Internal Representation of Simple Temporal Patterns

Dirk-Jan Povel
University of Nijmegen, Nijmegen, The Netherlands

In this study the imitation of several periodically repeating simple temporal patterns consisting of two or more intervals varying in their duration ratios has been investigated. The errors that subjects typically made in their imitations and the systematic changes that occurred during repeated imitations indicate that both musically trained and untrained subjects map temporal sequences onto an interval structure the nature of which is revealed by studying which patterns are correctly and which incorrectly reproduced. A "beat-based" model for the perception of temporal sequences is proposed. This model states that the first step in the processing of a temporal sequence consists of a segmentation of the sequence into equal intervals bordered by events. This interval is called the *beat interval*. How listeners select this beat interval is only partly understood. In a second step, intervals smaller than the beat interval are expressed as a subdivision of the beat interval in which they occur. The number of within-beat structures that can be represented in the model is, however, limited. Specifically, only beat intervals that are subdivided into either equal intervals or intervals in a 1:2 ratio fit within the model. The partially hierarchical model proposed, though in need of further elaborations, shows why the number of temporal patterns that can be correctly conceptualized is limited. The relation of the model to other models is discussed.

More than 3 decades ago, Fraisse (1946) discovered a remarkable phenomenon in the production and perception of durations. He found that subjects who were asked to produce by tapping temporal patterns consisting of 2–6 taps basically used only two durations. These two durations, called a *long duration* and a *short duration*, are distinct from each other; the longer duration is typically at least twice as long as the shorter one. Fraisse (1946) reported a range of long/short ratios that varied from 2.18 to 3.25 depending on the length and complexity of the patterns tapped. Subsequently, he described some experiments in which subjects

imitated simple temporal patterns. A most important finding from these experiments was that stimuli in which the longer interval is less than double the duration of the shorter interval tend to be reproduced with a long/short ratio of 2:1. Fraisse (1946) concluded from these findings that there exist favorite rhythmic structures. The number of such favorite structures, however, is restricted, since subjects seem to be able to conceptualize only two distinct durations that roughly relate as 2:1. Later, Fraisse (1956) added weight to this point by reporting that in a representative sample of western music, an average of 86% of the occurring tone durations related as 1:2. Although the two findings seem to fit remarkably well, the suggestion that the perception and production of rhythms could be understood by an internal representation that allows only two distinct durations seems too simple.

In the present article I investigate these questions: What limitations are there in the perception of temporal sequences? and Which sequences are easy, which are difficult to reproduce? By studying the latter question, I also hoped to better understand

the internal representation of temporal sequences. First, I will briefly review three models for the representation of the temporal aspects of sequential stimuli. The models differ in the amount of hierarchical connections that are assumed. They are based in part on the work of Cooper and Meyer (1960), who showed that a complete hierarchical description of certain aspects of the temporal structure of music is easily conceivable.

Martin (1972) proposed a complete hierarchical description of "natural" temporal patterns, which incorporates rules for the description of relative accent and of the relative timing of elements in a pattern. To describe the temporal structure of actual tunes, the concept of a "null branch," that is, the deletion of a tone or tap that is generated by the structure, had to be introduced. A "terminal rule" was proposed that describes the inversion of accent levels at the end of tunes or musical phrases. Three remarks concerning the model need to be made. First, if every "natural" pattern of whatever length were to be described by Martin's complex hierarchical model, complex descriptions consisting of trees with many null branches and many nodes (determined by the shortest relative duration in the sequence) would be necessary. Thus, Michon (1974), in applying Martin's accent rule to the "Vexations" by Erik Satie, needed a tree with five nodes from which not less than 32 accent levels resulted. One wonders what the psychological reality of such subtleties might be. Second, the model as described contains only binary trees and can therefore only describe sequences with a surface length of 2^n. Indeed Martin does not give one example of a tune or phrase in a 3/4 measure. The necessary change in the model can presumably be easily made, but the relation between double and triple divisions then becomes a problem, as I will show later. A last remark concerns the status of the model. The model seems to provide a theoretical description of the final representation of natural sequential patterns as stored in memory. In this respect it can be compared with coding models developed by Simon and Sumner (1968), Restle (1970), Vitz and Todd (1969), and Leeuwenberg

(1971) for other properties of tone sequences. Since this type of model states nothing about the process of encoding, the testing of the models is restricted to comparison of predicted and empirical complexity and sometimes to error analysis (Restle, 1970).

In an attempt to develop a coding language for rhythms, Povel (Note 1) proposed a "beat-based" model for the perception of temporal sequences. According to the model, the first step in the perception of temporal sequences is segmentation of the sequence into parts of equal length. The segmentation is based on the detection of accented events occurring at equal intervals in the sequence. These events are called *beats*, in accordance with the use of the term in music theory. An interval between beats can contain silence or one or more tones. In western music, intervals between beats have durations that range roughly from 250 to 1,500 msec. The effect of the beats, which can be seen in the regular movements that listeners make to music (foot tapping, swaying back and forth), can thus be considered a part of the process of perception. The psychological reality of the occurrence of beats for the perception of temporal sequences was demonstrated in a study which showed that subjects have great problems in imitating sequences with unequally spaced accented events (Povel, Note 2). The proposed model still has difficulties with the coding of specific event patterns within beats, such as the description of sequences that contain both triple and double time division like ♫♩♫ . I will return to this later. At this point I will only mention that the beat-based model has a considerably lower level of hierarchy than Martin's (1972) model. Among other things, a lower level of hierarchy implies that the coding can start before the whole sequence is presented.

The third model is an association model. It supposes that a subject codes a sequence of events as a chain of durations. The only relation between elements in the code is between adjacent ones. The theory seems unrealistic from the start: Not only does it encounter serious memory problems, but more importantly it cannot account for any of the organizational characteristics (grouping) that are found in the perception of sequential

patterns. Moreover, because the model does not take into account any relation between events beyond adjacent ones, it cannot account for any difference in the ease of remembering temporal sequences containing the same number of elements but differing in the actual durations involved, such as ♩♩♩♩♩♩♩ and ♩♩♫♫♩♩. Before discounting the theory, however, it should be realized that there are numerous temporal sequences that do not fit either of the two previous models, such as the sequence with intervals of 200, 760, 2,580, and 150 msec. Although such a sequence cannot be described in terms of the two former models, subjects are presumably able to imitate the sequence to some extent and distinguish it from other sequences.

Vorberg and Hambuch (1978) recently proposed three models for the production of continuously repeating groups of taps. The models seem closely related to the three models just described, since they also differ in the level of hierarchy. The models are an extension of the two-stage, timekeeper–motor–delay model for the production of isochronic series of taps developed by Wing and Kristofferson (1973a, 1973b). The authors proposed an interesting way of testing the models by deriving statistical predictions concerning the variance and covariance structure of the within-group interresponse intervals in their subjects' tapping. A major assumption, which also enables the straightforward derivation of the theoretical covariance matrices, is that the timekeepers are independent. The three models make different predictions. In an experiment that tested the models, none of the predictions of the partly and completely hierarchical models were confirmed by the data. Vorberg and Hambuch concluded, therefore, that the production of repeated groups of taps can be completely explained by a nonhierarchical chain model. However, they call this result paradoxical in view of another finding of theirs, namely, a positive serial autocovariance at a lag equal to the group length, which seems to suggest a higher organization.

To better understand the meaning of Vorberg and Hambuch's (1978) results, consider the way predictions from these models are made. Greeno and Simon (1974) pointed

out that any model of imitative behavior has to specify at least three distinctive processes that are prerequisites for such a task. The first process deals with the process of perception, which may entail the identification of structural characteristics. The second deals with the coding and storage of the perceived stimulus in memory. The third process bears on the regeneration of the sequence from the code stored in memory. Since Vorberg and Hambuch's predictions pertain to detailed characteristics of (re)productions, their model must at least include assumptions about the memory representation and the process that transforms the memory code into a tapping sequence in real time.

Two possible strategies of regenerating a temporal sequence from a memory code are theoretically feasible. Strategy 1 uses the memory code in an on-line fashion to generate the durations that define the tapped sequence. This means that for the subject who produces the same sequence repeatedly in a continuous fashion, as in Vorberg and Hambuch's (1978) experiment, the memory code is again and again used to generate the sequence of durations. It will be clear that if the memory code is of a hierarchical type, this strategy is not compatible with the assumption that the durations (timekeepers) are independent. Strategy 2 includes an additional memory stage. In this conceptualization, the memory code is not used on-line during the actual responding but instead is used once to generate a description of the sequence as a series of durations (timekeepers). This chainlike description is temporarily stored in memory and is the basis of the actual production of the sequence. This second strategy seems, in some respects, more economical. Moreover, it is compatible with an independency assumption. The three models proposed by Vorberg and Hambuch are models of memory codes. According to their description, the memory code or structural description is used to generate the subjects' tapping responses. That is, they assume that the subjects follow Strategy 1. But as mentioned, this strategy is incompatible with an assumption of independent timekeepers.

Thus, I conclude that it is not Vorberg

and Hambuch's (1978) results that are paradoxical but rather the paradigm used to test their models. The predictions assume that the timekeepers are independent, but the models themselves vary in the hierarchical organization of the timekeepers, that is, in the amount of dependency of the timekeepers. The question remains as to what Vorberg and Hambuch's negative results may mean. Certainly, the results are compatible with a production mechanism that includes Strategy 2, but Strategy 1 cannot be excluded at the moment. Note that if subjects use Strategy 2, no knowledge about the memory code can be gained by studying the variance and covariance structure in the subjects' responses.

This article describes a series of experiments that used a continuation procedure in which subjects imitated a number of different temporal patterns that varied in the relative duration of their component intervals. The main dependent variable was imitation errors. The nature of the errors formed the basis for development of a model for the internal representation of temporal sequences.

Experiment 1

From the work of Bartlett and Bartlett (1959), Treisman (1963), Michon (1967), Wagner (1971), and Wing and Kristofferson (1973a, 1973b), it is known that subjects can (re)produce an isochronic sequence of taps with considerable accuracy. The general finding is that accuracy (i.e., variance of the taps) is a slowly accelerating function of the duration of tap intervals. The absolute amount of the variance differs very much among the different authors, indicating that the measured accuracy is highly dependent on the method used and especially on the practice of the subjects.

The next question to ask concerns how well subjects can (re)produce sequences with two alternating durations. Within the present context, it is of special interest to determine whether the relation between the durations affects the imitation in some way. For instance, is imitation better when there is a simple relation between the durations? If the duration relation does not affect reproduction, then one might conjecture that a subject can perceive and store at least two

durations independently. The latter is an assumption made by Vorberg and Hambuch (1978), as mentioned before. If we do find that the duration relation plays a role in reproduction, then it presumably means that the subject represents the durations in terms of one another.

This question has been studied by Fraisse (1946) in the context of his general finding that subjects doing spontaneous tapping use two duration categories that roughly relate as 1:2. In one of Fraisse's (1946) experiments, the subjects imitated tapping patterns consisting of two intervals, t_1 and t_2; t_1 was fixed at 450 msec, and t_2 changed in steps from 210 to 690 msec. Note that this results in duration relations varying from roughly .5–.9. Throughout this article interval relations are indicated as the ratio of the shorter interval divided by the longer, so that all possible interval relations lie on a continuum from .1 (very large difference in duration) to 1.0 (equal durations). Fraisse (1946) reports that his subjects generally overestimated the duration difference, which means that the reproduced duration ratios show a tendency toward .5. He also reports that when the interval ratio in the stimulus approached 1.0, the subjects showed a tendency to equalize the durations rather than to make the distinctions bigger.

Since I want to determine whether or not durations are coded as related in the internal representation of temporal sequences, the imitation of patterns with duration ratios smaller than .5 must also be studied. I am particularly interested in whether tapping patterns comprised of durations with simple ratios, for example, 1:2 (.5), 1:3 (.33), and 1:4 (.25), are better imitated than patterns with more complex interval ratios like 2:3 (.66), 2:5 (.40), and 3:4 (.75).

Method

Twenty-five subjects, all undergraduate students at Indiana University, participated in the experiment. Ten of the subjects were musically trained, having played the piano for at least 5 yr; the other 15 subjects had no musical training whatsoever.

The stimuli presented to the subjects consisted of sequences of 150-msec beeps whose onset intervals were varied. The beeps were generated by a Mallory Sonalert (Model SC 628) DC tone generator that produced an almost sinusoidal waveform with a frequency of 2800 Hz. Typically, the sequence was presented in a cyclic

fashion. Subjects were asked to listen until they thought that they could imitate the sequence. (They were encouraged to tap on the table in synchrony with the beeps before starting the synchronization or imitation). The actual responses consisted of taps on one of two 10 × 10 cm² metal plates mounted on a small box. Because tapping the plate produced the same beep as in stimulus generation, the stimulus and response phase were perceptually continuous. When subjects made an error or wanted to start over because they thought they could do better, they tapped the second metal plate, which stopped the responding phase and gave rise to the presentation of the same stimulus. The subjects practiced the different tasks thoroughly before actual measurements were taken. Both stimulus generation and response collection were controlled by a PDP 11/04 computer.

The subjects performed the following tasks: First, subjects synchronized with seven single-interval sequences having intervals[1] of 200, 250, 340, 400, 580, 790, and 1,000 msec. For each subject 17 synchronization taps were recorded per sequence. The order of presentation was randomized across subjects.

Second, subjects continued for 17 taps the same seven sequences of Task 1, with a different random order. I included these tasks partly to warm up the subjects and partly to see whether the musically trained and untrained subjects would differ in accuracy. Moreover, these measurements provided a base rate with which to compare the (re)production of durations in other tasks.

The third task formed the kernel of the experiment. The subjects listened one after another to 16 sequences consisting of two durations (t_1–t_2 sequences) presented in a cyclic fashion. As soon as subjects felt ready, they imitated (continued) the sequence 17 times. The sequences, which were presented to the subjects in random order, are shown in Table 1. Only one cyle of the sequences is given. With respect to duration ratios, Stimuli 1–8 are identical to Stimuli 9–16. Stimuli 1–8 were chosen so that $t_1 + t_2 = 1,000$ msec; thus, the patterns are made in the same overall time. Stimuli 9–16 were chosen so that t_1 is always 250 msec, which is about the shortest time that can be tapped comfortably with one finger.

Results

The precision with which subjects tapped the seven single-interval sequences is pre-

sented in Figure 1 in terms of the mean of the individual standard deviations. Since no differences were found on this measure between the musically trained and untrained subjects, the data of both groups are combined. Figure 1a shows the standard deviations for the synchronization condition and Figure 1b for the continuation condition. The first two responses of each subject to each sequence were disregarded in computing the data. The relatively large imprecision for longer intervals in the synchronization condition was presumably caused by correction responses on the part of the subject.

The imitation results for the t_1–t_2 sequences are also based on 15 repetitions by the subjects, the first two imitations of each pattern having been excluded. Figure 2 shows the mean interval ratios of the imitations of the stimuli averaged over subjects. The results are displayed so that the actually produced duration ratio is shown as a deviation from the duration ratio of the presented stimuli. Since the stimulus duration ratio is marked as a point on a continuum of ratios that runs from .1 to 1, the mean deviation of each pattern is shown as a line length. The longer the line, the larger the deviation. If the deviation is toward .1, it means that the subjects increased the difference between the intervals (distinction); if the deviation is toward 1, it indicates assimilation. Table 2 contains per stimulus the interval ratio of the imitations averaged over subjects (also shown in Figure 2), the standard deviation, and the average drift. The standard deviation used here and in the re-

---

[1] Throughout this article, the term *interval* is used to indicate the interval between onsets of adjacent tones.

Table 1

t_1–t_2 *Sequences and Their Duration Ratios Used in Experiment 1*

Stimulus	t_1	t_2	t_1/t_2	Stimulus	t_1	t_2	t_1/t_2
1	200	800	.25 (1:4)	9	250	1000	.25
2	250	750	.33 (1:3)	10	250	750	.33
3	286	714	.4 (2:5)	11	250	625	.4
4	334	666	.5 (1:2)	12	250	500	.5
5	375	625	.6 (3:5)	13	250	417	.6
6	400	600	.66 (2:3)	14	250	378	.66
7	429	571	.75 (3:4)	15	250	333	.75
8	444	556	.8 (4:5)	16	250	312	.8

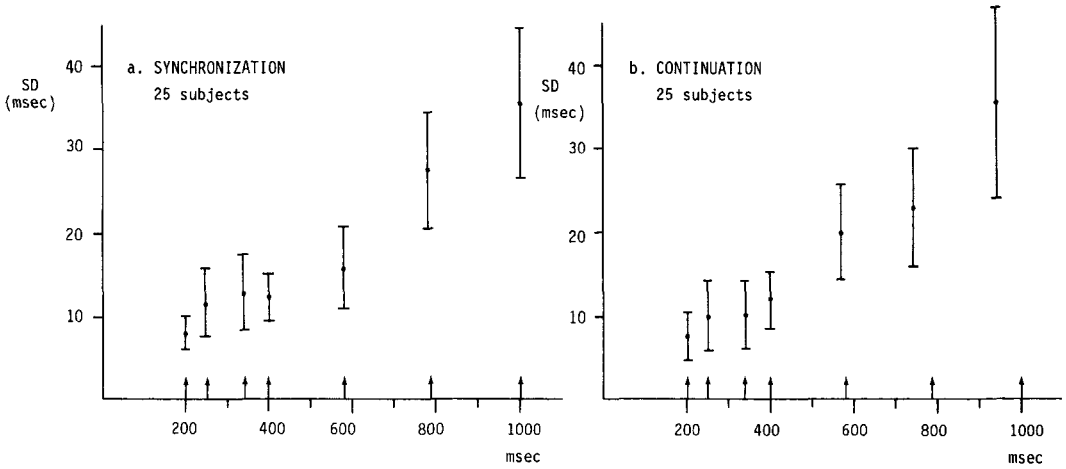*Note.* Only one period of each sequence is given. Durations are in msec.

Figure 1. Mean and standard deviation of individual standard deviations as a function of reproduced interval.

mainder of this article is, other than in the one-interval sequences, the standard deviation of the subject means and thus includes intersubject differences. *Drift* is defined here as the difference between the duration ratios of the first and last pair of intervals in the series of 15 imitations. A negative drift represents a tendency to make the difference between the two intervals smaller, whereas a positive drift indicates a tendency to enlarge the difference.

Table 2 shows that stimuli with interval ratios between .5 and 1 were reproduced with interval ratios that tended toward .5

(e.g., the interval ratios .66 [2:3] presented in Stimuli 6 and 14 was reproduced as .49 and .55, respectively), except for some subjects who showed a tendency toward 1 in their imitations of the .75 and .8 patterns. (This makes averaging the imitations of these patterns, as indicated in Table 2, unreliable.) The subjects, however, never produced a completely isochronic sequence showing that they did hear a difference between the two intervals. Apparently, they fought one of two strong tendencies: to make one interval twice as long as the other or to make the intervals equally long. Both solu-
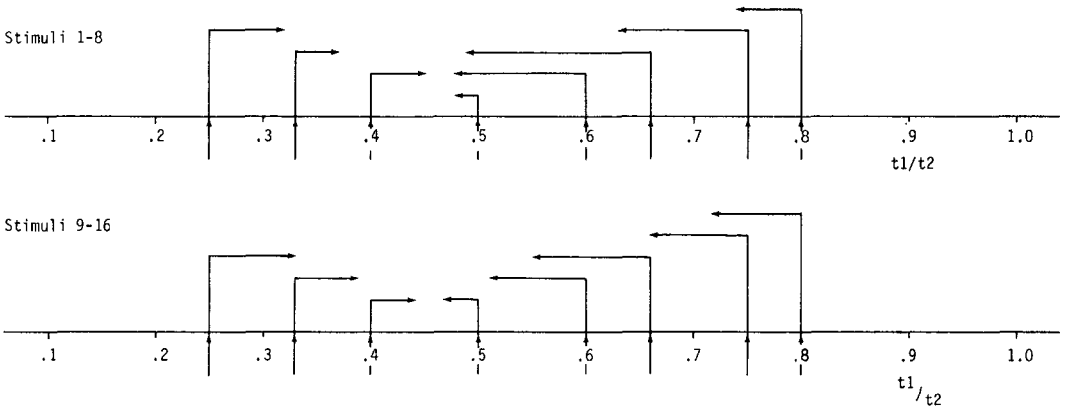

Figure 2. Distortion of duration ratios as found in the imitations of t_1–t_2 patterns. (Duration ratios of the stimuli are indicated by arrows pointing to the abscissa that displays the continuum of t_1/t_2 ratios. The end points of the horizontal arrows indicate the duration ratios in the imitations. See also Table 2.)

Table 2
Mean, Standard Deviation, and Drift of the Interval Ratios in the Imitation of t_1–t_2 *Patterns*

	Stimuli			Imitations				Stimuli			Imitations		
	$t_1 + t_2 =$ 1,000			t_1/t_2				$t_1 = 250$			t_1/t_2		
No.	t_1	t_2	t_1/t_2	M	SD	Drift	No.	t_1	t_2	t_1/t_2	M	SD	Drift
1	200	800	.25	.33	.06	−.03	9	250	1,000	.25	.33	.06	−.02
2	250	750	.33	.37	.07	−.03	10	250	750	.33	.39	.07	−.03
3	286	714	.40	.45	.04	−.03	11	250	625	.40	.44	.05	−.02
4	334	666	.50	.48	.04	.00	12	250	500	.50	.47	.05	.00
5	375	625	.60	.48	.04	.01	13	250	417	.60	.51	.04	.04
6	400	600	.66	.49	.04	.01	14	250	378	.66	.55	.04	.03
7	429	571	.75	.63[a]	.19	.04[a]	15	250	333	.75	.66[a]	.18	.06[a]
8	444	556	.80	.74[a]	.19	.04[a]	16	250	312	.80	.72[a]	.17	.07[a]

Note. N = 25. Durations are in msec.
[a] These values are unreliable, as they are the result of contrary tendencies. This is reflected in high *SD*.

tions were considered wrong, which made imitation of the patterns with ratios of .66, .75, and .8 particularly hard. Their difficulty is reflected by the remarks of the subjects during the experiment and by the high standard deviation. Note that the pattern ratios of .5, .6, and .66 were similarly reproduced with an average ratio of about .05. These findings are roughly in agreement with Fraisse's (1956).

Interval ratios smaller than .5 were also reproduced with a strong tendency toward .5. This was true both for intervals that had a simple relation, for example, 1:3 (.33) and 1:4 (.25), as well as for the interval that was more complexly related, for example, 2:5 (.4).

A negative drift is found in the imitations of all interval relations smaller than .5 (.25, .33 and .4), whereas a positive drift is found in the imitations of ratios bigger than .5 (.6, .66, .75, and .8), indicating that subjects moved toward a 1:2 ratio while responding.

Discussion

For the simple interval sequences, precision was virtually the same for synchronization and continuation. This is surprising because the underlying mechanisms for the two tasks are different. Continuation has an important memory component: Since the memory trace may decay during reproduc-

tion, drift in the sense of slowing down is typical for this condition (Wing & Kristofferson, 1973b). Synchronization does not have this drawback. Another possible advantage for synchronization is that it involves a sort of learning process because subjects receive continuous feedback on their performance. This would make synchronization more suited for assessing subjects' performance limits. But this same feedback, since it gives rise to correction responses, makes interpretation of the data difficult. Thus, in a pilot experiment, a systematic enlargement of the standard deviation of intervals that followed relatively long intervals was found , which suggests compensation for perceived errors in the long preceding interval. For this reason only a continuation method was used in subsequent experiments.

The general finding with reproduction of the t_1–t_2 sequences is that only patterns with intervals that relate as 1:2 are correctly imitated. Imitation of all other patterns have interval relations that deviate considerably from the interval relation in the stimuli. This deviation is systematic and can be described as a tendency toward an interval relation of 1:2. This tendency is reflected both in the average reproduced interval ratio as well as in the slow change in the ratio (drift) that occurs during the series of 15 imitations.

Before we attempt further interpretation of these findings, we should mention a per-

ceptual phenomenon that may have influenced the results. If a t_1–t_2 pattern is presented to a subject, beeps and not intervals will be perceptually grouped according to their proximity. Thus, t_1 becomes part of the group (figure) and t_2 part of the background. Both Fraisse (1956) and Bamberger (Note 3) have noticed that between-groups intervals are perceptually special. Subjects indicate that such intervals have hardly any reality for them. Indeed, in imitating these sequences one senses that much attention is directed toward imitating the within-group interval, whereas the between-groups interval hardly requires any attention.

Since the data may be greatly influenced by this difference between t_1 and t_2, Experiment 2 was conducted, using t_1–t_2–t_2 sequences in which t_1 and the first t_2 interval are within-group intervals. Fraisse (1956) did part of his experiments with t_1–t_2–t_2 patterns for the same reason. But again he only used ratios between .5 and 1.0, whereas my interest was also in ratios smaller than .5. The purpose of Experiment 2 was to determine whether the findings of Experiment 1 still obtain when t_1 and t_2 are both within-group intervals.

Experiment 2

Method

Twenty subjects, all undergraduates from Indiana University who had not been in Experiment 1, participated in the experiment. Nine subjects had had musical training, which averaged 5 yr. The method of presenting the stimuli, the responding of the subjects, and the collection of the data were the same as in Experiment 1. Partly as a check of subjects' aptitude and partly as a practice session, the subjects made 17 continuation taps for each of the seven one-interval sequences used in Tasks 1 and 2 of Experiment 1. Since the results were comparable to those presented in Figure 1b, I will not present these data here.

Table 3 shows the 12 temporal patterns used. The stimuli were always presented as shown, that is, beginning with the smallest interval. In this way the perceived grouping was stable from the beginning and coincided with the first three taps. A pilot study suggested that if the first interval is not the smallest one, it takes some time before the same grouping is perceived. The stimuli were presented in different random orders to the subjects. Stimuli 1–6 were constructed with $t_1 = 250$ msec and Stimuli 7–12 with $t_2 = 800$ msec.

Results

As in Experiment 1, the data of the two groups of subjects were pooled because no systematic differences between musically trained and untrained subjects were found. Figure 3 shows the mean ratios of the within-group intervals (t_1/t_2) for the imitation responses. Table 4 presents the means, standard deviations, and drifts for both the $t_1/t_{2(1)}$ and $t_1/t_{2(2)}$ ratios.

Imitation of the t_1–t_2–t_2 patterns shows roughly the same characteristics as imitation of the t_1–t_2 patterns in Experiment 1. The same general tendency toward a 1:2 relation of the intervals during reproduction is found for both the within-group intervals ($t_1/t_{2(1)}$) and for t_1 with the between-groups interval ($t_1/t_{2(2)}$). Still, there are some differences. First, the standard deviations of the patterns with interval ratios $\leqslant$ .05 (.25, .29, .33, and .5) are systematically higher than for the corresponding t_1–t_2 patterns, indicating that the subjects had trouble reproducing these patterns. During the experiment several subjects did complain about the difficulty of the sequences. Second, it is especially notable that the subjects had considerable difficulty reproducing the pattern with ratio .5, as reflected in the deviant reproduced ratios (except for the $t_1/t_{2(1)}$ ratio in the 250–500–500 stimulus) and in the extremely high standard deviation. Remember that the .5 ratio was the only one that was correctly imitated in a t_1–t_2 pattern.

Discussion

From a comparison of the imitation of t_1–t_2 patterns in Experiment 1 and the imitation of t_1–t_2–t_2 patterns in Experiment 2, it can be concluded that the tendencies found are typical for the reproduction of durations in this sort of pattern and cannot be attributed to an experimental artifact, namely, the difference between within- and between-groups intervals. The errors found in the reproductions show that the subjects have great trouble in perceiving and reproducing most of the presented duration relations correctly. The only duration ratio that is correctly reproduced is the ratio 1:2, and even that one appears to be difficult in a t_1–t_2 context.

Table 3
t_1–t_2–t_2 *Patterns Presented in Experiment 2*

Stimulus	t_1	t_2	t_2	t_1/t_2	Stimulus	t_1	t_2	t_2	t_1/t_2
1	250	1,000	1,000	.25	7	200	800	800	.25
2	250	862	862	.29	8	232	800	800	.29
3	250	750	750	.33	9	266	800	800	.33
4	250	500	500	.50	10	400	800	800	.50
5	250	379	379	.66	11	533	800	800	.66
6	250	333	333	.75	12	600	800	800	.75

Note. Only one period of each sequence is shown. Durations are in msec.

This result strongly supports Fraisse's (1956) idea that with temporal durations, subjects essentially only dispose of two time categories, a "short" and a "long" one that roughly relate as 1:2. In addition to substantiating Fraisse's own experiments, my data also show that if the interval ratios in the stimulus are smaller than .5 (.25, .29, .33, and .4), the subjects show a strong tendency toward a .5 ratio in their reproductions.

These findings imply that subjects do not conceptualize a temporal sequence as a chain of independent durations, but instead they regard the durations as related. In this context it is revealing that most subjects do not discover simple duration relations like 1:3 or 1:4. The data show clearly that these patterns are reproduced with the same tendency toward 1:2 as patterns with a more complex duration relation like 100:344 (.29).

The perception of the patterns studied in Experiment 1 and 2 can be described as fol-lows: The subject can spontaneously dispose of only two duration categories, which roughly relate as 1:2. For every sequence to be imitated, the subject tries to describe it as 1:2, but at the same time, the subject is aware that the stimulus deviates from the simple ratio but has difficulty assessing the extent of the deviation.

Suppose the subject's imitation lies between what he or she heard (the stimulus) and his or her schema, 1:2. As he or she produces imitations, he or she averages between his or her own imitation and 1:2. This would explain a progressive tendency toward the internal structure. In any case, the subjects' imitative behavior demonstrates their internally preferred way of coding temporal sequences and, at the same time, their seemingly limited capacity to reproduce even simple temporal patterns.

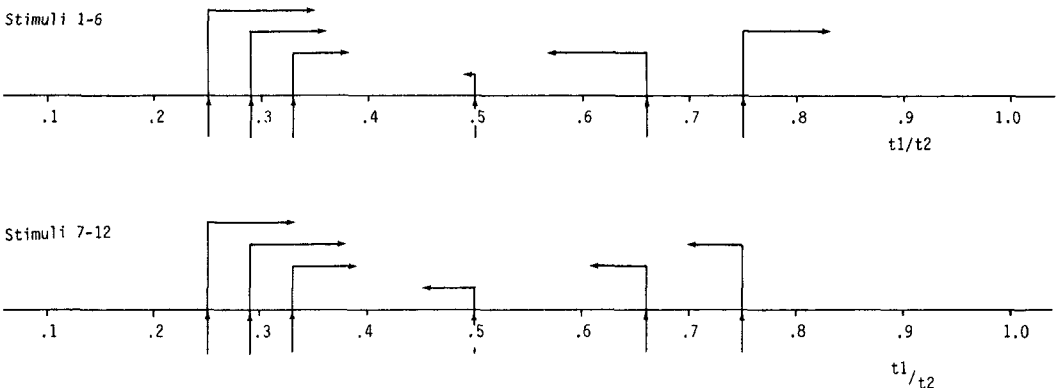There are two problems with this proposal. The problems subjects have in reproducing

Figure 3. Distortion of duration ratios in the within-group intervals of the imitations of t_1–t_2–t_2 patterns. (Duration ratios of the stimuli are indicated by arrows pointing to the abscissa that displays the continuum of t_1/t_2 ratios. The end points of the horizontal arrows indicate the duration ratios in the imitations. See also Table 4.)

Table 4
Mean, Standard Deviation, and Drift of the Interval Ratios in the Imitations of t_1–t_2–t_2 *Patterns*

	Stimuli				Imitations					
					$t_1/t_{2(1)}$			$t_1/t_{2(2)}$		
No.	t_1	$t_{2(1)}$	$t_{2(2)}$	t_1/t_2	M	SD	Drift	M	SD	Drift
					$t_1 = 250$					
1	250	1,000	1,000	.25	.35	.08	−.04	.32	.05	−.02
2	250	862	862	.29	.36	.07	−.05	.36	.09	.00
3	250	750	750	.33	.38	.06	−.02	.37	.08	.03
4	250	500	500	.50	.49[a]	.12	−.01[a]	.44[a]	.11	.04[a]
5	250	379	379	.66	.57[a]	.13	−.06[a]	.51[a]	.16	.00[a]
6	250	333	333	.75	.83[a]	.19	.05[a]	.81[a]	.26	.10[a]
					$t_2 = 800$					
7	200	800	800	.25	.33	.07	−.05	.32	.06	.02
8	232	800	800	.29	.38	.09	−.02	.36	.09	.01
9	266	800	800	.33	.39	.08	−.02	.37	.08	.02
10	400	800	800	.50	.45[a]	.10	−.01[a]	.44	.09	.02
11	533	800	800	.66	.61[a]	.19	.03[a]	.58[a]	.10	.06[a]
12	600	800	800	.75	.65[a]	.20	.02[a]	.62[a]	.15	.08[a]

Note. N = 20.
[a] These values are unreliable, as reflected by the high *SD*.

the t_1–t_2–t_2 pattern with t_1/t_2 ratio of .5 cannot be explained by the proposed way of coding temporal sequences; on the contrary, the current interpretation would predict no problems with the imitation of that pattern. Second, it is difficult to believe that the capacity to reproduce temporal sequences is as limited as implied by the description and by my data. It seems unlikely that even musically trained subjects would be unable to imitate simple duration ratios like 1:3 and 1:4, particularly since these duration relations occur frequently in adjacent intervals in music.

These considerations suggest that the internal representation of durations may be dependent on the context in which the durations occur. In other words, people may possess internal structures that fit some temporal sequences at the exclusion of others. This hypothesis was tested in Experiment 3.

Experiment 3

Method

Twenty-four subjects—9 with musical training who had participated in Experiment 2 and 15 new musically untrained subjects—imitated six temporal sequences in

the same way as described before. These six sequences were composed in a specific way, as shown in Table 5. Except for Stimulus 1, the different intervals forming ratios 1:2, 1:3, 1:4, and 2:3 were arranged in sequences so that when repeated in a cyclic fashion, sequences that occur frequently in music resulted.

Results

Table 6 summarizes the imitation responses separately for the musically trained and untrained subjects. As before, the mean ratios of several interval pairs in the stimuli and responses are given, as well as the standard deviation and drift.

Sequence 1 (200–200–600) was accurately imitated by the musically untrained subjects, but the musically trained subjects showed a strong tendency toward a 1:2 relation. Sequence 2 (250–250–500) produced precisely the reverse outcome, with good imitation by the musically trained subjects (note also the low standard deviation) and poor imitation by the untrained subjects. Sequence 3 (250–250–250–750) was correctly imitated by both groups of subjects. Sequence 4 (250–250–250–250–1,000) was well imitated by both groups of subjects. (Note the low standard deviation for the

Table 5
Temporal Patterns Used in Experiment 3

Stimulus	Pattern	Occurring ratios
1. 200–200–600	$t_1-t_1-t_2$	.33
2. 250–250–500	$t_1-t_1-t_2$	.5
3. 250–250–250–750	$t_1-t_1-t_1-t_2$	.33
4. 250–250–250–250–1,000	$t_1-t_1-t_1-t_1-t_2$	.25
5. 250–250–250–250–500–500	$t_1-t_1-t_1-t_1-t_2-t_2$	.5
6. 400–400–400–600–600	$t_1-t_1-t_1-t_2-t_2$	.66

Note. Only one period of each sequence is shown. Durations are in msec.

musically trained subjects.) Sequence 5 (250–250–250–250–500–500) was again well imitated by both groups of subjects, but Sequence 6 (400–400–400–600–600) was imitated poorly by both groups. (Note also the exceptionally high standard deviations.)

A general characteristic can be seen in all subseries of equal intervals (Stimuli 3, 4, 5, 6): Subjects tend to lengthen the first and last intervals (Povel, 1977).

Discussion

The most important finding of this experiment was that subjects were able to imitate interval relations of 1:3 and 1:4 accurately in the contexts used. Here I will try to define what the specific characteristics of the context were that enabled the subjects to conceptualize the sequences correctly. A first guess is that the sequences were correctly represented because the short interval (t_1) was repeated at least twice. (Remember that the interval pattern of the stimuli in Experiments 1 and 2 were t_1-t_2 and $t_1-t_2-t_2$, respectively.) This does not seem to suffice as an explanation, however, because the pattern 250–250–500 was well imitated by the musically trained subjects (and slightly less so by the untrained subjects), whereas the pattern 200–200–600 was poorly reproduced by the musically trained subjects (but, paradoxically, well imitated by the untrained subjects, a point I will return to later).

The explanation for subjects imitating some of these sequences correctly does not seem to lie in a specific pattern but instead seems to be determined by a specific combination of interval pattern and duration ratios. One characteristic of the correctly imitated sequences is that the shorter interval is repeated so many times that the total duration of the repetitions is equal to the longer interval. This configuration enables the subject to define the shorter interval as a subdivision of the longer one while this longer interval is repeated continuously. This supports the beat-based model proposed in the introduction.

According to the beat-based model, the first step in the process of perception is the detection of events (most likely accented events, which in the stimuli used arise from the temporal structure) that subdivide the sequence into equal intervals. The events are called *beats*. Thus, the pattern 250–250–250–750 could be heard as having recurrent beats at 750-msec intervals. Specifically, the pattern could be described as consisting of two intervals bordered by the beats, one interval filled with tones at one third the beat interval and the second empty. The most significant aspect of this model seems to be that the description of the sequence does *not* start with the smallest duration and try to define the total sequence in terms of that duration; rather, it starts at a higher level, the beat interval, and describes the sequence in terms of this interval. Any duration shorter than the beat interval is described as a subdivision of the beat interval. As defined before, the 250-msec intervals in the 250–250–250–750 msec sequence are coded this way: The first beat interval is filled with events at one third the beat duration, or similarly, the interval is filled with events that occur with a frequency three times as high as the beat frequency.

If the model proposed is valid, I should be able to show that the patterns used in Experiment 1 and 2 that for the greater part are imitated incorrectly do not fit in this model. It is true that these patterns did not evoke any typical beat reaction, as for in-

Table 6
Ratios of the Relevant Interval Pairs as Present in the Stimuli of Experiment 3 and as Imitated

Interval pair	Sequence 1[a] Musical subjects		Nonmusical subjects	
	t_1/t_3	t_2/t_3	t_1/t_3	t_2/t_3
Stimulus	.33	.33	.33	.33
Imitation	.46	.46	.35	.36
SD	.05	.04	.05	.06
Drift	.00	.04	−.04	.05

	Sequence 2[b]			
	t_1/t_3	t_2/t_3	t_1/t_3	t_2/t_3
Stimulus	.50	.50	.50	.50
Imitation	.48	.49	.44	.45
SD	.03	.03	.05	.05
Drift	.00	.00	.00	.03

	Sequence 3[c]					
	t_1/t_4	t_2/t_4	t_3/t_4	t_1/t_4	t_2/t_4	t_3/t_4
Stimulus	.33	.33	.33	.33	.33	.33
Imitation	.35	.33	.35	.34	.33	.35
SD	.05	.05	.05	.05	.05	.05
Drift	.01	.02	.04	−.01	.03	.02

	Sequence 4[d]							
	t_1/t_5	t_2/t_5	t_3/t_5	t_5/t_5	t_1/t_5	t_2/t_5	t_3/t_5	t_4/t_5
Stimulus	.25	.25	.25	.25	.25	.25	.25	.25
Imitation	.26	.26	.25	.27	.26	.25	.25	.27
SD	.02	.02	.02	.02	.04	.04	.04	.04
Drift	.00	.02	.02	.02	.01	.02	.03	.03

	Sequence 5[e]									
	t_1/t_5	t_2/t_5	t_3/t_5	t_4/t_5	t_5/t_6	t_1/t_5	t_2/t_5	t_3/t_5	t_4/t_5	t_5/t_6
Stimulus	.50	.50	.50	.50	1.00	.50	.50	.50	.50	1.00
Imitation	.50	.48	.49	.51	1.00	.49	.48	.48	.52	1.01
SD	.02	.02	.02	.02	.03	.04	.03	.03	.04	.04
Drift	.01	.00	.00	.01	.02	.00	.01	.03	−.02	.00

	Sequence 6[f]							
	t_1/t_5	t_2/t_5	t_3/t_5	t_4/t_5	t_1/t_5	t_2/t_5	t_3/t_5	t_4/t_5
Stimulus	.66	.66	.66	1.00	.66	.66	.66	1.00
Imitation	.62	.59	.61	1.01	.49	.48	.56	.89
SD	.14	.14	.13	.33	.12	.11	.21	.19
Drift	.02	.09	.07	.14	.04	.09	.09	.13

Note. From the imitations, the mean, the standard deviation, and the drift of the produced ratios are presented.
[a] $t_1 = 200$, $t_2 = 200$, $t_3 = 600$.
[b] $t_1 = 250$, $t_2 = 250$, $t_3 = 500$.
[c] $t_1 = 250$, $t_2 = 250$, $t_3 = 250$, $t_4 = 750$.
[d] $t_1 = 250$, $t_2 = 250$, $t_3 = 250$, $t_4 = 250$, $t_5 = 1,000$.
[e] $t_1 = 250$, $t_2 = 250$, $t_3 = 250$, $t_4 = 250$, $t_5 = 500$, $t_6 = 500$.
[f] $t_1 = 400$, $t_2 = 400$, $t_3 = 400$, $t_4 = 600$, $t_5 = 600$.

stance foot tapping, on the part of the subjects. On the contrary, subjects frequently reported experiencing a great irregularity in the presented sequences. It is possible, however, to conceive a beat in these sequences, coinciding with the first tone of each period. Note, however, that the description of the within-beat interval event(s), under the assumption that the beat interval coincides with the period, requires a subdivision into unequal parts in all these sequences. The t_1–t_2 pattern 250–750, for instance, would be described as consisting of a beat of 1,000 msec, which is subdivided into two intervals of 250 and 750 msec respectively. It will be clear that the introduction of the beat concept in this way does not in any way reduce the pattern. If such a coding is still applied, it follows from our results that a subdivision of the within-beat interval into unequal parts is not possible within the model, with the exception of a subdivision into parts that relate as 1:2.

We encounter another constraint of the model in Stimulus 6, the pattern 400–400–400–600–600. Note that this stimulus can easily be segmented into equal intervals of 1,200 msec and therefore be described as the first beat interval filled with events at three times the beat frequency and the second beat interval filled with events at two times the beat frequency. But if we consider the contorted imitation of this stimulus (even by rather experienced musicians), we must conclude that the proposed coding is not a natural one. Therefore, it is suggested that in a "natural" coding system, the first subdivision level below the beat interval can be of only one type, a subdivision into either two, three, or possibly more units.

Because of the conclusion that subdivisions of beat intervals of the same sequence must be of one type, we would expect that Stimulus 5, 250–250–250–250–500–500, would be badly imitated if it is coded as two beat intervals of 1,000 msec each, the first subdivided into four and the second into two durations. This code is drawn on the left-hand side of Figure 4. The fact, however, that this sequence is imitated perfectly suggests two other possible codings, both fitting within the beat-based model. The first one, shown on the right-hand side of Figure 4, also starts with two 1,000-msec beats. In this
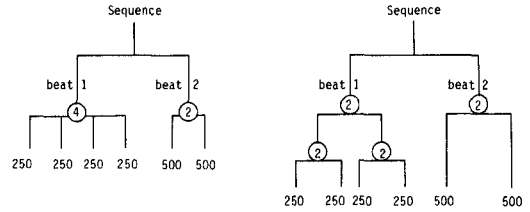

Figure 4. Two alternative codes for a temporal sequence. (The numbers at the nodes refer to the number of branchings at that level.)

description, the first subdivision of each beat interval is divided into two equal parts, but then the first beat is further subdivided on a lower level into two subparts. Another possible coding, involving only two levels, starts with four 500-msec beat intervals, the first two intervals subdivided into two equal parts of 250 msec each.

Two questions remain to be answered: What stimulus characteristics determine the selection of the beat interval? and How many hierarchical levels are conceivable in this system? The two questions are related because the selection of the beat determines the possible number of levels in the hierarchy: In general, the longer the beat interval, the more levels are possible. This can be demonstrated with the temporal sequence 600–600–200–200–200–200–200–200. Depending on whether beat intervals are 1,200- or 600-msec long, the two codes given in Figure 5 would result. A question that arises in the context of this example is whether a specific subdivision on one level also dictates the subdivision on a lower level. If that were the case, the left-hand code would not be possible because Level 2 is subdivided into two subunits and Level 3 into three. This question could be studied if one could manipulate selection of beat intervals by subjects.

Some exploratory work suggests that beat selection is influenced by at least the following stimulus characteristics: (a) accent level (if a temporal pattern allows the selection of two beat intervals, as in the sequences presented before, the beat interval will be chosen that is marked by consistently accented beats) and (b) tempo (if duration of the last example presented is doubled, selection of beats according to the left-hand diagram in Figure 5 becomes impossible). The
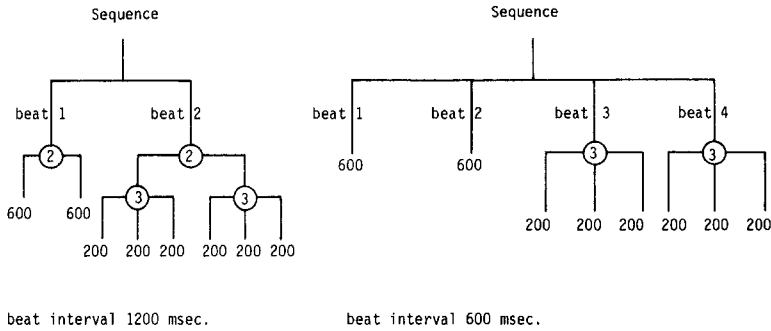
Figure 5. Two alternative codes for a temporal sequence. (The numbers at the nodes refer to the number of branchings at that level.)

preferred duration of a beat interval lies between approximately 250 and 1,500 msec. It is also possible that codability of the events within a beat-interval candidate may influence selection. Remember that we reported almost no beat responses to the stimuli of Experiments 1 and 2. How the mentioned factors interact in the final choice of the beat interval will have to be determined by further research.

Finally, it is of interest to try to understand the differences between the musically trained and untrained subjects against the background of the proposed model. First, it should be noted that there were no discernable differences between the two groups in imitation of the longer sequences (3, 4, 5, and 6); only the two short sequences (1 and 2) were imitated differently. This fact suggests that the difference in imitation behavior between the two groups can be understood if we assume that for musically untrained subjects, the two short stimuli (200–200–600 and 250–250–500) did not contain enough clues to trigger a beat-based coding strategy. On the other hand, the musically trained subjects are accustomed to sequences that fit that model and applied the model to both sequences, including the stimulus (200–200–600) that cannot be described by that model. Thus, the musically untrained subjects do not try to find a recurrent beat in the 250–250–500 sequence but instead perceive it as an unstructured group of three tones and enlarge the within-group intervals. The musical subjects, on the other hand, apply a regular recurrent beat notion to Stimulus 1 (200–200–600), which

is accomplished by reducing the long interval.

According to the description presented here, subjects have at least two possible ways to code temporal sequences. Specific characteristics of the stimulus determine which coding scheme is applied. The stimuli of Experiments 1 and 2 did not fit a beat-based coding and were therefore internally represented as rather unstructured groups of tones. Such an unstructured representation is imprecise and is strongly influenced by a tendency to conceptualize duration differences in a 1:2 relation. On the other hand, the stimuli used in Experiment 3, with the exception of Stimulus 1, did fit a beat-based description that was actually used by all subjects. It is understandable that the musically trained subjects also applied it to a stimulus that did not fit such a description.

General Discussion

The general conclusion drawn from the current experiments is that the perception of time, or more precisely the perception of temporal sequences, is determined in large part by an internal structure on which subjects try to map presented temporal sequences. Only when a temporal sequence completely fits this mental structure are subjects able to imitate the temporal sequence correctly. In all other cases severe distortions occur. The distortions are reflected in higher standard deviations but, more importantly, in systematic errors that can be viewed as attempts by subjects to fit the presented sequence into their internal structure. These

continuous attempts by subjects to fit the presented sequences into one or another internal structure, even when only two intervals are involved (Experiment 1), are enough ground to reject a chainlike description of the internal representation of durations. Our results clearly show that subjects are unable to store durations independently.

This study suggests that subjects try to apply a beat-based model and that imitation of actual sequences will be accurate to the degree that the structure of the sequence fits the structure of the model. The first step of coding consists of an attempt to divide the sequence into equal intervals bordered by events. The selection of this so-called beat interval is constrained by length—it cannot be longer than about 1.5 sec—and possibly by the structure of the events within the interval. If the complexity of the latter is too high, there will be a strong tendency to define a shorter beat interval. It appears from our data that only three fillings of a beat interval are possible within the model: empty; filled with events at equal intervals; and filled with events that unequally subdivide the beat interval, provided the subdivision is in two parts that relate as 1:2. This last constraint of the model is due to Fraisse's (1946, 1956) finding that in spontaneous tapping, subjects tend to use two durations that relate roughly as 1:2. The exact relation is subject to considerable change and caused by only partly understood factors.

There is an additional limitation to the point above which states that beat intervals in a sequence must all be subdivided in the same way. All of these constraints greatly limit the number of temporal sequences that fit the model. Even simple temporal patterns like the ones used in Experiment 1, and more so those in Experiment 2, give rise to serious coding problems because the preferred coding is not compatible with the structure of the sequence.

The difference between the model proposed here and Martin's (1972) model, described in the introduction, is that the current model is much less hierarchical. If Martin's model is in fact too hierarchical, as suggested in the introduction, the beat-based model may prove to possess too little hierarchy, especially for the coding of longer

sequences. Maybe a further attempt to apply and test both models may point to an intermediate solution.

The puzzling results of Vorberg and Hambuch's (1978) experiment seem to become understandable in light of the findings which suggest that only production models that assume dependency between so-called timekeepers are compatible with the cognitive structures involved.

A final remark is in order about the origin of the internal structure: From comparing the imitation of sequences of musically trained and untrained subjects, it can be concluded that the cognitive structures are not specific to musically active subjects. It is still possible, of course, that subjects acquire the internal structure by passive experience with music. The fact, however, that even trained musicians have problems with patterns that occur frequently in music (e.g., the 400–400–400–600–600 pattern) seems to point to a more universal origin. This can be verified by comparing the imitative behavior of subjects from different musical cultures.

Reference Notes

1. Povel, D. J. L. *The perception and production of auditory sequences and rhythms* (Tech. Rep. 79-1). Bloomington: Indiana University, Department of Psychology, March 1979.
2. Povel, D. J. L. *The effect of different types of feedback on the imitation of rhythms.* Paper presented at the 6th International Colloquim of Empirical Esthetics, Paris, July 1976.
3. Bamberger, J. *The development of musical intelligence I: Strategies for representing simple rhythms* (Tech. Rep. No. 342). Cambridge: Massachusetts Institute of Technology, Artificial Intelligence Laboratory, November 1975.

References

Bartlett, N. R., & Bartlett, S. C. Synchronization of a motor response with an anticipated sensory event. *Psychological Review*, 1959, *66*, 203–218.
Cooper, G., & Meyer, L. *The rhythmic structure of music.* Chicago: University of Chicago Press, 1960.
Fraisse, P. Contribution à l'étude du rythme en tant que forme temporelle. *Journal de psychologie normale et pathologique*, 1946, *39*, 283–304.
Fraisse, P. *Les structures rythmiques.* Louvain: Publications Universitaires de Louvain, 1956.
Greeno, J. G., & Simon, H. A. Processes for sequence production. *Psychological Review*, 1974, *81*, 187–198.
Leeuwenberg, E. L. J. A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, 1971, *84*, 307–349.

Martin, J. G. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 1972, *79*, 487–509.

Michon, J. A. *Timing in temporal tracking.* Assen, The Netherlands: Van Gorcum, 1967.

Michon, J. Programs and "programs" for sequential patterns in motor behavior. *Brain Research*, 1974, *71*, 413–424.

Povel, D. J. L. Temporal structure of performed music. Some preliminary observations. *Acta Psychologica*, 1977, *41*, 309–320.

Restle, F. Theories of serial pattern learning: Structural trees. *Psychological Review*, 1970, *77*, 481–495.

Simon, H. A., & Sumner, R. K. Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment.* New York: Wiley, 1968.

Treisman, M. Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs*, 1963, *77*(13, Whole No. 576).

Vitz, P., & Todd, T. C. A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, 1969, *76*, 433–449.

Vorberg, D., & Hambuch, R. On the temporal control of rhythmic performance. In J. Requin (Ed.), *Attention and performance VIII.* Hillsdale, N.J.: Erlbaum, 1978.

Wagner, C. The influence of the tempo of playing on the rhythmic structure studied at pianists playing scales. In E. Jokl & H. Hebbelinck (Eds.), *Medicine and sport VI.* Basel, Switzerland: Karger, 1971.

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

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

Intons-Peterson Appointed Acting Editor

Frank Restle, editor of the *Journal of Experimental Psychology: Human Perception and Performance*, died on October 25, 1980, after a brief illness. Margaret Jean Intons-Peterson, Restle's colleague at Indiana University, will serve as Acting Editor until the Publications and Communications Board conducts a search and names a permanent editor. Effective immediately, authors should submit manuscripts to Intons-Peterson at the Department of Psychology, Indiana University, Bloomington, Indiana 47405.

Candidates for journal editorships must be members of APA. To nominate candidates, prepare a statement of one page or less in support of each nomination, and submit no later than February 15, 1981, to the Chair of the Search Committee, David Zeaman, Department of Psychology, Box U-20, University of Connecticut, Storrs, Connecticut 06268.