Journal of Experimental Psychology:
Human Perception and Performance
1995, Vol. 21, No. 1, 3-18

0096-1523/95/$3.00

Perception and Production of Temporal Intervals Across a Range of Durations: Evidence for a Common Timing Mechanism

Richard B. Ivry and R. Eliot Hazeltine
University of California, Berkeley

Study participants performed time perception and production tasks over a set of 4 intervals ranging from 325 to 550 ms. In 3 experiments, variability on both the production and perception tasks was found to be linearly related to the square of the target intervals. If the perception and production of short temporal intervals use a common timing mechanism, the slopes of the functions for the 2 tasks should be identical. The results of Experiment 1 failed to support this prediction. However, when the 2 tasks were made more similar by providing a single (Experiment 2) or multiple (Experiment 3) presentations of the target interval per judgment or production, the perception and production functions were nearly identical. The results suggest that temporal judgments and productions are based on an integrated internal representation of the target interval rather than reference to an internal oscillatory process.

Many behaviors are dependent on the accurate processing of temporal information. For example, skilled movements reflect precise timing of activity across different groups of muscles. Many perceptual tasks similarly require processing changes in information over time. Anticipating the location of a moving object or determining the duration of a stimulus requires temporal information processing. Although the representation of temporal information may be an implicit, emergent property in some dynamic tasks, there are motor and perceptual tasks that appear to be dependent on the operation of an internal timing mechanism (Ivry, 1993). The focus of this article is whether motor and perceptual tasks that require precise timing are dependent on a common mechanism.

To investigate this issue, comparisons must be made across motor and perceptual tasks that require precise timing. If tasks use the same timing mechanism, then performance should vary in similar ways among the tasks. Keele, Pokorny, Corcos, and Ivry (1985) used a correlational method to compare variability on motor and perceptual timing tasks. Study participants were tested on two tasks, a repetitive tapping task and a time perception task. For each task, the primary dependent variable was the participants' temporal consistency. For the tapping task, consistency was evaluated by the standard deviation of the intertap intervals. For the perception task, an estimate of the standard deviation of the psychophysical function describing perceptual acuity was obtained by an adaptive psychometric procedure. A significant correlation of .52 was found between these two measures of temporal acuity, suggesting that the two tasks used a common timing mechanism. This interpretation was strengthened by the lack of correlation between performance on the perception task and a second motor task that did not require precise timing.

An alternative approach that is essentially correlational in nature has been used in experiments using neurological patients (Ivry & Diener, 1991; Ivry & Keele, 1989; Ivry, Keele, & Diener, 1988). Ivry and Keele (1989) tested patients with lesions of the premotor cortex, basal ganglia, and cerebellum as well as age-matched control individuals on the repetitive tapping task and time perception task. Only the patients with cerebellar lesions were significantly impaired on both tasks. This group was not impaired on a control perception task, leading Ivry and Keele to conclude that the patients' deficit was specific to the timing requirements of the tapping and time perception tasks. Indeed, they argued that this timing capability is exploited in a wide range of tasks, both motoric and perceptual, that require the representation of temporal information (see Ivry, 1993; Keele & Ivry, 1991).

In the previous discussion, variability on the tapping and time perception tasks was taken as a measure of the consistency of an internal timing mechanism. Although these tasks were selected for their simplicity, it would be fallacious to expect that all of the variability should be attributed to a timing mechanism. Other psychological operations are surely involved in the performance of these tasks (e.g., in perception, detecting a signal; in production, implementing a response), and these operations can be expected to contribute additional sources of variability. It would be useful to be able to divide the total variability into (at least) two components: a component that measures variability in the

Richard B. Ivry and R. Eliot Hazeltine, Department of Psychology, University of California, Berkeley.

This research was supported by National Institutes of Health Grant NS30256, a Sloan Fellowship in Neuroscience, and a National Science Foundation Predoctoral Fellowship.

We are grateful to Steven Keele, Simon Grondin, and Dirk Vorberg for their comments during the course of this project. Informed consent was obtained from all of the study participants, and the data were coded to protect the anonymity of the participants.

Correspondence concerning this article should be addressed to Richard B. Ivry, Department of Psychology, University of California, Berkeley, California 94720. Electronic mail may be sent via Internet to ivry@garnet.berkeley.edu.

timing mechanism and a component that measures variability attributed to the additional psychological operations excluding the timing mechanism.

Wing and Kristofferson (1973; Wing, 1980) developed a model that decomposes variability on a repetitive tapping task into two components. One component is identified with an internal timing mechanism or clock. The other component is associated with those processes involved in implementing a motor response. By assuming that these two components are independent and that the responses are generated in an open-loop manner, Wing and Kristofferson have shown that the correlation between successive intertap intervals provides a measure of implementation variability. By subtracting this estimate from the total variability, an estimate of the clock variability can be obtained.

Substantial support for the model has been obtained in studies with both normal and neurologically impaired populations (Ivry & Keele, 1989; Ivry et al., 1988; Wing, 1980). However, when applied in the study of Keele et al. (1985), the correlation between the clock estimates and time perception was not larger than the correlation between total tapping variability and time perception. There are at least three possible explanations for this result. First, clock variability is only indirectly estimated in the Wing and Kristofferson (1973) model. Any error in estimating the implementation variability will produce an error in the clock estimate. That is, although the clock and implementation processes are assumed to be independent, the estimates are not obtained independently. Second, Wing and Kristofferson labeled all of the nonimplementation variance as "clock" variance. This designation is probably inaccurate. The so-called clock variance can be expected to include all sources of variability that operate prior to the implementation process (Ivry & Corcos, 1993). Thus, it is more appropriate to consider the nonimplementation variability as "central" variance. If these other processes are not active in the perception task, then their presence would be expected to lower the correlation between the tapping and perception scores. A corollary to this last point is that the perception task can also be expected to contain sources of variance in addition to those attributed to an internal timing mechanism. For example, variance may arise in processes involved in detecting the onset and offset of the stimuli and, as with all behavioral tasks, general factors such as attention and arousal can affect performance. Unfortunately, there is no analog of the Wing–Kristofferson model for this task and thus a decomposition of the total variance is not possible.

We now turn to an alternative method for measuring variability on time production and perception tasks. This method will be referred to as the *slope analysis.* The basic assumption underlying the slope analysis is that the total variance on a task requiring precise timing is composed of two distinct parts: a duration-dependent source and a duration-independent source (see Getty, 1975). This general relationship can be expressed as

$$\sigma^2 \text{ Total} = \sigma^2 \text{ Duration-dependent} + \sigma^2 \text{ Duration-independent}. \quad (1)$$

By definition, duration-independent sources are invariant across different target intervals (the interval being measured). Correspondingly, duration-dependent sources are expected to grow with increases in the interval being produced or measured.

The function describing how this component is related to the target interval depends on certain assumptions made regarding the timing mechanism. If timing is dependent on the operation of a Poisson process, then total variance is expected to grow in proportion to the base interval (Abel, 1972; Creelman, 1962; Killeen & Weiss, 1987). That is, the total variance is linearly related to the target interval or

$$\sigma^2 \text{ Total} = kD + c, \quad (2)$$

in which k is the constant of proportionality representing the rate of increase in duration-dependent variability, D is the duration of the target interval, and c is a constant representing the duration-independent variability component.

Alternatively, as with many psychological processes, variability in a timing mechanism may conform to Weber's law in that the standard deviation will be a constant proportion of the target interval. The original form of this law does not include an intercept term, a term taken in the current discussion to correspond to the variance of the duration-independent component. Assuming that the two sources of variability in Equation 1 are independent, Getty (1975, Equations 1–3; see also Fetterman & Killeen, 1992, Equation 1) has derived a generalized form of Weber's law that includes a duration-independent component,

$$\sigma^2 \text{ Total} = k^2D^2 + c, \quad (3)$$

in which the terms are as defined for Equation 2. For values of D that are reasonably greater than zero (i.e., greater than 100 ms), the Weber fraction is approximated by k, the square root of the slope term in Equation 3 (Getty, 1975).

Numerous time perception and time production studies have been conducted using the basic logic of the slope analysis to compare the validity of Equations 2 and 3. Specifically, these experiments have measured changes in variability as a function of the target interval and evaluated whether the function is best described when the target interval is plotted on the abscissa or when the square of the target interval is plotted on the abscissa. For the most part, the evidence is supportive of the latter formulation, that is, the generalized form of Weber's law. This result has been obtained in time perception experiments with human participants at least with target intervals in the range of 100–2,000 ms, using a variety of psychophysical methods (Allan & Gibbon, 1991; Fetterman & Killeen, 1992; Getty, 1975; Wearden, 1991, 1992). Converging results have also been obtained in animal research with both rats (Church, Getty, & Lerner, 1976) and pigeons (Fetterman & Killeen, 1992). Fewer time production studies have been reported, but these have also tended to provide results conforming to the generalized form of Weber's law (Grondin, 1992; Wearden & McShane, 1988).

One reported exception comes from Wing (1980). In that study, participants performed the repetitive tapping task with the target interval varying between trials from 220 to 490 ms. The total variability was plotted as a function of the mean produced intervals (Equation 2), and the estimates of the clock and implementation components were derived from the two-process model of Wing and Kristofferson (1973). A strong linear fit was obtained for both the total variability and clock functions, with the percentage of variance accounting for 93% and 95%, respectively.[1] As predicted by the model, the implementation estimate did not change consistently across the different target intervals.

However, a slope analysis using Equation 3, the generalized form of Weber's law, can be performed on the data from Wing (1980; see Ivry & Corcos, 1993). When the results from this analysis are compared with those provided by Wing, two interesting results emerge. First, the strength of the linear fit as measured by the proportion of variance accounted for is actually slightly greater when the square of the mean produced intervals was plotted on the abscissa. The values are 95% and 96% for the total variability and clock functions, respectively. The similarity in the regression results when one of the variables is squared (as in Wing, 1980) in comparison to when that variable is not transformed, thus pointing out a weakness of this measure in evaluating these functions.

Second, and more interesting, the two analyses yield very different values for the intercept, the constant representing the duration-independent component. In Wing's (1980) original analysis (Equation 2), the intercepts for the total variability function and the clock function are -140.4 and -184.2 ms, respectively. These negative intercepts, if they are assumed to not represent measurement error, are nonsensical. A priori, one would expect the intercept for the clock function to be zero. The value from the total variability function should provide an estimate of the variance attributed to the duration-independent source, and variances cannot be negative. When the intercepts are calculated from Equation 3, the values are more meaningful. The clock function intercept of 19 ms approaches zero, and the intercept of the total variability function is 54 ms, a value that is reasonably close to the average implementation estimate obtained from the Wing and Kristofferson (1973) model (approximately 30 ms). Thus, contrary to the original interpretation of Wing, these data are also in accord with the generalized form of Weber's law. It should be noted that most of the studies examining how variability on time discrimination and production tasks is related to changes in the target interval have tended to focus on the slope component in Equations 2 and 3. An examination of the intercept component can also be informative as to the appropriate function.

In summary, the slope analysis method has two important features. First, it provides a means for partitioning the duration-independent component of the total variability for both time production and time perception tasks. The estimate of this component is obtained from the intercept values, and there is no a priori reason to expect these values to be equal for production and perception tasks. Second, the slope analysis provides a direct means for estimating a crucial property of the duration-dependent component of the total variability: Assuming a linear relationship (e.g., Equation 2 or 3), the slope parameter provides an estimate of the change in variability as a function of the base interval. This method stands in contrast to that provided by Wing and Kristofferson (1973), in which the clock component is obtained indirectly.

Previous studies that have used the logic of the slope analysis have tested performance on either perception tasks or time production tasks. As such, none allow a comparison of performance between these types of tasks. In the experiments reported below, we tested participants on both time perception and production tasks in which they were required to do the tasks across a range of target intervals. This design allows us to compare empirically the validity of Equations 2 and 3 on both tasks while using the same set of participants. Moreover, assuming that as in previous studies Equation 3, the generalized form of Weber's law, provides a better description of the data, it is predicted that the Weber fractions would be equal for the time perception and production tasks. If confirmed, this result would provide evidence that a common timing mechanism is used in both tasks, converging with the results obtained in correlational studies (Ivry & Keele, 1989; Keele et al., 1985). However, if the production and perception tasks use different timing mechanisms, there is no reason to expect a correspondence between the Weber fractions.

Experiment 1

In Experiment 1, participants were tested on two tasks. One task was the repetitive tapping task introduced by Wing and Kristofferson (1973). The primary measure of interest was the variability of the participants' intertap intervals. In the other task, time perception acuity was evaluated using an adaptive psychophysical procedure (Pentland, 1980), with the primary measure also being a measure of variability. For both tasks, target intervals of 325, 400, 475, and 550 ms were tested in separate blocks. After multiple test sessions, production and perception functions were calculated using either the target interval (Equation 2) or the square of the target interval (Equation 3). With the results of these calculations, a comparison was made between the production and perception functions to assess whether the tasks share a common timing mechanism.

Method

Participants

Eight right-handed undergraduate students from the University of California, Santa Barbara or Berkeley, participated in the experiment. Each was tested on four different days, with each session lasting approximately 1 hr. Each participant was paid $25.

[1] Calculations are based on extrapolations from the data shown in Figure 6 of Wing (1980).

Apparatus

The testing station consisted of a personal computer, a color monitor, and a response board. The computer was used to control the presentation of the stimuli and to collect all responses. The sounds for both the perception task and the pacing signals for the production task were generated by the internal speaker of the computer. Responses on the perception task were typed on the computer keyboard. Responses on the production task were made on the response board. This board was constructed of wood and measured 20 × 30 cm. A Plexiglas lever, 14 cm long, was mounted over a microswitch. Flexion of approximately 1 cm was required to activate the microswitch, and there was minimal resistance from the switch.

Tasks

Production task. The production task involved the repetitive tapping paradigm introduced by Wing and Kristofferson (1973; see also Ivry et al., 1988). On each trial, the participant was presented with a series of 100-Hz tones. Each tone was 50 ms in duration, and the stimulus onset asynchrony (SOA) was fixed at the target interval. In separate blocks, the target interval was set to 325, 400, 475, and 550 ms. The participants were instructed to begin tapping with the tones once they had identified the beat. Once the first response was generated, the tones continued until the individual had produced 12 intervals (13 taps). At that point, the tones ended and the participant continued tapping until he or she had produced an additional 31 intervals (32 taps). The end of the trial was signaled by a 1-s tone. Feedback was displayed on the screen, indicating the participant's mean intertap interval and standard deviation for both the paced and unpaced phases of the trial. The interval separating the paced and unpaced phases was not included in these calculations nor in any other analyses. A production block consisted of six trials.

Perception task. The Parameter Estimation by Sequential Testing (PEST) procedure (Pentland, 1980) was used to estimate difference thresholds for target intervals of 325, 400, 475, and 550 ms. Application of this psychophysical procedure to time perception is described in detail in Keele et al. (1985; see also Ivry & Keele, 1989). On each trial, four 50-ms tones (1000 Hz) were presented, grouped into two pairs of two tones each. The SOA between the first pair of tones was set to the target interval and was referred to as the *standard interval.* After a 1-s pause, the second pair of tones was presented. The SOA between the second pair was variable and was referred to as the *comparison interval.* The participants' task was to judge whether the comparison interval was shorter or longer than the standard interval. One second after the response, the first tone of the standard interval was played.

The PEST procedure is adaptive in that the duration of the comparison interval for each trial is set to the current estimate of the participants' difference threshold (defined as 1 *SD* from the standard interval). Separate estimates are made for the threshold corresponding to the comparison interval at which the participant responds longer on approximately 25% of the trials and for the threshold corresponding to the comparison interval at which the participant responds longer on approximately 75% of the trials. On the basis of the participant's response, the assumed psychometric function is reestimated after each trial to provide the comparison interval for the next trial. This procedure is repeated for 50 trials (25 trials per threshold). The initial values for the shorter and longer thresholds are based on our previous research (Keele et al., 1985), although simulation studies show the algorithm to be relatively insensitive to starting values.

At the end of the block, feedback was displayed on the screen, indicating a standard deviation score and the point of subjective equality (PSE). The standard deviation score was calculated as the difference between the two thresholds divided by 2 (each threshold was set to be equal to 1 *SD* from the PSE). The PSE score was defined as the average of the two thresholds.

Procedure

On each session, the participants completed one production block and one perception block with each of the four target intervals (325, 400, 475, and 550 ms) for a total of eight blocks. The blocks were arranged in pairs consisting of production and perception with a given target interval. For a given interval, the production block always preceded the perception block. The order of the intervals was set so that, on the four different sessions, each interval would be tested in each of the four positions (first pair, second pair, third pair, and fourth pair). The exact orders were set by constructing different Latin squares for each participant.

Each production block was preceded by one practice trial consisting of 12 paced and 31 unpaced intervals. Each perception block began with four practice trials, two in which the comparison interval was shorter than the standard interval and two in which the comparison interval was longer than the standard interval. The comparison interval for the practice trials was set to be highly discriminable (>2 *SD* shorter or longer than the standard interval). The practice trials were continuous with the test trials in the perception task.

Results and Discussion

Slope Analysis

For each participant, variance scores were obtained on each day for the production and perception tasks at the four different target intervals. No significant changes were observed in tapping and perception scores over the four test sessions, $F(3, 21) = 1.00$, $p > .40$.

The mean variances across participants and sessions are plotted in Figure 1 on axes generated to correspond to Equation 3. For the tapping data, the values on the abscissa are the mean intertap intervals squared rather than the target intervals squared. Participants tended to tap at a slightly faster rate for each of the target intervals, but the effect was not large. For the perception data, the values on the abscissa correspond to the mean PSEs squared for each target interval. The mean PSEs were close to the target intervals, although there was a trend for the PSE to be slightly greater than the target interval for standards of 325, 400, and 475 ms.

As can be seen, variance increased as a function of target interval. This finding supports a fundamental premise of the slope analysis: Measures of temporal variability should become larger with increases in the base interval being timed. The manner in which variability increased was evaluated by performing linear regressions for each participant, one conforming to Equation 2 and one conforming to Equation 3. The results of these analyses are given in Table 1. Although both equations provide a good account of the linear trend in the data, more of the variance is accounted for by the generalized form of Weber's law (Equation 3).

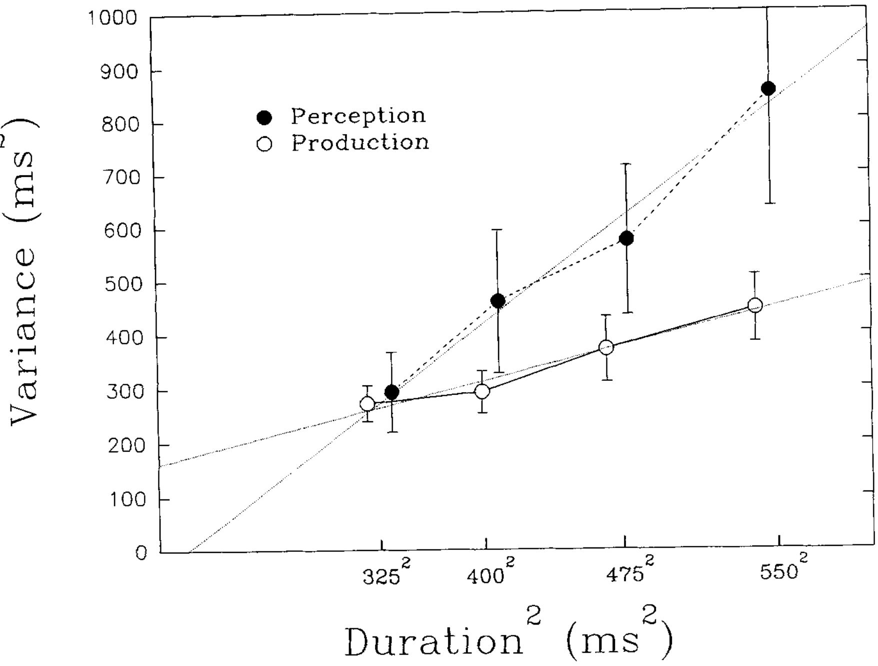

Figure 1. Mean variances on the perception and production tasks in Experiment 1, plotted as a function of duration squared. The plotted abscissa values for the production function are the square of the mean of the produced intertap intervals. For the perception task, the abscissa values are the square of the mean points of subjective equality. This transformation was chosen to provide a direct depiction of Equation 3. Error bars indicate one standard error of the mean.

This result is similar to that observed in previous perception experiments (e.g., Getty, 1975) and our reanalysis of the production data reported in Wing (1980). Note that the values listed in the table are the means calculated across individuals. Although this calculation indicates a reasonable linear component for both equations, the R^2 values increase considerably when calculated from functions based on aggregate data. When the analysis is calculated using the mean produced interval (Equation 2), the percentage of variance accounted for by a linear component is 93.4% for the production task and 97.9% for the perception task. When the mean produced interval is squared (Equation 3), these percentages rise to 96.7% and 98.7%, respectively.

Even stronger evidence that Equation 3 is a more accurate description of the data comes from a comparison of the estimates of the duration-independent variability component, the intercept values. Equation 2 yields a value of -5.01 ms^2 for the production task, a result suggesting that this second source of variance is negligible, a result at odds with previous findings (e.g., Ivry & Keele, 1989; Wing & Kristofferson, 1973). Even more problematic is the large negative intercept calculated from Equation 2 for the perception task. The intercept values from Equation 3 are more reasonable. The production intercept is positive and provides an estimate that converges with that obtained by the Wing–Kristofferson model (see below). The perception intercept is negative, although the value is considerably less negative than that obtained from Equation 2. We expect that the negative intercept for this function is due to measurement error, an inference supported by the large standard errors for the longer intervals. For each intercept value, we tested whether the observed value was significantly differ-

Table 1
Regression Statistics Derived From Equations 2 and 3 for the Data of Experiment 1

	Perception			Production		
Equation	Slope	Intercept	R^2	Slope	Intercept	R^2
Equation 2						
M	2.37	−523.57	0.63	0.81	−5.01	0.70
SE	0.80	266.01	0.13	0.19	52.88	0.12
Equation 3						
M	0.00266	−11.30	0.64	0.00096	160.61	0.72
SE	0.00090	109.95	0.12	0.00022	24.57	0.11

Note. *M* = mean calculated from the individual regressions; *SE* = standard error of the mean.

ent than zero. Assuming there are duration-independent sources of variance, these values should be greater than zero. This prediction was confirmed only for the production value from Equation 3, $t(7) = 6.17, p < .001$. The negative perception value from Equation 2 actually approached significance, $t(7) = 1.84, p < .12$.

Two analyses were conducted to evaluate whether the duration-dependent source of variability was the same in the time production and time perception tasks. First, a comparison was made of the Weber fractions calculated for the two tasks. The common timing hypothesis predicts that these fractions should be equal. Across the eight participants the mean Weber constants (square root of the slope calculated in Equation 3) for the production and perception tasks were 0.0287 and 0.0463, respectively. The difference between these two values is marginally significant, $t(7) = 2.14$, $p < .10$ (two-tailed). As can be seen in Figure 1, the perception function is generally steeper than that observed for the production task.

Second, a 2 (task) × 4 (duration) analysis of variance (ANOVA) was conducted on the estimated duration-dependent variances. With Equation 1, these estimates can be obtained by subtracting the duration-independent variability component from the total variance. These duration-dependent estimates were made for each participant by subtracting the intercept values from Equation 3 from the total variability scores obtained at each target interval. For two of the participants the intercept values from the perception regressions were negative; in these cases no subtractions were made and the total variability scores were used. Note that this second analysis is not independent of the comparison of the Weber fractions. However, the ANOVA is more stringent. The common timing hypothesis predicts that there will not be a significant interaction (i.e., there will be parallel functions) nor will there be a main effect of task.

As expected, there was a significant effect of duration, $F(3, 21) = 9.28, p < .001$, indicating that the duration-dependent estimate increased for the longer target intervals. Contrary to the prediction of the common timing hypothesis, there was a marginally significant effect of task, $F(1, 7) = 4.40, p < .10$, and a significant interaction, $F(3, 21) = 3.24, p < .05$. The duration-dependent estimates for the perception task were generally greater than the estimates for the production task, and the difference increased for the longer target intervals.

In summary, the comparison of the Weber fractions as well as the ANOVAs failed to support the hypothesis that the production and perception tasks use the same timing mechanism.

Analysis of the Production Data Using the Wing–Kristofferson Model

The production data can also be analyzed using the Wing–Kristofferson model. This model provides a means for estimating peripheral (duration-independent) sources of variability and then, by means of subtraction, attributes the remainder to duration-dependent sources. A complete discussion of the model can be found elsewhere (Ivry et al., 1988; Wing, 1980).[2] The production data can be analyzed to provide a new test of the Wing–Kristofferson model's prediction that the peripheral source of variability will be invariant across different target intervals and to compare the estimated component sources of variability from the Wing–Kristofferson model and the slope analysis.

Decomposing the total variability into separable estimates of central and implementation sources provided further support for the usefulness of the Wing–Kristofferson model (Figure 2). In accordance with the findings of Wing (1980), the implementation estimate was found to be relatively independent of the base interval, whereas the estimate of the central component was directly related to the base interval.

Across the four target intervals, the magnitude of the implementation variance from the Wing–Kristofferson model is 49.0 ms^2. This value can be compared with the duration-independent variance estimate of 160.61 ms^2 obtained with Equation 3 in the slope analysis. Logically, the duration-independent component from the slope analysis must be larger than the implementation estimate from the Wing–Kristofferson model. This is because the former must include all of the variance attributed to the implementation system plus any additional central sources of variance that are duration independent. This analysis provides additional support for Equation 3. Note that this comparison cannot be made with Equation 2 because the slope analysis here yielded a negative intercept.

Experiment 2

The results of Experiment 1 are problematic, although they do support the basic logic of the slope analysis method. In both the production and perception tasks, the data are in accord with a generalized form of Weber's law and therefore allow a test of whether a common timing process is involved in both tasks. However, the predicted similarity of the estimated duration-dependent functions for the two tasks was not evident. The duration-dependent estimates for the perception task were generally larger than the duration-dependent estimates for the production task.

The lack of correspondence between the perception and

[2] In brief, the model assumes that the central timing and implementation processes can be represented as independent random variables. Each interval is composed of one sample from the central timing distribution and two samples from the implementation distribution (corresponding to the taps marking the beginning and end of an interval). However, because successive intervals share a boundary, the implementation source of variability will produce a negative correlation between these intervals. For example, if the implementation process is delayed for a given tap, the interval ending with this tap will be lengthened and the following interval will be shortened. Thus, an estimate of the duration-independent source of variability is obtained from the covariance between successive intervals. The data from the current experiment are in accord with the basic predictions of the Wing–Kristofferson model. In 122 of the 128 (95.3%) production blocks, there was a negative correlation between neighboring intervals. No systematic patterns were observed for higher lags.

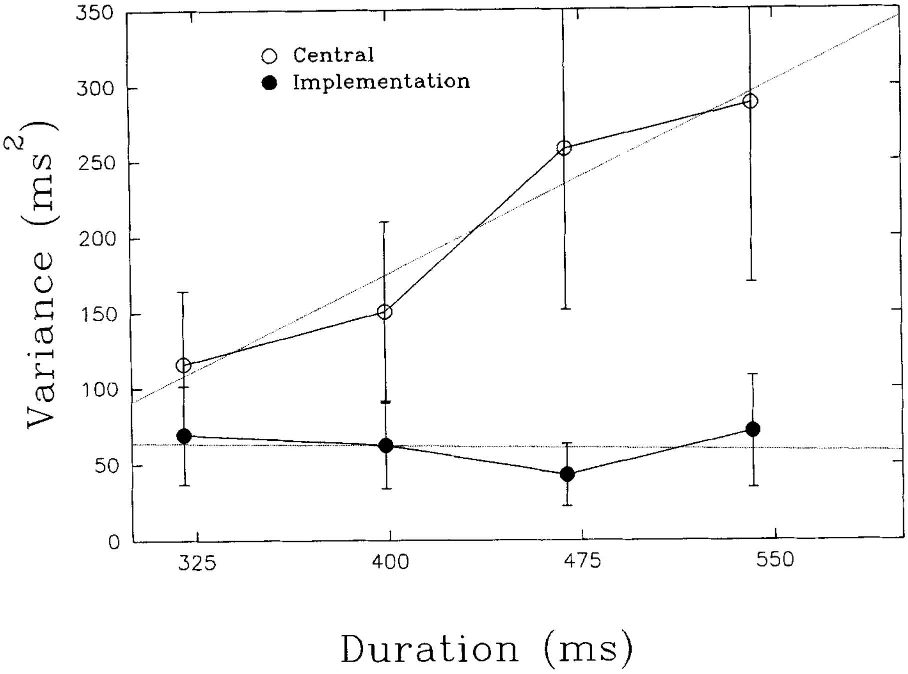

Figure 2. Mean estimates of the central and implementation sources of variance derived from the Wing–Kristofferson model.

production functions may suggest that these two tasks require the operation of different timing mechanisms. This conclusion would be at odds with our previous findings showing high correlations between these two tasks in healthy participants (Keele et al., 1985) as well as impaired performance on both tasks in patients with cerebellar lesions (Ivry & Keele, 1989).

Alternatively, the lack of correspondence in the duration-dependent estimates may be the result of differences in the temporal requirements for the two tasks. Specifically, the repetitive tapping task requires that the person generate a consecutive series of timed responses. In contrast, for the perception task, the participant is asked to judge the duration of isolated intervals. It is possible that the repetitive nature of timing in the production task might confer some advantage.

In Experiment 2, the production task was modified so that participants produced only isolated intervals. This modification precludes an analysis of the data using the Wing–Kristofferson model, but this analysis is not necessary for the slope analysis. However, by placing more similar constraints on the temporal component of each task, we should achieve a more reasonable comparison of the underlying mechanisms.

Method

Participants

The participants for Experiment 2 were the same eight students who had taken part in Experiment 1. Each was tested on four sessions and was paid $25. Four of the participants completed Experiment 1 prior to beginning Experiment 2; the order was reversed for the other four participants.

Apparatus

The same equipment was used as in Experiment 1.

Tasks

Production task. Unlike the repetitive tapping task of Experiment 1, the production task of Experiment 2 required the participants to produce isolated intervals. For the "paced" phase, the participant was presented with two tones, separated by the target interval. The word TAP would appear on the monitor 400, 600, or 800 ms after the offset of the second tone. The participant would then make two keypresses, attempting to create an interval that matched the target interval. Approximately 600 ms later, the computer would again present the target interval followed by the instructions to tap. This procedure was repeated 12 times per trial. After the participant produced his or her 12th interval, the computer would no longer present the target. Instead, the word TAP was displayed an additional 31 times, and the participant was instructed to produce two responses after each presentation, in an attempt to reproduce the target interval. The interval between the second tap of each interval and the presentation of the word TAP was randomly selected from three alternatives of 400, 600, or 800 ms.

After the participant had produced 12 paced and 31 unpaced intervals, a 1-s tone would indicate the end of the trial. Feedback was provided regarding the participant's mean intertap interval and standard deviation for both the paced and unpaced phases of the trial. The first unpaced interval was not included. As in Experiment 1, a production block consisted of six trials. Four production blocks were run on each session with target intervals of 325, 400, 475, and 550 ms.

In summary, the amount of data collected for each participant in Experiment 2 was identical to that in Experiment 1. Over the four sessions, each individual completed 24 trials at each target interval. The primary data of interest from the unpaced phase consisted of 30 intervals per trial. On each session, one practice trial was included for each target interval.

Perception task. The perception task was also modified in Experiment 2 in an attempt to make it more similar to the new production task. The PEST procedure was used, but without the presentation of a standard on each trial. That is, on each trial, the participant would hear a single pair of tones and have to judge if the interval between the pair was shorter or longer than an implicit standard. To help the participants establish an implicit standard, the first 10 trials of each block used intervals that were very discriminable (based on previous data, selected to be at least 2 *SD* from the target interval), and feedback was provided. For half of these trials, the interval was shorter than the implicit standard and for the other half, the interval was longer. After each response, feedback would indicate whether the correct response should have been shorter or longer. Following these training trials, 50 additional trials were run, with the duration of the comparison interval determined by the PEST procedure. As in Experiment 1, feedback was provided at the end of each block, indicating a standard deviation score and the PSE. There were four blocks of the perception task with target intervals of 325, 400, 475, and 550 ms.

Procedure

The procedure was identical to that used in Experiment 1. For each session, the two tasks with a given target interval were conducted successively with the production task always preceding the perception task. Note that the production task may also help the participants establish an implicit standard for the following perception task. The test order of the different target intervals within a session and between sessions was determined by a Latin square for each participant.

Results and Discussion

Unlike Experiment 1, a significant reduction in variability was found over the four sessions, $F(3, 21) = 8.85, p < .001$. This practice effect tended to be greater for the production task than for the perception task, $F(3, 21) = 3.05, p = .051$. Averaged over the four target intervals, the tapping standard deviation scores dropped from 25.4 to 19.7 ms, whereas the improvement on the perception task only slightly fell, from 23.7 to 22.3 ms.

For comparing production and perception, the data were averaged over the four sessions. The mean variances plotted across duration squared are presented in Figure 3 and the regression analyses for the two tasks are summarized in Table 2. As in Experiment 1, the regression analyses favor plotting variance against duration squared (Equation 3) over duration (Equation 2), indicating conformity to a generalized version of Weber's law. First, the R^2 values are higher when the abscissa corresponds to duration squared. (This also holds for aggregate functions: For perception, comparison is .99 vs. .98; for production, comparison is .98 vs. .94.) Second, when variance is plotted against duration (Equation 2), the intercepts are negative for both the production and perception functions, with the perception value being sig-

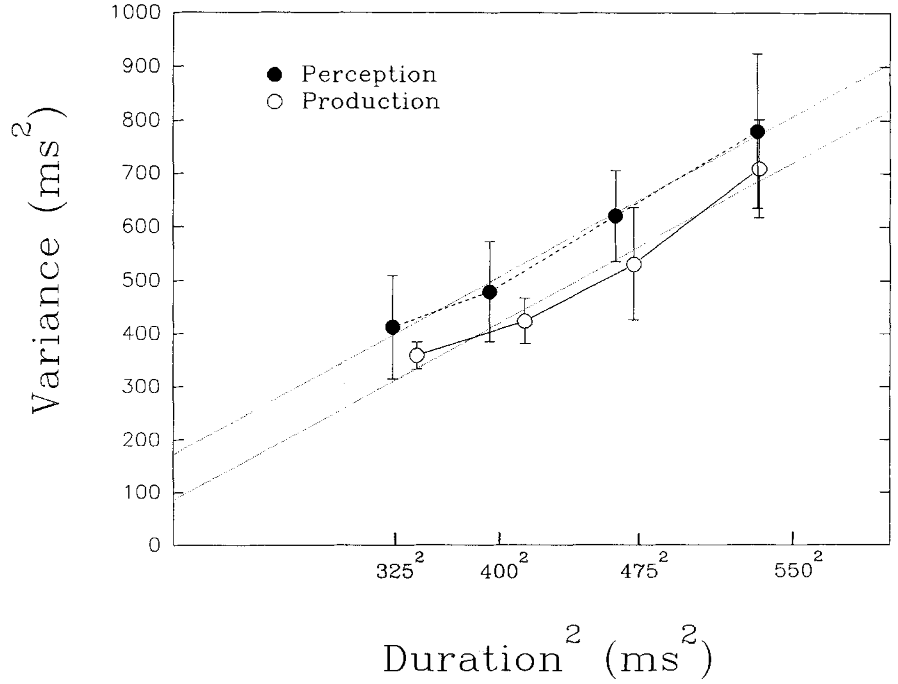

Figure 3. Mean variances on the perception and production tasks in Experiment 2, plotted as a function of duration squared. Abscissa values are the same as those in Figure 1.

Table 2
Regression Statistics for Experiment 2

Equation	Perception			Production		
	Slope	Intercept	R^2	Slope	Intercept	R^2
Equation 2						
M	1.82	−209.56	0.78	2.04	−394.30	0.66
SE	0.22	50.83	0.06	0.57	204.56	0.08
Equation 3						
M	0.00214	167.62	0.79	0.00237	34.96	0.69
SE	0.00027	59.25	0.06	0.00066	87.70	0.08

Note. M = mean calculated from the individual regressions; SE = standard error of the mean.

nificantly less than zero, $t(7) = 3.86$, $p < .01$. Again, negative intercepts are not meaningful in this situation. In contrast, the intercepts for both the production and perception tasks were positive when calculated from Equation 3. However, this result was only significant for the perception task, $t(7) = 2.65$, $p < .05$. By the logic of the slope analysis, these intercepts provide an estimate of the duration-independent sources of variability associated with each task.

Estimates of the duration-dependent component of the total variability were then computed as in the preceding experiment. In this experiment, negative intercepts were found for three participants on the production task and for two participants on the perception task. For these cases, total variability scores were used. A 2 (task) × 4 (duration) ANOVA of these estimates revealed a significant effect of duration, $F(3, 21) = 28.71$, $p < .001$. Unlike Experiment 1, there was no indication of a main effect of task, $F(1, 7) < 1.0$, nor of an interaction, $F(3, 21) < 1.0$. There was also no difference in the Weber constants (k in Equation 3) derived from the production (0.0457) and perception (0.0455) functions for each individual, $t(7) = 0.03$. In conclusion, the similarity of the two functions supports the hypothesis that a common timing mechanism is used in the production and perception tasks.

Given that the same individuals participated in both Experiments 1 and 2, it is possible to make some comparisons of the variability functions across the experiments. First, the perception functions are quite similar (compare Figures 1 and 3). The Weber constants in Experiments 1 and 2 were 0.0463 and 0.0455, respectively. This result might appear surprising given that a standard interval was presented on every trial in Experiment 1 and suggests that, at least with practiced participants, an explicit standard is superfluous. Second, the biggest difference between the two experiments is in the production data. The duration-dependent estimates were considerably larger in Experiment 2. Whereas the Weber constant was 0.0287 in Experiment 1, the comparable value in Experiment 2 was 0.0457, $t(7) = 3.69$, $p < .01$. From this, it can be inferred that the repetitive tapping task provides a reduction in duration-dependent variability in comparison to tapping isolated intervals.

Experiment 3

Experiments 1 and 2 differ primarily in terms of the tasks used to measure temporal variability in the production tasks. To account for the lower variability scores in Experiment 1, it is assumed that the repetitive nature of the tapping task confers an advantage, perhaps by introducing some stability into the timing mechanism. One reason why repetition might facilitate performance in the repetitive tapping task is that the pacing signal is sounded repeatedly, producing a series of at least 12 isochronous intervals. This source of repetition might lead to a more accurate representation of the target interval (Keele, Nicoletti, Ivry, & Pokorny, 1989; Schulze, 1978, 1989).

To explore this hypothesis, the production and perception tasks were modified in Experiment 3 to involve repetition of the target interval. In the production task, a modified version of the repetitive tapping task was used in which participants tapped with a pacing signal and then produced a short series of five unpaced intervals. In the perception task, a series of six tones was presented for each trial. The first five tones formed four isochronous intervals of the standard duration. The sixth tone completed an interval that was either shorter or longer than the standard interval. Thus, in both tasks, the target interval is presented multiple times on each trial. It is expected that this repetition will produce a more veridical representation of the target interval. In turn, this representation should lead to better performance on both the production and perception tasks. If this improvement is comparable across the two tasks, then the duration-dependent variability estimates should be similar for the two tasks.

Method

Participants

Seven individuals were tested in Experiment 3. Six were recruited from the student population at the University of California, Berkeley, and each was paid $25. The seventh participant was R. B. Ivry. Each participant completed four sessions. Four of the participants had participated in Experiments 1 and 2, and R. B. Ivry has extensive practice on these tasks. The remaining two recruits did not have experience in time production or perception experiments.

Apparatus

The same equipment was used as in Experiment 1.

Tasks

Production task. A modified repetitive tapping task was used in Experiment 3 that was intended to be a hybrid of the production tasks used in Experiments 1 and 2. Each bout began with the successive presentation of 50-ms computer sounds separated by the target interval. Participants were asked to tap in synchrony with the pacing signals. After the first tap, the computer generated four more tones, allowing the listener to produce five paced intervals. Then the tones ended, and the participant was required to produce seven more taps or six unpaced intervals. A 1-s tone signaled the end of the bout. This procedure was repeated six times per trial. The first unpaced interval was not included in the analysis. Thus, as in Experiments 1 and 2, a total of 30 unpaced intervals were recorded per trial and the variability was calculated across these 30 intervals. Feedback was provided at the end of the trial, indicating the mean intertap interval and standard deviation for the paced and unpaced phases.

Each block consisted of six trials. Production blocks with target intervals of 325, 400, 475, and 550 ms were tested on each session. A single practice trial was provided prior to each block. Thus, the primary data set for the unpaced phase was identical to that collected in the previous experiments. Note that the participants produced 30 paced intervals in Experiment 3 (5 intervals per bout × 6 bouts), a number that is larger than the 12 paced intervals of the previous experiments.

Perception task. The perception task was modified to provide multiple presentations of the target interval. On each trial, the participants heard a series of six tones. The first five tones created four isochronous intervals of the target duration. The sixth tone created a fifth interval that was either shorter or longer than the target interval. Each block was composed of 4 practice trials and 50 test trials. The duration of the comparison interval for the practice trials was set to values that were accurately judged by all of the participants. The duration of the comparison interval for the test trials was set according to the PEST procedure. Feedback was provided at the end of each block, indicating the difference thresholds required to report when the comparison interval was shorter or longer than the standard interval. There were four blocks of the perception task with target intervals of 325, 400, 475, and 550 ms.

Procedure

The procedure was identical to that used in Experiments 1 and 2.

Results and Discussion

The data for one participant were excluded from the analyses. Her responses on the perception tasks showed an overwhelming tendency to report that the comparison interval was longer than the standard. Because of this, her variability scores were underestimated, given that the test values selected by the PEST procedure reached the end of the range of test values. Although her tapping scores were comparable to those obtained with the others, all of her data were eliminated, because the critical comparison is between tasks.

Variance scores were obtained on each task at the four target intervals. The effect of session approached significance, $F(3, 15) = 2.57, p < .10$, with performance improving over sessions. However, this factor did not interact with task or the target interval (both F values < 1.0). A composite score for each of the six participants was computed across the four sessions. The means are presented in Figure 4 and the regression analyses are summarized in Table 3. The regression analyses do not clearly favor one equation over the other. Unlike the previous experiments, the percentage of variance accounted for by a linear component is slightly higher for Equation 2 on the perception task and is equal for both equations on the production task. (When calculated from aggregate data, the R^2 values are all greater than .92.) The intercept for the perception task is again negative (-70.33) when calculated from Equation 2. In contrast, the intercepts are both positive when the calculations are based on Equation 3, with the value being significantly greater than zero for the production task, $t(5) = 3.61, p < .02$. These intercepts provide an estimate of the duration-independent sources of variability for each task.[3]

Estimates of the duration-dependent components of the total variability scores were computed. For one participant on the perception task, the intercept from the regression analysis was negative, and thus, the total variability scores were used. A 2 (task) × 4 (duration) ANOVA of these estimates revealed a significant effect of duration, $F(3, 15) = 31.92, p < .001$. There was no effect of task, $F(1, 5) < 1.0$, nor was there any suggestion of an interaction, $F(3, 15) < 1.0$. The slope values for the two tasks were quite similar, yielding Weber constants of 0.0217 and 0.0213 for the tapping and perception tasks, respectively, $t(5) = 0.08$.

It is possible to make only tentative comparisons between the results of Experiment 3 and those obtained in Experiments 1 and 2. There is only a partial overlap of participants, and all of these participants completed Experiments 1 and 2 prior to beginning Experiment 3. Nonetheless, there were three participants who did complete all of the experiments, and the mean Weber constants for these participants are listed in Table 4. The tapping values are similar in Experiments 1 and 3, the two experiments that involved variants of the repetitive tapping task. In contrast, this value is larger in Experiment 2, in which the participants produced isolated intervals. On the perception tasks, the Weber constant is lowest in Experiment 3, in which multiple presentations of the standard were given.

In summary, Experiment 3 provides converging evidence for the hypothesis that a common timing mechanism is used in the production and perception tasks. As in Experiment 2, functions of the estimated duration-dependent variability are similar for the two tasks. Moreover, it appears that performance on both the tapping and perception tasks improves when participants are provided with multiple samples of the target interval.

[3] As in Experiment 2, the Wing–Kristofferson model cannot be applied because the 30 unpaced intervals of each trial are not produced continuously. However, the model can be applied to the five-interval runs of unpaced taps within each bout. The mean motor delay variance estimate is 88.4 ms^2. This value is comparable to the estimated duration-independent variance of 102 ms^2 obtained from the slope analysis.

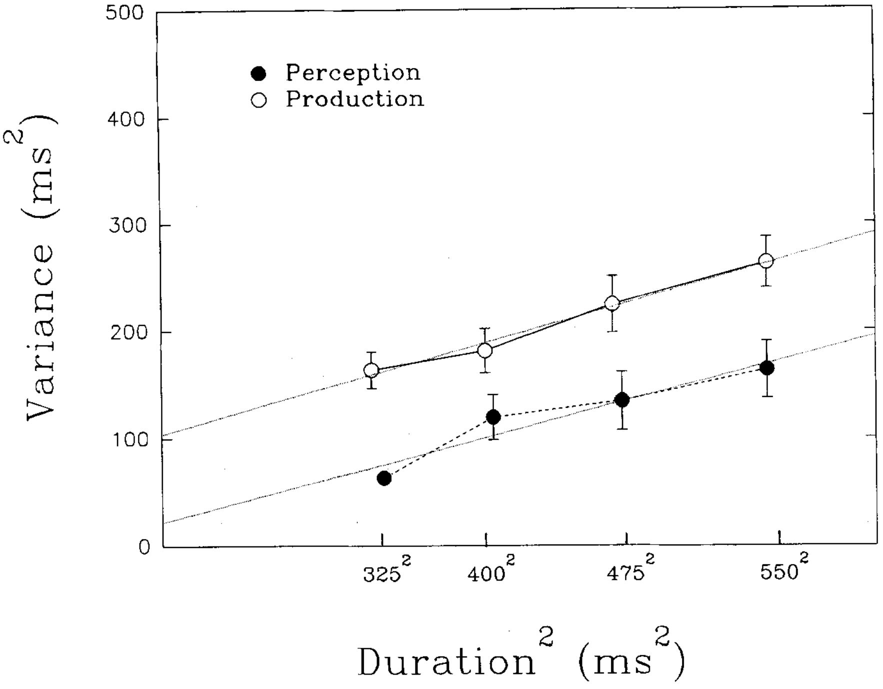

Figure 4. Mean variances on the perception and production tasks in Experiment 3, plotted as a function of duration squared. Abscissa values are the same as those in Figure 1. Note the difference in scale from Figures 1 and 3.

Experiment 4

In the preceding experiments, variability on the time production and perception tasks was lower when the participants were provided with multiple presentations of the standard. Presumably, this enabled the participants to form a more accurate representation of the target interval. This hypothesis was proposed on a post hoc basis to account for the smaller Weber constants on the production task in Experiment 1 and was used to generate a test of the common timing hypothesis in Experiment 3. However, differences in methodologies, participants, and experience preclude a direct assessment of this hypothesis from the preceding experiments.

Experiment 4 was designed to test directly the hypothesis that multiple presentations of the standard interval lead to a reduction in temporal variability. The standard interval was presented either once or four times on each bout. In the tapping task, the participants were then required to reproduce this interval; in the perception task, they judged whether a comparison interval was shorter or longer than the standard interval(s).

The current experiment was also used to evaluate a second factor that was confounded with the number of presentations of the standard in the previous studies. Specifically, in the preceding experiments, for conditions involving multiple presentations of the standard, the participants either produced a continuous series of intervals or judged a comparison interval that was continuous with the standard. For conditions involving the presentation of a single standard, the participants were required to produce an isolated interval or judge a comparison interval that was discontinuous

Table 3
Regression Statistics for Experiment 3

Equation	Perception Slope	Perception Intercept	Perception R^2	Production Slope	Production Intercept	Production R^2
Equation 2						
M	0.43	−70.33	0.75	0.46	6.26	0.79
SE	0.09	31.49	0.07	0.13	55.59	0.14
Equation 3						
M	0.00048	22.87	0.72	0.00054	102.38	0.79
SE	0.00011	13.29	0.08	0.00025	28.37	0.13

Note. *M* = mean calculated from the individual regressions; *SE* = standard error of the mean.

Table 4
Summary of Weber Fractions for Experiments 1, 2, and 3

Experiment and participants	Task: Production	Task: Perception
Experiment 1		
M for all participants	.0287	.0463
S.C.H.	.0082	.0505
G.R.I.	.3604	.0292
L.I.N.	.0130	.0134
Experiment 2		
M for all participants	.0457	.0455
S.C.H.	.0316	.0484
G.R.I.	.0394	.0431
L.I.N.	.0278	.0303
Experiment 3		
M for all participants	.0217	.0213
S.C.H.	.0107	.0277
G.R.I.	.0292	.0190
L.I.N.	.0110	.0223

with the standard. It might be argued that the continuous mode led to improved performance, perhaps because an internal beat was established by the standard intervals and was used as a reference in either producing or perceiving the target intervals. Also, there is less opportunity for the representation to decay in memory. To unconfound the number of repetitions of the standard and the mode of continuity, we also manipulated whether the standard and comparison intervals were continuous or discontinuous in Experiment 4.

The goal of this experiment was to evaluate factors that may affect temporal acuity. In doing so, the experiment did not require the slope analysis. Instead, the participants were tested on each of the tasks with a single duration of the standard interval, 500 ms. There were a total of eight conditions, created by combining two tasks (production and perception) with two levels of the number of presentations of the standard interval (one vs. four) with two levels of the mode of presenting the comparison interval with respect to the standard (continuous vs. discontinuous).

Method

Participants

Six undergraduate students at the University of California, Berkeley, were recruited. The participants were paid approximately $7.50 per hour based on a wage of $5 per hour plus a performance-determined bonus. Each individual completed four 1-hr sessions over a 10-day period. All of the participants were naive concerning the purpose of the experiment, although one had participated in Experiment 3.

Apparatus

The same equipment was used as in Experiment 1.

Tasks

Two independent variables were manipulated to create four conditions for both the time perception and time production tasks. The first factor was the number of presentations of the standard interval. The standard either was presented once on each bout or was repeated four times in succession. The second factor involved a manipulation of the temporal relationship between the standard and comparison intervals. The tones marking the comparison interval either were continuous with the tones marking the standard interval or were separated by a 1-s interstimulus interval. Note that for the perception task, the discontinuous, single-standard condition is identical to that used in Experiment 1, whereas the continuous, multiple-standard condition is identical to that used in Experiment 3. The remaining two perception tasks had not been used in the previous experiments nor had any of the production tasks.

Production tasks. In each condition, individuals produced a single, unpaced interval on each bout. Each bout began with the presentation of a series of tones separated by 500 ms. In the single-standard condition, the standard interval was presented once (two tones). In the multiple-standard condition, the standard interval was presented four times (five tones). For the discontinuous condition, the word TAP was presented 900, 1,000, or 1,100 ms after the last tone. The participant would then make two taps on the response key, attempting to create an interval that matched the standard interval. If the first response occurred prior to the onset of the word TAP, a tone sounded indicating a premature response and these bouts were repeated. For the continuous, multiple-standard condition, participants were required to begin tapping with either the second or third tone creating the standard intervals in the series. The participants were instructed to continue tapping when the tones ended to produce a single, unpaced interval. If the participant did not start tapping by the fourth tone or if he or she did not make two taps after the offset of the tones, the bout was repeated.

It was not possible to create a true continuous, single-standard condition, because the listeners could not synchronize with the first tone. Instead, they were instructed that after hearing a pair of tones separated by 500 ms, they were to attempt to "echo" the standard interval by making two taps. Given the echo instructions, we have observed that participants tend to make their first tap at approximately the time at which a tone would have occurred if the series of standard intervals had been continued. By using these instructions, we sought to make the response interval approximately continuous with the standard interval. Bouts were repeated when their first response did not begin within 800 ms of the offset of the second tone forming the standard. However, participants were not informed about this criterion, because we wanted to minimize changes in task requirements (e.g., we did not want to make this a reaction time task), and few violations of this criterion were observed.

Each trial consisted of 20 bouts, and the primary dependent variables were the mean and standard deviation calculated over the 20 bouts. These scores were provided as feedback at the end of each trial. There were five trials per day for each condition. Thus, participants produced 100 intervals per condition per day.

Perception tasks. In each condition, the participant's task was to judge if a comparison interval was shorter or longer than a standard interval. The PEST procedure was used to determine the duration of the comparison interval on each trial. The standard interval was presented either once (two tones) or four times (five tones). For the discontinuous condition, the comparison interval was presented 900, 1,000, or 1,100 ms after the offset of the last tone in the standard series. For the continuous condition, the comparison interval was continuous with the standard(s). Note that in these conditions, the second to the last tone marks the end of the last standard interval and the beginning of the comparison interval.

Five practice trials were provided at the beginning of each condition. The discrimination was set to be relatively easy for the first four practice trials. On the fifth practice trial, the comparison interval was equal to the standard interval. The test phase consisted of 54 trials per condition with half of the trials used to estimate each threshold. At the end of the block, feedback was provided concerning the PSE and the estimated standard deviation.

Procedure

The procedure was similar to that used in the previous experiments. Participants alternated between tapping and perception tasks, with each pair being drawn from the same condition. For example, after completing discontinuous, single-standard tapping, the analogous perception condition would be tested. Over the four days, the order of tasks was shuffled so that each pair occurred in all four positions within a session. Unlike the preceding experiments in which tapping always preceded perception, the order of the two tasks was counterbalanced.

Prior to the start of the experiment, participants were informed about the criterion used to calculate bonus payments. The amount of the bonus was inversely related to the standard deviation scores. In the tapping task, bonus payments were made only when the mean produced interval was within 30 ms of the standard interval.

Results and Discussion

The data for one participant were not included in the following analyses. This individual reported having great difficulty with the required delay in the discontinuous, multiple-standard tapping task. Indeed, his standard deviation scores for three of the days in this condition were more than 3 *SD* from the mean for the other five participants, and the score for his fourth day was only slightly better. In contrast, his data for the other three conditions were comparable to that produced by the rest of the participants.

Because temporal variability is dependent on the base duration, it is important to verify that the means of the produced intervals and PSEs are comparable across the different conditions. These data were compared in a four-factor ANOVA (Task × Number of Standards × Mode of Continuity × Blocks). None of the factors produced significant effects nor were any of the interactions significant. In the tapping tasks, the mean produced interval over all individuals was 489 ms, an 11-ms reduction of the standard interval. The mean PSE over all of the perception conditions was 486 ms.

Given these results, a comparison can be made of the variability scores. Figure 5 summarizes the data. The top panel shows the results for the production tasks and the bottom panel shows the results for the perception tasks. As before, a four-factor ANOVA was used to evaluate the data with mean variability scores used for the dependent variable. Of primary interest was the finding that participants were less variable following the presentation of four standard intervals in comparison to one standard interval, $F(1, 4) = 20.23$, $p < .02$.[4] There was no difference in performance between conditions in which the produced or comparison interval was continuous with the standard interval or was discontinuous with the standard interval, $F(1, 4) = 2.72$, $p > .15$. The factor of task approached significance, $F(1, 4) = 5.99$, $p < .10$, reflecting the fact that participants tended to have lower variances on the perception tasks than on the production tasks (17.9 vs. 24.0 ms). In addition, there was a reliable effect of block, $F(3, 12) = 8.19$, $p < .01$, suggesting that participants' performance improved over the course of the four sessions (mean standard deviations averaged over the production and perception tasks for Blocks 1–4, respectively, were 23.5, 22.1, 19.5, and 18.8 ms). However, none of the two-way or higher order interactions approached significance.

In summary, the results of this experiment indicate that temporal acuity is improved when participants are presented with multiple repetitions of the standard interval. Although the multiple standard advantage was larger for the perception condition than for the production condition (4.9 vs. 3.0 ms), this factor did not interact with any of the other variables. Producing or perceiving a target interval that is continuous with the standards did not produce a reliable reduction in variability in comparison to conditions in which the target interval was discontinuous with the standards. Thus, although the number of standards and mode of continuity were confounded in the previous experiments, the current results emphasize the importance of the first factor.

General Discussion

The experiments reported in this article were designed to answer two questions. First, we were interested in comparing different formulations of how temporal variability increases with duration. Second, we sought converging evidence that time production and time perception tasks use a common timing mechanism.

Temporal Variability as a Function of Duration

Previous research had clearly established that temporal variability is greater for longer intervals than for shorter intervals. However, the exact form of this relationship has been the subject of extensive debate, especially given the fact that divergent predictions can be derived from different models of timing (e.g., see Killeen & Weiss, 1987). As shown in the introduction, simple regression analyses that focus on the percentage of variance accounted for by a linear component may fail to prove analytic. Further insight may be provided by examining the second component of a regression analysis, the intercept. This fact has been overlooked in previous studies. For example, although Wing (1980) found a strong linear relationship between temporal variability and duration, the best fitting function also had a large negative intercept when the data were plotted as a function of duration. A reanalysis of Wing's data

[4] If the sixth participant is included in the analysis, the multiple repetition of the standard interval leads to reduced variability in the perception task only.

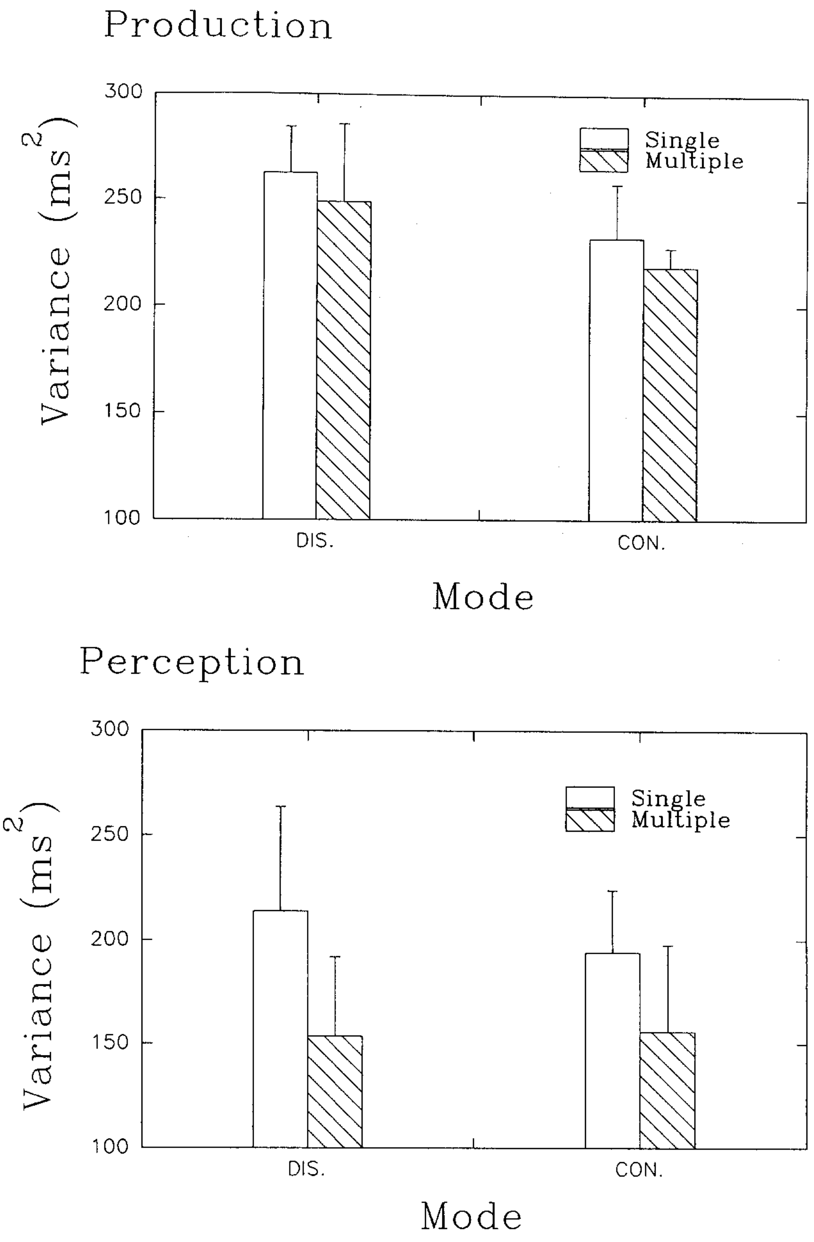

Figure 5. Mean variances for the eight conditions in Experiment 4. Top panel presents the data from the production task; bottom panel presents the data from the perception task. DIS. = discontinuous; CON. = continuous.

revealed an even better linear fit when variability was plotted as a function of duration squared, a finding that corresponds to Weber's law. Moreover, the intercept values from the transformed function were all positive and in a range that converged with alternative estimates of duration-independent variability.

Experiments 1–3 provide further support for the hypothesis that temporal variability conforms to Weber's law and emphasize the importance of evaluating the intercept terms when comparing models of timing. For five of the six functions (3 experiments × 2 tasks), the best linear fit was obtained when the data were evaluated with the equation based on Weber's law (Equation 3). More important, whereas negative intercepts were generally observed when variability was plotted against duration, the intercepts were positive for five of the six functions when duration squared was plotted on the abscissa. The consistency of these results argues strongly against models that predict the relationship described by Equation 2. Thus, simple clock-counter models in which the subintervals are generated by a Poisson process (Killeen & Weiss, 1987, Case 1) are not tenable given the current results.

Do Production and Perception Tasks Use a Common Timing Mechanism?

The slope analysis was developed to provide a novel test of whether performance on time production and time perception tasks reflects the operation of a common timing mechanism. The basic premise of the analysis is that the total variability on these tasks will reflect both variability in the timing mechanism and variability resulting from duration-independent processes. Thus, by varying the base duration, the growth of the variability function can be compared across tasks and is predicted to be equivalent for tasks using a common timing mechanism. The slope analysis thus provides a new method for comparing temporal acuity on perception and production tasks because duration-independent sources of variability can be factored out of each task.

The basic prediction of the common timing mechanism was not supported in Experiment 1. The slope for the production task was found to be smaller than that obtained for the perception task. Although this result might indicate that the two tasks access different timing mechanisms, an alternative hypothesis was that the demands on the timing mechanism were not equated for the two tasks. In particular, the standard interval was presented multiple times for each trial on the tapping (production) task but only once per trial on the perception task. In Experiments 2 and 3, the presentation of the target intervals was made to be approximately equal for the two tasks. Given this manipulation, the slope values were found to be remarkably similar for the two tasks. Indeed, the more stringent prediction that the time-dependent sources of variability will be equal was also confirmed (i.e., there was no main effect of task nor was there a significant Task × Duration interaction). Thus, these results suggest that time production and time perception use a common timing mechanism. These findings converge with correlational evidence obtained with normal (Keele et al., 1985) and neurologically impaired populations (Ivry & Keele, 1989).

It should be noted that the conclusions drawn from Experiments 2 and 3 regarding a common timing mechanism for production and perception are based on a null result. Although this is problematic, there are three aspects of the data that bolster these conclusions. First, the mean regression values for the production and perception tasks are remarkably similar (Figures 3 and 4). Second, providing a single standard for the production task was expected to make the two tasks more comparable in Experiment 2. Moreover, given the hypothesis that multiple standards reduce temporal variability, we expected an increase in the slope value for the production task for this experiment. Both of these predictions were confirmed. Third, as predicted in Experiment 3, providing multiple intervals produced a reduction in the slope values for both tasks. Thus, across experiments, we were able to manipulate the slopes for both tasks in a corresponding fashion.

Implications for Models of Timing Mechanisms

As discussed above, certain timing models can be rejected on the basis of the slope analysis. Moreover, the finding that temporal variability is reduced when study participants are presented with multiple repetitions of the target interval also has important implications for models of human timing. We hypothesize that repeating the target interval allows the participants to form a more accurate internal representation that will guide his or her productions or be used for comparison in perception tasks.

An alternative hypothesis is that repeating the target interval allows participants to entrain an internal oscillatory process to the target frequency. That is, the participant establishes an internal beat that can be maintained once the external stimulus is removed. Note that this hypothesis does not assume that the successive intervals are integrated to form a stable representation. Rather the successive standard intervals establish an oscillatory process (e.g., a limit cycle) that can form the basis for producing subsequent intervals or judging variations in the duration of comparison intervals.

Experiment 4 provided a comparison of beat-based timing and the hypothesis that repeating the standard led to the establishment of a stable standard interval. From the beat model, we would expect temporal variability to be smallest when the produced or judged intervals were continuous with the standards. In these conditions, the production task allows the participant to tap on the beat, and in the perception task, the comparison interval can be judged in reference to whether the final marker occurs on the expected beat. In contrast, if repeating the standard allows participants to form a stable representation of the target interval, then it should not matter if the comparison interval is continuous or discontinuous with an internal beat. The results supported the latter hypothesis. Temporal variability was lower when the standard was repeated four times, and this effect was found for both the continuous and discontinuous conditions. Thus, these experiments provide additional support for the hypothesis that timing is interval based rather than beat based.

These results are in accord with previous perception studies that have compared beat-based and interval-based models of human timing. Schulze (1978, 1989) found that temporal acuity varied with the number of repetitions of the standard interval. In his studies, the number of repetitions varied from two to six. For target durations ranging from 50 to 400 ms, there was a consistent reduction in the difference threshold as the number of repetitions of the standard interval was increased. Schulze (1978) had interpreted this effect as favoring beat-based models of timing over interval-based models. However, Keele et al. (1989) proposed a different interval-based model akin to that referred to above: Namely, that repeating the standard improves the fidelity of an internal reference interval. In two experiments, Keele et al. (1989) failed to find evidence that perceptual judgments of interval durations were improved when the comparison interval fell on beats established by a series of standard intervals. Instead, performance was best when there were multiple comparison intervals regardless of whether these comparison intervals became more discrepant from an internal beat or fell closer to the internal beats. Considered together with the present findings, these studies suggest that the timing mechanism used in these tasks does not operate as an internal oscillator, but rather that performance is dependent on the establishment of an internal reference interval.

References

Abel, S. (1972). Discrimination of temporal gaps. *Journal of the Acoustical Society of America, 52,* 519–524.

Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation, 22,* 39–58.

Church, R. M., Getty, D. J., & Lerner, N. D. (1976). Duration discrimination by rats. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 303–312.

Creelman, C. (1962). Human discrimination of auditory durations. *Journal of the Acoustical Society of America, 34,* 582–593.

Fetterman, J. G., & Killeen, P. R. (1992). Time discrimination in *Columba livia* and *Homo sapiens. Journal of Experimental Psychology: Animal Behavior Processes, 18,* 80–94.

Getty, D. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics, 18,* 1–8.

Grondin, S. (1992). Production of time intervals from segmented and nonsegmented inputs. *Perception & Psychophysics, 52,* 345–350.

Ivry, R. (1993). Cerebellar involvement in the explicit representation of temporal information. In P. Tallal, A. M. Galaburda, R. R. Llinas, & C. von Euler (Eds.), *Temporal information processing in the nervous system* (pp. 214–230). New York: Annals of the New York Academy of Sciences.

Ivry, R., & Corcos, D. (1993). Slicing the variability pie: Component analysis of coordination and motor dysfunction. In K. Newell & D. Corcos (Eds.), *Variability and motor control* (pp. 415–447). Champaign, IL: Human Kinetics.

Ivry, R., & Diener, H. C. (1991). Impaired velocity perception inpatients with lesions of the cerebellum. *Journal of Cognitive Neuroscience, 3,* 355–366.

Ivry, R., & Keele, S. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience, 1,* 136–152.

Ivry, R., Keele, S., & Diener, H. (1988). Dissociation of the lateral and medial cerebellum in movement timing and movement execution. *Experimental Brain Research, 73,* 167–180.

Keele, S., & Ivry, R. (1991). Does the cerebellum provide a common computation for diverse tasks: A timing hypothesis. In A. Diamond (Ed.), *Developmental and neural basis of higher cognitive function* (pp. 179–211). New York: Annals of the New York Academy of Sciences.

Keele, S., Nicoletti, R., Ivry, R., & Pokorny, R. (1989). Mechanisms of perceptual timing: Beat-based or interval-based judgments? *Psychological Research, 50,* 251–256.

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

Killeen, P., & Weiss, N. (1987). Optimal timing and the Weber function. *Psychological Review, 94,* 455–468.

Pentland, A. (1980). Maximum likelihood estimation: The best PEST. *Perception & Psychophysics, 28,* 377–379.

Schulze, H. H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research, 40,* 173–181.

Schulze, H. H. (1989). The perception of temporal deviations in isochronic patterns. *Perception & Psychophysics, 45,* 291–296.

Wearden, J. H. (1991). Human performance on an analogue of an interval bisection task. *Quarterly Journal of Experimental Psychology, 43B,* 58–81.

Wearden, J. H. (1992). Temporal generalization in humans. *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 134–144.

Wearden, J. H., & McShane, B. (1988). Interval production as an analogue of the peak procedure: Evidence for similarity of human and animal timing processes. *Quarterly Journal of Experimental Psychology, 40B,* 363–375.

Wing, A. (1980). The long and short of timing in response sequences. In G. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 469–484). Amsterdam: North-Holland.

Wing, A., & Kristofferson, A. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics, 14,* 5–12.

Received June 25, 1993
Revision received December 13, 1993
Accepted February 14, 1994 ■