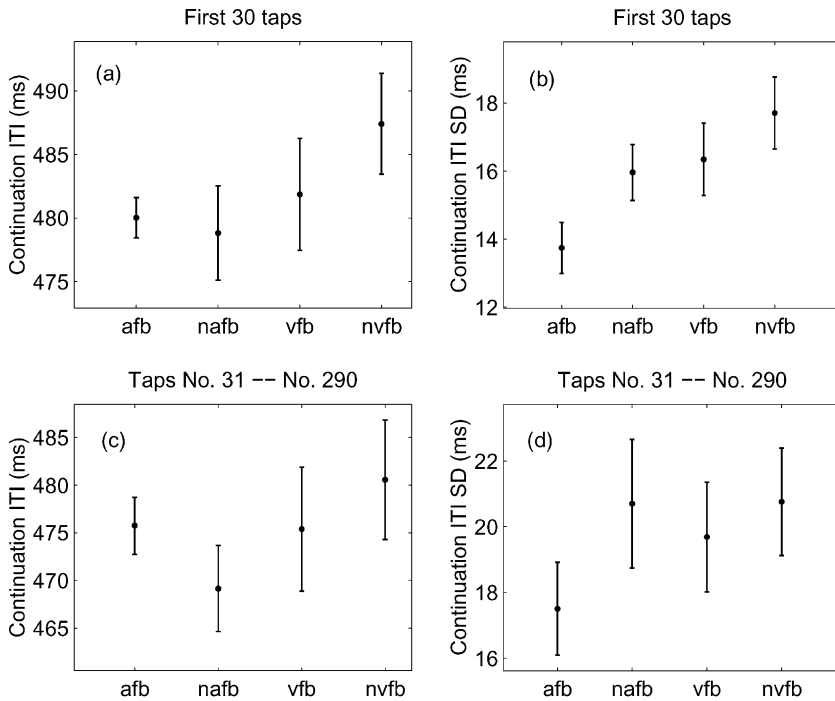


Fig. 2. Means and standard deviations of ITIs in the continuation conditions: (a) and (b): first 30 taps; (c) and (d): subsequent 260 taps. Error bars represent standard errors. afb = auditory feedback; nafb = no auditory feedback; vfb = visual feedback; nvfb = no visual feedback.

significant in the subsequent taps. The two-factor analysis suggested that the difference in variability is mainly a feedback effect (tapping with feedback produced smaller variances than without feedback), but the effect of modality (being visual or auditory) is only close to significance in the initial 30 taps, and not significant in the subsequent taps.

3.3. Power spectrum analysis

Turning to the principal focus of this study, we applied spectral analysis to ITI and asynchrony time series to study the time scale of variability in the synchronization and continuation conditions. Fig. 3(a) shows the power spectra (averaged across all eight participants) of the ITI time series in the synchronization tasks. The spectrum of the visual synchronization condition clearly shows greater power than the spectra of the other two conditions below a certain frequency value, indicating higher variability at low frequencies (i.e., slow drift). To estimate this “critical frequency” value, we calculated power differences between the visual condition and the other two conditions. This difference spectrum was calculated for each individual participant, then averaged across all eight participants. Plotted in Fig. 3(b) and (c) are averaged difference spectra plus or minus two standard errors of estimation. To find the

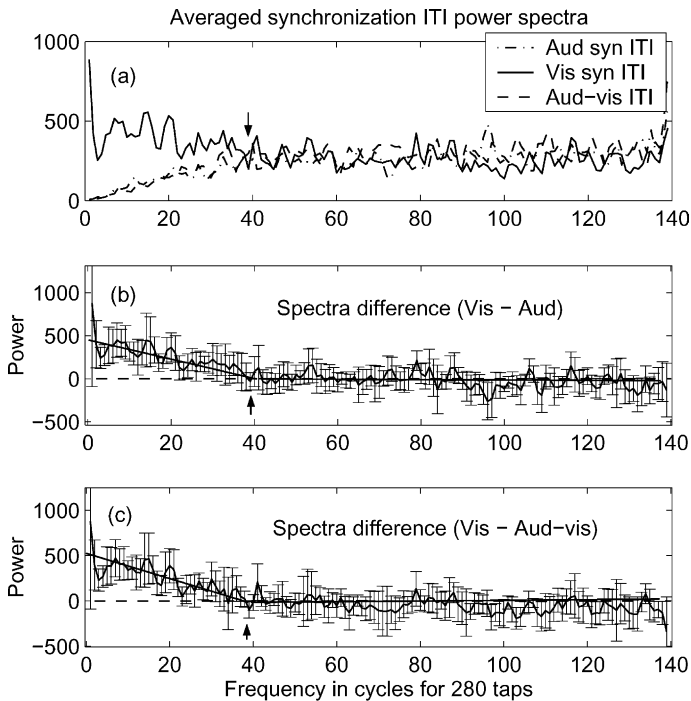


Fig. 3. (a) Averaged power spectra of ITI time series from the three synchronization tasks. (b) Power spectrum difference between visual and auditory conditions. (c) Spectrum difference between visual and auditory–visual conditions. In (b) and (c), error bars represent double standard errors. Arrows indicate the critical frequency which separates the visual condition from the other two conditions.

frequency where the differences reach zero, we fit two lines to the difference spectra. A linear function (with a negative slope) was fit to the low frequency part, and a straight line with slope zero was fit to the higher frequency part (because at higher frequencies power differences are not significantly different from zero). A parameter (frequency) needs to be chosen to separate these two regions of linear fit. The frequency giving the best least-squares fit is an estimate of the critical frequency. For the visual vs. auditory comparison (Fig. 3(b)), linear regression revealed that the critical frequency was 43.3 cycles (unit of frequency = number of cycles fitting into 280 taps; 95% confidence interval: 30.7–63.1, derived from the 95% confidence interval for the estimate of the low frequency slope). For the visual vs. auditory–visual comparison (Fig. 3(c)), the critical frequency was 39.5 cycles (95% confidence interval: 27.1–59.7). The mean of the two critical frequencies is 41.4 cycles, or about 6.8 taps (=280 taps divided by 41.4 cycles). This corresponds to an interval of about 3.3 s (488 ms times 6.8 taps). The analysis here suggests that the difference of variability between visual synchronization and the other two conditions may be largely due to variations at time scales longer than 3 s. Such variations represent slow drift.

As a comparison, Fig. 4 plots the autocorrelation functions of the ITI time series from auditory, auditory–visual and visual synchronization conditions. It is well

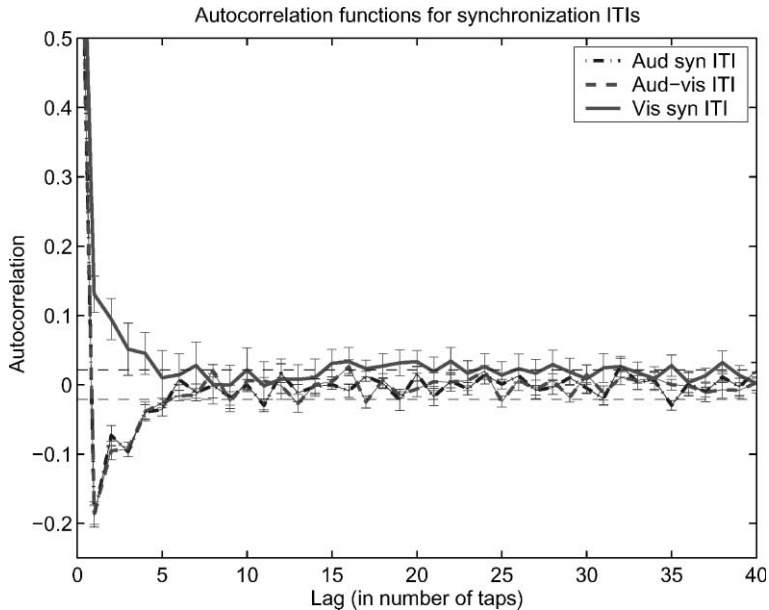


Fig. 4. Autocorrelation functions of ITI time series from auditory, auditory–visual, and visual synchronization conditions. Plotted are means across all participants and multiple trials and their standard errors of estimation. Between two dashed horizontal lines are levels of correlations which are not significantly different from zero.

known that the autocorrelation function is mathematically related to the power spectrum function through an inverse Fourier transform (Chatfield, 1987). But because of the finite length of the time series and a limited number of trials, the information revealed by the autocorrelation function and the power spectrum may not be the same. From Fig. 4, we can see that autocorrelations of auditory and auditory–visual synchronization are negative at small lags (<5) (negative lag-1 correlation was predicted in Wing & Kristofferson's (1973) model), while autocorrelations of visual synchronization are positive at the same lags. Conclusions, however, can not be drawn unequivocally that the differences between visual and auditory or auditory–visual synchronization are local and only occur at short lags. The reason is the following. From the appearance of ITI time series and power spectrum analyses, we know that visual synchronization ITI time series have slow and low frequency drifts. Because these drifts are deviations from the overall mean ITI and persistent over a number of taps, short-lag autocorrelations are dominated by the slow drift and variation around the local mean. Higher lag autocorrelations are small (not significantly different from zero) and may not provide reliable information about the long-range properties of the time series. So, the short-lag (local) autocorrelation function of a time series with slow drift is strongly influenced by the drift, instead of being determined by local correlations only. To avoid this, some detrending procedures need to be applied (see Ogden & Collier, 1999). But it might be arbitrary to decide which

trend or variations in which frequency ranges are to be removed. This is the reason why we chose spectral analysis to study the variations of rhythmic tapping in our current research. More detailed comparisons between the autocorrelation function and power spectrum analysis can be found in Priestley (1981) and Beran (1994).

The power spectrum of asynchronies was also calculated. Fig. 5 shows the averaged power spectra of asynchronies in the auditory and auditory–visual conditions from all eight participants, and the averaged asynchrony spectrum in the visual condition from the five participants who maintained synchronization successfully. Here the spectra are plotted on log–log scales to demonstrate the similarity to a $1/f$ type of power spectrum in the auditory and auditory–visual conditions (cf. Chen et al., 1997, 2001). Once again, there is more power in the visual condition at low frequencies. The critical frequency beyond which the spectra are no longer different is about the same (40 cycles) as for the ITIs. This figure demonstrates that the $1/f$ spectrum observed in synchronization tasks is a modality-dependent phenomenon. If the pacing stimulus is visual, the overall spectral shape of asynchronies is not $1/f$. This is due to low frequency power or slow drift.

In the continuation conditions, spectral analysis was applied to the ITI time series from the steady-state phase only (260 taps); the transient phase was not included and was not long enough to support a separate spectral analysis. Fig. 6(a) shows the averaged power spectrum for each of the four different conditions. Whereas the spectra roughly follow a $1/f$ slope at low frequencies, they have different slopes at higher frequencies (see also Gilden et al., 1995). The breakpoint where two slopes separate occurs around a frequency of 10 cycles. This frequency value corresponds to 26 taps or 12–13 s. (The average ITI in the continuation tasks was about 475 ms – see

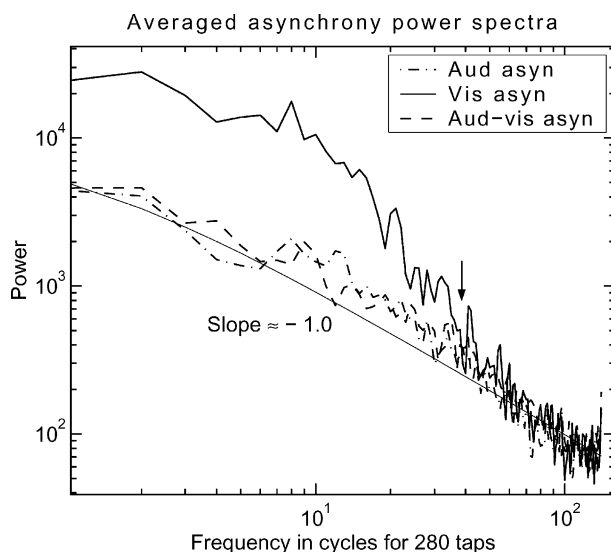


Fig. 5. Averaged power spectra of asynchrony time series from the three synchronization tasks.

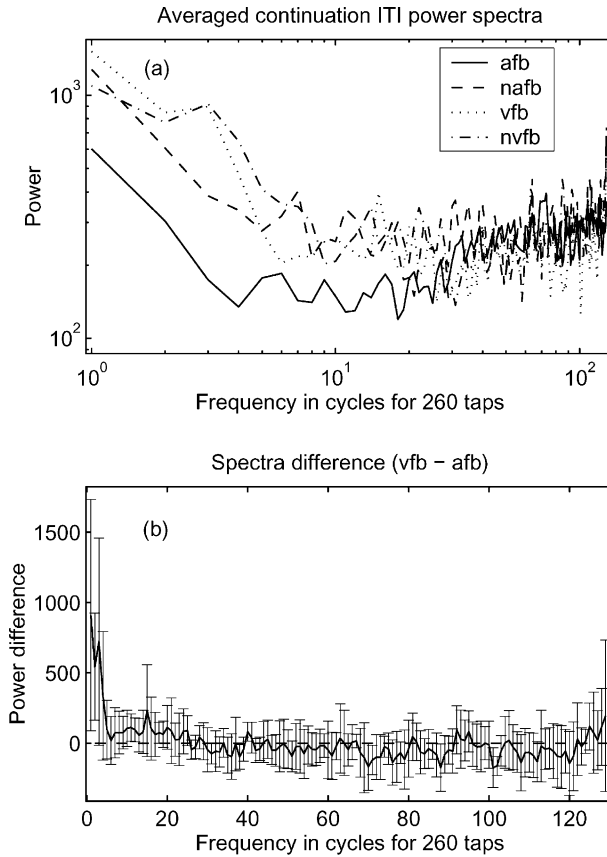


Fig. 6. (a) Power spectra of continuation ITI time series in four conditions. Abbreviations for experimental conditions are the same as in Fig. 2. (b) Power spectrum difference between continuation tapping with visual feedback and with auditory feedback. Error bars represent double standard errors.

Fig. 2(c) – so that $26 \times 475 \text{ ms} \approx 12 \text{ s}$.) This value is close to the range reported by Musha et al. (1985); M. Yamada (1996) and M. Yamada and Yonera (2001). Because there was only one tempo condition in our experiment, the question of whether the breakpoint corresponds to a fixed time scale or a fixed number of taps at different tapping rates cannot be resolved here. One thing that can be inferred from our results, however, is that the value of the breakpoint is not related to the modality of the metronome during the initial synchronization phase of the experiment or the type of feedback participants experienced during continuation tapping.

Fig. 6(a) also suggests that continuation tapping with auditory feedback generated lower power (smaller variance) at low frequencies. Statistical tests, however, demonstrated that the difference is not reliable. (The effect was largely due to one participant.) Fig. 6(b) shows the power differences between the visual and auditory feedback conditions. Plotted are differences averaged across eight participants ± 2

standard errors of the mean. None of the differences are significantly different from zero, except for very low frequency values (1 or 2 cycles). This result is consistent with the analysis of ITI mean and standard deviation in Fig. 2(c) and (d), which suggested that auditory feedback does not strongly reduce the variance of ITIs in continuation tapping.

4. Discussion

Isochronous tapping with a metronome is a seemingly simple task. Yet the behavioral simplicity of the task belies its biological complexity. (In this vein it is worth noting that there are, to our knowledge, no reports of other species being successfully trained to tap or peck in synchrony with an isochronous external stimulus.) Our study joins a tradition of research aimed at illuminating central timekeeping mechanisms via the study of timing variability in isochronous tapping (e.g., Chen et al., 2001; Schöner, 2002; Vorberg & Wing, 1996; Wing & Kristofferson, 1973). In particular, we were interested in the idea that overall variability may represent the aggregation of neural processes operating at different time scales, rather than the variability of a single timekeeper. The idea of processes at different time scales is suggested by the fact that isochronous tapping exhibits both a negative first-order correlation between successive ITIs, possibly due to motor variability or local phase-correction mechanisms (Mates, 1994; Vorberg & Wing, 1996; Wing & Kristofferson, 1973), and a positive higher-order correlation between ITIs over longer distances, due to drift (Madison, 2000), similar to what we observed in the visual synchronization condition.

We pursued the topic of multiple time scale processing by comparing the variability of taps synchronized to visual vs. auditory pacing stimuli. It has long been known that tapping is more variable in the former case (e.g., Bartlett & Bartlett, 1959; Kolers & Brewster, 1985; Repp & Penel, in press). Previous studies, however, have not provided a quantitative comparison of the time scales of variation in auditory vs. visual synchronization. Such information could provide valuable clues about the scales over which timing-related neural processes operate. To gather this information we applied spectral analysis to ITIs and asynchronies in synchronization tasks with an auditory vs. a visual metronome. We also included an auditory–visual condition, in which each metronome event consisted of a simultaneous sound and flash. This extended the work of Repp and Penel (in press), who showed that timing patterns in this mixed-modality condition were determined by the auditory stimuli and even when subjects are explicitly instructed to ignore the sounds and attend to the flashes.

In addition to synchronized tapping, we also examined continuation tapping, in which the participant attempts to continue tapping at the same tempo despite the absence of a metronome. In our study, the continuation phase either did or did not have feedback. If feedback was present, each tap generated a tone or a flash identical to those experienced during the pacing stimulus. Kolers and Brewster (1985) had found that differences in variability between auditory and visual tapping during continuation are just as large as during synchronization, as long as sensory feedback

accompanies each tap. Furthermore, a recent study of continuation tapping without feedback (Semjen & Ivry, 2001) had found larger variability of ITIs after a visual than after an auditory pacing sequence. We sought to replicate these intriguing findings, which suggest that central timekeeping mechanisms are more effectively entrained by auditory than by visual input, even when both consist of easily discernable isochronous events which are identical in tempo.

4.1. Synchronization experiment

Replicating the work of previous researchers, we found that overall timing variability is larger in visual than in auditory and auditory–visual synchronization, both in terms of ITIs and asynchronies. Variance in the auditory–visual condition was not significantly different from the auditory condition, indicating auditory dominance in rhythmic synchronization tasks. This result reinforces the findings of Repp and Penel (in press), who reported auditory dominance not only in the variability of asynchronies, but also in the direction and magnitude of error correction responses to small perturbations in auditory–visual metronome sequences.

Furthermore, in the visual condition, some participants were prone to shift away from the required tempo and to lose synchronization, whereas this never occurred in the auditory and auditory–visual conditions. Repp (submitted for publication) recently investigated the rate limits of synchronization with visual and auditory sequences similar to those used here. These average “synchronization thresholds” were at event IOIs of about 450 and 120 ms, respectively. Since we used an IOI of 488 ms in our study, participants were operating fairly close to (and in three cases, apparently below) their synchronization thresholds in the visual condition. Repp also found, however, that variability was much larger in visual than in auditory synchronization over a wide range of IOIs, which makes us confident that our results are not specific to the IOI we have chosen.

Moving beyond overall variance, power spectrum analysis of ITIs and asynchronies provided information about the time scale of the difference in variances across conditions. Results indicated that the differences in variability of ITIs and asynchronies are situated in the low frequency range, at time intervals longer than 3 s. That is, at periods longer than ~ 3 s the drift in the visual condition has substantially more power than the drift in the other conditions, whereas the variances at shorter time scale are similar across modalities. It is important to note that this finding is not the only conceivable one: Visual synchronization might have exhibited greater variability across all time scales, perhaps suggesting poorer overall entrainment of a central timekeeper. Instead, there is a distinct breakpoint at which the variability of visual and auditory (and auditory–visual) synchronization diverges (Fig. 3). This suggests that timing mechanisms operating on durational scales below 3 s are similar across domains, but that there exist longer-scale timekeeping mechanisms which differ across sensory modalities.

The fact that the point of divergence between the power spectra for the visual and auditory conditions is at ~ 3 s is interesting. This interval may be related to the transition from reaction to anticipation in rhythmic sensorimotor coordination

(Engström, Kelso, & Holroyd, 1996). That is, if an isochronous stimulus period is longer than 3 s, participants' responses lag behind the stimulus onset on average, indicating that participants are reacting to the stimulus. When the period is shorter than 3 s, participants typically anticipate the stimulus, tapping before or close to metronome onset. Elbert, Ulrich, Rockstroh, and Lutzenberger (1991) reported that brain evoked potentials are different when participants reproduce intervals above and below 3 s. Studies on patients with brain lesions have also suggested that different brain structures are involved in timing tasks requiring temporal intervals longer than 3 s (von Steinbüchel, 1998). Mates, Radil, Müller, and Pöppel (1994) argued that this interval (for which they provided a somewhat lower estimate of 2.4 s) is the upper limit of a temporal integration process in sensorimotor synchronization. This integration process may be mostly determined by a central mechanism, and less dependent upon different modalities. This could be the reason that we observed similar variances in auditory and visual synchronization within 3 s. Above this temporal integration interval, effects of modality dominate and produce higher variability in the visual condition, even though at the present stage the exact mechanism that causes higher variability in visual tasks remains unclear.

4.2. Continuation experiment

We did not replicate the finding of Kolers and Brewster (1985) that the difference in variability between auditory and visual continuation tapping (with feedback) was as large as that in synchronization tapping. Neither did we replicate the finding of Semjen and Ivry (2001) that in continuation tapping without feedback, variability was slightly greater following synchronization with a visual stimulus. There was significant effect of feedback on the variability of ITIs, but the difference between visual and auditory feedback conditions was only close to significance in the initial 30 taps, and not significant in the subsequent taps. The difference found by Semjen and Ivry was small and may have been specific to their more complex rhythmic sequences. As to the Kolers and Brewster study, it is unclear whether their participants were informed about the continuation phase. It is possible that they continued tapping in the belief that they were still synchronizing with a sequence, and that this caused the differences in variability. Fraise (1971) noted that uninformed participants tend to accelerate when tapping with auditory feedback, and this tendency may be enhanced with visual feedback. Also, our participants probably had more musical training than those of Kolers and Brewster and thus were better able to maintain a steady tempo, regardless of feedback. There was some suggestion in our data that auditory feedback may be able to stabilize the tapping rhythm, but the effect was small and largely due to one participant.

Spectral analysis revealed the characteristic "V" shape in (log–log scale) power spectra of continuation ITI time series. This phenomenon and the time scale of the break point (at about 12 s) are consistent with previous observations (Gilden et al., 1995; M. Yamada, 1996; M. Yamada & Yonera, 2001). The "V" shape can be understood as evidence of two different sources of variability in continuation tapping: one at low frequency ranges, perhaps of the $1/f$ type (Gilden et al., 1995); and

the other due to motor variability or sequential correlation in “central clock” mechanisms, as suggested by theoretical models (Gilden et al., 1995; Wing & Kristoffer-son, 1973; Wing, 1977). Our results suggest that these latter sources of variability are not strongly influenced by visual or auditory feedback.

4.3. Conclusion

Our study has provided evidence that timing variability of tapping in synchrony with an isochronous metronome is segregated into discrete time scales, with a breakpoint around 3 s. At this duration the variability of tapping to a visual metronome becomes significantly greater than the variability of tapping to an auditory (or auditory–visual) metronome. This time scale may represent a modality-independent temporal integration window which operates in sensorimotor integration (cf. Mates et al., 1994). In contrast to synchronization, continuation tapping (without an external pacing stimulus) has similar variance structure regardless of the modality of prior synchronization or the presence of feedback. Nevertheless, spectral analysis of continuation tapping reveals a characteristic nonmonotonic “V” shape, with the inflection point of the “V” around 12 s, suggesting a breakpoint in the control of self-paced tapping. The meaning and cognitive significance of this long time scale awaits future research.

Acknowledgements

This research was supported by NIH grant MH-51230 to Bruno Repp, and by the Neurosciences Research Foundation. We thank Dr. Gerald Edelman and Dr. W. Einar Gall for their encouragement and support.

References

- Bartlett, N. R., & Bartlett, S. C. (1959). Synchronization of a motor response with an anticipated sensory event. *Psychological Review*, *66*, 203–218.
- Beran, J. (1994). *Statistics for long-memory processes*. Boca Raton, FL: Chapman & Hall/CRC.
- Chatfield, C. (1987). *The analysis of time series. An introduction*. London: Chapman & Hall.
- Chen, Y., Ding, M., & Kelso, J. A. S. (1997). Long memory processes ($1/f^\alpha$ type) in human coordination. *Physical Review Letters*, *79*, 4501–4504.
- Chen, Y., Ding, M., & Kelso, J. A. S. (2001). Origins of timing errors in human sensorimotor coordination. *Journal of Motor Behavior*, *33*, 3–8.
- Dunlap, K. (1910). Reactions to rhythmic stimuli, with attempt to synchronize. *Psychological Review*, *17*, 399–416.
- Elbert, T., Ulrich, R., Rockstroh, B., & Lutzenberger, W. (1991). The processing of temporal intervals reflected by CNV-like brain potentials. *Psychophysiology*, *28*, 648–655.
- 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 (Auditory and visual rhythms). *L'Année Psychologique*, *49*, 21–41.

- Fraisse, P. (1971). Les repères du sujet dans la synchronisation et dans la pseudo-synchronisation (The subject's guide marks in synchronization and pseudo-synchronization). *L'Année Psychologique*, 71, 359–369.
- Gilden, D. L. (2001). Cognitive emissions of $1/f$ noise. *Psychological Review*, 108, 33–56.
- Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). $1/f$ noise in human cognition. *Science*, 267, 1837–1839.
- Granger, C. W. J. (1966). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238.
- Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear, limit cycle oscillatory processes: Theory and data. *Journal of Motor Behavior*, 13, 226–261.
- Klemmer, E. T. (1967). Sequences of responses to signals encoded in time only. *Acta Psychologica*, 27, 197–203.
- Kolers, P. A., & Brewster, J. M. (1985). Rhythms and responses. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 150–167.
- Madison, G. (2000). On the nature of variability in isochronous serial interval production. In P. Desain & L. Windsor (Eds.), *Rhythm Perception and Production* (pp. 96–113). Lisse: Swets & Zeitlinger.
- 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.
- Musha, T., Katsurai, K., & Terauchi, Y. (1985). Fluctuations of human tapping intervals. *IEEE Transactions in Biomedical Engineering*, 32, 578–582.
- Ogden, R. T., & Collier, G. L. (1999). On detecting and modeling deterministic drift in long run sequences of tapping data. *Communications in Statistics – Theory and Methods*, 28, 977–987.
- Priestley, M. B. (1981). *Spectral analysis of time series*. London: Academic Press.
- Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz, J. A., Cox, R. W., & Binder, J. R. (1997). Distributed neural systems underlying the timing of movements. *Journal of Neuroscience*, 17, 5528–5535.
- Repp, B.H. Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits/costs of interval subdivision. Submitted for publication.
- Repp, B.H., & Penel, A. Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, in press.
- Schöner, G. (2002). Timing, clocks, and dynamical systems. *Brain and Cognition*, 48, 31–51.
- Semjen, A., & Ivry, R. B. (2001). The coupled oscillator model of between-hand coordination in alternate-hand tapping: a reappraisal. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 251–265.
- Stevens, L. T. (1886). On the time-sense. *Mind*, 11, 393–404.
- von Steinbüchel, N. (1998). Temporal ranges of central nervous processing: clinical evidence. *Experimental Brain Research*, 123, 220–233.
- Vorberg, D., & Wing, A. (1996). Modeling variability and dependence in timing. In H. Heuer & S. W. Keele (Eds.), *Handbook of perception and action* (vol. 2, pp. 181–262). London: Academic Press.
- Wing, A. M. (1977). Effects of type of movement on temporal precision of response sequences. *British Journal of Mathematical and Statistical Psychology*, 30, 60–72.
- Wing, A. M. (1980). The long and short of timing in response sequences. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 469–486). Amsterdam: North-Holland.
- Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception and Psychophysics*, 14, 5–12.

- Yamada, M. (1996). Temporal control mechanism in equaled [sic] interval tapping. *Applied Human Science*, *15*, 105–110.
- Yamada, M., & Yonera, S. (2001). Temporal control mechanism of repetitive tapping with simple rhythmic patterns. *Acoustical Science and Technology*, *22*, 245–252.
- Yamada, N. (1995). Nature of variability in rhythmical movement. *Human Movement Science*, *14*, 371–384.