Discrimination of auditory temporal patterns

JAAN ROSS and ADRIANUS J. M. HOUTSMA

Institute for Perception Research, Eindhoven, The Netherlands

Two *same-different* discrimination experiments were performed for click patterns having a total duration of about 4 sec and interclick intervals of $n \times 250$ msec, with n a random integer. In Experiment 1, the influence of the physical click group structure on discrimination performance was investigated. In Experiment 2, the effect of the strength of an induced internal clock on discrimination performance was measured. Performance was poor if the group structure of clicks was maintained during a change in click pattern and also if the induced internal clock strength was low. The performance of about 70% of the subjects improved significantly if either a change in click grouping structure occurred or a strong internal clock could be induced. These results cannot be accounted for with simple models based on single-interval duration discrimination or between-pattern correlation statistics.

This paper deals with the cognitive representation of rhythmic patterns in music, particularly with factors that either enhance or inhibit our ability to detect small changes in such patterns. The sound patterns we studied are simple click patterns, which are quasimusical in the sense that clicks are separated by time intervals that are integer multiples of some basic time interval. Click patterns are devoid of any pitch, timbre, or dynamic variation. The use of such stimuli permits one to avoid possibly confounding influences of pitch, timbre, or loudness on perceived rhythm, as is easily the case when one is listening to real music. It nevertheless preserves a situation close enough to musical practice to ensure that results will be musically relevant.

Handel (1989) has defined rhythm as an interplay between meter and grouping. According to this viewpoint, we might consider rhythm as a sequence of acoustic events that (1) interact with the implemented periodical framework (i.e., meter) and (2) may be divided into a number of sound-event clusters or groups. His dichotomy between meter and grouping implies a relationship between serial and hierarchical processing of temporal auditory information. Grouping would apparently be associated with serial processing, while meter presumes complex multi-level coding of sound information.

The influence of both meter and grouping of events on rhythm perception has been confirmed by data from several investigations. One of the first among these, by Royer and Garner (1966), established the importance of groups of adjacent sound elements in a pattern (called runs) for its perceptual organization. Subjects had to reproduce the

time pattern of an indefinitely repeated eight-sound sequence from two different buzzers by pressing two push buttons in synchrony with the perceived pattern. The investigators found that subjects tended to start responses either at the longest possible run or with a multiple repetition of a subgroup of a sequence. When neither of these conditions could be met, the time pattern of the sequence was difficult to repeat.

If sound patterns are composed of single sounds and pauses, instead of two alternating types of sounds, the structure of pause groups (or gaps) seems to be of equal importance to that of runs. In an experiment performed by Preusser, Garner, and Gottwald (1970), subjects had to continuously reproduce repeated sequences composed of buzzer sounds and pauses, at the rate of three events per second. The reproductions revealed two organizing principles: preference for the longest gap at the end, or for the longest run of events at the beginning of a pattern.

Martin (1972) suggested that cognitive representation of auditory events has a hierarchical component rather than being a purely serial process, and tested this idea experimentally (Sturges & Martin, 1974). Subjects were exposed to two-element patterns of seven or eight events, which were immediately either repeated exactly or repeated with a small change. Patterns were rhythmically well structured or poorly structured. The seven-event patterns were derived from the eight-event patterns according to an ad hoc principle, aimed at maintaining the same number of runs and the same number and location of accents in both patterns. The double task of the subjects was to detect exact repetitions and to write down the pattern. Because the subjects performed generally better with the rhythmically structured and with the eight-event patterns in both tasks, Sturges and Martin (1974) concluded that metric organization of a sequence favors its cognitive representation.

To define the metric organization of a sequence, Povel and Essens (1985) introduced the concept of an internal

This study was made possible by a postdoctoral research fellowship grant from the Eindhoven University of Technology to J.R., who is affiliated with the Institute of Language and Literature, Tallinn, Estonia. Correspondence concerning this article should be addressed to A. J. M. Houtsma, Institute for Perception Research, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

clock. This concept can be described as follows. Exposed to a temporal sequence, the listener tries to induce a simple isochronous beat pattern (clock) that perceptually seems to fit the temporal structure of the sequence. Whether a clock will be induced and, if so, which clock is induced depends on the distribution of perceptually accented events in the sequence. For the results of Sturges and Martin (1974), a rather straightforward perceptual accentuation rule seemed to apply: only run-initial high tones are accented. Povel and Okkerman (1981) needed a slightly more complicated set of rules in order to account for the perceived accent structure in equitone sequences: perceptually accented sound events in a sequence are (1) single tones, (2) the second tones in two-tone clusters, and (3) the initial and final tones of clusters consisting of three or more tones.

Both the metric organization and the run structure of sound sequences appear to have influence on our ability to discriminate or recognize such sequences, as has recently been shown by Palmer and Krumhansl (1990). Two other recent papers, however, seem to show that one can account for temporal pattern discrimination performance solely on the basis of single-interval duration discrimination thresholds as known from traditional psychophysics. In the first of these papers, Sorkin (1990) investigated perception of short (629- to 1,311-msec) temporal sequences consisting of 1-kHz tone bursts. He varied inter-onset time intervals (gaps) in a random manner, at the same time controlling the number of gaps in a sequence, the mean gap duration, and their standard deviation. Subjects had to discriminate between the cases of two identical sequences and the cases of two random sequences having a correlation coefficient between zero and .8. Performance could be explained with an intertone gap correlation model that was limited by an internal variability (jitter) of about 15 msec. This is at the high end of the range of values obtained in duration discrimination experiments with one-interval paradigms (Abel, 1972).

In the other study, Monahan and Hirsh (1990) investigated discrimination of pairs of sequences consisting of six 1-kHz tone bursts each. The intertone gaps in both sequences were either short or long, the long gaps being twice as long as the short gaps, in the overall range between 50 and 400 msec. Both sequences in a pair were identical, except for a delay of one tone either in the first or in the second sequence, the length of the delay being varied. It was found that listeners could detect this delay about as successfully as in single-interval discrimination paradigms.

In this paper, we have attempted to combine the stochastic pattern generation procedure used by Sorkin (1990) with the use of quasimusical temporal sequences of events that are separated by time intervals with simple-integer relationships. The former seems attractive from a heuristic point of view and contrasts with the ad hoc principles of pattern construction in some earlier studies such as Sturges and Martin's (1974). The latter is associated with a more

traditional paradigm of rhythm reproduction experiments, being closely related to Western music practice (Povel & Essens, 1985).

The first question addressed by this study is how discrimination performance is affected by changes (1) in grouping and (2) in pattern-inherent potential for metric organization. Two experiments are reported in which detection of a small change in pattern was measured under two experimental conditions. In the first experiment, one condition implied that a change in a sound pattern did not cause any changes in the sound event grouping, whereas under a second condition it mostly did. In the second experiment, the numerical estimate of metric strength developed by Povel and Essens (1985) was applied in order to vary internal clock strength induced by the patterns.

The second question investigated in this study is whether Sorkin's (1990) correlation model, which involves internal noise with a magnitude similar to the single-interval duration discrimination threshold, is still valid for patterns with longer inter-onset time intervals and longer overall durations than the ones he used. It is likely that patterns of different complexity are processed in different ways and that the apparent similarity with single temporal-gap discrimination holds only for short and/or simple sound sequences. Suggestions of this kind have been expressed by Sorkin, who found that discrimination performance was degraded if the time span of the sequences to be compared was longer than 1 sec, or if listeners were required to compare sequences containing more than 12 auditory events. Monahan and Hirsh (1990) found interactions of a pattern with position of the delayed tone to be dependent on the tempo, which leads to further suggestions that different performance models are needed to explain results obtained at different tempos.

GENERAL METHOD

Design

Three sets of 50 temporal random click patterns were generated, one to be used for the first and two for the second experiment. The generation algorithm was such that every 250 msec the occurrence of a click or a rest (the absence of a click) was determined with a probability of .5. In addition, the minimum possible number of clicks in a pattern was set at five in Experiment 1, and at seven in Experiment 2, in order to avoid the occurrence of completely silent patterns. Finally, the occurrence of a click at the very last position of a pattern was not allowed (the reason for that restriction is discussed below).

Random click patterns can conveniently be described as series of ones and zeros, a one marking a click and a zero marking a rest, with every digit corresponding to a time interval of 250 msec. For example, a pattern 10100011 would consist of four clicks with interclick time intervals of 500, 1,000, and 250 msec, respectively.

Half of the patterns in a set were subsequently modified in the following two ways. A pair of two successive events, a click immediately followed by a rest (a 1-0 combination), occurring somewhere in the pattern, was randomly selected. For a type A modification, the temporal order of the click and the rest was simply reversed. For a type B modification, an extra 250-msec rest (zero) was inserted after the rest. For instance, if in an original 15-step

pattern represented as (111001100001110) the 1-0 pair at the third and fourth positions was randomly selected, the type A modification would yield the pattern (110101100001110) and the type B modification would yield (111000110000111). Note first that the presence of a zero at the last position of an original pattern is necessary to make a type B modification possible without either reducing the total number of clicks in a pattern or increasing the length of a pattern. Note further that if the randomly selected pair of click and nonclick happens to be at the very end of the pattern, the addition of an extra pause for a type B modification does not change the pattern.

The principal difference between the type A and type B modifications is that the former almost always changes the click group structure, whereas the latter never does. In the example given above, a click structure with three groups of three, two, and three successive clicks each has been transformed into a structure with four groups of two, one, two, and three clicks, respectively, for type A, but has remained the same after the type B modification. This is true in most cases. It can happen, however, that a type A modification does not change the click group structure. If the click in the selected pair forms a one-element group (i.e., if it is preceded and followed by a nonclick), an interchange of the positions of the elements in a pair affects neither the number of click groups in the pattern nor the number of clicks in each group. The probability of this happening is .25.

After some pilot experiments, the length of a pattern was set at 15 units of 250 msec, making the total pattern length equal to 3,750 msec. For the particular group of subjects used, this length appeared to keep scores within reasonable limits—that is, between 50% and 100% correct.

Procedure

The general experimental procedure was that each subject was given all stimulus sets once, with one half of the stimuli being pairs of an original pattern followed by a modification, and another half being pairs of an original and an exact repetition. The silent interval between patterns of a pair was 5 sec. The entire stimulus, including the interpattern silent interval, was stored as a single buffer in computer memory and converted into 16-bit sound at a sampling rate of 1 kHz. The waveform of each click was a positive 1-msec pulse followed by a negative 1-msec pulse, with a peak amplitude equal to that of an 85-dB SPL sinusoid.

Subjects were seated in a sound-insulated chamber and received sound stimuli via TDH-49 headphones. They were instructed to respond whether the two patterns in a pair were identical or not by pressing one of two buttons on a response box. The time allotted for a response was unlimited. Feedback about the correct answer was provided immediately following a response.

Subjects

There were 14 subjects, all recruited from the research staff of the Speech and Hearing Group of the Institute for Perception Research in Eindhoven. They had various degrees of musical experience, ranging from professional musical training to no musical training at all. All subjects did have experience with psychoacoustic experiments. Before the first session, they received written instructions as well as a few oral comments on the task. They were exposed to 10–20 pairs of patterns as training, with responses not being recorded. They were then exposed to 50 pairs of patterns (one stimulus set), which typically lasted about 20 min. Each subject did one stimulus set for each experimental condition, usually on the same day. Those participating in both experiments did no more than two sessions on 1 day.

EXPERIMENT 1

This experiment was performed to see whether click patterns would be easier to discriminate if a change in

pattern also implies a change in the physical click group structure. The experiment was conducted in two parts. In the first part, pulse patterns were either repeated exactly or followed by a type A modification (inducing a physical group structure change in 75% of the cases). In the second part, the same original pulse patterns were either repeated exactly or followed by a type B modified pattern, in which the group structure was always preserved. The a priori probability of an exact repeat was .5 for both parts.

If we consider a pattern modification (type A or type B) as a target, as in a classical signal detection experiment, we can distinguish *hits* (correct identification of a change), *false alarms* (reporting a change if pulse pattern is an exact repeat), *misses* (reporting a repeat while a change occurs), and *correct rejections* (reporting a repeat for an exactly repeated pattern). The usual way to present the data is then to plot the probability of a hit, P_h , against the false alarm probability, P_f . This determines a point on the so-called receiver operating characteristic (ROC) curve. The advantage of such a display is that the sensitivity index (d') and response bias are shown independently of one another. Deviation of data points from the negative diagonal is a measure of response bias, whereas the sensitivity index is determined by the distance measured along the negative diagonal from the center ($P_h = P_f = .5$) to the intersection with the ROC curve. The percentage of correct responses can, of course, be computed from P_h and P_f with the simple formula:

$$P_c = .5 P_h + .5 (1 - P_f), \quad (1)$$

given that the a priori probabilities of *same* and *different* trials were equal.

Figure 1 shows two sets of ROC data points, one set corresponding to each type of pattern modification, and each point representing the results of 1 subject. Squares are for type A, and crosses for type B modifications. From the 14 subjects who participated, the data of 1 subject were rejected because for both modification conditions that subject's performance was consistently at chance level.

One can see, first, that for both conditions the response bias is rather small, given the clustering of points around the negative diagonal. Second, one can see that, on the average, the sensitivity index d' is larger for pattern changes of the A type than for the B type. Statistical analysis done on individual d' values (paired t test) shows that this was significant at the 99% level. Given the limited response bias, the data may be pooled to obtain average percentages of correct responses for both conditions with the aid of Equation 1. This yields 85% and 64% correct for the modifications of types A and B, respectively.

We will next investigate whether it is possible to explain the obtained results if we consider the experiment as a single-interval discrimination task, by assuming that somehow the subject is able to concentrate only on time intervals in the sequence that are actually changing. Let us first consider the type B pattern modification in which a lengthening of an interclick interval by 250 msec had

Figure 1. Hit versus false alarm rates for individual subjects in Experiment 1, with targets being type A modifications (squares) and type B modifications (crosses). Receiver operating characteristic curves with d' values of 1.1 and 0.6, computed with Sorkin's model, are shown for comparison.

to be detected. Abel (1972) found that duration JNDs for time intervals greater than about 200 msec were reasonably well described by a constant Weber fraction of about 0.20. As long as the 250-msec increment is added to an interval that is equal to or smaller than 1,000 msec, the lengthened interval will be discriminable from the original one. That is, lengthened runs initially composed of one, two, and three zeros are easily discriminated from unlengthened runs. Only the lengthening of a run of four or more zeros will lead to discrimination failures. The probability of such a run is theoretically $.5^4 + .5^5 + .5^6 + \dots = \frac{1}{8}$ for an infinite sequence. Given the fact that patterns were limited to 15 events and the randomly selected 1-0 pair could be anywhere in this sequence, the actual probability of four or more successive silent events around the inserted silent event is closer to $1/10$. Consequently, we expect 90% of the stimuli with type B modifications to be above discrimination threshold, which, given the discrete nature of the interval change, should yield about 90% correct answers. With the type A modifications, in which a 1-0 combination is changed into a 0-1 combination, the situation is basically the same except that two silent intervals, the one preceding and the one following the click, will change simultaneously. The added information, in comparison with type B modifications, should therefore result in a score of even more than 90% correct.

Actually measured scores of 85% and 64% correct for the two conditions are much lower than predictions made on the basis of single-interval Weber fractions, especially for the type B modifications. This is not totally surpris-

ing. In the sound sequences used, there are many silent intervals and there is considerable uncertainty about the interval in which a change may occur. This uncertainty translates into poorer performance if a comparison is made with single-interval discrimination, where such uncertainty does not exist.

Let us next investigate how the results of Experiment 1 are accounted for by Sorkin's (1990) correlation model. In this model, discrimination between random pulse patterns is based on a simple correlation statistic, which is compared with a decision criterion. If it is larger, the subject decides that the patterns were different and, if smaller, patterns are considered to have been the same. The only free parameter of the model is internal noise (jitter), which consists of a constant term (about 12 msec) and a term that is proportional to the average interclick gap size.

Although Experiment 1 was not explicitly designed to test Sorkin's (1990) model, we have, in principle, all input parameters for the model available. Taking sample statistics of the original (unmodified) click patterns that were actually used, the average number of interclick gaps per pattern is found to be 5.4, and the average gap duration 497 msec, with a standard deviation of 312 msec. Sample correlation coefficients between original and modified patterns were found to be .75 and .90, respectively, for type A and type B modifications. For internal noise a value of 20% of the average gap duration was assumed, equal to the appropriate Weber fraction for such gap durations (Abel, 1972).

The d' values computed from Sorkin's model were 1.1 and 0.6 for the modifications of types A and B, respectively. Corresponding ROC curves are shown in Figure 1 and can be compared with the actually obtained data points. Average d' values, obtained by computing sensitivity indices for each subject and averaging across subjects, were found to be 2.3 ($SD = 1.0$) and 0.8 ($SD = 0.5$) for the two conditions, which is much higher than model predictions. In particular, the difference in sensitivity between the two conditions is greatly underestimated by the model. Also, if we use population statistics rather than sample statistics as model input, the predicted d' values increase to 1.3 and 0.7, respectively. Population statistics are obtained by generating many pulse patterns according to the rules explained in the General Method section, and computing their averages and standard deviations. This, however, still underestimates discrimination performance for type A modifications. One can even force the model predictions to match empirical findings for type B modifications by, for instance, decreasing the gap discrimination Weber fraction from 0.20 to 0.18. This raises predicted d' from 0.7 to 0.8, but raises d' for the other condition from 1.3 to only 1.5, still quite short of the actually found value of 2.3. This discrepancy appears to be consistent with Sorkin's own data, which suggest that his model breaks down for sound sequences that are larger than about 12 events or longer than about 1 sec.

One could also try to interpret the results in terms of a simple waveform correlation model that computes the

cross-correlation coefficient of both click waveforms for each trial. (Note that such a model is quite different from the gap-pattern correlation model of Sorkin discussed earlier). Correlations between an original click pattern and a type A modification should be larger than correlations between originals and type B modifications, since the former involves only an interchange of two elements whereas the latter involves a one-element shift of part of the pattern. One would therefore expect the type B modification to be easier to distinguish from the original than the type A modification. The data, however, show that the opposite is true.

On the basis of remarks by subjects about the experimental tasks, we believe that some subjects adopted a special strategy for distinguishing type A modifications from originals. By coding the click group structure of each pattern in the form of a small set of digits and by subsequently comparing the two codes, subjects should in theory be able to reach an 87.5% correct score level. For instance, the original pattern (111001100001110), mentioned in the General Method section, would be coded as (323), representing the run lengths of successive ones (clicks separated by 250-msec intervals), whereas a type A modification of this pattern (110101100001110) would be coded as (2123). Incorrect responses (12.5%) would occur if a type A modification did not result in a group structure (digit code) change, which is expected to happen in 25% of half the number of trials. This predicted level is very close to the actually observed score level of 85% correct. Consistent and successful use of such a simple counting strategy would also imply, however, that all response mistakes would be "misses" and that there would not be any false alarms. Figure 1 shows that this is not quite the case.

Because the counting of pulses and the coding of the numbers of each group is a very special memory strategy, known as *figural grouping* (Bamberger, 1978; Handel, 1992; Povel & Essens, 1985), we decided to prevent the use of such a strategy in the next experiment by employing only pattern modifications of the B type. In that case, the click group structure would always remain unchanged, rendering a strategy based on figural grouping rather useless.

EXPERIMENT 2

In Experiment 2, discrimination performance was investigated for patterns of 15 elements that differed from each other in their potential for hierarchical organization. This potential was quantified by the strength of an induced internal clock according to a procedure developed by Povel and Essens (1985). The procedure, which enables computation of clock strength for any pulse pattern in which inter-onset time intervals are integer multiples of some common unit, is illustrated in Table 1. For every pattern (in this example, a 12-element pattern shown in the upper row), accents must be determined first which, according to Povel and Essens, always fall on (1) isolated

Table 1

Input pattern	110101110100			
Accents provided	120202120200			
Possible clocks	unit	E_1	E_0	C
	2	2	4	18
	3	2	0	2
	4	1	2	9

Note—See Experiment 2 for further explanation.

clicks (Positions 4 and 10), (2) the second click of a cluster of two clicks (Position 2), and (3) the initial and final clicks of a cluster consisting of three or more clicks (Positions 6 and 8). Next, all elements of a pattern obtain a numerical strength estimate: accented clicks a value of 2, nonaccented clicks a value of 1, and nonclicks a value of 0 (see second row of Table 1).

The procedure of finding the best clock involves minimizing C , the amount of counterevidence a clock meets in an actual sequence, given by the expression

$$C = 4E_0 + E_1, \quad (2)$$

where E_0 is the number of coincidences between clock ticks and pauses and E_1 is the number of coincidences between clock ticks and unaccented clicks. The lower rows of Table 1 show examples of three generated clocks of two, three, and four units, respectively. Computed counterevidence C has its minimum value ($C = 2$) for the three-unit clock, which is therefore the best of the three.

Some additional restrictions were put on the clock-generation procedure in this study. First, unlike in the procedure adopted by Povel and Essens (1985), who allowed patterns to be repeated many times, every pattern was repeated only once in the present study. Consequently, the distinction made by Povel and Essens between divisor and nondivisor clocks has no meaning in this study. Second, the beginning of a clock unit (a *bar*) in our study always coincided with the beginning of the first click of the pattern; no upbeats were allowed. Third, only clocks of two, three, or four units were generated and compared with each other, since they are the most commonly used clocks in musical practice.

Two sets of 50 stimulus pairs were generated, differing in metric strength according to the parameter C described above. In Condition A, the counterevidence C was zero for all original patterns or their modification—at least one of the patterns had maximum metric strength. In Condition B, the value of C was 9 or more, causing at least one of the patterns in each stimulus to be metrically weak. Just as in Experiment 1, stimuli in a set of random pulse patterns were programmed to be repeated exactly with an a priori probability of .5 or, alternatively, to be followed by a modified version of type B (with an inserted 0 after a randomly chosen 1-0 pair). Both sample sets of stimuli are shown in the Appendix, together with the total number of erroneous responses to each stimulus by 10 subjects.

It is evident that the insertion of an extra pause into a pattern will in most cases diminish its metric strength.

Figure 2. Hit versus false alarm rates for individual subjects in Experiment 2 with targets being type B modifications and with metrically strong (squares) and metrically weak (crosses) pulse sequences. Receiver operating characteristic curves, with d' values of 0.5 and 0.7, computed from Sorkin's model, are also shown.

In this experiment, we therefore investigated whether subjects could use the difference in the metric strength between original and modified patterns as a discrimination cue. The order of presentation of the original pattern (with a controlled metric potential) and its modified version (with an uncontrolled one) in a pair was randomized over the experiment. Subjects were expected to perform better in Condition A than in Condition B. If original pulse patterns have maximum metric strength, a pattern modification should be easier to detect to the extent that it weakens the metric strength. If original patterns are already metrically weak, a pattern change should only be detected if, by chance, insertion of an extra rest would significantly increase its metric strength.

Figure 2 displays the subjects' performance as points on the ROC curve, where squares correspond to Condition A and crosses to Condition B. Thirteen subjects participated, but the data of only 10 are plotted. Three subjects were found to be merely guessing. As in Experiment 1, the d' values for the two conditions were found to be significantly different at the 99% level (paired t test), with subject-averaged values of 2.3 and 0.8, respectively, for Conditions A and B. For cases in which the detection rate was 1.0 or the false alarm rate 0.0, rates were artificially changed to 0.99 and 0.01 to keep d' values finite.

In Condition A, there was a very strong response bias, where mistakes were mostly misses rather than false alarms. Five subjects even made no false alarms at all. A similar response bias was observed by Handel (1992), who used 16-unit patterns. This bias could be interpreted

as evidence for use of the clock sense as a discrimination clue. If the original pattern is metrically strong (Condition A), inducing a strong clock sense, an exact repeat of the pattern will (almost) always be recognized as such because the induced clock will also fit the second pattern. This suppresses or even eliminates false alarms. Misses can be attributed to those occasions when, by chance, a modification by inserting an extra rest in a metrically strong pattern results in a pattern that has about the same metric strength. If original patterns are metrically weak, as is the case in Condition B, misses and false alarms should be more evenly distributed, as indeed appears to be the case.

One could, of course, also argue that the observed response bias does not imply use of an internal clock, but rather reflects a general reluctance of the listeners to respond "different" unless stimuli are clearly different. The patterns were, after all, presented only once, which might have been insufficient to induce a clock sense in all cases of a metrically strong sequence. Such a general bias hypothesis, however, cannot explain why under Condition B little or no response bias was observed.

The finding of Monahan and Hirsh (1990) that discrimination of pulse patterns can be considered as single-gap discrimination is apparently not generalizable to the conditions of Experiment 2. Because modifications consisted of the insertion of an extra rest, a score of about 90% correct would be expected on the basis of the Weber fraction of the silent gap that was changed by insertion of the extra rest, as was already explained in the section on Experiment 1. This holds for both Conditions A and B. The actually obtained scores, expressed as percentages of correct responses, were 81% and 65% for Conditions A and B, respectively. The discrepancy for Condition A could still be accounted for by pointing at the large response bias, which caused the score of 81% correct to be low compared with the sensitivity d' of 2.3. Monahan and Hirsh's model, however, cannot explain the drop to 65% correct ($d' = 0.8$) found for Condition B, since it predicts the same performance levels for Conditions A and B.

A similar argument applies with respect to Sorkin's (1990) model. Computation of sensitivity d' from the statistics of the pattern samples used for Conditions A and B yields predicted values of 0.5 and 0.7, respectively, and corresponding ROC curves are shown in Figure 2 for comparison.¹ Given the actually measured average d' values of 2.3 and 0.8, the prediction appears poor. Again, similar to the Monahan and Hirsh model, it is the significant improvement of performance observed with increase of induced clock strength that both models seem to miss.

CONCLUSIONS

In this study, we have investigated some holistic and analytic discrimination mechanisms for distinguishing different rhythmic pulse patterns. One of the analytic models, that of Monahan and Hirsh (1990), was found to give a poor account of our experimental data, since it predicts

much better performance than was actually obtained. It appears that concentration on only the portion of the temporal pattern where the change may occur, an essential assumption of the model, is only possible if there is no uncertainty about the moment of possible occurrence. If there is uncertainty, as there was in our experiments, the model appears to break down.

Also, Sorkin's (1990) discrimination model, which is based on comparison of a single pattern-correlation statistic with an internal criterion, seems not to work for the conditions we investigated. Model predictions were lower than actually observed performance levels, and the model was especially unable to account for the large differences in performance level observed between different experimental conditions. It appears that Sorkin's model, which is essentially holistic, works only for sequences of not too many events and of limited time span. For sequences of 10 events or more and spanning several seconds, subjects appear to do more than compute single correlation statistics. They seem to be able to isolate, concentrate on, and operate on details of elements or groups of events in the sequence.

One such strategy that we found in Experiment 1 is an analytic way of counting event runs and coding these runs in terms of simple digits. Such "figural grouping" is a form of verbal coding, similar to what Durlach and Braida (1969) called "context coding." It has been found to be very robust against degrading influences of time delays or acoustic interference between the stimuli that are to be compared. Although such counting strategies may not be very representative of typical musical behavior, they were shown to be effective clues in a pattern discrimination task and might even be occasionally useful as memory aids for percussionists performing rhythmically complicated passages.

Finally, a very powerful holistic percept that allows perceptual discrimination between pulse sequences was found to be the metric strength or "clock sense." Significant effects of metric strength on perception and reproduction of pulse sequences have been shown earlier by Povel and Essens (1985) and by Palmer and Krumhansl (1990). In Experiment 2, it was shown that a disturbance of a strong internal clock is detected much better than a disturbance in a weak internal clock, with sensitivities (d') differing by a factor of about three. A strong response bias was

found, however, in the sense that sequences with the same metric strength were much more likely to be perceived as the same than sequences with a different metric strength were likely to be perceived as different.

REFERENCES

- ABEL, S. M. (1972). Discrimination of temporal gaps. *Journal of the Acoustical Society of America*, **52**, 519-524.
- BAMBERGER, J. (1978). Intuitive and formal musical knowing: Parables of cognitive dissonance. In S. S. Madeja (Ed.), *The arts, cognition and basic skills* (pp. 173-209). St. Louis: CEMREL.
- DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.
- HANDEL, S. (1984). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.
- HANDEL, S. (1992). The differentiation of rhythmic structure. *Perception & Psychophysics*, **52**, 497-507.
- MARTIN, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, **79**, 487-509.
- MONAHAN, C. B., & HIRSH, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, **47**, 227-242.
- PALMER, C., & KRUMHANS, L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 728-741.
- POVEL, D.-J., & ESSENS, P. (1985). Perception of temporal patterns. *Music Perception*, **2**, 411-440.
- POVEL, D.-J., & OKKERMAN, H. (1981). Accents in equitone sequences. *Perception & Psychophysics*, **30**, 565-572.
- PREUSSER, D., GARNER, W. R., & GOTTFELD, R. L. (1970). Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *American Journal of Psychology*, **83**, 151-170.
- ROYER, F. L., & GARNER, W. R. (1966). Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception & Psychophysics*, **1**, 41-47.
- SORKIN, R. D. (1990). Perception of temporal patterns defined by tonal sequences. *Journal of the Acoustical Society of America*, **87**, 1695-1701.
- STURGES, P. T., & MARTIN, J. G. (1974). Rhythmic structure in auditory temporal pattern perception and immediate memory. *Journal of Experimental Psychology*, **102**, 377-383.

NOTE

1. The d' values computed with Sorkin's model from the two stimulus sets used for Conditions A and B are rather typical, as was confirmed by subsequent computations on other similarly generated click sequences. The slightly larger d' value for metrically weak click sequences is mostly due to a consistently larger value of the standard deviation of interclick gaps.

(Continued on next page)

APPENDIX Stimulus Sets for Experiment 2: Sequences and Responses

Condition A, Strong Clock																	Condition B, Weak Clock																			
1	1	0	1	1	0	0	1	0	0	1	0	1	1	0	0	D	5	1	0	<u>1</u>	0	1	0	0	0	1	1	1	1	1	0	0	0	D	1	
2	0	0	1	0	1	1	<u>1</u>	0	0	1	0	1	0	1	0	D	7	2	0	0	1	1	1	0	0	1	0	1	1	1	0	0	0	S	2	
3	0	0	0	0	0	1	1	1	<u>1</u>	0	1	0	1	1	0	D	8	3	1	0	1	1	1	1	0	0	0	0	0	0	0	1	0	S	5	
4	1	0	1	0	<u>1</u>	0	0	1	0	1	1	1	0	0	0	D	5	4	1	0	1	1	1	<u>1</u>	0	0	0	1	0	0	0	1	0	D	4	
5	0	0	0	0	0	1	1	1	1	1	0	1	0	1	0	0	S	2	5	1	1	0	0	0	0	1	1	0	1	<u>1</u>	0	0	1	0	D	8
6	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	S	0	6	0	1	1	0	1	1	0	0	0	0	1	1	1	0	0	S	3	
7	0	0	0	1	1	1	1	0	1	<u>1</u>	0	1	0	0	0	D	3	7	<u>1</u>	0	0	0	0	1	1	1	1	1	1	0	0	0	0	D	6	
8	1	1	1	1	1	0	0	0	1	0	0	0	1	0	0	S	0	8	<u>1</u>	0	1	0	0	0	0	1	1	1	1	0	0	0	0	D	1	
9	1	0	1	1	1	0	0	0	1	0	0	1	1	0	0	S	1	9	0	1	1	1	1	1	<u>1</u>	0	0	0	1	0	0	0	0	D	4	
10	0	1	0	1	0	1	0	0	0	1	1	1	0	1	0	S	2	10	1	1	0	0	1	0	1	0	0	1	<u>1</u>	0	0	1	0	D	4	
11	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	S	1	11	1	0	0	0	0	1	1	0	0	0	1	1	1	1	0	S	5	
12	<u>1</u>	0	0	0	1	0	1	1	1	0	1	0	1	0	0	D	1	12	1	1	0	0	0	1	0	1	1	<u>1</u>	0	0	0	1	0	D	4	
13	<u>1</u>	0	0	1	0	1	1	0	1	1	0	0	1	0	0	D	0	13	1	1	<u>1</u>	0	0	0	0	0	0	0	0	1	1	1	1	0	D	2
14	0	1	0	1	0	1	0	1	1	1	0	0	0	0	1	S	0	14	1	1	0	0	0	1	0	1	0	1	1	1	0	0	0	S	3	
15	0	1	0	0	0	1	0	1	1	1	0	0	1	0	1	S	1	15	1	0	0	1	1	1	0	1	0	0	0	<u>1</u>	0	1	0	D	0	
16	0	0	0	1	0	1	1	<u>1</u>	0	0	1	1	0	1	0	D	1	16	1	0	1	<u>1</u>	0	0	0	0	0	0	0	1	1	1	1	0	D	2
17	0	0	0	1	1	<u>1</u>	0	1	0	1	1	1	0	0	0	D	2	17	0	1	1	1	0	0	1	1	0	0	1	1	0	0	0	S	1	
18	0	1	0	0	1	0	1	1	0	0	1	0	1	1	0	S	1	18	0	1	0	0	1	1	1	1	1	0	0	1	0	0	0	S	3	
19	0	0	0	0	0	1	1	1	1	1	0	0	1	1	0	S	2	19	1	0	0	0	0	1	1	<u>1</u>	0	1	1	0	0	1	0	D	3	
20	0	0	0	1	0	0	1	1	1	1	<u>1</u>	0	1	0	1	D	3	20	0	1	1	0	0	1	1	0	0	0	1	1	1	0	0	S	2	
21	1	0	0	1	<u>1</u>	0	1	1	1	0	0	0	0	1	0	D	0	21	1	<u>1</u>	0	0	0	0	1	0	1	0	1	1	0	0	1	0	D	1
22	1	0	0	1	1	0	1	0	1	0	1	0	1	0	0	S	1	22	1	0	1	1	0	1	0	<u>1</u>	0	1	0	0	1	0	0	1	D	1
23	0	0	<u>1</u>	0	0	1	0	0	0	1	1	1	1	1	0	D	6	23	0	0	1	1	0	1	1	1	1	0	0	0	0	0	1	0	S	6
24	0	0	0	1	1	1	1	0	0	1	0	1	1	0	0	S	3	24	1	0	1	1	0	1	0	1	0	0	0	1	0	1	0	S	4	
25	1	0	1	1	1	0	0	0	1	0	1	0	1	0	0	S	2	25	1	1	0	1	0	0	0	1	1	1	0	0	0	0	1	0	S	4
26	0	0	0	0	1	1	1	1	0	1	<u>1</u>	0	1	0	0	D	3	26	0	1	1	0	1	1	0	1	0	0	0	1	1	0	0	S	2	
27	1	0	0	<u>1</u>	0	1	1	0	1	1	0	0	1	0	0	D	1	27	1	0	0	1	0	0	0	0	1	1	1	1	0	0	1	0	S	3
28	1	0	0	0	<u>1</u>	0	0	1	1	0	1	1	0	1	0	D	3	28	1	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	S	3
29	0	0	0	0	1	0	1	0	1	0	1	1	1	1	0	S	1	29	1	0	<u>1</u>	0	1	1	0	0	0	1	1	0	0	1	0	D	1	
30	0	1	0	0	0	1	0	1	1	1	0	0	1	1	0	S	0	30	0	1	1	0	1	0	1	0	0	1	0	1	1	0	0	S	5	
31	1	0	0	1	1	0	0	1	1	0	<u>1</u>	0	1	0	0	D	0	31	0	1	1	0	0	1	1	<u>1</u>	0	0	1	1	0	0	0	D	9	
32	1	0	1	1	0	0	<u>1</u>	0	0	1	0	1	1	0	0	D	0	32	1	0	1	0	0	1	1	1	1	1	0	0	0	0	0	S	2	
33	0	0	0	0	1	1	1	1	0	1	<u>1</u>	0	0	1	0	D	5	33	1	1	0	0	0	1	1	1	0	0	0	1	1	0	0	S	1	
34	0	1	0	0	1	<u>1</u>	0	1	1	0	1	0	1	0	0	D	5	34	1	1	0	1	0	1	0	1	0	0	1	1	0	0	0	S	3	
35	1	0	1	<u>1</u>	0	1	0	0	1	0	0	1	0	0	1	D	2	35	1	0	0	0	0	0	1	1	0	1	1	0	1	0	1	S	8	
36	0	0	0	1	0	1	1	1	0	0	1	1	0	1	0	S	2	36	1	1	<u>1</u>	0	0	1	0	0	0	0	0	1	0	1	1	0	D	0
37	0	0	1	1	1	0	<u>1</u>	0	1	1	0	0	0	1	0	D	2	37	1	1	0	0	0	1	1	<u>1</u>	0	1	0	0	0	1	0	D	4	
38	0	1	1	<u>1</u>	0	1	0	1	0	1	0	0	0	1	0	D	2	38	1	0	0	1	0	0	0	1	1	1	<u>1</u>	0	0	1	0	D	5	
39	1	0	0	0	1	0	1	0	1	0	1	1	1	0	0	S	1	39	1	1	0	1	1	1	1	0	0	0	1	0	0	0	0	S	4	
40	0	1	0	0	1	1	1	<u>1</u>	0	1	0	0	1	0	0	D	2	40	1	1	1	1	0	0	0	0	0	1	1	0	0	0	1	0	S	6
41	1	0	0	1	1	0	0	<u>1</u>	1	0	0	1	1	0	0	D	1	41	0	1	<u>1</u>	0	1	0	1	0	0	1	1	1	0	0	0	D	2	
42	1	1	<u>1</u>	0	1	0	0	0	1	0	1	0	1	0	0	D	1	42	1	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	S	6
43	0	0	0	1	0	<u>1</u>	0	1	1	1	1	1	1	0	0	D	0	43	0	1	1	1	0	1	1	0	1	0	0	0	1	0	0	S	3	
44	0	1	0	0	1	0	1	1	0	1	<u>1</u>	0	1	0	0	D	4	44	1	1	1	1	0	0	0	1	0	1	1	0	0	0	0	S	3	
45	0	0	0	1	0	0	1	<u>1</u>	0	0	1	1	1	1	0	D	0	45	1	0	1	0	0	1	0	1	0	0	1	0	1	1	0	S	4	
46	0	1	0	<u>1</u>	0	1	0	0	0	1	0	1	1	1	0	D	0	46	1	1	0	0	0	1	0	1	0	0	1	1	1	0	0	S	6	
47	0	0	0	1	0	0	0	1	1	1	1	1	1	0	1	S	2	47	1	1	<u>1</u>	0	0	1	0	0	0	1	0	0	1	1	0	D	5	
48	0	0	0	1	0	1	<u>1</u>	0	1	0	0	1	1	1	0	D	1	48	1	1	0	1	0	1	0	1	1	1	0	0	0	0	0	S	4	
49	0	0	0	1	0	1	0	<u>1</u>	0	1	1	1	1	0	0	S	0	49	1	1	<u>1</u>	0	0	1	1	0	0	0	0	0	1	1	0	D	3	
50	0	1	0	1	0	1	1	1	0	1	0	0	0	1	0	S	2	50	1	0	1	1	1	1	0	1	0	1	0	0	0	0	0	S	3	

Note—If the column following the sequence indicates S (*same*), the sequence is repeated exactly during the second half of the trial. If D (*different*), an extra zero is inserted right after the underlined numeral one. In the last column, the numbers of errors (out of 10 subject responses) are given.