In D. Deutsch (Ed.)
The Psychology of Music.
New York: Academic Press, 1982. Pp. 181-239.

7

Timing by Skilled Musicians

Saul Sternberg, Ronald L. Knoll, and Paul Zukofsky

I. Introduction: Perception, Production, and Imitation of Fractions of the Beat 182
A. Previous Studies of the Psychophysics of Time 183
B. Procedures and Notation 184
C. Subjects 185
D. Caveats 186
E. Principal Findings 186
II. Perceptual Judgment of Beat Fractions 187
A. Single-Fraction Perceptual Judgment (Experiment 1) 188
B. Multiple-Fraction Perceptual Judgment (Experiment 2) 194
C. Judgment Precision 196
III. Production of Beat Fractions 198
A. Use of Tap-Click Synchronization to Correct for Differential Subjective Delays 198
B. Expectations from a Feedback Model of Production 200
C. One-Response Production (Experiment 3) 200
D. Rejection of the Feedback Model 201
E. Repeated-Response Production (Experiment 4) 202
F. Production Precision: Evidence Against a Reaction-Time Explanation of the Production Error 203
G. Implications of Other Analyses of Psychophysical Scaling for the Production-Perception Disparity 205
IV. Imitation of Beat Fractions 207
A. Four Simple Alternatives for Imitation 207
B. One-Response Imitation (Experiment 5) 209
C. Choice Among Alternative Imitation Functions: Rejection of Accuracy and Full-Concatenation Models 209
D. Implicit Scaling of Beat Fractions from Imitation and Production Data 211

[1]The research described in this chapter was conducted in the Human Information-Processing Research Department of Bell Laboratories, Murray Hill, N.J. Much of it was described in a doctoral dissertation in psychology submitted by R. L. Knoll to Princeton University.

ISBN 0-12-213560-1

V. A Shared-Process Model of the Perception, Production, and Imitation of Beat Fractions .. 212
A. Definition of the Model 212
B. Restrictions on the Four Processes 214
VI. Further Analysis of Perceptual Judgment 215
A. Attention Shifts and Delays: Effect of Marker-Click Pitch (Experiment 6) 215
B. Time Marking by Onset versus Offset: Invariance of Judgment with Prolonged Markers (Experiment 7) 216
C. A Test of the Rate Constancy of Subjective Time Between Beats: Effect on Fraction Perception of Fraction Location Relative to the Beat (Experiment 8) 217
D. A Constraint on the Precision of Dual Time Judgments and Its Implications for Timing Models and the Use of Feedback 218
E. The Dependence of Perceptual Judgment on Duration versus Fraction: Effects of the Beat Interval (Experiment 9) 220
VII. Further Analysis of Production 224
A. Production Errors with Musical Instruments (Experiment 10) 224
B. Production with Marker-Click Feedback (Experiment 11) 225
C. Production of Single versus Multiple Subdivisions of the Beat (Experiment 12) 226
VIII. Summary 229
Glossary 231
Appendices 231
A. Staircase and Constant Stimulus Methods 231
B. Additional Details of Design and Procedure 233
C. Measures of Location of the Psychometric Function 234
D. Measures of Spread of the Psychometric Function 236
E. Stimulus-Averaging versus Response-Averaging Methods for Deriving a Psychophysical Scale 236
References 237

I. INTRODUCTION: PERCEPTION, PRODUCTION, AND IMITATION OF FRACTIONS OF THE BEAT

In this chapter we report 12 experiments that explore how skilled musicians perceive time and time performance in contexts similar to those in music. We discuss some of the perception and performance constraints we found and their implications for underlying timing mechanisms. Our emphasis is on the short time intervals—fractions of a second—that are among the shortest durations specified by musical notation. In our experiments, as in Western music, these intervals occur in the context of a train of periodic beats and are defined as fractions of the beat interval, or duration ratios. Because the beats are provided externally, our experimental tasks are probably most directly analogous to aspects of ensemble playing or of solo playing with a metronome or conductor. Musician subjects are among the best for answering questions about the relation between notation and performance, the constraints on the precision of ensemble playing, and temporal illusions in listening. We hope, however, that the constraints and mechanisms we uncover are relevant to human timing in general; our choice of musicians as subjects in timing experiments is based on our belief that a fruitful approach to the understanding of any human function is the study of skilled practitioners of that function.

We have examined the performance of skilled musicians in three laboratory tasks designed to capture aspects of musical practice: perception, production, and imitation of fractions of the beat interval. All three functions are required of musicians during ensemble rehearsal and performance, for example. It is plausible that because of the requirement that players "keep together," performing experience would cause the three functions to become at least consistent with each other and probably "correct" (consistent with the notation) as well. Neither of these expectations was borne out by our experiments; instead, we observed surprisingly large systematic errors and inconsistencies across these laboratory tasks. In agreement with our observations, the one study of the temporal coordination of ensemble playing of which we know (Rasch, 1979) reveals considerable inaccuracy even in unison attacks.[2] Our findings suggest that other studies of timing in actual musical performance would be of great interest, but unfortunately too little is known at present for us to comment on the extent to which our findings apply outside the laboratory.

The study of performance in our three different but related tasks, together with the analysis of relations among performances, permits some surprisingly strong inferences about these timing mechanisms. Some of our experiments may also be regarded as first steps in establishing a relation (a psychophysical function) between traditional Western musical notation—a notation that specifies time ratios—and the corresponding perceived and produced time ratios among people highly trained in use of the notation.

A. Previous Studies of the Psychophysics of Time

There is a vast, conflicting literature on the psychology of time that we shall only touch on here; we refer the interested reader to Fraisse (1963, 1978), Poppel (1978), Sternberg & Knoll (1973), Woodrow (1951), and Zelkind & Sprug (1974) for reviews and references.

For time intervals greater than a second, there is a long history of experiments focused on determining the psychophysical function relating subjective to actual duration. Most investigators have assumed that estimation and production are consistent, in the sense that they reflect the same psychophysical function. (We shall see this assumption fail dramatically in our experiments.) Investigators have agreed less well on how to account for reproduction (imitation) performance.[3] Based on his review of the data from over 100 studies of the reproduction, production, and estimation of

[2]The limitation in Rasch's analysis of trio playing to measurement of unison attacks probably implies that note sequences representing small fractions of a beat—where we found the largest errors and inconsistencies—were underrepresented.

[3]For example, according to Eisler's (1976) analysis, the subject in a reproduction task produces an interval whose subjective duration is half of the subjective duration of the sum of itself and the presented interval. This leads to a nonlinear relation between the presented and reproduced intervals. According to other models (e.g., Carlson & Feinberg, 1968), the subject produces an interval that is subjectively equal to the presented interval; this implies a linear reproduction function of unit slope, regardless of the underlying psychophysical function.

time intervals, Eisler (1976) concluded that subjective duration D is a power function of objective duration d with an exponent of about .9: $D = ad^{.9}$. (An exponent less than unity implies that the subjective ratio of two time intervals is smaller than the corresponding objective ratio; for example, 2 sec would appear less than twice as long as 1 sec.)

For intervals smaller than one second, the data are both more sparse and less consistent. Both Michon (1967) and Svenson (1973) report magnitude-estimation data supporting a change in the exponent of the power function at about .5 sec; whereas the exponent is approximately 1.0 for intervals greater than .5 sec, its value decreases to about .5 for smaller intervals. However, other investigators report either no such change in the function (e.g., Steiner, 1968) or changes at different points (Nakajima, Shimojo, & Sugita, 1980; see also Zwicker, 1969, and Fastl, 1977).[4] According to one view of human timing that has attracted interest, there exists a central timing process (or "clock") that functions similarly in the judgment, production, and reproduction of duration; results from experiments with time intervals greater than a second provide some evidence favoring such a common central process.[5] The discontinuities in judgment revealed in the experiments of Michon and Svenson suggest that a more sensitive test of a common central timing process might be obtained with time intervals less than a second; this suggestion provides one framework for the present investigation.

Despite the importance of time ratios in music, we know of no substantial studies in which judgments or productions of a range of *ratios* have been systematically examined.[6]

B. Procedures and Notation

We have examined performance in three basic tasks (illustrated on the left side of Fig. 1) and variations thereof. All three involve a train of beats specified by *beat clicks;* in most experiments the time from one beat click to the next (*beat interval b*) was 1.0 sec. In describing these tasks we denote stimuli by lower-case letters and responses by upper-case letters.

In the *perceptual judgment* task, one (or more) of the beat clicks was followed by a *marker click*, to form a *time-pattern stimulus* (see Fig. 1A). The time interval between beat and marker clicks (called the *fractional interval*) was to be judged in relation to the beat interval. (For example, if the *stimulus fraction* is $f = 1/8$ and the beat interval is $b = 1000$ msec, the fractional interval is $bf = 125$ msec.) In musical terms our fractions corresponded to note values between a 32nd note (𝅘𝅥𝅯) and a quarter note (♩), in which the quarter note equals one beat and the rate (in most experiments) was 60 beats

[4]Data from the three subjects in the study by Nakajima *et al.* (1980) are too inconsistent to justify firm conclusions.

[5]See Treisman (1963), .25–6 sec; Carlson & Feinberg (1968, 1970), 1–10 sec; and Adam, Castro, & Clark (1974), 3–40 sec.

[6]Except in one small study (Richards, 1964), the only ratios studied in production or perception tasks have been 1:2 (halving) and 2:1 (doubling). Povel (1981) has studied a range of ratios, but only in an imitation task.

per minute. The subject's response in this task was made in terms of *fraction names*, N. (For example, he or she might be asked to decide whether the stimulus fraction appeared to be less than or greater than N = 1/8th of a beat.) The fraction name was specified both in musical notation and as a numerical fraction; for example, the subject in this case would be asked whether the marker click was early or late relative to the pattern ♩ 𝅘𝅥𝅯♩.. . (The *correct* or *target* fractional interval for N = 1/8, given b = 1000 msec is, of course, bN = 125 msec.) The outcome of this procedure is the determination, for each of a set of fraction names, which stimulus fraction f corresponds to it. The relation between f and N defines a *judgment function*, $f = \mathbf{J}(N)$.[7]

In the *production* task a train of beat clicks was presented, but no marker click (see Fig. 1C). The "stimulus" here was a *fraction name* n. The subject made a *timed response* by tapping his or her finger after a beat click, with the aim of producing a fractional interval between click and tap whose duration was appropriate for the specified fraction name. The ratio of fractional interval to beat interval gives the *produced fraction* F; the fractional interval is bF. The outcome of this procedure is the determination, for each of a set of fraction names n the average value of the fraction F produced to correspond to it. The relation between F and n determines a *production function* $F = \mathbf{P}(n)$.

In the *imitation* task the stimulus was a *time-pattern stimulus*, as in the judgment task (defined by the beat interval b and the fractional interval bF) and the response was a timed response F, as in the production task (see Fig. 1E). The subject attempted to equate the produced fraction F to the stimulus fraction f. The outcome of this procedure is the determination—for each of a set of stimulus fractions f—the average value of the fraction F produced to correspond to it. The relation between F and f determines an *imitation function*, $F = \mathbf{I}(f)$.[8]

Note that the marker click and the timed response of our basic procedures were single offbeat events. Although not frequent in earlier music, the playing of a note after the beat without playing a note on the beat is not unusual in the music of the past 60 years.

C. Subjects

Our principal subjects were three professional musicians: Susan Bush, flutist (SB); Pamela Frame, 'cellist (PF); and Paul Zukofsky, violinist and conductor (PZ). PZ, who had substantially more musical experience than SB or PF, produced data that were more consistent, both within and across experiments. For this reason (and because results from other subjects usually agreed with his), we tend to weight his data more heavily. We obtained a small amount of corroborative data from Pierre Boulez (PB),

[7]Note that we have expressed the stimulus f as a function $\mathbf{J}$ of the response N for convenience in later discussion. We shall use the opposite convention for production and imitation.

[8]In some instances in which there is no ambiguity and the beat interval is 1 sec, we shall use the fraction symbols (n, N, f, and F) also to denote fractional intervals (bn, bN, bf, and bF). Note also that we do not distinguish notationally between quantities such as n, N, f, and F, and their means.

composer and conductor. In Experiment 12 we used two experienced amateur players (JM and SS) as well as PZ.

We carefully avoided informing subjects (including PZ and SS, coauthors) in any way about their performance until after the series of experiments in which they participated was complete, and deliberately provided no trial-to-trial feedback.

D. Caveats

Most of our 12 experiments made use of three subjects, a relatively small number, expecially given instances of inconsistency. We are confident of our major conclusions, however, especially because we found the same trends in more than one experiment. Nonetheless, we suggest caution in generalizing from our findings. We used only two subjects in Experiments 6 and 8, and only one subject in Experiments 7 and 11; results from these experiments should therefore be treated with special caution. Our most stable subject (PZ) served in all 12 experiments, permitting useful comparisons. Some of our findings are clear as well as surprising; nonetheless, they should be regarded as starting points to be confirmed and extended.

E. Principal Findings

In Experiments 1–5 we employed two variants each of the judgment and production tasks, and one variant of the imitation task. All three procedures resulted in proportional errors that are large (20–50%) for small fractions.

As described above, we define beat fractions (and values of N, n, F, and f) in terms of the interval between a tap or marker click and the *previous* beat click: we call these *forward fractions.* It should be noted, however, that a large forward fraction (such as 7/8) corresponds to a small *reverse fraction* (1/8) measured from tap or marker click to the *next* beat. Data from all three procedures hint at systematic errors associated with small reverse fractions that are qualitatively similar to the errors we observe for small (forward) fractions. In this report we emphasize performance for small values of n and f, however, because it is more reliable in all three tasks, and we have more data in that region.

Subjects tend to "overestimate" small fractions (Section II). The fraction names N associated with stimulus fractions f are too large: $f = \mathbf{J}(N) < N$. From this result one might expect that the fraction $F = \mathbf{P}(n)$ produced in response to a stimulus name n would be too small. Instead, produced fractions are too large: $F = \mathbf{P}(n) > n$ (Section III). For example, though a stimulus fraction had to be shorter than 1/8 to be called "1/8," when subjects tried to produce a fraction of "1/8" they produced an interval greater than 1/8.

This inconsistency between judgment and production performance for small fractions requires us to reject feedback models of production (Sections III,B and III,D), in which produced fractions are adjusted by judging them.

Imitation performance (Section IV) is very similar to production performance: for

small fractions, $F = \mathbf{I}(f) > f$. The existence of systematic errors in imitation argues against models in which the same transformations (psychophysical functions) relate stimulus fractions and produced fractions to their internal representations. Quantitative comparison of the imitation error to the judgment and production errors argues against a concatenation model of imitation, in which a fraction name produced by a covert judgment then serves as input to the production process.

Taken together, results from the three tasks suggest an information-flow model containing four processes (Section V) with an input process shared by judgment and imitation, and an output process shared by production and imitation. Nothing quantitative is assumed about the four processes, yet properties of the data permit some surprisingly strong inferences about them. The model is outlined in Fig. 7; readers may find it helpful to examine this figure before reading further.

In Section VI we explore and dismiss three potential sources of errors in judgment: the time to shift attention from beat to marker (Section VI,A), the possible importance of stimulus offsets ("releases") as well as onsets ("attacks") (Section VI,B), and the possibility that the rate at which subjective time elapses varies with location within the beat interval (Section VI,C).

We report evidence of a special difficulty associated with concurrent time judgments (Section VI,D), and by varying the beat interval we demonstrate that the judgment error can be described neither in terms of the fraction f alone or the fractional interval bf alone (Section VI,E).

In Section VII we explore and dismiss five potential sources of the errors in production: the use of finger-tap responses rather than notes played on musical instruments (Section VII,A), the absence of adequate response feedback (Section VII,B), the possibility of a distortion of subjective time near the beat (Section VII,C), and the use of single isolated responses that do not fill the beat interval and of off-beat responses not accompanied by any on-beat response (Section VII,C). We also note a tendency for errors in production to be accompanied by displacement (phase shift) of the subjective beat.

Details of experimental method and analysis are given in five appendices. We recommend that readers not interested in technical details omit these appendices, as well as the footnotes and Sections II,C, III,F, III,G, and IV,D.

II. PERCEPTUAL JUDGMENT OF BEAT FRACTIONS

Our principal aim in Experiments 1 and 2 was to determine the stimulus fractions f that were judged to be equivalent to various fraction names N; we would thereby have a psychophysical scale $f = \mathbf{J}(N)$ for fractions of a beat. (Note that in this chapter the term "scale" never denotes a musical scale.) A secondary aim was to measure the *precision* of expert judgments of beat fractions—the sensitivity of judgment probabilities to changes in f. We explored two different methods that permitted us to determine, for each of a set of fraction names, the stimulus fraction f that was subjectively equivalent to it; values of N ranged from 1/8 of a beat to 1 (a full beat). In

Experiment 1 (single-fraction judgment), a fraction name N was specified and the subject then judged, for each of a set of stimulus fractions, whether it was larger or smaller than N. In Experiment 2 (multiple-fraction judgment), the subject selected a response from a set of eight categories (such as "between 1/8 and 1/7 of a beat") whose boundaries were defined by fraction names. Our use of both single- and multiple-fraction procedures was motivated partly by a desire to assess the invariance over experimental methods of the systematic perceptual errors we discovered. Other differences between the procedures are discussed below.

A. Single-Fraction Perceptual Judgment (Experiment 1)

The stimulus patterns in Experiments 1, 3, and 5 are represented on the left side of Fig. 1; the pattern of beat clicks was held constant across these experiments to minimize stimulus differences among the three procedures. Two preliminary beat clicks were followed by a pause ("rest") of one beat (symbolized by a broken line) and then by two more beat clicks. (We used the pause so as to separate the stimulus and response components of each trial in the imitation task.) On each trial in Experiment 1 (Fig. 1A) the final beat click was followed by a marker click. Subjects judged whether the beat fraction appeared too large or too small relative to a specified fraction name N.[9]

Subjects judged fractions in relation to the fraction names 1/8, 1/6, 1/4, 1/2, 3/4, 5/6, 7/8, and 1. The name stayed the same for 75 consecutive trials as the stimulus fraction was varied by an "up-and-down" or "staircase" procedure. (The effect of the staircase procedure is to concentrate the stimuli close to the fraction that is judged to be neither too large nor too small relative to the specified name—i.e., subjectively equivalent to it. See Appendix A for more information about our use of this procedure.)

For each fraction name the resulting data permitted us to estimate the stimulus fraction f subjectively equal to it, which we call the "PMF mean". They also provided a measure of judgment variability discussed in Section II,C, which we call the "SD (standard deviation) of the PMF." Readers not interested in details of method need not understand how these estimates are determined. For each fraction name the method starts with the estimated psychometric function (PMF) provided by our data: a function, usually S-shaped, that associates with the value of each stimulus fraction the proportion of trials on which that fraction appeared "too large." The *location* of the PMF on the f-axis for a specified fraction name is, roughly, the stimulus value where judgment probabilities change most rapidly as the stimulus fraction is changed. This location separates two intervals on the f-axis: a "small-f" region where f tends to be judged too small relative to the name N and a "large-f" region where f tends to be judged too large. The location therefore corresponds to a fraction f that appears

[9]Subjects actually selected responses from six alternatives, representing three degrees of confidence for "larger" and three for "smaller." For the present report, however, we have pooled responses from each of the two sets of three to produce two response classes.

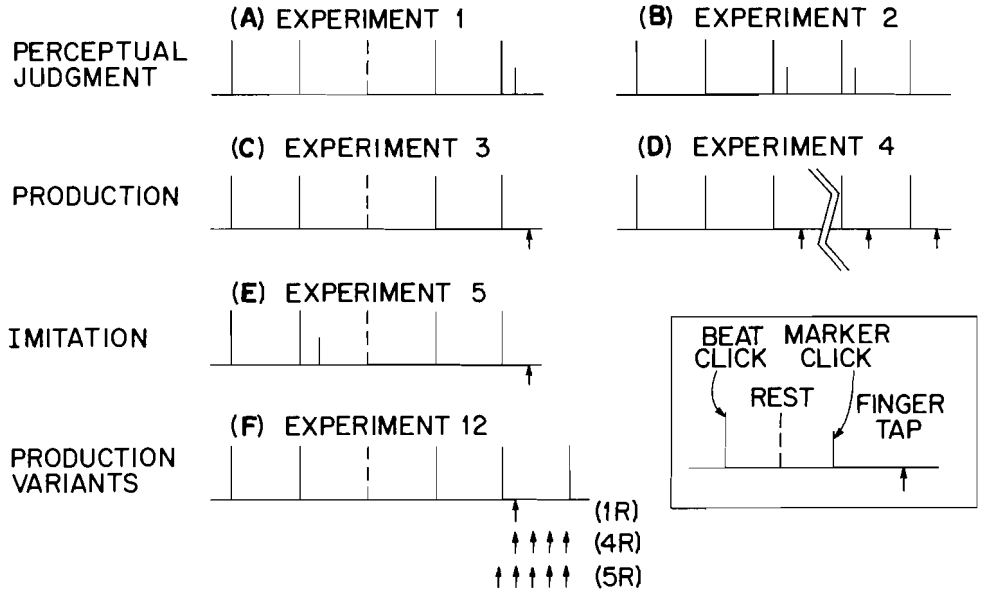

Fig. 1. Time-pattern stimuli and finger responses in six experiments. Beat and marker clicks were 5-msec tone bursts of 3000 and 2500 Hz, respectively, presented over headphones at approximately 30 dB above threshold. The tone bursts started at a zero-crossing and were gated by a voltage controlled amplifier so as to have a 1-msec rise time, a 3-msec steady level, and a 1-msec fall time. Judgment responses were made with a multiple response keyboard. Finger-tap responses were made by tapping on a plate and thereby completing an electric circuit. The interval between successive beat clicks was 1 sec in Experiments 1–5 and .5 sec in Experiment 12. Presented fractions were defined by the time interval between beat click and marker click. Produced fractions were defined by the time interval between beat click and tap, corrected for differential delays (see Section III,A). The train of events on each trial in Experiment 4 included ten finger-tap responses.

subjectively equal to N. A conventional measure of location is the estimated 50% point, or median, of the PMF (i.e., the f-value for which the judgments are equally divided between "smaller than N" and "larger than N"). Instead we report the estimated *means* of PMFs as location measures; our preference for the mean over the median—which differ little in these experiments—is explained in Appendix C, together with our estimation method.

A set of such PMF means establishes a *judgment function*, $f = \mathbf{J}(N)$, a psychophysical scale that associates with each fraction name N, its subjectively equal fraction.[10] (The inverse function, $N = \mathbf{J}^{-1}(f)$, to be used in Section IV,A, therefore gives the value on the name scale associated with a specified stimulus fraction.)

The results are shown in Table I; column labels give the fraction name and its equivalent fractional interval in msec, and row 1 shows the mean stimulus fraction for the three subjects (SB, PF, and PZ). If judgments were free of systematic error, entries in this row would equal the column headings. Instead, as shown by the

[10]Note that this procedure for establishing a psychophysical function, in which averaging is performed in the *stimulus* domain, differs from more common procedures in magnitude and category scaling in which *responses* are averaged. In Appendix E we discuss a comparison of the two methods applied to data from Experiment 2 (multiple-fraction judgment) in which, unlike Experiment 1, both methods can be applied.

TABLE I

Results from Experiments 1–5 and Six Critical Contrasts[a,b]

Experiment or contrast	Fraction name N, n (in judgment, production) or fraction f (in imitation) (and corresponding fractional interval)											
	1/8 (125)	1/7 (143)	1/6 (167)	1/5 (200)	1/4 (250)	1/3 (333)	1/2 (500)	3/4 (750)	5/6 (833)	7/8 (875)	1 (1000)	SE (df)
1. Single-fraction judgment, $\mathbf{J}_1$	85.3 (z)	—	125.5 (fz)	—	265.2	—	455.3 (z)	787.3	835.4 (b)	883.9 (b)	977.8	23.9 (14)
2. Multiple-fraction judgment, $\mathbf{J}_2$	59.3 (*bfz)	79.7 (*bfz)	105.4 (*bfz)	154.4 (*bfz)	207.3 (*bf)	303.6	451.7 (*bz)	—	—	—	—	10.1 (12)
3. One-response production, $\mathbf{P}_3$	139.1	—	197.8 (*b)	—	259.3	—	486.3	759.5	812.3	865.7 (z)	1023.3	11.7 (14)
4. Repeated-response production, $\mathbf{P}_4$	156.8 (bz)	181.4 (bz)	190.4 (bfz)	—	256.7 (z)	—	500.1	743.9	814.2 (b)	853.7 (b)	—	18.3 (14)
5. Imitation, $\mathbf{I}_5$	158.2 (*z)	—	185.2 (z)	—	254.5	—	491.0	773.0	800.8 (*f)	853.3	986.6	14.6 (14)
6. $\mathbf{P}_3 - \mathbf{J}_1$	53.8 (z)	—	72.3 (*fz)	—	−5.9	—	31.0 (z)	−27.8	−23.1 (z)	−18.2 (bz)	45.5 (z)	28.2 (14)
7. $\mathbf{P}_4 - \mathbf{J}_2$	97.5 (*bfz)	101.7 (*bfz)	85.0 (*bfz)	—	49.4 (*b)	—	48.4 (*bz)	—	—	—	—	10.0 (8)
8. $\mathbf{P} - \mathbf{J}$ (mean)	75.7 (*bz)	—	78.7 (*bz)	—	21.8 (bz)	—	39.7 (*bz)	—	—	—	—	10.0 (6)
9. $\mathbf{J}_1^{-1} - \mathbf{I}_5$ (Alternative 2)[c]	34.2 (f)	—	17.0	—	11.0 (z)	—	43.2 (z)	−57.3 (fz)	7.3 (fz)	−4.4 (z)	—	26.0 (6)
10. $\mathbf{P}_3 - \mathbf{I}_5$ (Alternative 3)	−19.1	—	12.6	—	4.8	—	−4.7	−13.5	11.5 (z)	12.4 (z)	33.7 (b)	17.2 (14)
11. $\mathbf{P}_3\mathbf{J}_1^{-1} - \mathbf{I}_5$ (Alternative 4)[c]	52.5 (f)	—	29.6	—	10.8	—	36.0 (z)	−62.0 (fz)	−1.7 (fz)	35.9 (z)	—	29.3 (6)

[a] Entries in rows 1 and 2 (3 and 4) are mean presented (produced) fractional intervals associated with the fraction names specified by column headings. Entries in row 5 are mean produced fractional intervals associated with presented fractions specified by column headings. Entries in rows 6–11 are the mean differences indicated. All entries are in msec.

[b] The letters b, f, z are the last initials of our three principal subjects SB, PF, and PZ. A letter is placed next to an entry if the corresponding subject's value of that entry differs significantly ($p < .05$) from the column heading (rows 1–5) or from zero (rows 6–11); the tests for individual subjects were based on the between-replications variance pooled over fractions. The asterisk indicates that the mean over subjects is significantly ($p < .05$) different from the column heading (rows 1–5) or from zero (rows 6–11). Each condition in Experiments 1, 3, and 5 had two replications of the procedure; each condition in Experiments 2 and 4 had from one to three replications. Questions about characteristics of the "population" of subjects from which our "sample" of three was drawn depend on tests of the mean cell entries over subjects. For Experiments 1, 3, and 5 such tests were based on fractions × subjects × replications analyses of variance in which replications was regarded as a fixed effect. For Experiments 2 and 4, results from different replications were averaged and subjected to fractions × subjects analyses of variance. SE estimates used for t-tests of cell entries are based on subjects × fractions interaction mean squares, whose *df*s are also indicated.

[c] Data from subject SB were excluded from the means in rows 9 and 11 because the between-replications variances for these contrasts are greater than those of PF and PZ by a factor of about 19. (This exclusion has the effect of producing means that are heavily weighted by degree of precision.) None of SBs values of the contrasts associated with alternatives 2 and 4 differed significantly from zero.

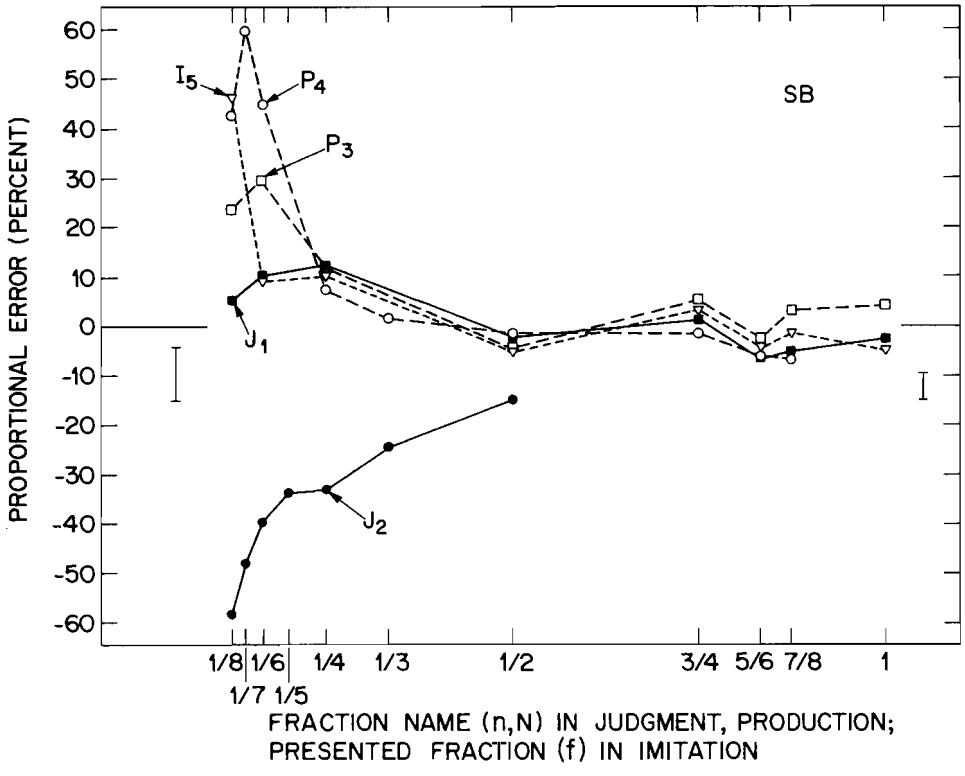

Fig. 2. Results for subject SB from five judgment (**J**), production (**P**), and imitation (**I**) experiments, Ordinate values denote signed proportional error: $(f-N)/N$ for judgment, $(F-n)/n$ for production, and $(F-f)/f$ for imitation, in percentage units. Corresponding abscissa values are N, n, and f, respectively, expressed as fractions. Subscripts are experiment numbers.

accompanying letters, six of the 24 tests of individual data indicated significant departures from equality. Judgment variability tends to increase with fraction size. (See Section II,C, and Getty, 1975, for examples.) This is one reason for our representing the data for individual subjects in Figs. 2, 3, and 4 as signed proportional error (in percent) versus N-value. Our second reason is the importance in music of time ratios and of the rates at which notes occur. The value of $\mathbf{J}(1/8)$ from a brief session with our fourth subject, PB, is included in Fig. 4.[11]

A numerical example of the proportional error measure for $N = 1/8$ may be helpful.

[11] As noted in Section I,E, a large fraction (such as 7/8) defined from the *previous* beat (as described by the instructions to subjects) corresponds to a small fraction (1/8) measured relative to the *next* beat. If such small reverse fractions were overestimated we would expect that just as we tend to find $\mathbf{J}(N) < N$ for $N < 1/2$, so we would find the symmetric relation $[1 - \mathbf{J}(N)] < 1 - N$, equivalent to $\mathbf{J}(N) > N$, for $N > 1/2$. For PZ and PF both relations tend to obtain, indicating symmetry, although the effect for large fractions is not significant; for SB both relations tend to be reversed, also indicating symmetry. Since the beat following the response was not represented by a click in this procedure, as was the previous beat, any such symmetry suggests that the beat click itself may not be an important determinant of performance and that there is indeed an internal event associated with the final beat. One deficiency of our proportional error plots, of course, is that they obscure systematic irregularities associated with large fractions.

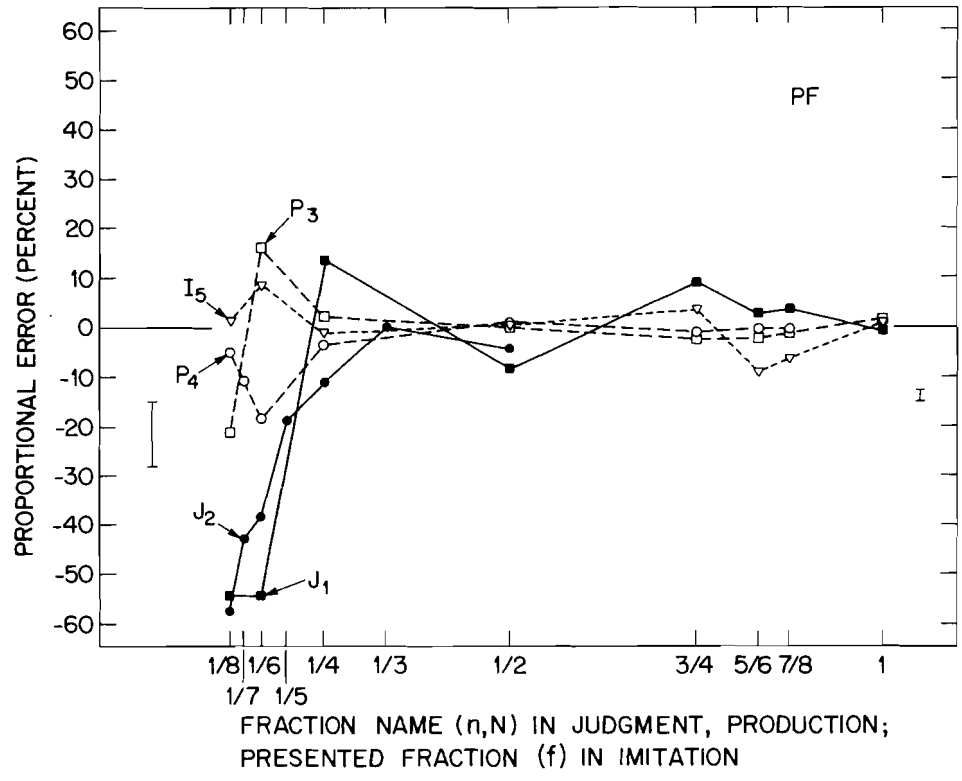

Fig. 3. Results for subject PF from five experiments. See caption of Fig. 2.

PB's judgment data implied that relative to a beat interval of 1 sec, 62 msec was subjectively equal to 1/8 of a beat (125 msec): $\mathbf{J}(1/8) = 62$. Since $[\mathbf{J}(1/8) - 125]/125 = (62-125)/125 = -.50$, there was a -50% error: the fraction that was judged subjectively equal to 1/8 of a beat was 50% too small. Put another way, the name (1/8) assigned to $f = 62$ was 100% too large: an instance of surprisingly radical overestimation.

Figure 5 shows a different representation of the data from PZ, the most stable of our three principal subjects. Here, $\ln(bf)$ is plotted against $\ln(N)$. If $f = \mathbf{J}(N)$ were a power function, the judgment data in this figure could be well fitted by a straight line. (The slope of such a line is the exponent of the corresponding power function.) Clearly, no one power function can describe these data; if separate linear segments were fitted to small-N and large-N ranges, exponents would be about 1.81 for small fractions ($N \leq 1/4$), and close to unity for larger fractions.[12] We defer further discussion of the judgment data to the next section.

[12]Note that the more conventional judgment function derived from "magnitude estimation" procedures, in which the experimenter specifies f and the subject provides N, would give an average N as a function of f; $N = \mathrm{M}(f)$. M can be regarded as the inverse of J; if they were power functions, their exponents would be reciprocals. (The exponent of M for small fractions is about .55 for the data in Fig. 5.) The change in the exponent of J as n is increased therefore conforms approximately to the findings for magnitude scaling of subjective duration reported by Michon (1967) and Svenson (1973), mentioned in Section I.

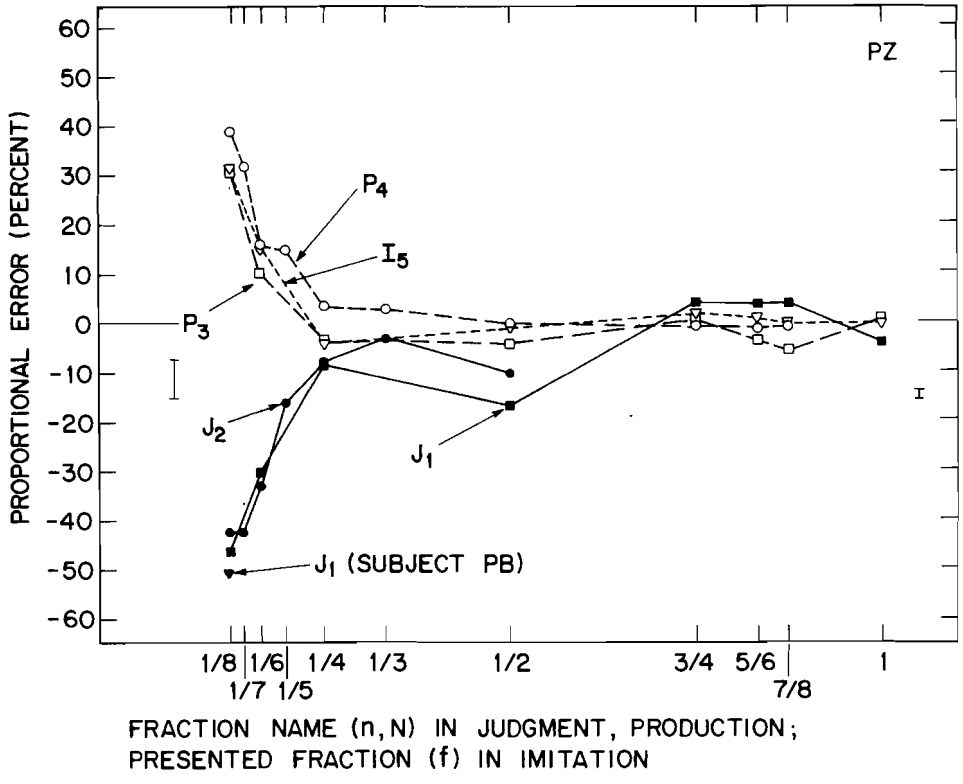

Fig. 4. Results for subject PZ from five experiments and one datum from subject PB. See caption of Fig. 2.

B. Multiple-Fraction Perceptual Judgment (Experiment 2)

The stimulus pattern used in Experiment 2 is represented in Fig. 1B. On each trial the subject heard five beat clicks, with marker clicks following the third and fourth; subjects therefore had two opportunities to observe the beat-marker interval before each judgment. This interval was varied from trial to trial over a wide range (from a minimum of 43 msec to a maximum of 891 msec) by a constant-stimulus method (see Appendix A). The subject selected a response from a set of eight categories, each denoting a range of fraction names and bounded by "simple" fractions (involving small integers): "less than 1/8 of a beat," "between 1/8 and 1/7," "between 1/7 and 1/6,"..., "between 1/3 and 1/2," and "greater than 1/2." The eight ordered categories define seven between-category boundaries on a hypothetical response continuum. For each boundary and each stimulus fraction f we determined the proportion of responses in all categories above that boundary. Regarded as a function of f, this proportion for a specified boundary defines an estimated PMF; this procedure produces seven such PMFs. (Consider, for example, the boundary $N = 1/7$ between the second and third category. Responses to a fraction f in categories above that boundary—categories 3 through 8—are all associated with judgments that f appears

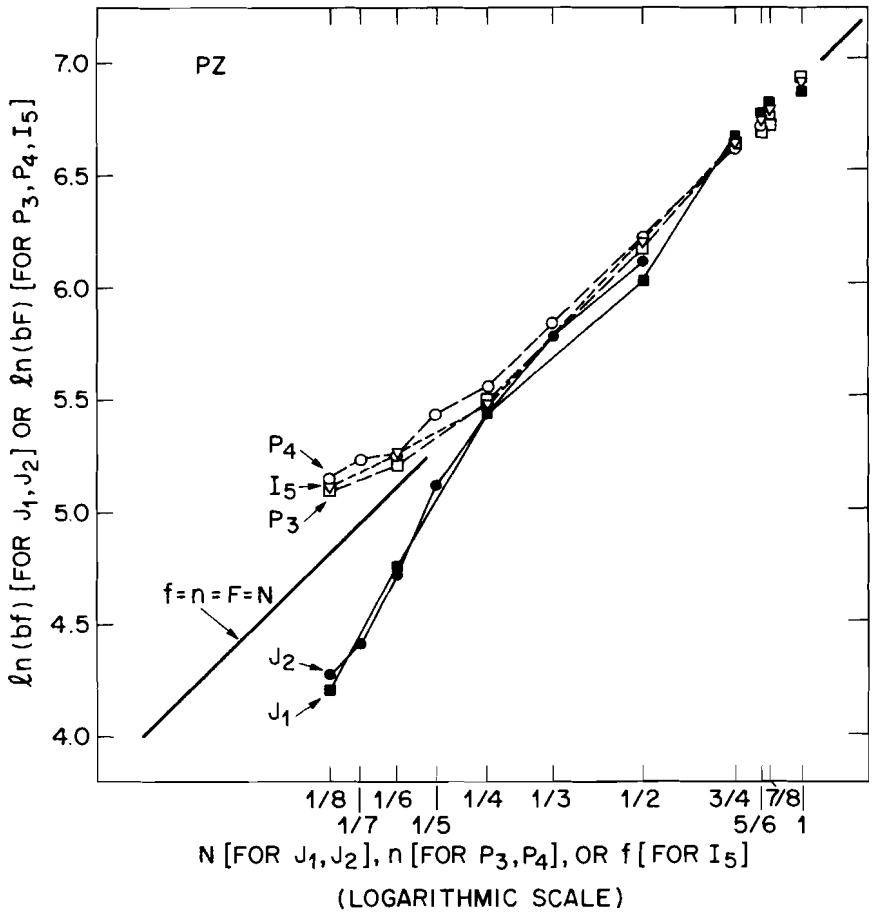

Fig. 5. Results for subject PZ from five judgment (**J**), production (**P**), and imitation (**I**) experiments. Ordinate values are natural logarithms of the fractional intervals *bf* (for judgment) and *bF* (for production and imitation) in msec. Abscissa values represent *N* (for judgment), *n* (for production), and *f* (for imitation), also on a logarithmic scale. Subscripts are experiment numbers. To make units the same on abscissa and ordinate, multiply abscissa values by the beat interval (1000 msec). Power functions are represented as straight lines on this kind of graph.

greater than 1/7 of a beat. As f is increased, the proportion of judgments in this "supercategory" increases, defining an estimated PMF associated with $N = 1/7$.)

As in Experiment 1, means of the resulting set of PMFs were used to establish a judgment function, $f = \mathbf{J}(N)$, for each subject.[13] The mean function for three subjects

[13]This judgment function can be regarded as associating an average stimulus value with each of a set of values (category boundaries) on the response continuum. For each partitioning of the eight categories into the pair of "supercategories" defined by a particular boundary, we treated the data in the same way as in Experiment 1 (see Appendix C). Experiment 2 can also be treated by the more conventional procedure in which an average *response* value is associated with each of a set of *stimuli*. To permit such averaging in the present data, the value of a category response could be taken to be the geometric mean of the values of its two boundaries, for example. That the judgment functions from the two procedures are similar is shown in Appendix E, which also explains our preference for the PMF method.

is presented in row 2 of Table I, individual proportional error data are shown in Figs. 2, 3, and 4, and $\ln(f)$ versus $\ln(N)$ for PZ is shown in Fig. 5.

In Experiment 1, PF and PZ (corroborated by PB) showed large and systematic overestimation of small stimulus fractions, while SB did not. In Experiment 2, however, all three principal subjects showed this effect; quantitative agreement between Experiments 1 and 2 was excellent for PZ, good for PF, but poor for SB. Of the 21 tests based on individual data in Experiment 2, 16 indicated significant departures from equality of N and $\mathbf{J}(N)$, and each of the mean differences also proved significant. Taken together, our data show that musicians radically overestimate small fractions of a beat: $N > f = \mathbf{J}(N)$.

Although we cannot explain the anomalous results from SB in Experiment 1, we are more impressed by the consistency of PZs mean data between experiments than by the inconsistency of SBs: PZ is the most experienced musician among our principal subjects, and his data *wthin* experiments are by far the most consistent. In Experiment 1 each trial included only one presentation of the fraction to be judged; the fraction size was varied by a staircase procedure over a narrow range of equally spaced values, and subjects had to judge stimulus fractions relative to only a single "target" fraction name during a group of trials. In Experiment 2, on the other hand, each trial included two successive presentations of the fraction to be judged; the fraction size was varied over a wide range and over unequally spaced values by a constant stimulus method; and subjects had to judge stimulus fractions relative to an array of seven fraction names (category boundaries). We conclude that the mean values of the criteria that subjects employ in making these perceptual judgments are affected little by either the number of observations per trial, the range of fractions to be judged, or the number of fractions with respect to which each judgment was made. We shall see in the next section, however, that the choice of procedure does influence the precision of these judgments.[14]

C. Judgment Precision

Insofar as a subject is more sensitive to the stimulus fraction, his judgment probabilities will change more rapidly as the fraction is changed, and the PMF will rise more steeply, or have less *spread.* It is convenient to regard the PMF as a (cumulative) distribution function characterized by a standard deviation (SD) as well as a mean. The SD of the PMF is one measure of its spread, and therefore of the imprecision or variability of judgments.[15]

[14]On each trial in Experiment 1, both N and f were provided to the subject; there is some question whether the value of N associated with the stimulus f by the function $f = \mathbf{J}(N)$ should be regarded as a response to f. In Experiment 2, where the subject explicitly selected an interval on a continuum of N-values, there is less uncertainty in identifying N as the response. Because of this, together with similarity of the two judgment functions for each of two subjects, we shall regard f as "input" and N as "output" for both experiments.

[15]An alternative and more traditional measure of precision is the difference threshold (DL), which is defined as half of the interquartile range or, roughly, the change in the stimulus fraction required to change

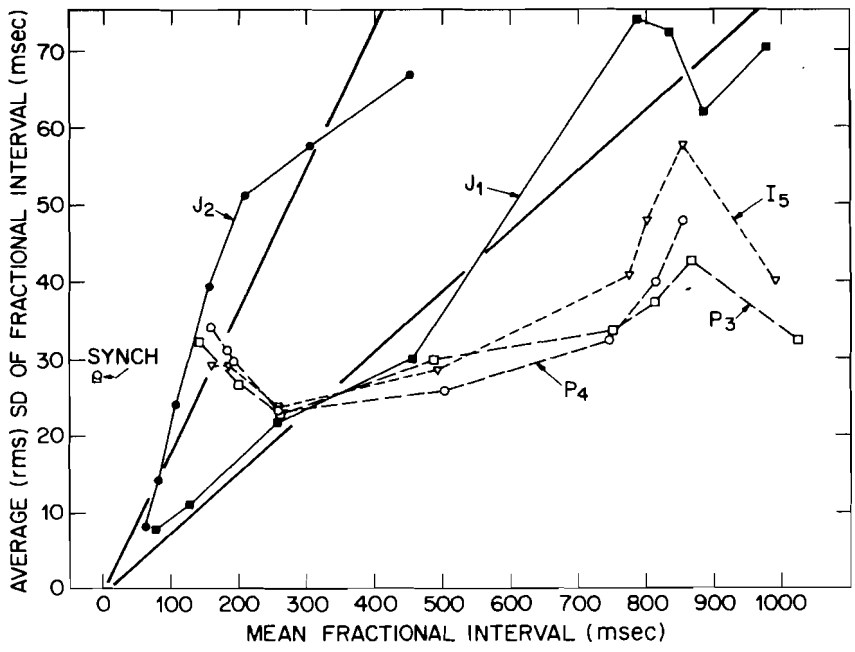

Fig. 6. Performance variability in five experiments. Root mean squared SDs (in msec) over replications and subjects. For judgment experiments $\mathbf{J}_1$ and $\mathbf{J}_2$, SDs were estimated from PMFs regarded as cumulative distributions of the fractional interval *bf*. (See Appendix D.) For production ($\mathbf{P}_3$ and $\mathbf{P}_4$) and imitation ($\mathbf{I}_5$) experiments, SDs were calculated from the distributions of the fractional interval *bF*. Also shown is the rms SD from the synchronization condition (Section III,A). In all cases, the SD(x) of a quantity x is plotted as a function of the mean M(x) of that quantity. Linear functions passing through the origin—which represent Weber laws SD(bf) = kM(bf) with k=.08 and k=.19 for judgment Experiments 1 and 2, respectively—were fitted by eye.

Fig. 6 includes average SDs of the PMFs in Experiments 1 and 2.[16] In both experiments the SD increases approximately in proportion to f, consistent with a Weber law; proportionality constants are about .08 and .19 for Experiments 1 and 2, respectively. (Sampling error in these data preclude a powerful test of the Weber law, however.) Getty (1975) discusses related data and some of the implications of Weber's law for timing models. Insofar as a fraction has been better learned than others or is easier to "compute" given a beat interval, one might expect it to show greater precision (smaller SD) and therefore fall below the Weber law line; there is no dramatic evidence supporting this conjecture, however. Judgment precision is systematically greater (smaller SDs) in the one-fraction than the multiple-fraction procedure, despite

the proportion of "larger than" judgments from .50 to .75. See Appendix D for a discussion of our preference for the SD and our method of estimating it.

[16]The figure actually shows square roots of mean variances (root mean square SDs). Throughout this paper we have chosen to average variances rather than SDs because different sources of variability (such as fraction and subject, in perceptual judgments, or timing and response mechanisms, in production) are more likely to be additive in variance units. We have shown SD rather than SD^2 in the figure, however, because the expression of Weber's law, SD = kf, where k is a constant, is then simply a straight line through the origin.

the similarity of judgment means discussed above. (This finding suggests that the range of stimuli f or of fractions N with respect to which judgments are made has a large effect on the variability of subjects' criteria, but only a small effect on their means.)

Despite the systematic errors shown by the judgment function, Fig. 6 reveals the precision of the judgments to be high. For example, in Experiment 2, the f-value that was subjectively equal to $N = 1/8$ (125 msec) was 59.3 msec, and the SD was 7.9 msec. This implies that an f-value of only 72 msec would be judged "larger than 1/8" on 95% of the trials,even though this f-value is 53 msec (or 42%) smaller than the "correct" value.

III. PRODUCTION OF BEAT FRACTIONS

The systematic errors in perceptual judgment discussed above, which are proportionately very large for small fractions of a beat, make it particularly interesting to examine musicians' accuracy in *producing* brief time intervals defined as beat fractions. We used a method of *timed response*. The subjects' task in Experiments 3 and 4 was to use a finger tap to terminate a time interval that started with a beat click and thereby produce a beat fraction F that corresponded to a specified fraction name n. (Subjects could hear as well as feel themselves tapping, since the earphones that delivered the clicks provided negligible attenuation of other sounds. A reader who taps the hard surface of a desk top will hear a "thump" similar to what our subjects heard.)

Our aim was to determine the relation between a set of fraction names and the set of corresponding fractions, and thereby establish a production function, $F = \mathbf{P}(n)$. A secondary aim was to measure the precision of such expert timed responses—the variability of the time intervals they defined. Again we used two methods, to assess the invariance of the systematic timing errors we discovered. One method required a single timed response on each trial; the other required a repeated series of responses corresponding to a fixed-beat fraction, thereby permitting more immediate adjustment to perceptual feedback.

A. Use of Tap-Click Synchronization to Correct for Differential Subjective Delays

We wished to compare the timing mechanisms used in perceptual judgment with those used in production and imitation, partly to test the idea that they are the same. It is possible, however, that the subjective delays associated with events that mark the ends of the relevant intervals—beat click, marker click, or finger tap—are different. By correcting for any such differences we can examine the timing mechanisms more directly.

This need for correction seems especially acute for finger taps. Even abrupt taps are extended in time, which makes it unclear how to associate a single time point with a

response. Our equipment measured the time at which the finger first contacted a metal plate, but the subjective time of the response might be equally well described by the time when the "command" to make the response is issued, the time when maximum pressure is achieved, the time when the finger breaks contact with the plate, or some other feature of the response, possibly adjusted by perceptual delays.

Suppose that the beginning and end of an interval are marked by events b and e, respectively, T_b and T_e are their physical occurrence times, and D_b and D_e are the delays in registering the events internally. (For a tap, the mean "delay" may be negative.) Then the registration times are $T_b + D_b$ and $T_e + D_e$; and whereas the objective interval between events is $T_e - T_b$, the interval between registration times is $(T_e - T_b) + (D_e - D_b)$. Thus, to correct the measured interval for internal delays, we must estimate the delay difference, $D_e - D_b$.

If the two delays are equal, the difference is zero, and no correction is needed. For the perceptual judgment experiments, in which both ends of the relevant interval were marked by the same class of events (clicks), we felt that equality of delays was a plausible starting assumption. (In Section VI we report some findings favorable to this assumption.)

Our solution to this problem in the production experiments, where the critical interval begins with a beat click and ends with a finger tap, was to measure the difference between the two subjective delays by using a special condition in which subjects were asked to synchronize their responses with beat clicks. This synchronization task can be thought of as a production task with $n=0$, except for the occurrence of a beat click at the time when the response should occur. Since we shall be using $\mathbf{P}'(n)$ to denote the raw (uncorrected) mean production time for fraction n, we use $\mathbf{S}'(0)$ to denote the mean measured response "delay," $\mathbf{S}'(0) = T_t - T_c$, where t and c denote tap and click, respectively. Suppose that the subjects succeed in locating the mean of their distributions of subjective occurrence times of responses coincident with the mean of the subjective beat times. The registration-time difference defined above is then zero: $(T_t - T_c) + (D_t - D_c) = 0$. It follows that $-\mathbf{S}'(0) = D_t - D_c = T_c - T_t$ provides the desired estimator of the delay difference.

The raw (uncorrected) mean production times $\mathbf{P}'(n)$ were corrected by subtraction: $\mathbf{P}(n) = \mathbf{P}'(n) - \mathbf{S}'(0)$. In the synchronization conditions, subjects responded slightly before the beat click [$\mathbf{S}'(0) < 0$], implying that the subjective delay associated with the tap response was greater than the perceptual delay associated with the beat click. The corrections, then, slightly increase the measured values of $\mathbf{P}(n)$. Application of the synchronization correction depends on the assumption that the difference between these beat-click and tap delays in the production task is the same as in the synchronization task and is independent of the interval between beat click and tap (which in turn depends, in production, on the specified fraction name).[17]

[17]One objection to the use of the synchronization correction is based on the possibility that this assumption will be violated. For example, constraints on attending simultaneously or in close succession to click and tap might produce special differential delays ("prior-entry" effects). See Sternberg and Knoll (1973), Section VI.

B. Expectations from a Feedback Model of Production

What relation might we expect between performances in the judgment and production tasks? One appealing hypothesis is a simple feedback process in which a subject (1) judges the size of each produced fraction F with respect to the target-fraction name n using the same perceptual mechanisms in this judgment as in the judgment task itself, and (2) adjusts subsequent productions accordingly.[18] One would then expect systematic errors in judgment and production tasks to be equal to that $\mathbf{P}(n) = \mathbf{J}(N)$ when $n = N$. For example, if an interval of 62 msec is *judged* to correspond to 1/8 of a 1-sec beat [$\mathbf{J}(1/8) = 62$], the same interval (after correction for differential subjective delays) should be *produced* for 1/8 of a beat [$\mathbf{P}(1/8) = 62$].

C. One-Response Production (Experiment 3)

On each trial in Experiment 3 (Fig. 1C), the subject attempted to respond with a single finger tap following the last beat click so as to produce a fraction F that corresponded to a specified fraction name n. The name remained the same for 25 consecutive trials. We used the same fraction names as used in Experiment 1 to define a set of experimental conditions. For each condition we calculated a mean raw response time, $\mathbf{P}'(n)$.

In a synchronization condition, the subject attempted to synchronize the finger tap with a beat click added at the end of the normal stimulus pattern. The resulting values of $\mathbf{S}'(0)$ were -13.8 msec, -16.3 msec, and -9.0 msec, for subjects SB, PF, and PZ, respectively.

The mean production function, $\mathbf{P}(n) = \mathbf{P}'(n) - \mathbf{S}'(0)$ for the three subjects is presented in row 3 of Table I, individual proportional error data are shown in Figs. 2, 3, and 4, and $\ln(F)$ versus $\ln(n)$ for PZ is shown in Fig. 5. PZ and SB show large positive proportional errors (overproduction) for fractions 1/8 and 1/6. Because PF does not show this effect, the error shown by the mean production time for small fractions is somewhat smaller; for only one of the eight means (for $n = 1/6$) is the inaccuracy significant. Proportional errors for larger fractions are small and of varying sign, with a slight tendency for produced intervals associated with the largest fractions (5/6, 7/8) to be too small (underproduction).[19]

[18]Carlson and Feinberg (1968) and Adam, Castro, and Clark (1974) present data that favor an "internal clock" or counter model that applies to both judgment and production of time intervals in a range from 1 to 40 seconds. According to Carlson and Feinberg's model, systematic errors in production and judgment result from changes in the counter's rate between learning the count to be associated with a time interval specified by name and performing the experimental tasks. The resulting relation between errors in production and judgment is consistent with a feedback model. Furthermore, because the counter rate is assumed not to change systematically between stimulus and response in an imitation (reproduction) task (Section IV), this kind of model requires imitation to be accurate. Given that fractions are being judged in our experiments, however, and the beat interval (which can "calibrate" the counting rate) is presented on each trial, the counter mechanism does not easily lend itself to explaining any systematic errors in our tasks.

[19]This tendency can also be described as overproduction of the small "reverse fractions" 1/6 and 1/8 that are defined by the intervals between tap and subsequent beat. A tendency toward such symmetry in

D. Rejection of the Feedback Model

All three subjects show large discrepancies between $\mathbf{P}(n)$ and $\mathbf{J}(N)$ for the two smallest fractions. In row 6 of Table I are shown values of the difference $\mathbf{P}(n) - \mathbf{J}(N)$ for the matched procedures of Experiments 1 and 3. The feedback model requires this difference to be zero for all values of $n=N$; for the two smallest fractions it is substantially greater than zero, especially in relation to the precision (Fig. 6) of perceptual judgments. The two largest fractions show smaller but directionally symmetric discrepancies.[20]

A numerical example based on mean data from Experiments 1 and 3 (rows 1 and 3 of Table I) may help to clarify the failure of the feedback model. The mean interval *produced* to correspond to $n = 1/8$ was $\mathbf{P}(1/8) = 139.1$ msec. To what fraction name N would this interval correspond if the same perceptual judgment mechanisms were used as in Experiment 1? Since $\mathbf{J}(1/8) = 85.3$ msec and $\mathbf{J}(1/6) = 125.5$ msec, N is clearly greater than 1/6; since $\mathbf{J}(1/4) = 265.2$ msec, it is clearly less than 1/4. To determine N exactly, we must solve $\mathbf{J}(N) = 139.1$ msec for N. If we regard fraction names as lying on a continuum, and use linear interpolation, we find the corresponding fractional interval to be 174.8 msec or slightly greater than 1/6 of a beat. Thus, if subjects perceived intervals between click and tap in the same way as intervals between two clicks, the average subject would *perceive* the interval he or she *produced* for 1/8 as larger than 1/6.

Differences between judgment and production procedures could, in principle, result from "constant errors" based on differential subjective delays for which we have not corrected. (We have not attempted to estimate or correct for any delay difference between beat click and marker click in the judgment task, partly because of findings discussed in Section VI. The synchronization-based correction that we did apply to the production data might be inappropriate.) But on the most straightforward view, the absolute effects of differential delays should be of the same size for all fraction values. Hence, insofar as the inconsistency between perception and production depends on fraction size (which it does reliably, as shown below), we have to search elsewhere for an explanation.

One possibility is that although discrepant feedback was available, the time interval between successive trials and the number of stimulus events between one response and the next prevented subjects from making appropriate corrections. Because our subjects said neither that they were dissatisfied with their productions nor that they "came in late," this possibility seems unlikely. Nonetheless, it was tested in Experiment 4, which called for timed responses to 10 successive beat clicks on each trial.

---

judgment performance was noted in footnote 10. Note, however, that when the *precision* of performance (Fig. 6) is considered, rather than its *mean*, performance at the two ends of the beat interval appears far from symmetric.

[20] $\mathbf{J}(1)$ and $\mathbf{P}(1)$ can be regarded as alternative measures of the *subjective* beat interval. That they differ (significantly for PZ) suggests that this interval might be task dependent (also suggested by findings in Experiment 12). Deviations of $\mathbf{P}(1)$ and $\mathbf{J}(1)$ from 1000 msec are sufficiently small, however, so that our use of the actual rather than subjective beat interval to define beat fractions makes little difference, especially for small fractions.

E. Repeated-Response Production (Experiment 4)

On each trial in Experiment 4, we presented 12 beat clicks. As shown in Fig. 1D, subjects attempted to make 10 consecutive finger-tap productions so as to produce a fraction that had been specified by name. The first response was produced after the third beat click. We used the fraction names 1/8, 1/7, 1/6, 1/4, 1/2, 3/4, 5/6, and 7/8; the name remained the same for 25 consecutive trials. For each fraction name n and each position k in the sequence of 10 responses, we determined the mean raw response time, $\mathbf{P}_k'(n)$.

In a synchronization condition, subjects attempted to make tap responses that coincided with each of the last 10 beat clicks, thereby generating values of $\mathbf{S}_k'(0)$ for $k = 1,\ldots, 10$. The corrected response time was obtained by subtraction: $\mathbf{P}_k(n) = \mathbf{P}_k'(n) - \mathbf{S}_k'(0)$.

Trend over Repeated Responses

Subjects produced large positive errors ("too late") for the smallest three fractions (1/8, 1/7, 1/6). Values of $\mathbf{P}_k(n)$ averaged over these fractions and the three subjects, for $k = 1,\ldots, 10$, were 168.4, 172.9, 176.9, 180.0, 175.4, 181.4, 176.3, 184.5, 176.6, and 169.8 msec, respectively. There is no evidence of a decrease in the size of the error over repetitions. (The slope of a line fitted to these values is .43 msec/repetition.) It is also instructive to compare the time of the first response $\mathbf{P}_1(n)$ (with synchronization correction $\mathbf{S}_1'(0) = -13.4$ msec) with the time of the mean response $\mathbf{P}(n)$ (with $\mathbf{S}'(0) = -24.9$ msec), separately for the three fractions. For 1/8, 1/7, and 1/6, values for the first response (mean response) are, respectively, 151.9 (156.8), 169.5 (181.4), and 183.9 (190.4) msec.

There is no evidence that subjects used feedback from one response to the next in the repeated response procedure to reduce the size of their production errors.

Means over Repeated Responses

Further analyses were based on mean response times over the 10 positions. These were adjusted by the mean synchronization corrections; $\mathbf{S}'(0)$ was -22.6, -35.2, and -16.9 msec for SB, PF, and PZ, respectively.

The mean of the production functions, $F = \mathbf{P}(n)$, for the three subjects is presented in row 4 of Table I, individual proportional error data are shown in Figs. 2, 3, and 4, and ln (F) versus ln(n) for PZ is shown in Fig. 5.[21] As in Experiment 3, PZ and SB show large positive proportional errors for small fractions, whereas PF does not; in addition, PZ and SB show negative errors for large fractions, whereas PF does not, providing further evidence for the symmetry in performance at the two ends of the

[21]Like the judgment function, ln(F) versus ln(n) cannot be fitted by a single power function (straight line). A two-limb fit gives an exponent of about .57 for the range $1/8 \leq n \leq 1/4$ and about 1.0 for $n > 1/4$.

beat interval mentioned in Sections II,A and III,C. Agreement between Experiments 3 and 4 is best for PZ. Again, performance for large fractions is relatively accurate.

Values of the difference $\mathbf{P}(n)-\mathbf{J}(N)$ appropriate for another test of the feedback model are provided in row 7 of Table I. We used judgment data from Experiment 2 because it provides values that are independent of Experiment 1, because it was run in close temporal proximity to Experiment 4, and because the stimulus conditions (Figs. 1B and 1D) were similar. All three subjects show large discrepancies between judgment and production of small fractions. Each of the five mean values of the difference in row 7 of Table I (values for all the fractions $n = N$ that were common to the two experiments) is significantly and substantially greater than zero; the difference diminishes reliably as fraction size increases. (This implies, of course, that the proportional difference decreases to an even greater extent.) The failure of the feedback model is even more dramatic for these data than for the initial test (row 6 of Table I).[22]

Our best estimates of production-perception discrepancies are obtained by combining the two tests; means for the fraction values that are common to the four experiments are given in row 8 of Table I. The mean contrast is significantly greater than zero ($t_7 = 4.41$; $p < .005$), and the difference depends significantly on fraction size ($F_{3,6} = 7.56$; $p < .025$).

F. Production Precision: Evidence against a Reaction-Time Explanation of the Production Error

Average SDs of production and synchronization times are included in Fig. 6 for Experiments 3 and 4. Data from the two experiments are in good agreement; data from each of the three subjects averaged over the two experiments produced a U-shaped function with a minimum SD (maximum precision) between $n = 1/4$ and $n = 1/2$.[23] The precision of synchronization performance, with mean $\mathbf{S}'(0)$, is slightly but not reliably greater than the precision of production of a whole beat, with mean $\mathbf{P}'(1)$.

For fractions $n \geq 1/4$, production precision exceeds judgment precision, but this

[22]It remains a puzzle how production performance might be calibrated (or learned) by players without feedback about the timing of individual responses. One possibility is that the association of produced fractions with fraction names is learned through attempts to produce extended sequences of temporally regular notes that fully occupy the interval between one beat and the next. Counting, rather than the timing of individual responses, could then provide a measure by which to adjust the rate. Given this possibility (and also to assess the generality of our findings), it is interesting to ask whether the first response by a subject attempting to produce a temporally regular sequence displays the same pattern of errors as that obtained with a single response. This question is considered in Section VII,C, in a study of multiple divisions of the beat.

[23]A U-shaped function with a minimum in this range has also been obtained from one subject in a time production task by A. B. Kristofferson (1976; Experiment II). Experimental conditions differed considerably from ours, with feedback provided on each trial, and the one subject had about 10 times as much practice in the experimental task as ours did. Nonetheless, his minimum SD (about 12.7 msec) did not differ significantly from the minimum (22.2 msec) of the mean SD produced by our subjects.

relation is reversed for small fractions. Insofar as subjects do not experience their productions of small fractions as highly variable, we therefore have further evidence against the feedback model.

One explanation of the production errors for small fractions (but not the similar errors for small "reverse fractions") attributes them to a combination of musical training and the existence of a minimum reaction time (RT). The minimum voluntary RT to auditory stimuli is between 100 and 150 msec. Furthermore, there are delays (which differ across instruments) between excitation and acoustic response. The combination of these two effects makes it virtually impossible to produce a note 125 msec after a signal to respond such as a beat, when the event that triggers the response is the beat itself. Players could try to "anticipate" the beat (by timing their responses from the penultimate beat), but since the occurrence of the beat in musical performance is variable, this might be risky. Players therefore appear to be in a situation in which they are musically required to produce discriminably different response delays, some of which are less than the minimum RT.

One possible solution would be to time responses from the final beat, but bias the productions so that the intervals for small fractions are both greater than the minimum RT and distinctive. For example, if to respond later than the minimum reaction time a subject produces an interval of 150 msec (rather than 125 msec) for 1/8 of a beat, then an interval greater than 150 msec must be produced for 1/7 of a beat, etc.

However, the variability of the production-time distributions suggests that responses associated with small fractions may be timed from the penultimate beat. (Synchronization responses presumably *must* be timed from this beat.) Hence, the production errors may *not* be due to a constraint imposed by a minimum RT. The argument (whose impetus and conceptual framework is provided by Snodgrass, Luce, and Galanter, 1967) is as follows.

We start by assuming that the variance of a distribution of response delays increases monotonically (or at least does not decrease) as the mean of the distribution increases, where the mean is measured relative to the reference signal from which subjects time their responses.[24] Therefore, if all responses were timed from the final beat click, we would expect the variance to increase monotonically with fraction size. As we have seen, this expectation is violated by our data. One interpretation of the increased variability for small fractions is that subjects were timing their responses from the penultimate beat click in these conditions. The most salient errors in production are then not the result of a constraint imposed by the minimum RT.

If subjects timed small fractions from the penultimate beat click but large fractions from the final beat click, the argument above, in its simplest form, implies that no large-n productions should have variances greater than the small-n productions. One difficulty is the suggestion in the data (Fig. 6) of a peak in the SD-function when n = 7/8. (Because this difference between the SD at n = 7/8 and n = 1/8 is shown only by

[24]See, e.g., Snodgrass, Luce, and Galanter (1967) for intervals $\geq$ .6 sec, Treisman (1963) for intervals $\geq$ .25 sec, and references cited therein.

SB and PZ, the mean difference is not reliable, however.) A slight elaboration of our account can deal with this difficulty. Suppose that the response for small fractions can be triggered either by perception of the final beat click *or* by a timing process initiated by the penultimate beat click, whichever occurs first. This could shorten production delays that would otherwise be exceptionally long, thereby reducing the variance. (See Kornblum, 1973.) Evidence favorable to such a facilitation effect of the final beat click is provided by a comparison we made with PZ between one-response production of $n = 1/8$ and the same procedure with the final beat click replaced by a "rest" (or an "imaginary beat"). Omission of the click increased the mean response delay by about 17% (from 149.5 to 175.5 msec), thereby increasing the mean error, but more than doubled the SD (from 26.0 to 70.2 msec).[25]

However, our argument for the idea that subjects timed their productions of small fractions from the penultimate beat click depends on the assumption, introduced above, that the variance of a produced time interval cannot decrease as its mean increases. Even if true for time intervals defined in isolation, this assumption may be false for intervals that are defined as different fractions (or multiples) of a standard interval, as in our production task. Perhaps fractions that are "simpler," or more practiced, or that require less "computation" (such as $n = 1$) are produced more reliably. In models in which timing is accomplished by counting a stream of internal events until a criterion is reached, the reliability with which the criterion count is *set* may have to be considered as well as the variability of the inter-event intervals. This possibility is supported in our data by the (statistically significant) reduction in variability from **P**(7/8) to **P**(1).

G. Implications of Other Analyses of Psychophysical Scaling for the Production–Perception Disparity

Readers with a special interest in psychophysical scaling may find it interesting to consider our perception and production experiments in relation to the "magnitude estimation" and "magnitude production" methods used to investigate many perceptual domains. These methods often produce power-function relations between stimulus and response values; such psychophysical scales are often summarized by power-function exponents. The exponent β determines how ratios of stimuli ($\phi_1 < \phi_2$) are mapped onto ratios of the numerical magnitudes or names ($\psi_1 < \psi_2$) associated with them: $\psi_2/\psi_1 = (\phi_2/\phi_1)^\beta$. An exponent larger (smaller) than 1.0 implies that the name ratio is larger (smaller) than the stimulus ratio.

Partly because there is a "correct" relation between n and F that musicians are presumably trained to achieve, we are interested in the relation between n-values and F-values and not merely in the relation between *ratios* of pairs of n-values and pairs of F-values. (The latter relation, but not the former, is captured by the exponent of a fitted power function.) Sizes of the exponents are nonetheless useful to consider. We

[25]This comparison also suggests that the existence of a positive production error for small fractions does not depend on the presence of an actual beat click.

have already noted that both $\mathbf{J}(N)$ and $\mathbf{P}(n)$ deviate dramatically from power functions if the full range of fractions is considered. However, power functions fitted only to the data for small fractions fit reasonably well and do capture one aspect of the discrepancy between perception and production. A power function fitted to the data for small fractions in our experiments has an exponent greater than 1.0 for production but less than 1.0 for perception.

Under some conditions with other perceptual continua, magnitude production exponents are larger than those obtained in magnitude estimation. There are at least three reasons why our finding may not be an instance of the same phenomenon. First, Teghtsoonian and Teghtsoonian (1978) have shown that the difference between exponents depends on the stimulus range and, indeed, is reversed for narrow ranges. Stimulus ranges in our two perception experiments differed greatly; their endpoints were about 50 and 850 msec in Experiment 2, but only about 90 and 140 msec (for PZ and $n = 1/6$), for example, in Experiment 1. Nonetheless, we obtained good agreement between experiments for PZ and PF. In both production experiments, the stimulus (fraction name) was fixed for a long series of responses (25 and 250 responses in Experiments 3 and 4, respectively). The best description to assign to the stimulus range in this case would therefore appear to be "narrow."

Second, the Teghtsoonians argued that the effect depends on the avoidance by subjects of extreme response ratios (i.e., either much larger or much smaller than unity). This analysis seems inapplicable in a straightforward way to Experiment 1, where the overt responses were "larger than N" and "smaller than N." (To make it applicable, one could assume that subjects produce covert responses of particular N-values that they then categorize in terms of the specified target N-value to determine the overt response.)

Third, if mechanisms of the kind discussed by the Teghtsoonians were responsible for the difference we observed between $\mathbf{J}(N)$ and $\mathbf{P}(n)$ for small fractions, we would expect a difference in the same direction for large fractions, contrary to what we observed.

In the discussion above we have assumed that time *ratios* F or f (between produced or stimulus intervals and the beat interval) are the objects to be produced or judged. A less obvious alternative is to consider these objects to be time *intervals* bF or bf. In that case, a procedural difference between our experiments and many others becomes important. The Teghtsoonians argue persuasively that in choosing a response on one continuum to associate with a stimulus on another, subjects refer to the prior stimulus and prior response and choose a response that generates a response ratio equal to the subjective stimulus ratio. With traditional methods the prior stimulus and response are those from the previous trial and usually vary from one trial to the next. In contrast, with our methods the (large) beat interval b—corresponding to fixed prior values ("standards") $f = 1$ and $N = 1$ on the two continua—is presented on each trial, becomes a prior stimulus, and can perhaps be regarded as generating a prior response. Suppose that we accept this alternative analysis (together with the idea that subjects produce covert N-responses in Experiment 1). Then, if subjects tend to avoid extreme response ratios, they would both overproduce and overestimate small fractions but

not large ones, as observed. The Teghtsoonians' analysis can therefore provide one viable account of the perception-production difference, if we combine it with rejection of the feedback model.[26] [27]

Two findings incline us against the interval alternative. The first, by Michon (1967), is the absence of effects of stimulus range in an experiment on magnitude estimation of time intervals. Given this finding, one would have to argue that the psychophysics of beat fractions and of time intervals differ, with the former more like other sensory domains, or that although the standard interval (like a beat interval) was not presented on each trial in Michon's experiment, subjects presented a fixed standard to themselves.

The second finding is our own, emerging from a comparison of perceptual judgments of beat fractions across different beat intervals (described in Section VI,E). This experiment permitted us to remove the confounding of the duration of the interval being judged with its fraction value. A straightforward application of the interval alternative, based on avoidance of extreme response ratios, implies that the judgment error depends only on the response (the N value) and therefore on the fraction f rather than the interval bf; in contrast, we found that the error depends on both fraction value and interval duration.

IV. IMITATION OF BEAT FRACTIONS

In both the judgment and production tasks, subjects must associate beat fractions and their names. In the imitation task (sometimes called the method of reproduction), this association is not called for, at least not explicitly: the "input" is the same stimulus f, as in the judgment task, and the "output" is a timed response F, as in the production task. The relation, $F = \mathbf{I}(f)$, between f and F in the imitation task may therefore tell us whether the systematic errors found in the other tasks depend on the requirement to associate names with beat fractions. More generally, by exploring the imitation task and its relation to the judgment and production tasks, we hoped to explain or describe the production-perception disparity in terms of characteristics of internal transformations, some of which may be shared by pairs of the three tasks and some of which may be task-specific.

A. Four Simple Alternatives for Imitation

There are four simple and interesting possible outcomes of the imitation experiment, two of which have been explicitly considered in studies of time-interval perception.

[26]The possible inconsistency in conventional procedures between magnitude estimates (of stimuli controlled by the experimenter) and feedback from magnitude productions (controlled by the subject) is also a puzzle, of course.

[27]Insofar as symmetric errors are found for large fractions—as discussed in Sections II,A; III,C; and III,E—this account would, of course, have to be elaborated.

Alternative 1: Accurate Imitation

The first possibility is that despite errors in the other tasks, imitation (after the synchronization correction) will be accurate: $F = \mathbf{I}(f) = f$. The imitation function is thus the identity transformation for which output equals input and which we denote "**E**", and we have $\mathbf{I} = \mathbf{E}$. This outcome would follow from any model in which the errors found in the judgment and production tasks result from the processing of fraction names: Carlson and Feinberg discuss an example of such a model based on a "clock" or event-counting process.[28]

Alternative 2: Imitation Consistent with Perceptual Judgment

A second possibility is that the relation between f and F in imitation is the same as the relation between f and N in the perceptual judgment task so that $\mathbf{I}(f) = \mathbf{J}^{-1}(f)$, or $\mathbf{I} = \mathbf{J}^{-1}$. This consistency could arise if the stimulus pattern, which is common to the two tasks, is processed in the same way, leading to the same internal representation, and if the further transformations of this representation (that lead to F in imitation and N in judgment) are equivalent.

Alternative 3: Imitation Consistent with Production

A third possibility is that the relation between f and F in imitation is the same as the relation between n and F in production so that $\mathbf{I}(f) = \mathbf{P}(n)$ when $f = n$. The relation $\mathbf{I} = \mathbf{P}$ could arise, for example, if f and n generate a common internal representation when $f = n$, which is then transformed by the same processes in the two tasks to produce F.

Alternative 4: Imitation Combines the Errors of Judgment and Production

The fourth possibility is easiest to motivate by considering a model of imitation performance that would generate it. We describe this as a *full-concatenation* model because it calls for the application of all of the internal transformations used in the other two tasks. According to this model, the subject covertly assigns to f a value on a continuum containing fraction names $N = \mathbf{J}^{-1}(f)$ (just as in the judgment task) and then produces the timed response $F = \mathbf{P}(N)$ that corresponds to that N-value (just as in the production task). We ignore, for the present, the possibility that special difficulties would be introduced by N-values that did not correspond to simple fractions. The result is $F = \mathbf{I}(f) = \mathbf{P}[\mathbf{J}^{-1}(f)]$. (We represent this by $\mathbf{I}(f) = \mathbf{P}\mathbf{J}^{-1}(f)$ or $\mathbf{I} = \mathbf{P}\mathbf{J}^{-1}$.) Note that since $\mathbf{J}^{-1}(x) > x$ and $\mathbf{P}(x) > x$ for small x (on average), the two combined errors are in the same direction. Thomas and Brown (1974, Section V)

[28]See footnote 18 for a description of the model.

assumed a full-concatenation model in their study of the filled-duration illusion in the perception of time intervals.[29,30]

B. One-Response Imitation (Experiment 5)

In Experiment 5 (Fig. 1E) the first pair of beat clicks was followed by a marker click. (In the corresponding judgment task of Experiment 1, the marker followed the second pair of beat clicks, so the contexts for the time-pattern stimuli in the two experiments were not precisely the same.) The subject attempted to respond with a single finger tap after the final beat click (as in the corresponding production task of Experiment 3) to imitate the presented fraction defined by the marker click. The fractions to be reproduced were the objectively correct fractions that correspond to the fraction names used in Experiments 1 and 3.[31] In the imitation task, however, no name was specified to the subjects. The fraction to be imitated remained the same for 25 consecutive trials. We determined the raw mean response time, $\mathbf{I}'(f)$ for each fraction; we then corrected this value by subtraction: $\mathbf{I}(f) = \mathbf{I}'(f) - \mathbf{S}'(0)$, with the same synchronization correction used in Experiment 3.

Mean values of $\mathbf{I}(f)$ are given in row 5 of Table I, proportional error curves for individual subjects are shown in Figs. 2, 3, and 4, $\ln(F)$ versus $\ln(f)$ is plotted for PZ in Fig. 5, and mean SDs are shown in Fig. 6. As in the production task, imitations of the small fractions 1/8 and 1/6 tend to be too large, and imitations of the complementary large fractions 7/8 and 5/6 tend to be too small. Symmetric distortion in the direction of 1/2 has also been described by Fraisse (1956, Chapter IV) and Povel (1981). Note, however, that the effect is absent in our data for the fractions (1/4, 3/4) closest to 1/2.

C. Choice among Alternative Imitation Functions: Rejection of Accuracy and Full-Concatenation Models

To test the four alternative possibilities for imitation performance outlined in Section IV,A, we calculated deviations between the observed imitation function and tbe function expected from that alternative for each replication within each subject's

[29]If we add to this model the assumption that the component operations are stochastically independent, it follows that the variance of F in imitation must be at least as great as the variance of F in production. It must also be at least as great as the variance that would be induced in the production of F by virtue of variability in the N-values on which responses are based in the judgment task. Given plausible assumptions, the SD that measures this induced variability can be estimated by multiplying the SD of the appropriate PMF from the judgment task by the derivative of the production function $\mathbf{P}(n)$ at the appropriate n-value. [Since $\mathbf{P}(n) \cong n$ for $n \geq 1/4$, this derivative is close to 1.0 for $n \geq 1/4$.]

[30]Imitation would also combine the errors of judgment and production if, for example, it shared just an input process with the former and just an output process with the latter *and* if these two processes were fully responsible for the errors in their respective tasks.

[31]It is a limitation of the experiment that other fraction values, such as those *judged* to be equivalent to simple fraction names, were not used as stimuli for imitation.

data, based on results from the matched procedures of Experiments 1, 3, and 5. For example, for Alternative 2 we calculated the contrast $\mathbf{J}_1^{-1} - \mathbf{I}_5$ for each fraction; insofar as this alternative is valid, these contrasts (whose means over subjects are displayed in row 9 of Table I) should be close to zero.[32]

Numerical examples for Alternatives 2 and 4 may be helpful; we shall base them on second replication data from PZ, for $f = 1/8$ ($bf = 125$). These examples are clearer when the argument of **J** is expressed in fractional intervals (in msec) rather than fractions. Thus, for Alternative 2, $\mathbf{J}(1/6) = \mathbf{J}(166.7) = 96.6$, and $\mathbf{J}(1/4) = \mathbf{J}(250) = 206.2$. Linear interpolation gives $\mathbf{J}(188.6) = 125$, or $\mathbf{J}^{-1}(1/8) = \mathbf{J}^{-1}(125) = 188.6$. According to Alternative 2, this value should be equal to $\mathbf{I}(1/8) = 161.1$. The contrast is $\mathbf{J}^{-1}(1/8) - \mathbf{I}(1/8) = 27.5$ msec.

For Alternative 4 we need $\mathbf{P}[\mathbf{J}^{-1}(1/8)]$, and from above we have $\mathbf{J}^{-1}(1/8) = 188.6$. We therefore need $\mathbf{P}(188.6)$. Second replication data from PZ give $\mathbf{P}(1/6) = \mathbf{P}(166.7) = 170.3$ and $\mathbf{P}(1/4) = \mathbf{P}(250) = 242.0$. Linear interpolation gives $\mathbf{P}(188.6) = 189.2$. According to Alternative 4, this value should be equal to $\mathbf{I}(1/8) = 161.1$. The contrast is $\mathbf{P}[\mathbf{J}^{-1}(1/8)] - \mathbf{I}(1/8) = 28.1$ msec.

We have used three methods to compare the relative goodness of fit of the four alternatives to our data. Since none of these methods is ideal, but taken together they point clearly in one direction, we mention results from all three. We restrict our attention to the seven fractions $1/8 \leq f \leq 7/8$ for which we were able to calculate contrasts for all four alternatives. For Alternatives 1 through 4, respectively, the numbers of individual subject contrasts (of 21 possible) that reach significance are 3, 8, 2, and 7, respectively, favoring Alternatives 1 and 3. The numbers of tests of means over subjects that reach significance are 2, 0, 0, and 0, however, indicating more consistency over subjects in the failures of Alternative 1 and thereby favoring Alternative 3. The mean squared deviations (contrasts) for the four alternatives are 514, 972, 148, and 1454, respectively, clearly favoring Atlernative 3; the same ordering is observed for the mean squared deviations associated with the three smallest fractions, which fall within the range of our most interesting and surprising findings.

Taken together then, our results favor Alternative 3 (imitation consistent with production) for the range of fractions we examined and permit us to reject the two alternatives considered in Section IV,A (accurate imitation, and the full-concatenation model) that have been previously considered for longer durations.

Further evidence bearing on the choice among the four alternatives can be found in relations among the variabilities of performance in the three tasks (Fig. 6). First, the SD functions for $\mathbf{I}_5$ and $\mathbf{P}_3$ are strikingly similar in form, again favoring Alternative 3. The increasing divergence of the two functions with size of the produced fraction is statistically significant, however ($t_2 = 8.4$; $p < .02$). In the context of the mechanism proposed in Section IV,A, for Alternative 3, this divergence could arise if the value of the common internal representation is more variable when it is derived from f than from n

[32]Note that nonlinearities in the computation for Alternatives 2 and 4 result in discrepancies between the *mean contrasts* displayed in rows 9 and 11 of Table I and *contrasts of the means* of the components in rows 1, 3, and 5.

and if this variability difference grows with f. Second, although the SD of imitation is no smaller than the SD of production (as required by the full-concatenation model together with a stochastic independence assumption; see footnote 29), the SD of imitation *is* smaller than the corresponding variability measure associated with the judgment task, which violates an additional requirement of that model. This observation argues further against Alternative 4.

D. Implicit Scaling of Beat Fractions from Imitation and Production Data

Performances in the judgment, production, and imitation tasks interest us primarily because of the light they and the relations among them can shed on underlying timing mechanisms. As already discussed, however, the judgment and production tasks can also be regarded as two different methods for establishing a psychophysical scale—a function that relates the beat fraction f and its subjective magnitude. In each method the scale is established by identifying the subjective magnitude with an n-value. Both methods are explicit in that the subject's response is identified directly with one of the terms in the function. If we believe there is one "true" scale, then the fact that the two methods disagree implies that the scale derived from at least one of them is incorrect. As discussed in Section III,G, this difficulty also arises in other perceptual domains and has been attributed to effects on response-generation processes that distort the inferred association between stimuli and their internal representations. In Section III,F, we considered one such explanation (the existence of a minimum RT) for the systematic error we found in production.

The assumption required to use judgment and production tasks as explicit scaling methods—that responses accurately reflect magnitudes of the internal representations that are the objects of interest in psychophysics—is therefore subject to question. We can replace this assumption by a much weaker one if we use an *implicit scaling procedure* in which the scale is derived by combining data from production and imitation. The weaker assumption permits response biases or other distorting effects associated with responses to exist and requires only that those that operate in the generation of the timed response (F) in production also operate in the generation of the timed response in imitation. Under this assumption (which we use in developing the model described in Fig. 7) if a stimulus n_0 in production leads to the same timed response as does a stimulus f_0 in imitation so that $\mathbf{P}(n_0) = \mathbf{I}(f_0)$, then the internal representations of n_0 and f_0 that are used by the (common) response generation processes must have equal magnitude. The equation $\mathbf{P}(n) = \mathbf{I}(f)$ therefore establishes a scale relating n and f that is free of response effects; it may be written $n = \mathbf{P}^{-1}\mathbf{I}(f)$.

Let us consider what the four simple alternatives for imitation (Section IV,A) imply about the resulting implicit scale:

1. *Accurate imitation*. Since $\mathbf{I}$ is the identity transformation in this case, $n = \mathbf{P}^{-1}\mathbf{I}(f) = \mathbf{P}^{-1}(f)$ or $f = \mathbf{P}(n)$, so that the implicit scale is the same as the scale based on production.

2. *Imitation consistent with judgment.* Here $\mathbf{I} = \mathbf{J}^{-1}$. Hence $n = \mathbf{P}^{-1}\mathbf{I}(f) = \mathbf{P}^{-1}\mathbf{J}^{-1}(f)$; or $f = \mathbf{JP}(n)$. Since $\mathbf{P}(x) > x$ and $\mathbf{J}(x) < x$ for small x, the relation between f and n specified by the implicit scale depends on relative magnitudes of the errors in $\mathbf{J}$ and $\mathbf{P}$.

3. *Imitation consistent with production.* Here $\mathbf{I} = \mathbf{P}$. Hence $n = \mathbf{P}^{-1}\mathbf{I}(f) = \mathbf{I}^{-1}\mathbf{I}(f) = f$, so the implicit scale is free of systematic error (veridical).

4. *Imitation combines errors.* Here $\mathbf{I} = \mathbf{PJ}^{-1}$. Hence, $n = \mathbf{P}^{-1}\mathbf{I}(f) = \mathbf{P}^{-1}\mathbf{PJ}^{-1}(f) = \mathbf{J}^{-1}(f)$, so that the implicit scale is the same as the scale based on perceptual judgment.

We have seen above that results of the imitation experiment favor Alternative 3. One implication is that despite the inaccuracy (and inconsistency) of the explicit scales based on judgment and production data, the implicit scale based on combining results from the two tasks with a common response (imitation and production) is free of systematic error.

V. A SHARED-PROCESS MODEL OF THE PERCEPTION, PRODUCTION, AND IMITATION OF BEAT FRACTIONS

In this section we present an information-flow model of performance in our three tasks. It is a minimal model in that we make as few assumptions as we can and limit ourselves to accounting for major features of the data. We think of each task as involving processes that perform input, translation, and output functions, and a principle of parsimony leads us to assume that different tasks share whatever processes they can. Given this starting point, the model incorporates the minimum possible number of constituent processes.

A. Definition of the Model

The processes in the model responsible for judgment performance are represented by the two upper boxes in Fig. 7. A time-pattern stimulus f generates an internal uncategorized or "analog" representation by a transformation $\mathbf{T}_{fa}$. (We call the representation "analog" only to indicate that it does not reflect a categorization of the stimulus that maps 1-1 onto fraction names.)[33] This representation must then be converted by a transformation $\mathbf{T}_{aN}$ into an internal "name" representation to generate the required fraction-name response N. The resulting compound transformation is denoted $\mathbf{T}_{aN}\,\mathbf{T}_{fa}$.

The processes in the model responsible for production performance are represented by the two lower boxes in the figure. A fraction-name stimulus n is converted into an internal analog representation by a transformation $\mathbf{T}_{na}$, which is then used to generate

[33]One possibility is that this representation encodes both fraction (normalized marker interval) and beat interval. An alternative is that the beat interval is reflected by the rate of an internal clock or accumulator so that the analog representation has to encode only the marker interval in terms of the count or value accumulated.

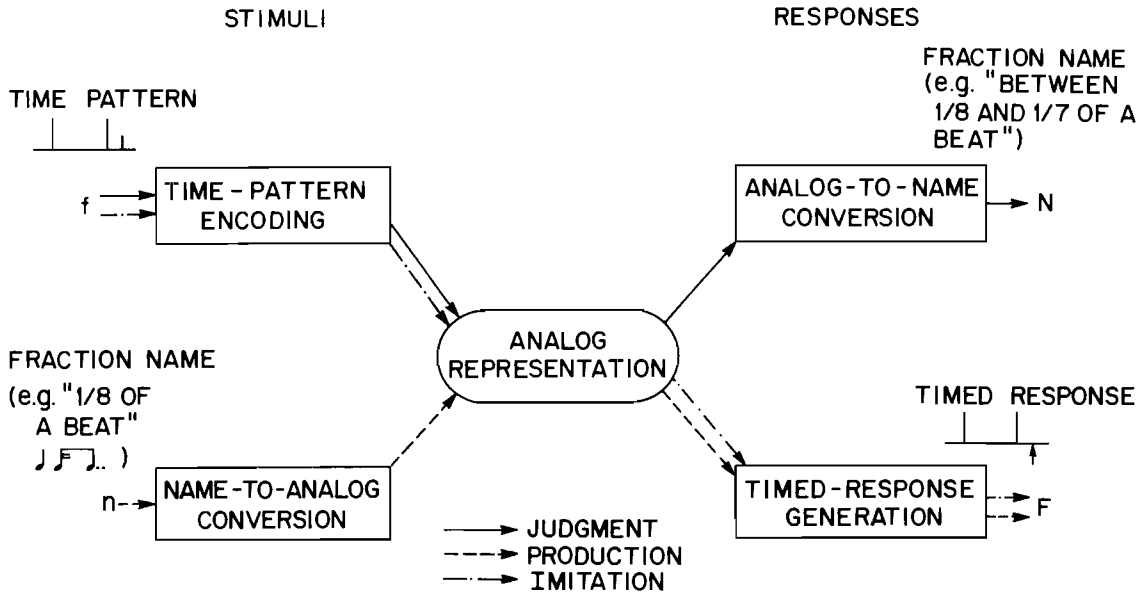

Fig. 7. A shared-process information-flow model of the perception, production, and imitation of beat fractions. The model incorporates four processes that convert time-pattern (*f*) or fraction-name (*n*) stimuli into time-pattern (*F*) or fraction-name (*N*) responses and that make use of a common intervening representation. Transformations carried out by the four processes are symbolized by $\mathbf{T}_{\text{input,output}}$. Paths of information flow for the three tasks are represented by unbroken, broken, and dotted arrows, respectively.

the required timed response by a transformation $\mathbf{T}_{aF}$.[34] Since no feedback process has been incorporated in this account and judgment and production share no common processes, the inconsistency between perception and production is not paradoxical.

A full-concatenation model of imitation (Alternative 4) would most naturally be represented by a system in which the upper and lower pair of processes had separate intervening representations, instead of the common analog representation shown in Fig. 7. Information could then not flow directly from time-pattern encoding to timed-response generation. Instead, a covert response output of the pair of processes used in judgment would become the input for the pair of processes used in production; the resulting compound transformation converting *f* to *F* would be $\mathbf{T}_{aF}\,\mathbf{T}_{na}\,\mathbf{T}_{aN}\,\mathbf{T}_{fa}$. Because such a model can be rejected, we adopt a *partial-concatenation model* of imitation, which shares only the encoding process of the judgment task and the response-generation process of the production task and makes use of an internal representation that is common to the two tasks. The resulting compound transformation converting *f* to *F* is $\mathbf{T}_{aF}\,\mathbf{T}_{fa}$.[35,36]

[34]Again there are several ways in which the (subjective) beat interval might be represented to provide the information that must be incorporated with the fraction name to define the response.

[35]Note that it is only because production and perception errors are not compensatory (i.e., do not conform to the feedback model) that we can discriminate a partial- from a full-concatenation model of imitation.

[36]An alternative two-process model of imitation in the same spirit would separate the information flow in judgment and production into phases that precede and follow the establishment of internal *name* repre-

B. Restrictions on the Four Processes

Experiments to be described in Sections VI and VII help further to elucidate performance in the judgment and production tasks, and will eventually help to flesh out the skeleton shown in Fig. 7. Even with the results presented thus far, however, if we assume the structure of the model some interesting and surprising inferences can be made about the relations among the transformations $\mathbf{T}_{fa}$, $\mathbf{T}_{aN}$, $\mathbf{T}_{na}$, and $\mathbf{T}_{aF}$ carried out by its four component processes. Given these four transformations, there are six transformation pairs; our data permit an inference about the relation between the members of each pair.

The starting points for these inferences are idealizations of four of the properties that appear to characterize performance in the three tasks. The four properties are as follows:

a. $\mathbf{J} \neq \mathbf{E}$. (There are systematic errors in judgment.)
b. $\mathbf{P} \neq \mathbf{E}$. (There are systematic errors in production.)
c. $\mathbf{P} = \mathbf{I}$ (When $n = f$, response times in imitation and production are the same.)
d. $\mathbf{I} \neq \mathbf{J}^{-1}$. (The response fraction F in imitation is not equal to the name N associated with the same f in judgment.)

To make the inferences, we start by using the model to write each of the functions $\mathbf{J}$, $\mathbf{P}$, and $\mathbf{I}$ in terms of the pair of transformations they reflect: $\mathbf{J}^{-1} = \mathbf{T}_{aN}\mathbf{T}_{fa}$, $\mathbf{P} = \mathbf{T}_{aF}\mathbf{T}_{na}$, and $\mathbf{I} = \mathbf{T}_{aF}\mathbf{T}_{fa}$. The inferences are as follows:

1. From property (c) we have $\mathbf{T}_{aF}\mathbf{T}_{na} = \mathbf{T}_{aF}\mathbf{T}_{fa}$, or $\mathbf{T}_{na} = \mathbf{T}_{fa}$. The two input transformations are therefore the same, and hence the internal (analog) representations of stimuli n and f have the same magnitude when $n = f$. This corresponds to the observation that the implicit scale relating f to n (Section IV,D) is free of error. Identity of the input transformations of n and f suggest that performance is not an accidental property of input processes; changes in details of the time-pattern stimulus should therefore not have major effects on performance. Evidence favoring this suggestion is presented in Sections VI,A, and VI,B.
2. From property (a) we have $\mathbf{J}^{-1} \neq \mathbf{E}$ or $\mathbf{T}_{aN}\mathbf{T}_{fa} \neq \mathbf{E}$ and hence $\mathbf{T}_{aN} \neq \mathbf{T}_{fa}^{-1}$. (Not surprisingly, given errors in judgment, its input and output transformations are not inverses.)
3. Combining (1) and (2), we find $\mathbf{T}_{aN} \neq \mathbf{T}_{na}^{-1}$. Thus, the transformations analog to name (in judgment) and name to analog (in production) are not inverses.
4. From property (b) we have $\mathbf{T}_{aF}\mathbf{T}_{na} \neq \mathbf{E}$, or $\mathbf{T}_{aF} \neq \mathbf{T}_{na}^{-1}$. (Not surprisingly, given errors in production, its input and output transformations are not inverses.)
5. Combining (1) and (4), we find $\mathbf{T}_{fa} \neq \mathbf{T}_{aF}^{-1}$. Thus, the transformations time-pattern stimulus to analog (in judgment and imitation) and analog to timed response

---

sentations rather than (the earlier) internal *analog* representations. We prefer our alternative because it seems less likely to us that an interesting or plausible transformation (other than the identity transformation) would relate stimulus or response *names* to their internal representations than that such a transformation would relate stimulus or response *times* to their internal representations.

(in production and imitation) are not inverses. This could also have been inferred from $\mathbf{I} \neq \mathbf{E}$.

6. From property (d) we have $\mathbf{T}_{aF}\mathbf{T}_{fa} \neq \mathbf{T}_{aN}\mathbf{T}_{fa}$, or $\mathbf{T}_{aF} \neq \mathbf{T}_{aN}$. In other words, the two output transformations are distinct (unlike the two input transformations): values of the N and F derived from the same internal (analog) representation are distinct. A difference between the output transformations for N and F makes it plausible that changes in response details might influence performance in production and imitation; some tests of this possibility are presented in Section VII.

VI. FURTHER ANALYSIS OF PERCEPTUAL JUDGMENT

In Sections VI and VII we report results of our search for explanations of the errors associated with small fractions in the judgment and production tasks; we describe four variations of the judgment task and three variations of the production procedure. Our aim in most of these experiments was to determine not whether there was *any* effect of changes in experimental conditions, but whether there were any effects large enough to suggest major sources of the performance errors.

A. Attention Shifts and Delays: Effect of Marker-Click Pitch (Experiment 6)

The presented fraction is defined by the difference between the onset times of the beat and marker clicks. The *subjective* occurrence time of a click, however, may differ from its objective time by an amount that depends on perceptual delay (possibly influenced by the amount of processing required to mark its occurrence). To the extent that the perceptual delays of the beat and marker clicks differ, the presented fraction that is judged subjectively equal to a fraction name will differ from the objective fraction that corresponds to that name, even in the absence of other perceptual distortions. As mentioned in Section III,D, we felt that equality of perceptual delays was a plausible starting assumption for beat and marker clicks. In this section and the next we report results that bear on its validity.

For the perceptual judgment data described thus far, the beat and marker clicks had different pitches. In one possible explanation of the judgment errors, perception of the marker click is assumed to be delayed by the shift of attention from the pitch of the beat clicks to the pitch of the marker click. (For example, findings by Van Noorden, 1975, suggest that the delay might increase with the pitch difference by about 100 msec/octave.) Suppose that this attention shift can be initiated, and possibly completed, after the beat click but before the marker click, if there is enough time between them. (The marker pitch could be learned from earlier trials.) Suppose further that if the shift has not been completed before the marker click, a time interval is required for the marker to attract attention, whose duration decreases with time after the beat click; perception of the marker is delayed until the attention shifts. This

would explain both the judgment error for small fractions and its decrease in magnitude for larger fractions.

This hypothesis implies that the errors associated with small fractions should be influenced by any manipulation that alters the time to shift attention, such as variation of the pitch difference between the beat and marker clicks.

A second reason to suspect that the pitch difference may be implicated in the judgment errors is based on its possible influence on perceptual organization of the series of clicks into sequential groups (Woodrow, 1909, 1951) or simultaneous streams (Bregman, 1978).

To investigate the effect of pitch differences, we had two subjects (PZ and SB) perform in the procedure of Experiment 1 with $N = 1/8$, one second beat intervals, and marker clicks of 1700, 2500, and 3000 Hz; the beat-click frequency was always 3000 Hz. The frequency of the marker click remained the same for 75 consecutive trials, and **J**(1/8) was derived from the last 50 trials of the staircase procedure. According to the attention-shift hypothesis, **J**(1/8), the f-value associated with $N = 1/8$ should be greater (and closer to 1/8) when marker and beat clicks are closer in pitch.

An analysis of variance failed to show a significant effect on **J**(1/8) due to the frequency variation: for marker-click frequencies of 1700, 2500, and 3000 Hz, **J**(1/8) had mean values of $f = 79.2$, 69.6, and 69.9 msec, respectively, with a standard error (based on 2 *df*) of 2.1 msec (a nonsignificant effect in the wrong direction). These results make unlikely an explanation of the estimation errors in terms of the time to shift attention along the pitch continuum.

B. Time Marking by Onset versus Offset: Invariance of Judgment with Prolonged Markers (Experiment 7)

In general, one might expect the internal response to any brief stimulus to differ from the stimulus itself in both shape and duration (see Sternberg & Knoll, 1973, Sec. IV; Fastl, 1977). Furthermore, the subjective occurrence time of a stimulus should depend on the particular feature of the internal response used to mark it. If the internal responses produced by the beat and marker clicks were different or if different features were used to mark their occurrence times, these differences by themselves could produce the observed judgment errors. For example, if the subjective duration of a time interval delimited by a pair of clicks corresponded to an interval delimited by the *onset* of the internal response produced by the first and the *offset* of the internal response produced by the second, the subjective duration would be greater than the objective duration, defined as the difference between click onset times.[37]

To test this possibility we conducted a small perceptual judgment experiment

[37]If auditory signals are presented in close temporal proximity, the internal representation of one (especially the second) is probably affected by the presence of the other (e.g., Duifhuis, 1973; Fastl, 1977; Penner, 1974). Thus, forward masking causes the first of two clicks to elevate the detection threshold of the second. Such effects become negligible with delays of at most 100 msec, however, and are therefore unlikely to be important in determining the judgment error.

(with PZ as the only subject). We used the procedure and stimulus values of Experiment 2 and compared the normal time-pattern stimuli (with all clicks 5 msec in duration) with stimuli in which the marker duration was 62 msec. Let us assume that relative to its onset, the perceived offset time of a tone burst is delayed by about the same amount as its duration is increased. Given the hypothesis, then, we are led to expect an increase in $f = \mathbf{J}(N)$ of about 57 msec in the prolonged marker condition. Instead, we obtained no change: over six fraction names the measured mean increase was a negligible .6 ± 1.9 msec. (The SE is based on variability among the effects on PMF means for N = 1/7, 1/6, 1/5, 1/4, 1/3 and 1/2. For N = 1/8 there were insufficient data to generate a PMF.) There was neither a main effect of marker duration nor an interaction of marker duration with fraction size.

It is reasonable to suppose that any feature of an internal response whose occurrence time is invariant with changes in stimulus duration is located at or near the beginning of that response. The absence of an effect of marker duration therefore argues that the onset rather than the offset of the marker response is the critical feature that determines its subjective occurrence time, and suggests that the systematic judgment errors cannot be attributed to different features of the internal response being used to define the occurrence times of beat click and marker click. (If the offset rather than onset of the *beat* click were used by subjects, the resulting judgment error would be in the wrong direction.)[38]

C. A Test of the Rate Constancy of Subjective Time between Beats: Effect on Fraction Perception of Fraction Location Relative to the Beat (Experiment 8)

In the perceptual judgment experiments reported thus far, the conditions for which systematic errors are largest have two features in common: first, the interval to be judged is small, and, second, it occurs in close temporal proximity to (indeed, is bounded by) the beat click. Suppose that subjective time during the beat interval was inhomogenous in the sense that relative to physical time it elapsed faster near the beat and more slowly elsewhere in the beat interval. Then small fractions defined by intervals near the beat would be overestimated, as observed, but the same small fractions elsewhere in the beat interval, and large fractions initiated by the beat, might not be. To determine how proximity to the beat of the interval being judged affects perceptual judgment, we instructed subjects PZ and PF to judge whether the interval between a pair of marker clicks was larger or smaller than 1/8 of a beat for marker pairs at six different locations within the beat interval.

The beat and marker clicks were 5-msec tone bursts of 3000 and 2500 Hz, respectively. We used four intervals between markers (50, 60, 70, and 80 msec) chosen based on earlier results to permit us to estimate PMFs for judgments relative to 1/8 of

[38]We conjecture that this finding reflects a general property of the perception of musical timing and rhythm: the dominance of the sequence of time intervals between the *onsets* of successive notes (attacks) and the relative unimportance of offset times, which probably serve articulative rather than timing functions.

a 1-sec beat interval. (If effects of proximity to the beat in this initial experiment had been large, of course, this set of intervals might not have provided sufficiently complete functions at all proximities.) On each trial we presented clicks for three beats—B_1, B_2, and B_3; subjects were asked to imagine the next two beats, B_4 and B_5. A marker-click pair, M_1 followed by M_2, defined one of the four intervals, and was located at one of six positions within the sequence of beats, as follows: (1) symmetric about the midpoint of B_3 and B_4, (2) such that M_2 preceded B_4 by 100 msec, (3) such that M_2 was simultaneous with B_4, (4) such that M_1 was simultaneous with B_4, (5) such that M_1 followed B_4 by 100 msec, (6) symmetric about the midpoint of B_4 and B_5.

Position 4 is, of course, the arrangement that had been used in our previous experiments in which the interval to be judged is initiated by the beat, except that the event that marks the beat is a marker click rather than a beat click. (For this position, then, beat B_4 is signaled by a click and is not imaginary, unlike the other positions.) At position 3 the interval to be judged is terminated by the beat. At positions 1 and 6 the interval to be judged is as far as possible from any beat. The six positions combined with four marker intervals defined 24 stimuli that were presented in random sequence (method of constant stimuli).

We found no systematic effects of proximity to the beat on either PMF means [values of J(1/8)] or SDs. Over the two subjects, average PMF means for the six positions are 57.5, 61.7, 55.9, 60.0, 58.3, and 56.7 msec, respectively; (rms) average PMF SDs are 9.4, 10.9, 7.0, 5.5, 12.5, and 4.4 msec, respectively. Over the six positions the average PMF mean is 58.3 msec, and the (rms) average SD is 8.8 msec.[39] Again, the judgment error is surprisingly large: 58.3 msec is about 53% smaller than the correct value of 125 msec.[40] These results show that the judgment error depends neither on the judged interval being bounded by a beat click nor on the proximity of the judged interval to the beat. The rate at which subjective time elapses during the beat interval appears to be constant.

D. A Constraint on the Precision of Dual Time Judgments and Its Implications for Timing Models and the Use of Feedback

A further variation of the perception task revealed an interesting and unexpected limitation in the judgment of time intervals. In an extension of Experiment 8, we

[39]It is instructive to compare these results to findings for the same two subjects in Experiments 1 and 2. Like Experiment 2 the present experiment involved a method of constant stimuli rather than a staircase procedure; like Experiment 1 the present experiment called for a narrow range of f-values. Since results of Experiments 1 and 2 at $N = 1/8$ for PZ and PF were similar, we have combined them to obtain a PMF mean of 62.5 msec and an SD of 7.7 msec. The present experiment produced almost identical values, suggesting that uncertainty from trial to trial about the position of the interval to be judged (which was much greater in the present experiment) is an unimportant factor in judgment performance.

[40]Because the two marker clicks had the same frequency (2500 Hz), these results also provide further evidence against the notion that the estimation errors result from a pitch difference between the clicks bounding the interval to be judged.

instructed one subject (PZ) first to estimate the duration of the brief interval bounded by the marker clicks (relative to 1/8 of a beat), as in the main experiment, and then also to judge whether the longer interval between the last beat click and the marker pair was less than or greater than one beat interval (i.e., marker clicks before or after B_4). The subject was instructed to perform accurately in judging the brief interval (primary task) and, having done so, to judge the long interval (secondary task) as accurately as he could.

The need to make the long-interval judgment did not substantially alter either the mean or the SD of the duration PMF: without the secondary task these parameters were 54.4 and 6.0 msec, respectively; with the added task they were 58.1 and 7.9 msec, respectively. On the other hand, the subject's precision in judging the long interval appears to have been greatly impaired by having also to judge the brief interval. One measure of the loss in precision is obtained by comparing performance in the secondary task to earlier performance (Experiment 1, $N = 1$), judging only the position of a single marker click relative to the beat; this comparison reveals that the SD of the PMF from the secondary task is more than *10 times* the SD obtained in the single-judgment, single marker-click procedure.[41] (In Experiment 1 the mean and SD of the PMF were 962.5 and 48.8 msec, respectively, versus 1114.5 and 502.6 msec, respectively, in the secondary task.)[42]

It is helpful to consider this observation in relation to a particular class of mechanisms that may underlie the timing process. One candidate for the analog representation in the information-flow model of Section V is the value attained by an internal clock or accumulator. (See Creelman, 1962; Treisman, 1963; Wing, 1973; Eisler, 1975; and Getty, 1976, for particular realizations of this idea.) In the judgment task, for example, the hypothesized clock starts with the initial event defining the interval and stops with the terminal event. (Alternatively, the current value of the clock is "saved" when the terminal event is detected.) Results of the dual-task variation of Experiment 8 require elaboration of such clock models to explain why the two successive intervals (the long interval from the beat to the first marker and the short interval from the first to the second marker) could not both be accurately judged.

One possibility is that the timing process permits only intervals that are similar in duration to be accurately classified in quick succession or concurrently. For example, the clock might have an adjustable rate: a slow rate for accurately judging large intervals and a fast rate for accurately judging small intervals.[43]

[41]Strictly speaking, to control for the possibility that the poor performance in the secondary task might be due to the physical nature of the stimuli (the end of the long interval was defined by two marker clicks rather than one), performance in the secondary task should be compared for conditions that have identical physical stimuli and that differ only by the presence or absence of the primary duration judgment. Although this control is logically necessary, and therefore should be used in further investigation of the phenomenon, it would surprise us if such enormous changes in performance could be explained by such minor variation of the physical stimuli.

[42]Thus, a marker pair in position 1 (500 msec before B_4) was judged to have occurred *after* B_4 with probability about .11; a pair in position 6 (500 msec after B_4) was judged to have occurred *before* B_4 with probability about .22.

[43]This limitation could explain our finding that the multiple-fraction procedure (Experiment 2) elicited

According to a second possibility, the clock cannot be started and stopped rapidly, permitting precise timing of only one of two adjacent intervals and requiring use of an alternative, less precise mechanism for timing the interval from the beat click to the first marker click.[44]

Either of the above possibilities could also account for failure of the feedback model of production (Section III,D), if we assume that the same timer is used for production as for perception. To produce a fraction appropriate for a specified name, subjects must time an interval from the beat click to the *initiation* of the response. Since there is a delay between the start of a response and its actual occurrence, it is possible that having accurately timed when to initiate the response, subjects cannot also accurately judge when the response occurs. Alternatively, suppose that timing for the production of small fractions is initiated by the penultimate beat (Section III,F) but that judgment of fractions of all sizes depends on timing from the final beat. Again, the constraints on timing discussed above would prevent the perceptual mechanisms used in the judgment task from being used to evaluate feedback for small fractions in the production task.

E. The Dependence of Perceptual Judgment on Duration versus Fraction: Effects of the Beat Interval (Experiment 9)

In this section we examine two simple alternative ways to characterize perceptual judgment performance and the mechanisms responsible for judgment errors. One is a *duration model*, according to which the fundamental variable is the duration of the fractional interval. For a specified fraction name N the correct value of this interval can be represented as bN, where b is the beat interval and N is the fraction name; the obtained (matched) value is then bf, the absolute error is $bf - bN$, and the relative error is $(bf - bN)/bN$. According to the duration model, judgment error depends only on the correct duration bN; once that is specified there is no further effect of beat interval on either the absolute or relative error. For example, the mean judgment error for $N = 1/8$ at $b = 1000$ msec should be the same as the mean error for $N = 1/6$ at $b = 750$ msec, since in both cases the correct duration is $bN = 125$ msec. The attention-shifting mechanism considered in Section VI,A exemplifies such a model.

The second alternative is a *fraction model*, according to which the fundamental variable is the fraction, or duration *ratio*. For a specified fraction name N the correct fraction is N itself, the obtained value is f, the absolute error is $f - N$, and the relative error is $(f - N)/N$. According to the fraction model, judgment error depends only on the correct fraction; once that is specified, there is no further effect of beat interval on either the absolute or relative error. For example, the mean judgment error (expressed

---

lower judgment precision than the single-fraction procedure (Experiment 1). It could also explain the finding by Vorberg and Hambuch (1978) that subjects attempting to produce precisely timed rhythmic patterns control the timing with a set of *chained* "timers" that produce approximately equal durations rather than hierarchically *nested* (concurrent) "timers" that produce highly disparate durations.

[44]In developing this possibility, the accuracy with which subjects judge the regularity (equality) of trains of successive intervals (Schulze, 1978) would have to be considered.

TABLE II

Experiment 9: Design and Mean Data[a]

Beat interval (b) (msec)	Fraction name (N) 1/8		1/6		1/4		1/2	
750	.078 (.125)	58 (94)	.108 (.167)	α 81 (125)	.203 (.250)	β 152 (188)	.475 (.500)	δ 356 (375)
1000	.075 (.125)	α 75 (125)	.098 (.167)	98 (167)	.238 (.250)	γ 238 (250)	.450 (.500)	450 (500)
1500	.102 (.125)	β 153 (188)	.158 (.167)	γ 236 (250)	.250 (.250)	δ 374 (375)	.456 (.500)	683 (750)

[a] Left-hand cell entries are relevant to the fraction model. Upper left entry is the obtained mean fraction f; lower left entry (in parentheses) is the target fraction N (equal within each column). Right-hand cell entries are relevant to the duration model. Upper right entry gives the mean obtained duration bf in msec. Lower right entry (in parentheses) gives target duration bN. Pairs of cells marked with the same Greek letter have equal target durations.

as a fraction) for $N = 1/8$ should be the same for all beat intervals. A mechanism in which the beat interval has its effect by controlling the rate of an internal clock exemplifies such a model.

On the basis of the judgment experiments considered above, the two models cannot be distinguished because the beat interval was constant (1 sec). In Experiment 9 we used the procedure of Experiment 1 to compare judgment performance by SB, PF, and PZ for each of four fractions ($N = 1/8$, 1/6, 1/4, and 1/2) at three different beat intervals ($b = 750$, 1000, and 1500 msec). The fractions were chosen so that we could examine performance with the same target *fraction* N at each of three beat rates (permitting a test of the fraction model) and also with the same target *duration* bN at two beat rates each (permitting a test of the duration model). These two possibilities are most easily seen by examining cell entries in Table II. The design is orthogonal with respect to N and b; each of the four columns represents the same target *fraction* (left value in parentheses) for different beat intervals. Greek letters indicate those cells that represent the same target *duration* (right value in parentheses) at different beat intervals. For example, for both of the cells marked β ($N = 1/4$, $b = 750$, and $N = 1/8$, $b = 1500$) the target duration is 188 msec.

Tests of both models involved the examination of row (beat-interval) effects in an appropriate two-way table. Let us consider the fraction model first. Here the two-way table has three rows (beat interval) and four columns (fraction name). If the fraction model is correct there should be neither a row effect nor an interaction of rows with columns: error measures associated with the three cells in each column should be equal. (Means over subjects of the obtained fraction values, shown at the

upper left to each cell in Table II, appear, in contrast, to change systematically with beat interval.)

For the duration model the full design is not orthogonal; tests are made possible by reducing the design. Four fractional intervals—with durations 125, 188, 250, and 375 msec—each appear at a "smaller" and "larger" beat interval in Table II in cells designated by Greek letters. The reduced two-way table therefore has two rows ("smaller" and "larger" beat interval) and four columns (one per fractional interval). If the duration model is correct there should be neither a row effect nor an interaction of rows with columns in the reduced table: Error measures associated with the two cells in each "column" (marked by the same Greek letters in Table II) should be equal. (That pairs of beat intervals differ from column to column does not invalidate the test of this null hypothesis.) The pairs of mean obtained interval values for the same target interval, shown in Table II, do appear to depend little on beat interval. For example, for the two cells marked α, for which the correct duration is 125 msec, we obtained 81 msec at the smaller (750 msec) beat interval and a similar 75 msec at the larger (1000 msec) beat interval. (The analysis will show this independence of beat interval to be an artifact due to averaging over subjects, however.)

We were able to use the same error measure in tests of both models, based on analysis of variance. This was possible, first, because for each model absence of row effects and interactions for its absolute error would imply their absence for its relative error[45] and, second, because the two relative error measures [$(bf-bN)/bN$ for duration and $(f-N)/N$ for fraction] are equal.

Results of the analyses of variance for the two models are shown in the upper and lower halves of Table III. For the group analyses neither test shows a significant main effect of rows (b) nor a significant row-column interaction. Both analyses, however, reveal significant interactions of beat interval and subjects, indicating that there are row effects for individual subjects and thereby violating the models; differences among these effects for individuals are apparently large enough so that they cancel each other or otherwise render the main effects insignificant.

Results of the group analyses are consistent with the possibility that each subject's behavior conforms with one of the two models but that the same model does not apply to all three subjects. We tested this possibility by performing the same analyses for each subject separately; results are shown in the right-hand section of Table III. Both models are rejected by these individual analyses, with all three subjects providing evidence against the fraction model and two subjects providing (somewhat weaker) evidence against the duration model. For PZ, duration accounted for a larger percentage of variance than did fraction (76 versus 27%, respectively). For PF, the ordering was the same but the difference was small (80 versus 75%, respectively). Thus,

[45]Relative error can be obtained from absolute error in each case by dividing by the value of the column factor (correct duration or correct fraction). Suppose a two-way table of absolute errors has no row effects or row-column interactions. Then transforming its cell entries in a way that depends only on the column factor produces a new two-way table that also can have no row effects or row-column interactions. (If there is no row effect within any column before the transformation then there can be none after the transformation.)

TABLE III
Results of Analyses of Variance for Duration and Fraction Models

Model	Factor	p-value from group analysis	p-values from individual analyses		
			PZ	SB	PF
Fraction	Beat interval (b)	n.s.[a]	<.001	<.005	<.01
	Fraction (N)	n.s.	<.001	n.s.	<.001
	Subjects (s)	<.001	—	—	—
	$b \times s$	<.001	—	—	—
	$N \times s$	<.001	—	—	—
	$b \times N$	n.s.	<.001	n.s.	<.025
	$b \times N \times s$	<.005	—	—	—
Duration	Beat interval (b)	n.s.	<.005	n.s.	<.05
	Duration (bN)	n.s.	<.001	n.s.	<.001
	Subjects (s)	<.001	—	—	—
	$b \times s$	<.001	—	—	—
	$bN \times s$	<.001	—	—	—
	$b \times bN$	n.s.	n.s.	n.s.	n.s.
	$b \times bN \times s$	<.01	—	—	—

[a] n.s. = not statistically significant ($p > .05$).

although both simple models can be rejected convincingly, the data favor the duration over the fraction model.

Experiments 1 and 2 had a fixed beat interval of 1 sec and are therefore insensitive to the distinction between fraction and duration. To generalize from the results of Experiment 9, we must demonstrate that the findings do not depend on its special conditions, in which the beat interval was changed within sessions. One test of the invariance of performance is to compare the results of Experiments 1, 2, and 9 for those conditions common to all of them (i.e., a beat interval of 1 sec and fraction names of 1/8, 1/6, 1/4, and 1/2). An analysis of variance with the factors experiment, fraction name, and subject resulted in a significant effect of fraction name $F_{3,5} = 7.64$, $p < .025$, but a nonsignificant effect of experiment $F_{2,4} = .73$. Thus, performance (the size of mean judgment errors) at one beat interval appears not to be influenced by the subject having recently worked with other beat intervals. (This finding is consistent with a common belief about musical performance.)

In summary, although performance is better described in terms of the duration being judged than the fraction, the consistent effects of beat interval require us to reject a model in which duration is the sole determinant: beat interval as well as fractional interval influences the size of the judgment error.[46]

[46] If one considers the beat interval to be a standard against which each fractional interval is compared, this conclusion bears on the typical magnitude estimation paradigm in research on time perception. The

VII. FURTHER ANALYSIS OF PRODUCTION

In this section we describe results from three variations of the production task, aimed at testing hypotheses about major sources of the systematic error for small fractions.

A. Production Errors with Musical Instruments (Experiment 10)

One potential source of the production errors we observed is our choice of finger tapping as a response; our subjects are skilled players but not necessarily skilled finger tappers. It was possible that the errors would disappear if subjects used their instruments instead. We tested them in the one-response production task of Experiment 3 (Fig. 1C) using their instruments: flute (SB), cello (PF), and violin (PZ). On each trial a subject played a single note after the final beat click, attempting to produce an interval that corresponded to a specified fraction between the click and the beginning of the note. We measured the occurrence time of the note with an acoustic energy detector; its threshold was adjusted to be relatively low so that it would be exceeded near the start of the acoustic signal. The fraction name (n = 1/8, 1/2, or 1) remained the same for 25 consecutive trials. In Experiment 3 mean response times (corrected by synchronization) for these three fractions were 139.1, 486.3, and 1023.3 msec, respectively. Corresponding response times in Experiment 10 were 195.0, 480.0, and 1018.3 msec, respectively.[47] One reason the mean difference between experiments is so large at n = 1/8 is that whereas PF differed from the other two subjects in Experiments 3 and 4 in not showing a positive production error at n = 1/8, she did produce a substantial positive error in Experiment 10, bringing her into conformity with the other subjects. SDs of the response delays did not differ systematically between experiments; (rms) average SDs over fractions and subjects were 31.5 msec for finger taps and 34.0 msec for instrument notes. Mean values of **P**(1/2) and **P**(1) were almost identical between experiments, but **P**(1/8) was substantially larger in Experiment 10. Each of the three subjects shows this interaction between n-value and response mode, and an analysis of variance shows it to be significant.[48] That the mean response delay at n = 1/8

existence of a beat-interval effect suggests that results of magnitude estimation tasks that employ an explicit standard interval may depend on the size of the standard. In the absence of an explicit standard, subjects may use an implicit standard. This could increase variability within or between experiments if, for example, the implicit standard varied between subjects or depended on experimental manipulations such as the distribution or range of intervals to be judged.

[47]Values from Experiment 10 are raw response times with no synchronization correction applied. Unfortunately, we collected synchronization data with musical instruments only for PZ; for him, **S**′(0) was 4.2 msec (as compared to −9.0 msec for one-response finger taps). Results from the present experiment are sufficiently clearcut so that the absence of these small corrections is unlikely to affect our conclusions. Note that a synchronization correction would change all three response times by the same amount and would therefore have no effect on differences among **P**(n) values.

[48]The differences between absolute errors at n = 1/8 and n = 1, [**P**(1/8)−1/8]−[**P**(1)−1], for SB, PF, and PZ are 47.6, 49.8, and 57.8 msec, respectively.

associated with the musical instrument production was much larger than that of the tap response suggests that the errors measured with the tap response may *underestimate* the size of errors that occur in a more natural context.

B. Production with Marker-Click Feedback (Experiment 11)

The disparity between performances for small fractions in the perceptual judgment and production tasks led us to reject the feedback model of production, but we did not scrutinize the feedback itself. The feedback available from tapping the finger in the production task probably includes tactile, proprioceptive, and auditory cues, but did not include the marker click we used in the judgment task.[49] Could this difference between the events that terminated the critical intervals in the two tasks explain the perception-production disparity?

We asked this question in a small production experiment with one subject (PZ), in which each finger tap produced a marker click identical to the markers in Experiments 3 and 4, thereby providing augmented feedback.[50] With this procedural change the sequences of clicks in the judgment and production tasks (but not necessarily their timing) become identical. The critical difference between tasks is limited to how the timing of the final (marker) click is controlled: in one case the subject controls it directly by choosing when to tap his finger, attempting to make the click time correspond to a specified fraction; in the other case the experimenter controls the click time, and the subject judges it relative to that fraction. We used both one-response and repeated-response procedures.

For this subject, the mean value of **J**(1/8) over perceptual judgment Experiments 1 and 2 together with the corresponding conditions in Experiments 6 and 9 is 64.3 ± 3.5 msec (SE based on between-experiment variation). For production without marker-click feedback, the mean value of **P**′(1/8) over Experiments 3 and 4 is 158.5 ± 2.9 msec. For production under the new condition with augmented feedback, the mean **P**′(1/8) was 149.0 ± 2.9 msec.[51] (With click feedback the one-response and repeated-response procedures produced values that differed by only 4 msec.) Thus, the effect on **P**′(1/8) of adding marker-click feedback is small and not statistically significant, whereas **P**′(1/8) values from experiments both with and without augmented feedback differ reliably from **J**(1/8). Again, despite the large disparity between production and judgment, the subject appeared satisfied with his performance.

That the disparity was maintained—even when the cues in production that were

[49]The cues comprising finger-tap feedback may have been less salient and punctate than marker clicks, and might therefore have limited the precision of subjects' knowledge of the occurrence times of their responses. That the production error generalizes to played notes (Section VII,A) seems to argue against this as a major basis for the error, however.

[50]We actually ran SB as well, but her data were too few and too variable to be conclusive.

[51]We report uncorrected response times because the validity of any correction based on the tap-click synchronization procedure is suspect under these conditions; with an additional tap-produced click, special perceptual mechanisms may be available on which to base simultaneity judgments of marker click and beat click– mechanisms not available for judgments of the timing of taps versus clicks.

available for use as feedback included the same marker click as in perceptual judgment—supports our suggestion (Section VI,D) that subjects process stimulus information differently in the judgment task from the way they process feedback in the production task.

C. Production of Single versus Multiple Subdivisions of the Beat (Experiment 12)

The existence of large production errors for small fractions suggests a paradox: How can this finding be reconciled with our belief that musicians are able to produce response sequences that fill the beat interval accurately and evenly? Could the production error depend on our use of a single, isolated response? Suppose that beats are salient and relatively precise periodic time references. Then the task of accurately producing a sequence of evenly spaced responses, one or more of which are coincident with beats (certainly a more frequent pattern in music than the production of an isolated offbeat response) might be easier because better information might be available for error correction.

Experiment 12 was designed partly to explore this possibility. The three conditions are shown in Fig. 1F and are designated 1R, 4R, and 5R in accordance with the number of tap responses required. For descriptive purposes it is convenient to define potential responses in five positions, R_1, R_2,...,R_5. Accurate performance would require R_1 and R_5 to be coincident with beat clicks and would require the five tap times to be evenly spaced, dividing the beat interval into quarters. In Condition 1R only one of the responses (R_2) was executed; this is the condition studied in our other production experiments. In Condition 4R the first executed response is R_2, as in 1R, but the three following responses are executed as well; accurate performance would require the last response to be coincident with the beat. In Condition 5R all five responses are executed.

We used a beat interval of .5 sec instead of the 1.0 sec interval used in our other production experiments. The fractional interval associated with $n = 1/4$ therefore corresponds to 1/8 in previous experiments. Pilot work had demonstrated the existence of a production error in the 1R condition with the shorter beat interval; the experiment incidentally tests the generality of the production error at a different beat interval.

If the production error depends on planning and performing an isolated response shortly after the beat, then R_2 should be delayed only in Condition 1R. An alternative explanation of the error is similar to one we considered (and rejected) for the judgment error in Section VI,C: the possible inhomogeneity of subjective time during the beat interval. Because judgment and production appear to depend on timing mechanisms that are at least partially distinct, our rejection of such a possibility for the judgment task does not preclude it as an explanation of the production error. If subjective time elapsed relatively slowly near the beat, then R_2 would be delayed in all three conditions of Experiment 12.

The subjects were PZ and two experienced amateur players, JM and SS. Conditions were held constant for sequences of 25 trials, providing opportunities for detected errors to be corrected on later trials. Subjects tapped with alternate index fingers, chosen so that they used the same finger to produce R_2 in all conditions. We treated R_1 (in Condition 5R) as synchronization data and adjusted all response times by subtraction (as discussed in Section III,A). R_1 occurred early, on average, with the mean $\mathbf{S}'(0) = -29.5$ msec.[52] Results are displayed in Fig. 8.

Condition 1R produced a large and reliable positive mean error of 59.1 ± 16.6 msec, or 47%, thereby generalizing our finding to additional subjects and a shorter beat interval.[53] When R_2 was the initial response in the 4R condition, it was produced with as much delay as in the 1R condition. This result shows clearly that the production error is not a consequence of our requiring an isolated response. The mean occurrence time of R_2 in the 5R condition was 119.7 msec, which corresponds to a negligible negative error. Thus, it is the withholding of R_1 that causes the error in R_2. The accuracy of the interval between R_1 and R_2 in Condition 5R argues against any explanation that depends on a distortion of subjective time near the beat during the production task. Rather than being associated with the first subdivision *after* the beat, the productionerror for small forward fractions is associated with the absence of a response *on* the beat (i.e., with "coming in" shortly after the beat).[54]

Let us consider the occurrence times of the remaining responses in Conditions 4R and 5R. The displaced parallel lines in Fig. 8 fit well, indicating that the mean times

[52]The use of this correction is supported, for the average data, by the fact that it places the mean time of R_5 in Condition 5R within 3 msec of the final beat click. (Indeed, the time between R_1 and R_5 can be regarded as a measure of the subjective beat interval.) As we shall see, however, the remarkable accuracy of the mean response rate (or subjective beat interval) implied by this result is an artifact of averaging and does not apply to the data from individual subjects. Using the notation $\mathbf{P}(n;b)$, for PZ we have $\mathbf{P}(1;500) = 437 \pm 20$ msec, indicating a subjective beat interval that is too short. In contrast, from Experiment 3 for PZ we have $[\mathbf{P}(1;1000)]/2 = 506 \pm 2$ msec. Since the new condition is distinguished by "filling" the interval with repeated taps, the reliable difference may be an instance in production of a "filled-duration illusion" (Michon, 1965; Ornstein, 1969) that has frequently been observed in judgment tasks. Since the synchronization correction depends on a different finger from the finger used for R_2, in evaluating the timing of R_2 we must consider the possibility of a timing difference between fingers that would produce a sawtooth pattern of production times. Figure 8 shows that any such difference is small.

[53]It is important to determine whether the production error depends exclusively on either the target interval bn (duration model) or the target fraction n (fraction model). (We asked this question about perception in Section VI,E.) Comparison of these results to production performance in Experiment 3, in which the beat interval was twice as long, provides a small amount of evidence (from PZ only) bearing on this question. Using the notation $\mathbf{P}(n;b)$, we have for PZ, $\mathbf{P}(1/4;500) = 151 \pm 8$ msec, $\mathbf{P}(1/8;1000) = 164 \pm 6$ msec, and $\mathbf{P}(1/4;1000) = 242 \pm 6$ msec. Whereas the interval error difference between the first two (equal bn) is small and not significant, the fraction-error difference between the first and third (equal n) is large and significant. These results argue against a fraction model for the production error and provide (weak) evidence favoring a duration model.

[54]One possibility is that withholding a response on the beat (R_1) requires the establishment of an inhibitory state that takes time to dissipate and thereby delays R_2. This mechanism could not also lead to the enlargement of short intervals *initiated* by finger taps before the beat, however, that we observed in Experiments 3 and 4. (See Sections III,C and E.)

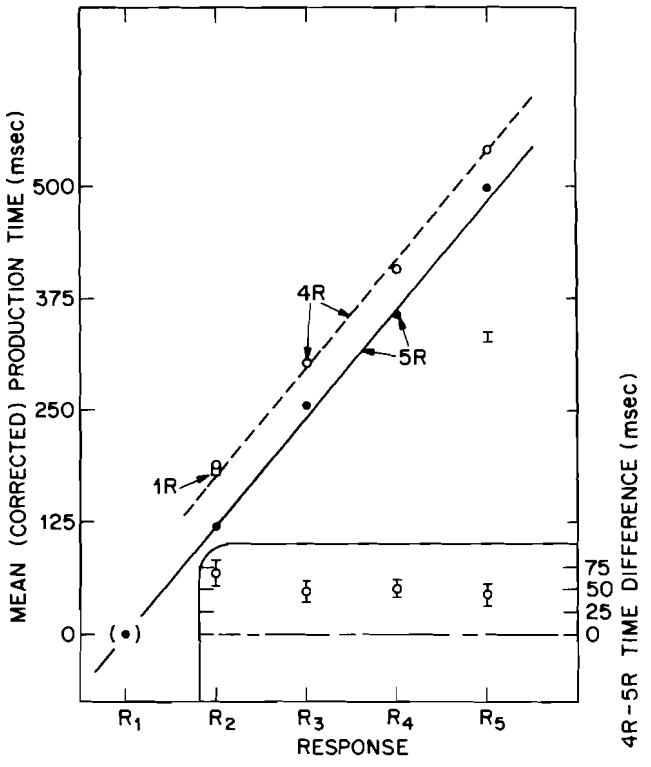

Fig. 8. Mean production times and production-time differences in Experiment 12 (multiple divisions of the beat). The main panel (left-hand ordinate) shows mean production time for each response in the three conditions, after a synchronization correction (−29.5 msec) based on R_1 is applied. (This forces production time for R_1 to be zero.) Parallel lines were fitted by least squares and then adjusted such that the 5R line passes through the R_1 point. The displayed ± 1 SE bar is appropriate for assessing adequacy of the fitted lines. The inset (right-hand ordinate) shows differences between production times of corresponding responses in 4R and 5R conditions together with SEs appropriate for assessing deviations from zero.

between successive responses are close to being equal, both within and between conditions.[55] The remarkable implication for Condition 4R is that the production error in R_2 is propagated through the three succeeding responses. This phenomenon seems best described as *displacement of the subjective beat*. Having produced R_2 with a delay, subjects do not "catch up" because their perception of the train of beats has also been shifted.[56] The lower panel of Fig. 8 displays the occurrence-time differences

[55]If subjects' response rates were accurate, the slope of the fitted lines (mean interresponse time) would be 125 msec/response. The actual mean interresponse time for Condition 5R alone is remarkably close to this value: 123.0 msec. This is an accident of averaging three very different values, however: for JM, SS, and PZ, the mean interresponse times in Condition 5R were 142.6, 117.2, and 109.3 msec, respectively. Such rate inaccuracies imply that even in Condition 5R the final response is far from coincident with the final beat click, again suggesting a surprising insensitivity to the time relation between a response and an external stimulus. (For example, in Condition 5R, PZ's mean R_1 occurred 34.7 msec before its beat click, and his mean R_5 occurred 88.3 msec before its beat click. If R_1 was subjectively coincident with its beat click, then R_5 should have subjectively anticipated the next beat by 53.6 msec, on average.)

[56]That there is a final *actual* beat click that is *not* shifted (and that should be coincident with the final response) raises the same question about beat displacement in 4R as it does about the incorrect response rate in 5R.

of corresponding responses in Conditions 4R and 5R; they range from 67.8 ± 13.9 msec for R_2 to 44.1 ± 12.3 msec for R_5. All three subjects show a decline in this difference with response number, suggesting that the amount of beat displacement is smaller than the full production error. The decline is not statistically significant, however; it remains for another experiment to determine whether all or only some of the production error is transformed into displacement of the beat. In either case, if subjects judge their response delays relative to the subjective beat, then the beat-displacement effect may help to explain why subjects seem satisfied with their delayed responses in the production task, and why the feedback model fails.

VIII. SUMMARY

In a series of experiments we have explored the judgment, production, and imitation (reproduction) of time ratios by three professional musicians. In five initial experiments we found that for small fractions in all three procedures our subjects exhibited systematic and substantial errors. In the judgment task they associated small stimulus fractions with names that were too large (*overestimation*). In both the production task (targets specified by fraction name) and the imitation task (targets specified by fractional interval) our subjects produced intervals that were too large (*overproduction*).

The relation between judgment and production errors requires us to reject feedback models of production, in which a subject uses judgment of the time interval from perceived beat click to perceived response (response feedback) to adjust produced fractions. The approximate equality of production and imitation errors, together with the existence of systematic errors in judgment, argues that imitation is not accomplished simply by concatenating the processes used in judgment and production. Instead, we propose a model containing four internal transformations, in which judgment and production share no transformations and imitation shares one transformation with judgment and another with production. Our data permit us to infer relations among the four transformations. Since production and imitation share the same response, data from these two tasks implicitly define a psychophysical scale for fractions of a beat (a function that relates stimulus fractions to their names) that is independent of potential distortions due to response generation; our results indicate that this implicit scale is free of systematic error, unlike the scales for fractions of a beat that are defined (explicitly) by our judgment and production data.

Results from seven additional experiments increase the generality of our findings and help discriminate among alternative explanations of the judgment and production errors. (1) In the judgment task, changing the pitch of the marker click had little effect, indicating that the overestimation we observed is not a consequence of delays in shifting attention from beat click to marker click. (2) Because performance remained approximately invariant as we altered marker duration, we conclude that subjective onsets rather than offsets of marker and beat clicks were used to mark their occurrence times, discrediting an explanation based on the use of marker offsets. (3) When we varied proximity to the beat of the fractional interval being judged and found no effect on the judgment error, we rejected the possibility of a distortion of

subjective time near the beat. (4) The judgment error depends more on the absolute size (duration) of the judged interval than on its size relative to the beat interval (fraction), but both factors have systematic effects; this finding argues against any model in which either interval alone or fraction alone determines the error. (5) The unexpectedly poor judgment precision we encountered when a subject had to judge two disparate intervals within the same stimulus pattern suggests reasons for failure of the feedback model of production; they are based on the possibility that the timing of two such intervals may be required in the production task if a subject attempts both to time the initiation of a response and to judge when it actually occurs.

In the production task, we found that (a) enriching the potential perceptual cues (feedback) that mark responses had little effect on performance, and that (b) when subjects used their musical instruments to perform the task they produced even larger mean errors for small fractions than in productions with finger taps. (c) Evidence from production variability suggests that the overproduction we observed is not a consequence of the existence of a minimum reaction time. In an experiment requiring multiple subdivisions of the beat interval, we found that (d) the production delay did not depend on the number of responses produced within a beat interval, but did depend on withholding a response on the beat that initiates that interval (so the first response is required to occur after the beat). (e) Responses that followed such an initial production, including one that was supposed to be coincident with the following beat, were delayed almost as much as the initial response, suggesting that displacement of the subjective beat accompanies the production error. If subjects judged their response delays relative to the subjective beat, this would provide another explanation for failure of the feedback model of production.

Among the issues that should be addressed in future research is the effect of musical training on performance in these tasks, and the basis of the inconsistencies we observed within and among skilled performers. To help understand the extent to which our results reflect properties of human timing in general rather than musical training it would be desirable to modify our paradigms to examine the performance of subjects unskilled in the use of musical notation. Studies of additional musicians would perhaps illuminate the differences we observed among our three subjects. More measurements are needed that would permit comparison of production and perception of small fractions versus large fractions (small "reverse fractions"). Our results raise the question whether similar errors are manifested in contexts that are more musical and in performance of actual music. If they are, we need to know how musicians reconcile the perception-production conflict suggested by our experiments, and exactly what role is played by response feedback in human performance requiring precise timing.

ACKNOWLEDGMENTS

We are especially indebted to Max V. Mathews for encouraging our collaboration, to Stephen Monsell, Marilyn L. Shaw, and Robert Teghtsoonian for helpful discussions and advice, and to A.S. Coriell for technical support. We also thank R.D. Luce, D.L. Noreen, R.T. Ollman, S. Roberts, D.A. Rosenbaum, D.L. Scarborough, M.S. Schoeffler, and J. Sloboda for useful comments on earlier drafts, and W.J. Kropfl and M.J. Melchner for technical help.

GLOSSARY

Here we provide a list of the main symbols used in order of appearance in the text, with brief definitions and numbers of the sections in which they are introduced.

b	Beat interval: time interval between one beat and the next. I,B
f	Stimulus fraction: ratio of stimulus interval duration to beat interval duration. I,B
n	Fraction name specified in numerical and musical notation. I,B
N	Fraction-name response in the judgment task. I,B.
$\mathbf{J}(N)$	Judgment function: defines relation between stimulus fraction and fraction-name response, $f = \mathbf{J}(N)$. I,B
F	Produced fraction in production and imitation tasks: ratio of delay of timed response to beat interval. I,B
$\mathbf{P}(n)$	Production function: defines relation between fraction-name stimulus and produced fraction, $F = \mathbf{P}(n)$. I,B
$\mathbf{I}(f)$	Imitation function: defines relation between stimulus fraction and produced fraction, $F = \mathbf{I}(f)$. I,B
PMF	Abbreviation of the term "psychometric function," a function that associates with each stimulus fraction the proportion of "too large" judgment. II,A
$\mathbf{G}^{-1}$	If $y = \mathbf{G}(x)$ denotes an operation that converts x to y, the operation that converts y to x is denoted $\mathbf{G}^{-1}$: $x = \mathbf{G}^{-1}(y)$. (Requires that there is only one value of x for each y.) II,A
$\ln(x)$	Natural logarithm of x. II,A
SD	Standard deviation. II,C
DL	Difference threshold: half the change in stimulus fraction required to change the proportion of "larger than" responses from .25 to .75. II,C
$\mathbf{S}'(0)$	Produced fraction in synchronization task. III,A
$\mathbf{P}'(n)$	Production function uncorrected for response delays: $\mathbf{P}(n) = \mathbf{P}'(n) - \mathbf{S}'(0)$. III,A
$\mathbf{I}'(f)$	Imitation function uncorrected for response delays: $\mathbf{I}(f) = \mathbf{I}'(f) - \mathbf{S}'(0)$. IV,C
RT	Reaction time. III,F
$\mathbf{T}_{fa}$	Transformation of stimulus fraction to analog representation. IV,B
$\mathbf{T}_{aN}$	Transformation of analog representation to fraction-name response. IV,B
$\mathbf{T}_{na}$	Transformation of fraction-name stimulus to analog representation. IV,B
$\mathbf{T}_{aF}$	Transformation of analog representation to produced fraction. IV,B
$\mathbf{E}$	Identity (or equality) transformation, $\mathbf{E}(x) = x$. IV,B

APPENDICES

A. Staircase and Constant Stimulus Methods

1. Collection of Staircase Data

The "staircase" or "up-and-down" procedure used in Experiments 1, 6, and 9 can be regarded as a method of collecting observations in which the stimulus presented on each trial depends on both the stimulus and the response of the immediately previous trial. On the first trial in this procedure a fractional interval (bf) either longer or shorter than the "correct" or "target" interval (bN) is presented. If the subject presses the key indicating that he or she judged the presented fraction to be "too small" relative to the fraction-name target, a longer interval is presented on the next trial. If the subject presses the "too large" key, a shorter interval is presented on the next trial. The subject is required to make one of these two responses; no "equal" response is available. The amount by which the interval is changed is called the *step size*. Use of the procedure depends on the assumption that the probability of a "too large" response increases monotonically with interval duration.

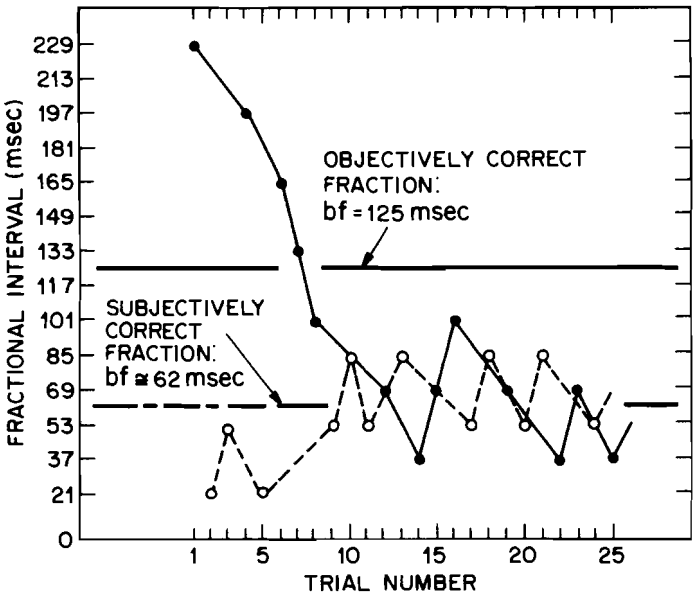

Fig. 9. Example of staircase procedure used in Experiment 1. The subject was PB. The fractional interval presented on a trial is shown versus trial number. Two staircases with regularly interleaved stimulus values were randomly interleaved across trials. Value of the stimulus presented on each trial depended on the history of the staircase selected on that trial. The line leaving a point rises if the response was "too small" and falls otherwise.

The immediate result of applying this stimulus selection rule is to choose an interval for each trial so as to reduce the likelihood of the previous response; the long-term result is to increase the number of presented intervals in the region where the subject is maximally uncertain (proportions of the two judgments approximately equal to each other and to .50), and where the proportion of "too large" judgments increases from being less than .50 to more than .50. We assume that in this region the presented fraction is *subjectively* close in value to the target fraction. In practice more reliable data are obtained if two or more independent staircases are randomly interleaved over trials and contain interleaved arrays of stimulus values. Under these conditions a large step size (e.g., 2.5σ, where σ is the SD of the PMF) has the virtue of producing rapid convergence with minimal bias. (See Kappauf 1967, 1969; Levitt, 1971; and references therein.)

Figure 9 illustrates 25 trials of the initial staircase procedure of Experiment 1 with data collected from PB when he was judging fractions relative to $N = 1/8$. On each trial the subject had to judge whether the presented fraction was too small or too large relative to 1/8 of a beat. The fractional interval (bf) is represented on the ordinate; a horizontal line marks the correct value ($bf = 125$ msec) for $N = 1/8$ and $b = 1000$. Trials are represented on the abscissa. Two staircases started at intervals much smaller and larger than 125 msec. Staircase 1 (broken line) started at 21 msec, and staircase 2 (solid line) started at 229 msec. For each staircase the step size was 32 msec, and the arrays of stimulus values were interleaved, giving a pooled array of stimulus values with 16-msec spacing. The selection of which staircase to use on a trial was random. If an interval was judged too small (large) relative to the target fraction, the next time that staircase was selected a larger (smaller) fraction was presented. In this manner, the staircases converged to a region where the subject was maximally uncertain. In the case shown in Fig. 9, the interval subjectively equivalent to $N = 1/8$, estimated by the mean of the PMF, was 62 msec.

We used two successive staircase procedures in Experiments 1, 6, and 9. First we ran 25 trials with only two interleaved staircases (as exemplified in Fig. 9) to obtain rapid convergence. The starting values (one smaller and one larger) were chosen to be symmetric about bN; 50 additional trials were then run with four interleaved staircases to provide the data we used to estimate the PMF. The new array of stimulus values and the four starting points were determined from the estimated mean and standard deviation ($\hat{\sigma}$) of the PMF for that fraction, based on the first 25 trials. The step size of each staircase was adjusted to be $2.5\hat{\sigma}$,

the pooled array of stimulus values had a spacing of $2.5\hat{\sigma}/4$, and the four starting points were separated from the mean by -9, -2, $+4$, and $+9$ step values. By widely distributing the starting points of the 50-trial series about the mean based on the initial 25-trial series we hoped to minimize bias. The proportion of "too large" responses was calculated for each fractional interval that had been presented,to obtain an empirical PMF.[57]

2. *Psychometric Function*

The plotted points at each stimulus (ordinate) value in Fig. 9 can be used to estimate a PMF. Thus, the fractional interval $bf = 69$ msec was presented on four occasions (trials 12, 15, 19, and 23). The response was "too large" on all but one (trial 15). (Only on that trial is the next bf value in that staircase increased.) The estimated PMF value at $bf = 69$ msec is therefore the proportion .25.

Although the sample based only on the illustrative data shown in Fig. 9 is very small (with only 21 judgments in the range from $21 \leq bf \leq 101$ msec) the estimated PMF shows no reversals: stimulus values bf (in msec), together with their associated sample sizes k and proportions p of "too large" responses, in the form (bf, k, p), are (21, 2, .00), (37, 3, .00), (53, 6, .17), (69, 4, .75), (85, 4, 1.00), and (101, 2, 1.00).

3. *Constant-Stimulus Method*

In the method of constant stimuli of Experiments 2, 7, and 8, the fractional interval on each trial was selected randomly without regard to the subject's prior responses from a set of intervals chosen in advance. In Experiments 2 and 7, for example, each set contained 24 different intervals, covering a wide range and with a spacing that increased with interval size.

B. Additional Details of Design and Procedure

In this section we mention some details of design and procedure of our 12 experiments not discussed elsewhere.

Experiment 1. The order of the judgment, production, and imitation procedures in Experiments 1, 3, and 5 was balanced in a 3×3 Latin square design within each of the two replications: each procedure was studied first, second, and third for some subject. For each subject the order of the three procedures was reversed between the first and second replications. (The Latin squares in the two replications were therefore mirror images.) Three distinct orders of fraction names were paired with the procedures so that the complete design for each replication was a Graeco–Latin square. In each replication for each fraction name, an initial 25 practice trials with only two interleaved staircases provided approximate estimates for the mean and variance of the PMF. Following a brief pause, 50 additional trials were run with four interleaved staircases.

Experiment 2. Fractional intervals were varied by the method of constant stimuli. On each trial the interval between beat click and marker click was determined by random selection without replacement from a set of 24 intervals. The subject was given a brief rest after each cycle through the 24 intervals; 7 such cycles defined a session. To reduce the impact of any effects associated with specific intervals, four different sets of 24 intervals were used in the course of the experiment. (The smallest and largest interval differed from set to set; over all sets intervals ranged from 43 to 891 msec.) The spacing between successively larger intervals increased approximately in the same way (harmonically) as the spacing between successively larger fraction names (1/8, 1/7, ..., 1/2). The first cycle of each session was considered

[57] A staircase procedure is typically designed to produce data that estimate a particular quantile of the PMF, such as the 50% point (the median), rather than the entire PMF. However, results from a Monte Carlo study (Sternberg & Knoll, unpublished) show that the empirical PMF, derived from the application of the staircase procedure to a known PMF, shows little distortion when independent staircases contain interleaved grids of step values, the step size of individual staircases is large, and the underlying PMF is symmetric.

practice and was excluded from the analysis; a session therefore contributed 6 test trials for each of 24 intervals to the analysis. The number of sessions run for SB, PF, and PZ was 4, 3, and 5, respectively.

Experiment 3. Each of two replications of each of the eight production conditions involved 25 trials and was run as part of the balanced Latin square design described for Experiment 1. Eight replications of the synchronization condition and four additional replications of the production condition with $n = 1$ were run after the main experiment; performance in the production condition indicated no systematic change relative to the main experiment.

Experiment 4. Principal conditions included synchronization and eight n-values. Each replication of a condition included 25 trials, or 250 responses. Number of replications per condition per subject varied from 1 to 4 and averaged approximately 2; conditions were run in an irregular order. Plots of results for individual subjects (Figs. 2–4) also show results for conditions (n = 1/5, 1/3) run on only a subset of the subjects.

Experiment 5. Each of two replications of each of the eight imitation conditions involved 25 trials and was run as part of the balanced Latin square design described for Experiment 1.

Experiment 6. The staircase procedure was run as in Experiment 1. Either one or two replications were run per condition; the order of conditions differed across subjects.

Experiment 7. The constant stimulus method of Experiment 2 was used with one session run for each of the two conditions. Intensity of the prolonged marker was reduced from 30 dB above threshold to about 21 dB to make the loudness of brief and prolonged markers more similar.

Experiment 8. In each cycle of 24 trials each stimulus was presented once. Seven cycles defined a session; one session was run per subject. The first cycle of a session was considered practice; each stimulus is therefore represented six times in the data for an individual subject.

Experiment 9. The order of the three beat intervals was balanced across subjects in a 3×3 Latin square design. In the first replication the order of the four N-values differed across subjects, but for an individual subject the order was the same for all three beat intervals. The four N-values within each beat interval were run consecutively before the beat interval changed. In the second replication the order of the 12 conditions was reversed for each subject. Use of the staircase procedure was similar to that in Experiment 1 except that in the second replication the initial 25-trial series was omitted and starting points were based on performance in the first replication.

Experiment 10. We asked subjects to choose a note that they felt they could produce consistently: PZ bowed an open A string on the violin, SB played C above middle C on the flute, and PF bowed D below middle C on the cello. All subjects started with a single replication of 25 trials with n = 1/4 for practice and then completed two replications each for n = 1/8, 1/2, and 1; the order of fraction names was varied between subjects and was reversed from the first to second replication within each subject.

Experiment 11. Two replications (of 25 trials) were run, first in the repeated-response procedure and then in the one-response procedure.

Experiment 12. Each condition was studied in three replications (of 25 trials). Within each subject the order of conditions was balanced. All three subjects are right-handed; SS and PZ performed response R_2 with their right index fingers, and JM performed R_2 with her left index finger.

C. Measures of Location of the Psychometric Function

Both staircase and constant-stimulus methods in our perceptual judgment experiments gave rise to PMFs. Among our considerations in choosing a location measure were (1) our desire to reduce nonadditive distortions in estimates of the effects of interest on underlying processes, (2) our desire to compare the location measure to corresponding measures in production and imitation with the idea that a common underlying process might contribute to all three, (3) our reluctance to assume a functional form for the PMF (such as cumulative Gaussian), (4) the smallness of sample sizes at each stimulus value for individual subjects, and (5) our interest in a spread measure as well.

It is convenient to regard the "true" PMF as a cumulative distribution function and the observed PMF as

an estimate of this function. The most common location measure used in psychophysics is the *median* (the 50% point) of the PMF. Instead, we used an estimate of the *mean;* comparisons of the two revealed extremely good agreement.

It may be helpful first to consider why we favor the mean over the median for the production and imitation experiments. For production, suppose that the time between beat click and timed response is the sum of a timing component $X_t(n)$ that interests us and that varies with fraction name n and other components such as input and output delays, X_s and X_r, that vary from trial to trial but do not depend on target fraction: $X(n) = X_t(n) + X_s + X_r$; $\mathrm{Var}\,(X_s + X_r) > 0$. Then changes with n in the *mean* of $X(n)$ accurately reflect changes in the mean of the process of primary interest X_t, whatever the distributions of X_t and $X_s + X_r$. (The mean of a sum of random variables equals the sum of the means.) This property does not characterize the median, however. It follows, for example, that if an estimate of the mean of $X_s + X_r$ is available, it can be used to "correct" the observed $X(n)$; this is not possible, in general, for the median. The same argument, with f substituted for n, applies to imitation. Similarly, the PMF may reflect processes such as X_s and X_r as well as a timing process that interests us; given plausible assumptions the mean is preferable there also. Furthermore, if we use the mean for production and imitation experiments and wish to compare results across experiments, the mean becomes the favored statistic for judgment experiments.

We used the Spearman-Kärber (S-K) method to estimate PMF means (Spearman, 1908; Epstein & Churchman, 1944; Church & Cobb, 1973). Let stimulus fractions be f_i, $i = u, u+1, \ldots, v$, and let p_i be the proportion of "too large" responses for stimulus f_i, which estimates a corresponding response probability $Pr\,\{L;f_i\}$. Let f_u be a stimulus such that we can assume $Pr\,\{L;f\} = 0$ for $f < f_u$, and let f_v be a stimulus such that we can assume $Pr\{L;f\} = 1.0$ for $f \geq f_v$. Then the S-K estimate of the r^{th} raw moment of the PMF is

$$m_r' = \sum_{i=u}^{v} (p_i - p_{i-1}) \left(\frac{f_i^{r+1} - f_{i-1}^{r+1}}{(r+1)(f_i - f_{i-1})} \right) ;$$

the estimated mean m_1' is obtained by setting $r = 1$.[58,59] The PMF mean can be regarded as a weighted stimulus average, where the weighting function is the derivative of the PMF, a measure of the sensitivity of the response probability to changes in the stimulus. Stimulus values in a region where the PMF rises more steeply contribute more to its mean.

We prefer to avoid strong assumptions about the form of the true PMF, or even about its symmetry. (If the PMF were symmetric then the sample mean and median would estimate the same quantity, of course.) We do feel justified in assuming that the true PMF is nondecreasing. Because our sample sizes are small, however, the *empirical* PMFs are occasionally nonmonotone. In such instances, we have used monotone regression to estimate the best-fitting (least squares) set of nondecreasing proportions $\{p_i^*\}$ from the empirical PMF $\{p_i\}$ before estimating parameters. (See Ayer, Brunk, Ewing, Reid, & Silverman, 1955; Kruskal, 1964; and de Leeuw, 1977.)[60,61]

For the data from Experiment 1, we compared and found excellent agreement between conventional median estimates obtained by applying linear interpolation to the $\{p_i^*\}$, and S-K mean estimates. The two replications for each of three subjects provided six sets of PMFs, with 8 PMFs per set. For the six sets of data, linear correlations of means versus medians ranged from .9995 to .9998, and slopes of the linear regression of means on medians ranged from .999 to 1.018.

[58]This is actually a modified S-K estimator, appropriate for a continuous distribution function (a piecewise linear integrated histogram) rather than a discrete distribution function.

[59]In our application of this method, as shown by the estimation equation, the proportions p_i enter with equal weights. An alternative would be to weight them by considering differences in sample size and binomial variability. Fortunately, for PMFs derived from staircase data, the distribution of observations is approximately triangular and is centered where $p_i \cong .5$; this compensates approximately for differences in binomial variability and justifies the use of equal weights.

[60]This "monotonizing" procedure can probably be improved upon by regarding the $\{p_i\}$ as estimates of values of an underlying continuous rather than a discrete distribution function.

[61]Note that at least one advantage of applying monotone regression is that without it the more conventional quantile measures of location and spread may not be uniquely defined. If only the S-K estimate of location is desired, the monotonizing transformation is not necessary, since it does not alter the estimated mean. We have used it because it influences the spread measure (SD estimate) based on the S-K method and is necessary to permit comparison of quantile with moment estimators.

D. Measures of Spread of the Psychometric Function

The conventional measure of spread (or precision) of the PMF is the DL, defined as half of the interquartile range. We obtained DLs for the PMFs in Experiments 1 and 2 by applying linear interpolation to the $\{p_i^*\}$. We compared these quantile measures to a set of corresponding SD estimates; these were based on m_1' and m_2' values obtained by applying the S-K method to the same $\{p_i^*\}$. (See Epstein & Churchman, 1944; Chmiel, 1976.)

Six sets of PMFs were used for the comparison, obtained from the three subjects in each of the two experiments. For each set of PMFs we determined a constant k such that $\overline{SD} = k\overline{DL}$, where $\overline{SD}$ and $\overline{DL}$ are means over the set of PMFs. (The relation between SD and DL depends on the shape of the PMF. For a Gaussian PMF, for example, $k = 1.48$. The mean of the six k-values we obtained is 1.53, suggesting that our PMFs are approximately Gaussian.) Our aim was to compare the variabilities of the two spread measures. This requires first adjusting them to have the same mean; we did so by multiplying the DLs in each set by the k-value obtained for that set. For each set of PMFs, we determined the variance of each spread measure SD and kDL across replications of the same condition (fraction name) and pooled these variances over fractions. The ratios of Var(kDL) to Var(SD) ranged from 1.5 to 11.0 over the six sets of PMFs, with a mean of 5.7 ± 1.7. This surprising finding indicates convincingly that, at least for our small-sample PMFs, the conventional quantile measure of spread (the DL) is far less reliable than the S-K estimator of the SD, and supports our use of the SD in the present report.

E. Stimulus-Averaging versus Response-Averaging Methods for Deriving a Psychophysical Scale

Our method of determining the judgment function, $f = \mathbf{J}(N)$, was to associate with each value of N the mean of the corresponding PMF. For each of a set of response (N) values this method can be regarded as providing a weighted mean stimulus, where the weighting function is given by the steepness (derivative) of the PMF. Stimulus averaging has at least three virtues relative to the more conventional response-averaging method (to be discussed below): (1) it can be applied to data from single-fraction as well as multiple-fraction experiments; (2) it does not require us to treat response values as being defined on an interval scale, as do response-averaging methods; and (3) each PMF mean presumably reflects only one subjective criterion because it is associated with only one N-category boundary. A response-averaging method depends, in general, on associating a set of more than two N-values with each f-value; the response average therefore reflects more than one criterion on the subjective N-scale. Such a method seems less comparable with the straightforward analysis of production and imitation, in which performance in any condition presumably depends on only one criterion, as in the single-fraction judgment experiment. Oyama (1969) described a similar stimulus-averaging method applied to magnitude estimates of loudness, and argued in its favor by noting virtue (2) above.

Despite these advantages it seemed important to check whether the results we obtained in the judgment experiment could be attributed to the unorthodox method we used to construct the judgment function. Data from the multiple-fraction judgment procedure of Experiment 2 permitted comparison of the two methods. Each of the six central response alternatives in that experiment were defined in terms of two category boundaries, such as "between 1/8 and 1/7." To average responses, we defined the value of each response to be the geometric mean of its two boundaries; the set of such response values associated with each stimulus fraction f was then averaged, again using the geometric mean.[62]

Seven stimulus intervals were defined, containing approximately equal numbers of fractions; the set of average responses within each interval were themselves averaged (using geometric means) and associated with the geometric mean stimulus fraction. The $\mathbf{J}(N)$ functions are truncated in the response-averaging

[62]Use of the geometric mean is common in response-averaging methods; it is equivalent to applying the arithmetic mean to logarithms of response values and is thought to be appropriate when the psychophysical scale is linear on logarithmic coordinates, as in the case of a power function.

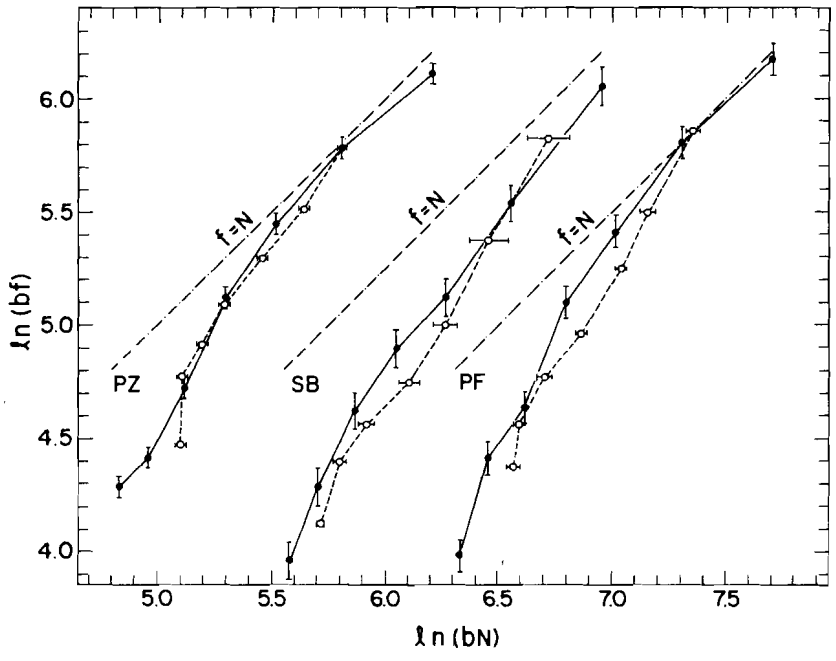

Fig. 10. Comparison of psychophysical scales derived by two different averaging methods from the multiple-fraction judgment experiment. The scale relates fractional intervals (*bf*) specified by time-pattern stimuli to fraction names expressed as intervals (*bN*). Data for SB (PF) are displaced .75 (1.5) log units to the right. Values derived from PMFs (weighted stimulus averaging for specified *N*) are shown by filled circles and unbroken lines. Values derived by response averaging (for specified *f*) are shown by open circles and broken lines. Lines representing $f=N$ are included for reference. Bars represent ±1 SE; SEs are based on variability across replications in the quantity averaged.

method because a stimulus must be excluded if any response to that stimulus falls in one of the end categories ("less than 1/8" or "greater than 1/2"), for which an acceptable response value cannot be defined.

Results from this procedure are shown in Fig. 10, together with results from the stimulus-averaging (PMF) method. For the PMF method, the eight response categories provide seven category boundaries and thus seven values of $\mathbf{J}(N)$.

Estimates shown of ±SE are based on between-replications SDs. These were pooled over fractions for the PMF method. For the response-averaging method, the SDs were smoothed by linear regression of SDs on means. The difference in SE estimates between the two methods can be regarded primarily as a result of the $\mathbf{J}(N)$ slopes being considerably greater than unity.

The scale based on the PMF method for each subject can be seen to be very similar to the scale based on response averaging. Oyama (1968) drew the same conclusion when he compared stimulus- and response-averaging methods for deriving a scale of loudness.

REFERENCES

Adam, N., Castro, A., & Clark, D. Production, estimation, and reproduction of time intervals during inhalation of a general anesthetic in man. *Journal of Experimental Psychology*, 1974, *102*, 609–614.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., & Silverman, E. An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, 1955, *26*, 641–647.

Bregman, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: L. Erlbaum Associates, 1978. pp. 63–75.

Carlson, V. R., & Feinberg, I. Individual variations in time judgments and the concept of an internal clock. *Journal of Experimental Psychology*, 1968, 77, 631-640.

Carlson, V. R., & Feinberg, I. Time judgment as a function of method, practice, and sex. *Journal of Experimental Psychology*, 1970, *85*, 171-180.

Chmiel, J. J. Some properties of Spearman-type estimators of the variance and percentiles in bioassay. *Biometrika*, 1976, *63*, 621-626.

Church, J. D., & Cobb, E. B. On the equivalence of Spearman-Kärber and maximum likelihood estimates of the mean. *Journal of the American Statistical Association*, 1973, *68*, 201-202.

Creelman, C. Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, 1962, *34*, 582-593.

de Leeuw, J. Correctness of Kruskal's algorithms for monotone regression with ties. *Psychometrika*, 1977 *42*, 141-144.

Duifhuis, H. Consequences of peripheral frequency selectivity for nonsimultaneous masking. *Journal of the Acoustical Society of America*, 1973, *54*, 1471-1489.

Eisler, H. Subjective duration and psychophysics. *Psychological Review*, 1975, *82*, 429-450.

Eisler, H. Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, 1976, *83*, 1154-1171.

Epstein, B., & Churchman, C. W. On the statistics of sensitivity data. *Annals of Mathematical Statistics*, 1944, *15*, 90-96.

Fastl, H. Subjective duration and temporal masking patterns of broadband noise impulses. *Journal of the Acoustical Society of America*, 1977, *61*, 162-168.

Fraisse, P. *Les structures rythmiques*. Louvain: Publications Universitaires de Louvain, 1956.

Fraisse, P. *The psychology of time*. New York: Harper & Row, 1963.

Fraisse, P. Time and rhythm perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception, vol. VIII: Perceptual coding*. New York: Academic Press, 1978. pp. 203-254.

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

Getty, D. Counting processes in human timing. *Perception & Psychophysics*, 1976, *20*, 191-197.

Kappauf, W. E. *Empirical modifications in the up-and-down method and its estimates of* μ and σ. Report No. 2, National Institutes of Health Research Grant NB-105576-01, University of Illinois, 1967.

Kappauf, W. E. An empirical sampling study of the Dixon and Mood Statistics for the up-and-down method of sensitivity testing. *American Journal of Psychology*, 1969, *82*, 40-55.

Kornblum, S. Simple reaction time as a race between signal detection and time estimation: A paradigm and model. *Perception & Psychophysics*, 1973, *13*, 108-112.

Kristofferson, A. B. Low-variance stimulus-response latencies: Deterministic internal delays? *Perception & Psychophysics*, 1976, *20*, 89-100.

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, *29*, 115-129.

Levitt, H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 1971, *49*, 467-477.

Michon, J. A. Studies on subjective duration II: Subjective time measurement during tasks with different information content. *Acta Psychologica*, 1965, *24*, 205-212.

Michon, J. A. Magnitude scaling of short durations with closely spaced stimuli. *Psychonomic Science*, 1967, *9*, 359-360.

Miller, R. G., & Halpern, J. W. Robust estimators for quantal bioassay. *Biometrika*, 1980, *67*, 103-110.

Nakajima, Y., Shimojo, S., & Sugita, Y. On the perception of two successive sound bursts. *Psychological Research*, 1980, *41*, 335-344.

Ornstein, R. E. *On the experience of time*. Harmondsworth: Penguin Books, 1969.

Oyama, T. A behavioristic analysis of Stevens' magnitude estimation method. *Perception & Psychophysics*, 1968, *3*, 317-320.

Penner, M. J. Effect of masker duration and masker level on forward and backward masking. *Journal of the Acoustical Society of America*, 1974, *56*, 179-182.

Pöppel, E. Time Perception. In R. Held, H. Leibowitz, & H. Teuber (Eds.), *Handbook of sensory physiology, volume VIII: Perception*. Berlin: Springer-Verlag, 1978. pp. 713–729.

Povel, D. J. Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, 7, 3–18.

Rasch, R. A. Synchronization in performed ensemble music. *Acustica*, 1979, *43*, 121–131.

Richards, W. Time estimates measured by reproduction. *Perceptual and Motor Skills*, 1964, *18*, 929–943.

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

Snodgrass, J., Luce, R. & Galanter, E. Some experiments on simple and choice reaction time. *Journal of Experimental Psychology*, 1967, *75*, 1–17.

Spearman, C. The method of 'right and wrong cases' ('constant stimuli') without Gauss's formulae. *British Journal of Psychology*, 1908, *2*, 227–242.

Steiner, S. Apparent duration of auditory stimuli. *Journal of Auditory Research*, 1968, *8*, 195–205.

Sternberg, S., & Knoll, R. L. The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973. pp. 629–685.

Svenson, O. Magnitude estimation of time intervals delimited by light flashes of different spatial separation. *Reports from the Psychological Laboratories* (*Report No. 394*). Stockholm: University of Stockholm, 1973.

Teghtsoonian, R., & Teghtsoonian, M. Range and regression effects in magnitude scaling. *Perception & Psychophysics*, 1978, *24*, 305–314.

Thomas, E., & Brown, Jr., I. Time perception and the filled-duration illusion *Perception & Psychophysics*, 1974, *16*, 449–458.

Treisman, M. Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs: General and Applied*, 1963, 77 (Whole No. 576).

Van Noorden, L.P.A.S. *Temporal coherence in the perception of tone sequences.* Doctoral thesis, Eindhoven University of Technology, The Netherlands, 1975.

Vorberg, D., & Hambuch, R. On the temporal control of rhythmic performance. In J. Requin (Ed.), *Attention and performance VII*. New Jersey: Lawrence Erlbaum Associates, 1978. pp. 535–555.

Wing, A., & Kristofferson, A. B. The timing of interresponse intervals. *Perception & Psychophysics*, 1973, *13*, 455–460.

Woodrow, H. A quantitative study of rhythm: The effect of variations in intensity, rate and duration. *Archives of Psychology*, 1909, *14*, 1–66.

Woodrow, H. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. pp. 1224–1236.

Zelkind, I., & Sprug, J. (Eds.) *Time research: 1172 studies.* Metuchen, N.J.: The Scarecrow Press, Inc., 1974.

Zwicker, E. Subjektive und objektive Dauer von Schallimpulsen und Schallpausen. *Acustica*, 1969, *22*, 214–218.