Coding position in a sequence by rhythmic grouping*

DAVID N. PERKINS†

Project Zero, Harvard University, Cambridge, Massachusetts 02138

It has been suggested that a perceived rhythmic organization may mediate remembering musical and other stimuli. This experiment examined whether or not Ss could register and remember position in a sequence using rhythmic grouping. Sixteen Ss heard tapped sequences 24-63 beats long, accented to encourage grouping by 4s. Ss tapped responses revealing whether or not they remembered the sequence length—the last tap's position. Significantly more incorrect responses were off by multiples of 4 beats than by adjacent amounts; these frequent errors of whole rhythmic groups of 4 showed that Ss coded sequence length rhythmically. Ss proved 53% accurate over four response conditions, with individual's scores ranging from 8% to 87%. It was concluded that Ss could count with rhythmic hierarchies, essentially equivalent to counting with a nonstandard number base, to code sequential position.

Rhythmic grouping is a prime example of the organizing impulse in human perception. Conventional Western music is apprehended as a cyclic structure of beats grouped into measures, which are in turn grouped into phrases, and so on. The particular hierarchical grouping of a piece, the number of beats per measure or measures per phrase, is signaled by a variety of accentual, melodic, and harmonic features (Cooper & Meyer, 1960; Bamberger & Brofsky, 1972). Rhythm perception will accommodate, but does not require, such cues. It is a commonplace that the listener will impose rhythmic groupings on the physically undifferentiated drips of a leaky faucet. Such observations place rhythm perception with other phenomena such as the grouping of dots in visual displays (Wertheimer, 1958) as part of the Gestalt tendency in human cognition to find, or create, order in a stimulus.

The question remains: Does rhythm perception—specifically the perception of meter—have an instrumental role in larger tasks? Is rhythmic grouping only a way of registering stimuli, a dead end that does not feed into further activities? A contrary view seems plausible. Neisser (1967, Chap. 9) stresses the role of rhythmic grouping in short-term memory, reviewing experiments demonstrating that grouped stimulus sequences are better remembered. In music, the metric pattern of a work may provide a temporal framework for recalling the piece's various harmonic and melodic events (Perkins & Howard, 1973). Thus, a particular chord might be coded as occurring on the third beat of the second measure of the second phrase.

If rhythmically coded position in a sequence could serve as a reference system for remembering other events, Ss should be able to use rhythmic grouping to

register and remember sequential position itself. An experiment was designed to test this hypothesis. Ss heard sequences of taps, differing in number of taps and accented so as to promote grouping by 4s and 16s (4 groups of 4). After the stimulus sequence stopped, Ss tapped back responses indicating whether they had correctly registered and remembered the length of the stimulus sequence, that is, the position of the last tap.

The analysis was informed by a formal parallel between rhythmic grouping and arithmetic with a nonstandard base. A music instructor might count beats and measures aloud in four-four time as follows: "one-two-three-four, two-two-three-four, three-two-..." This corresponds to the Base 4 count "00 01 02 03, 10 11 12 13, 20 21..." To preserve the formal parallel, "0" Base 4 stands for a "one" count, "1" Base 4 for a "two" count, etc.

Noting the percentage of Ss' accurate responses would not reveal whether Ss were indeed using Base 4 rhythmic grouping as a mnemonic code to mediate the task. For this purpose, Ss' errors were analyzed: How often were their responses just right, one beat short or long, two beats short or long, and so forth? Target shooting (with a one-dimensional target) provided a natural null hypothesis model for Ss' performances: One would expect a certain number of correct responses—"bull's eyes"—fewer responses exactly one beat short and fewer one beat long, fewer yet two beats short or long, and so forth. The curve representing number of errors n beats off would be monotone decreasing.

But suppose Ss were encoding stimulus length by Base 4 rhythmic grouping. A single error in the 16's place—the number of whole groups of 16 beats—would put the response off by 16 beats at once; an error in the 4's place would put the response off by 4 beats at once. Instead of being monotone decreasing, the curve representing number of errors n beats off might have peaks at 4s and multiples of 4. It was hypothesized that this would be the case.

The experiment also employed four different response conditions with varying memory demands. In "simple duplication," Ss attempted to tap back the same number

*This research was conducted at Project Zero of the Harvard Graduate School of Education, operating under National Science Foundation Grant GB-31064.

†The author thanks Roy Rudenstine for his assistance in executing this experiment. The treatment of rhythm in this paper and the parallel between rhythmic grouping and counting with a nonstandard number base derive from a collaboration with Vernon Howard to develop a notation for rhythm perception (Perkins & Howard, 1973).

of beats they heard. This required them to encode somehow the number of stimulus beats and then, remembering that "target" number, count up to it again as they responded. "Duplication with trick" demanded the same response, but Ss were instructed in a mnemonic trick, founded on the Base 4 model. It was hypothesized that duplication with trick would yield significantly fewer errors than simple duplication.

"Simple completion" required Ss to commence tapping as the stimulus stopped and to continue until the stimulus plus response beats summed to the "target" number 64, a number of simple form (1000) in Base 4 and constant for all stimulus sequences. It was hypothesized that simple completion would yield significantly fewer errors than simple duplication. "Completion with delay" required Ss to delay for 30 sec before commencing their completion responses. It was predicted that simple completion would yield significantly fewer errors than completion with delay.

METHOD

Subjects

Ss were eight male and eight female undergraduates who responded to ads placed around the campus. All Ss reported having normal hearing. Ss were paid \$2.00 for participating in this and another unrelated experiment.

Stimuli

The stimuli employed were 40 sequences of taps presented to Ss with a tape recorder. The taps were produced simply by rapping on a resonant box with a pencil at a regular rate of 4 taps/sec. The sequences ranged in length from 24 to 63 taps. In each sequence, the first of every 4 taps was stressed, and the first of every 16 was doubly stressed. The 40 sequences were randomly arranged in 4 groups of 10 sequences.

Response Conditions

In "simple duplication," a S was to tap back the same number of taps as he heard in the stimulus sequence. "Duplication with trick" was always introduced immediately after a S had finished "simple duplication" so that the instructions would make sense to the S. E remarked that the S had probably been perceiving the sequences as terminating with a small group of taps—one, two, three, or whatever. E noted that many people had difficulty remembering this "tail end" while tapping back the rest of the sequence, and requested that S try tapping back the tail end first and only then the "body" of the sequence. Ss uniformly appeared to understand the instructions, and guided by them proceeded to attempt just the same task as in "simple duplication," tapping back the same number of times as there were taps in the stimulus sequence.

In "simple completion," as a S heard the stimulus sequence stop, he was to commence tapping himself to continue the sequence up to a total number of 64 taps. In other words, the number of stimulus taps plus the number of response taps should equal 64. In preparation for the task, each S was given practice on a complete sequence of 64 taps from the tape recorder until the S duplicated the 64 taps correctly twice in a row. The fourth task, "completion with delay," was always attempted immediately after "simple completion." The procedure was exactly the same as with "simple completion," except that after the stimulus sequence ended, the S was to wait for the word

"go" from E before completing the sequence. E gave this signal 30 sec after the termination of the stimulus sequence.

Procedure

Ss were tested individually. Each S performed under the four different response conditions, and the basic procedure for each condition was the same. A S heard a sequence of taps from the tape recorder. When the sequence terminated, E stopped the recorder with a silent switch. In accord with the response condition, the S tapped a certain number of times on an electric typewriter, using his dominant hand. The page in the typewriter provided a permanent record of the number of taps in the S's response. After S finished typing, E turned on the tape recorder to present the next sequence.

Pilot work revealed that Ss were likely to employ various deliberate strategies in attempting the tasks. A S might count the taps in the decimal number system. One pilot S reported running through a familiar tune in his mind as the stimulus taps continued and duplicating the stimulus by running through the same tune to the same point as he tapped his response. Interesting though they were, such strategies were not the focus of the experiment and might have obscured the phenomena under investigation. Therefore, Ss were instructed not to count to themselves nor to employ any other memorized sequence like a song as a prop. Ss were encouraged to attend closely to the stimulus sequences and to do the best they could without such props. Of course, telling Ss what not to do did not guarantee that they would succeed in coding sequence length by rhythmic grouping; this was left for the error analysis to reveal.

The testing was conducted according to a three-way factorial design. One factor was sex. The second was whether the duplication tasks or the completion tasks were administered first. The third concerned the order of employing the four stimulus groups of 10 sequences each: the groups were designated 1, 2, 3, 4 and employed in rotating order on successive Ss so that over all Ss, each of the four tasks employed stimuli from each of the four groups.

Preliminary Analysis

The data were examined for any influence of the three factors—sex, duplication vs completion first, and stimulus groups. For each S, four scores were computed corresponding to the four response conditions. Each score recorded the S's number of correct responses, from 0 to 10, under one condition. Within each condition, sex differences were probed by *t* tests matched by the other two factors. No significant differences were found. A Friedman two-way analysis of variance tested whether any of the four stimulus sets led to consistently lower scores over all the conditions. No significant effects emerged.

A matched *t* test over all tasks compared performance on each task when it was presented first or second with performance when it was presented third or fourth. A consistent training effect emerged: when a condition occurred third or fourth, scores averaged 1.28 higher ($p < .05$, one-tailed). In addition, the data revealed substantial individual variation in Ss' capacities to deal with the tasks. A Friedman two-way analysis of variance compared the 16 Ss' scores on the four conditions. Results were significant ($p < .01$), showing that some Ss tended to perform better than others across all conditions. In fact, the number of correct responses out of 40 ranged from 3 to 35.

RESULTS

The Base 4 Code

Confirmation was found for the hypothesis that mistaken responses would be off by multiples of four

(including four) beats more often than by adjacent numbers. The graphs of Fig. 1 demonstrate the effect, showing for each condition, over all Ss, how many responses were exactly right, one beat off, two off, etc. Rather than decreasing monotonically, the graphs tend to peak at 4, 8, 12, and 16. To check for the significance of this effect, the number of responses off by a multiple of four was compared with the number of responses off by one less or two less than a multiple of four. Under the null hypothesis that the graphs were monotone decreasing, the first number would be expected to be no greater than half the second. This was refuted for duplication with trick, simple completion, and completion with delay (chi square, $p < .01$, $.02$, $.01$, respectively), though not for simple duplication.

The peaks at multiples of 16 were of special interest: they could indicate errors in the 16's place of a Base 4 coding. Since such large errors were few, the data from all four conditions were combined for statistical tests. The number of responses off by a multiple of 16 was compared with the number of errors off by one less or two less than a multiple of 16. Again, a monotonic decreasing curve would require that the first number be no greater than half the second; this was refuted (chi square, $p < .001$). Also, the peak at 16 was not simply to be taken as one of the peaks at multiples of 4. Although the peaks at 4, 8, and 12 decreased in height, the peak at 16 broke the trend: a comparison of the number of responses 12 off with the number 16 off revealed that the number 16 was significantly larger (chi square, $p < .001$). In sum, the data argued that Ss encoded the number of beats in a Base 4 number system.

Although the simple duplication condition did not exhibit statistically significant peaking, it is doubtful that simple duplication was somehow an exception to the coding practices demonstrated in the other conditions. More likely, a large number of unit's place errors in comparison with four's place errors obscured the peaking. The data supported this interpretation: the number of beats in each stimulus and response were written in Base 4 and for each condition, the number of responses wrong only in the unit's place and the number wrong only in the four's place were determined. In simple duplication, there were over twice as many one's place only as four's place only errors; in duplication with trick, there were slightly more four's place only errors than one's place only errors, a significantly different distribution (chi square, $p < .02$).

The Response Conditions

Confirmation was found for the hypothesis that Ss would perform better on duplication with trick than on simple duplication. This test employed the number of correct responses, from 0 to 10, that Ss made under each condition. Ss averaged 3.00 correct responses on simple duplication, but 6.00 correct responses on duplication with trick. The difference proved highly significant

Fig. 1. Number of responses each number of beats off from a correct response.

(t tests matched by Ss, $p < .005$, one-tailed).

Duplication with trick was always administered immediately after simple duplication so that the "trick" could be readily explained; could the difference in performance be due to the learning effect noted earlier? A graph of the duplication tasks was prepared, displaying for each stimulus sequence in its order of presentation the number of Ss correctly responding; the graph revealed an abrupt drop between simple duplication and duplication with trick. A straight line representing the learning effect was fitted by the eye to the simple duplication points and extended across duplication with trick. The slope was shallow, and the mean point of this line within duplication with trick differed significantly from the mean given by the data (t test, $p < .005$, one-tailed). It was concluded that the difference reflected not a learning effect, but a contrast in difficulty between the two tasks.

The contrast can plausibly be ascribed to a reduction in memory load under duplication with trick. In terms of the Base 4 theory, Ss in simple duplication tapped back first the groups of 16 and 4 and finally the units; their responding mirrored the order in which they registered the stimulus beats. In duplication with trick, Ss were in effect instructed to tap back the unit's place first. They no longer needed to remember the proper number of units while tapping back the lengthy groups of 16s and 4s, and their performance improved. A

finding mentioned above further supported this interpretation: unit's place only errors outnumbered 4's place only errors by over two to one in simple duplication, but under duplication with trick improved to become slightly fewer than 4's place only errors.

The hypothesis that simple completion would prove easier than simple duplication was confirmed. Mean correct responses was 6.62 on simple completion in contrast with 3.00 on simple duplication, again a highly significant difference (*t* test matched by *Ss*, $p < .005$, one-tailed). Again, differences of memory load provided a reason. In simple duplication, a *S* needed to count the stimulus sequence until it stopped and then remember that unique number while tapping back. In simple completion, a *S* always counted up to the same number of beats—64, and a number easily remembered because of its simple form in Base 4 arithmetic—4 groups of 16 or 1000.

The hypothesis that completion with delay would prove harder than simple completion was not confirmed. Though the mean correct responses for completion with delay, 5.69, was slightly smaller than the simple completion mean of 6.62, the difference was not significant. Since completion with delay was always presented after simple completion, the training effect identified earlier would tend to reduce the expected difference. But a graphical analysis like that conducted for the duplication tasks revealed no increase in difficulty as *Ss* moved from simple completion to completion with delay.

Clearly, the latter task demands somewhat more of memory than the former, and presumably if sufficiently many *Ss* were run, a significant difference would emerge. The principal point remains that *Ss* could hold a count in memory with little decrement for 30 sec when they had no other task to do, in contrast with their problems, characteristic of short-term memory studies (Neisser, 1967, pp. 235-240), when they needed simultaneously to remember a target and tap back in simple duplication.

Error Sources

The support found for a Base 4 model of *Ss*' cognitive code encouraged a closer examination of the errors in terms of that model. One obvious conjecture was that some stimulus sequences would prove less troublesome than others. Unit's place values of zero or three, reflecting the initiation and completion of a group of four, might be more easily remembered or might serve as "default options" when a *S* forgot the correct value.

To check this possibility, the data from all four conditions were combined to yield the number of stimuli with unit's place zero, one, two, or three responded to correctly or incorrectly. Analogous figures were computed for the four's place. For the four's and unit's places separately, the error distributions for zero, one, two, and three were examined to determine whether they could have been drawn from one

underlying distribution. This possibility could not be rejected; chi-square tests revealed no significant differences. It would be imprudent here to accept the null hypothesis outright, but safe to conclude that different values in the unit's or four's places did not differ grossly in their stability in memory.

All four tasks required a *S* to remember both a "count" of beats in the stimulus and his response and a "target" number where his response would terminate; did errors originate in the target, the count, or both? The comparisons of simple duplication with the other tasks revealed errors of forgetting the target; then could all errors be so interpreted? This possibility was tested. If forgetting the target were the primary error mechanism, forgetting one place would not encourage forgetting another also. One would expect that accurate memory for one place in the target would be independent of accuracy in another place, or perhaps even at its expense, just as happened in the four's and unit's places of simple duplication.

A chi-square test probed the statistical independence of errors in the unit's and four's places under each of the four response conditions. In all cases, there was significant departure from independence ($p < .02$, .001, .01, .01 for simple duplication, duplication with trick, simple completion, completion with trick, respectively). Furthermore, errors were nonindependent in a way contrary to the forgetting-the-target hypothesis: errors in the different places tended to occur together rather than alone. The target was probably not the one principal error source.

In addition, the findings made sense if the count was a major error source. Suppose a *S* were duplicating a stimulus of Length 213 Base 4 and miscounted anytime during his response, tapping an extra beat without mentally "registering" it. His typed response would then be 220 even though he remembered the target correctly, the single error affecting both unit's and four's places. In general, errors made in counting would spread from one place to the next leftward via the arithmetic carry, explaining the finding that errors in different places tended to occur together rather than alone.

DISCUSSION

The data argued strongly that *Ss* indeed employed a Base 4 code founded on rhythmic grouping to store stimulus sequence length—the target—and to keep count of their response taps. Logically, the count and target numbers could be remembered in various ways. Certainly *Ss* could employ the decimal number system, although here they were asked not to do so. But evidence was found that *Ss* in fact coded the count and target as predicted; errors differed from the correct responses significantly more frequently by multiples of 4 and 16 beats than by neighboring numbers, reflecting mistakes in the 4's and 16's places of a Base 4 code.

The different response conditions designed to vary

memory load further implicated a rhythmic grouping strategy. Simple completion employed the same target—an easily coded 1000 in Base 4—for each stimulus sequence, and yielded significantly fewer errors than simple duplication. Performance was significantly more accurate on duplication with trick than on simple duplication, and the trick—tapping back the unit's place first rather than last—only made sense in a Base 4 context.

The trick's effectiveness also illuminated the count-target comparison process. Target and count were compared not all places at once, but place by place and at different times during the response sequence. In simple duplication Ss first dispatched the 16's and 4's places, afterwards turning to the unit's place, which they had often forgotten by then. That a verbal hint permitted Ss to reverse this strategy indicated not an unconscious sequential comparison process such as Sternberg (1970) has identified in pattern recognition, but one quite accessible to deliberate manipulation.

Finally, remembering the count proved to be as much a problem as remembering the target. If all errors were target errors, one would expect mistakes in different places of the Base 4 code to occur independently; or perhaps accuracy in one place would encourage errors in another. On the contrary, the data revealed just the opposite trend: errors in different places were not independent but tended to occur together. This plausibly reflected mistakes made during the count, which would be propagated by carries from the unit's to the 4's and from the 4's to the 16's places as the count continued.

This study was introduced by questioning whether perceived rhythmic grouping was merely a way of registering the stimulus without further purpose. Contrary suggestions were cited: rhythm might provide a mnemonic framework for lists of items in short-term memory (Neisser, 1967) and for locating musical events in time (Perkins & Howard, 1973). This study demonstrated that rhythmic grouping could be used to code temporal position over sequences of beats up to 64 beats long. Positions could be remembered and reproduced as in duplication or counted beyond as in completion even after a delay. The coding procedures were accessible to conscious manipulation. Though some

Ss performed consistently poorly, some were very accurate, one being correct on 35 of the 40 stimuli. In all, rhythmic grouping could provide a reasonably accurate and stable framework for coding the position of musical and other events in time. The present study does not demonstrate such use, but shows that rhythmic grouping satisfies some conditions appropriate for this role.

Another factor fitting rhythmic grouping for such a role is the generality of the human capacity to perceive rhythmic hierarchies. For instance, a typical waltz pattern involves three beat measures organized into four-measure phrases, two of which form yet a larger grouping. This too can be viewed as a numbering system, but the base is variable, depending on the place: the unit's place counts up to three, the next place counts up to *four* groups of three, and the next place counts up to *two* of those groups of four. The present findings only demonstrate that Ss can employ a Base 4 cognitive code for certain counting tasks. But these further considerations suggest that humans have an immensely flexible coding system at their disposal, one which can count beats by employing whatever sequence of number bases best matches periodic features of the stimulus. The "new math" of arithmetic to nonstandard bases, as with so many developments, has played an unrecognized role in a living system for a long time.

REFERENCES

- Bamberger, J., & Brofsky, H. *The art of listening: Developing musical perception*. New York: Harper & Row, 1972.
- Cooper, G., & Meyer, L. *The rhythmic structure of music*. Chicago: Phoenix Books, University of Chicago Press, 1960.
- Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- Perkins, D., & Howard, V. Toward a notation for rhythm perception. Unpublished paper, Harvard Project Zero, 1973.
- Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. In J. Antrobus (Ed.), *Cognition and affect*. Boston: Little, Brown, 1970.
- Wertheimer, M. Principles of perceptual organization. In D. Beardslee and M. Wertheimer (Eds.), *Readings in perception*. Princeton: Van Nostrand, 1958.

(Received for publication July 9, 1973;
accepted August 17, 1973.)