

# Modality differences in short-term memory for rhythms

GEOFFREY L. COLLIER

*South Carolina State University, Orangeburg, South Carolina*

and

GORDON LOGAN

*University of Illinois, Urbana-Champaign, Illinois*

Prior research has established that performance in short-term memory tasks using auditory rhythmic stimuli is frequently superior to that in tasks using visual stimuli. In five experiments, the reasons for this were explored further. In a same–different task, pairs of brief rhythms were presented in which each rhythm was visual or auditory, resulting in two same-modality conditions and two cross-modality conditions. Three different rates of presentation were used. The results supported the temporal advantage of the auditory modality in short-term memory, which was quite robust at the quickest presentation rates. This advantage tended to decay as the presentation rate was slowed down, consistent with the view that, with time, the temporal patterns were being recoded into a more generic form.

The source of the well-established advantage of auditory over visual stimuli in short-term memory has long been controversial. This paper provides new data, following up on the theory of Glenberg and associates. They posited that auditory stimuli have better temporal encoding and that time of occurrence provides a good retrieval cue (Glenberg, Eberhardt, & Belden, 1987; Glenberg & Fernandez, 1988; Glenberg & Swanson, 1986). Therefore, the advantage that auditory material has is really due to superior temporal coding, rather than to a special auditory short-term store.

Visual and auditory rhythms provided Glenberg and associates with an attractive set of stimuli to test this theory, inasmuch as they are devoid of linguistic content. Glenberg, Mann, Altman, Forman, and Procise (1989) performed a series of experiments in which the stimuli were created from visual flashes or auditory beeps consisting of patterns of long and short durations. In most of the experiments, subjects had to recall ordinal information from each pattern, by tapping sequences on two keys, one indicating *long* and the other *short*. The results showed an advantage for auditory rhythms that was not due merely to stimulus clarity or salience, to the alerting advantage of audition, or to people's greater experience with auditory stimuli. Rather, audition enjoyed an advantage in working memory in all serial positions.

In follow-up experiments, Glenberg and Jona (1991) were able to eliminate the auditory advantage by manipulating rhythmic structure. They showed that the auditory advantage occurred only with rhythms in which the components bore a simple, integral relationship with each other (as in musically notated rhythms), but not when the rhythmic structure was complex. In a second experiment, the auditory advantage disappeared when the rhythms were slowed down sufficiently that they no longer cohered as a musical gestalt.

Dissenting voices have argued that linguistic and rhythmic modality effects could have different causes. Consistent with this, Crowder and Greene (1987) and Schab and Crowder (1989) performed experiments in order to demonstrate that there really are no temporal differences between the modalities when linguistic materials are employed. More germane to rhythms are the data of Watkins, LeCompte, and Fish (1992), whose experiments showed an auditory temporal advantage only when concurrent silent mouthing of the word *blah* suppressed subvocal recoding of visual input stimuli and only when the stimuli consisted of the same item repeated. It did not matter whether the material was linguistic or any of a variety of nonlinguistic stimuli. These differences between the rhythmic and linguistic modality effects led Watkins et al. to conclude that the two do not stem from the same cognitive sources, which, in turn, raises questions about the pertinence of Glenberg et al.'s rhythm data as support for a temporal theory of an auditory short-term memory advantage.

## THE PRESENT EXPERIMENTS

In the present research, we used a new experimental paradigm to replicate and further understand the advantage of auditory rhythms over visual ones. Two rhythms

---

This work was performed while the first author was a guest in the laboratory of the second author. The first author acknowledges support for this research from ADAMA Training Grant MH 14257, "Quantitative Methods for Behavior Research," Department of Psychology, University of Illinois. Correspondence concerning this article should be addressed to G. L. Collier, Box 7094, Department of Psychology, South Carolina State University, Orangeburg, SC 29117 (e-mail: collier@scsu.scsu.edu).

were presented sequentially, separated by a brief inter-stimulus interval (ISI). On half the trials, the rhythms were identical, and on half they were not, and the subject had to select *same* or *different*. All four permutations of the modality of the first and second rhythm were used: auditory–auditory (AA), visual–auditory (VA), AV, and VV. In contrast to prior procedures, this did not yield individual responses to individual stimulus elements (and thus no serial position curves) but did enable study of cross-modal same–different comparisons. The other innovation was the inclusion of faster tempos than had been used in prior research.

**PREDICTIONS**

We made three predictions. First, we predicted that, replicating prior research, auditory rhythms would have a general advantage over visual ones.

Second, we predicted that within the two mixed modality conditions, performance on condition AV would be better than that on condition VA. This follows from an *on-line comparison* model, in which the first stimulus is stored and replayed in tandem with the second, the subject responding *different* if, at any point, the two stimuli mismatch. Because only the first stimulus needs to be stored, it is more critical, so that putting the weaker visual stimulus in the initial position would be more harmful than putting it in the less critical, terminal position.

Third, we used the mixed modality conditions to contrast two models for the source of the auditory advantage. According to the *proprietary code* hypothesis, each modality is represented in its own code. Comparing visual and auditory stimuli would cause some sort of code-crossing overhead, and thus the mixed modality conditions would actually display inferior performance to that of the purely visual condition (VV). In contrast, the *single-code* hypothesis proposes that comparisons are made on the basis of a single code. This could occur if visual rhythms are “hummed in one’s head” (i.e., converted to an auditory code) so rapidly and automatically as to not tax the system or if all the stimuli are immediately converted to a common amodal temporal code. In either case, there would be no code-crossing disadvantage, and the mixed modality conditions would display performance somewhere between those for the strictly auditory and the strictly visual conditions.

**GENERAL METHOD**

The general design described here was shared by all five experiments, which differed only in the stimulus sets used.

**Subjects**

From 21 to 30 subjects per experiment (Table 1) were obtained from the University of Illinois introductory psychology subject pool. Each initially filled out a questionnaire on musical background.

**Stimuli**

Rhythmic patterns were displayed as empty ISIs delimited by brief beeps or flashes. Ratio patterns, such as 1–1–3–1–1–2, were

**Table 1**  
Summary of Stimuli Used in all Experiments

Experiment	No. of Elements	Unique Ratios	Minimum Duration	Mean Duration	Maximum Duration	Mean Trial Duration	Interstimulus Interval	Typical Different Rhythm
1	7	1, 2, 3, 4	1,000	1,203	1,500	3,406	1,000	3 1 2 1 1 2 1 2 1 1 2 1 1 2 1 1 3 1 1 2 1 1 2 1 1 2
2	6	1, 2, 3	1,000	1,125	1,250	3,250	500	1 1 3 1 1 2 1 1 2 1 1 2
3	7	1, 1.33, 1.6, 2.1, 2.25, 2.5, 2.63	1,443	1,644	1,848	4,288	1,000	1.33 1.6 1.6 1.6 1.6 1.6 1.6 1.33 1.6 2.1 1.6 2.25 2.1 1.33 1.43 1.6 1.6 2.25 2.1 1.33 1.43 2.1 1.6 2.25
4 many	6	1, 1.33, 1.43, 1.5, 1.6, 1.75, 2.1, 2.25, 2.5, 2.75	1,191	1,266	1,450	2,532	1,000	1.75 1 1.75 1.75 1.75 1 1.75 1 1.75 1.75 1.75 1
4 few	8	1, 1.43, 1.5, 1.75, 2.33, 2.5, and 2.75	1,036	1,234	1,531	3,468	1,000	1.75 1 1.75 1.75 1.75 1 1.75 1 1.75 1.75 1.75 1
5	Same as Experiment 1		1,000	1,203	1,500	3,000	1,000, 2,500, 7,250	Same as Experiment 1

Note—Column 2 shows the number of intervals in each rhythm (i.e., for Experiment 1, each rhythm had seven elements, delimited by eight flashes or beeps. Column 3 shows the unique ratio components used, where 1 = 125, 250, or 500 msec, depending on the tempo condition. Column 4 shows the shortest single rhythm at the fastest rate; these times should be doubled and quadrupled for the two slower conditions. Column 5 shows the average single rhythm at the fastest rate. Column 6 shows the longest single rhythm at the fastest rate. Column 7 shows the mean trial duration at the fastest rate, equal to twice the average rhythm duration plus the interstimulus interval. Column 9 shows a characteristic rhythm pattern, where the actual rhythms were created by multiplying these ratios by 125, 250, or 500 msec.

created and multiplied by the three base rates (125, 250, and 500 msec) to generate the actual patterns, such as 125–125–375–125–125–250 msec.

The patterns created for the five experiments all had between six and eight intervals, delimited by seven to nine pulses. Across all experiments, the rhythms ranged in duration from 1,000 to 1,848 msec at the fastest base rates, for a total trial duration of twice that plus an ISI. The slower base rates doubled and quadrupled these durations (see Table 1 for details).

Each experiment had 12 different rhythms. A trial consisted of a rhythm, an ISI, and then a second rhythm. For the *same* trials, 1 of the 12 patterns was repeated twice. For the *different* trials, the 2nd pattern was altered by changing one or more elements, the stimuli typically differing by two or three elements. Both rhythms were always at the same base rate; base rate only varied between trials. All the patterns for a given experiment had the same number of elements, and total stimulus durations were similar for both stimuli in each pair.

An example of a representative *different* trial in Experiment 1 is

Pattern 1:	3	1	2	1	1	2	1
Pattern 2:	2	1	1	2	1	1	2

The three stimulus rates and four modality conditions were applied to the 12 *same* and 12 *different* stimulus pairs, for a total of 288 trials in each experiment.

### Design

The experimental factors of base rate (125, 250, and 500 msec) and modality (VV, AA, VA, and AV) were within subjects, with the presentation order of the trials completely randomized separately for each subject.

### Equipment

The experiment was performed on an IBM AT computer, with timing controlled by a clock card. The auditory stimulus was a 1000-Hz. square wave beep. Visual stimuli were presented with a red light-emitting diode (LED) suspended in front of the monitor, embedded in a small black box with a hole cut in the middle, whose purpose was to focus attention on the LED and to mask peripheral reflections of the LED. The beeps and flashes were 3 msec for the first two experiments and 4 msec for all of the other experiments.

### Procedure

The subjects filled out a brief questionnaire on musical background, received instructions and 5–10 practice trials, and then were left alone for the 288 experimental trials.

On each trial, the subjects saw the trial number in the upper left-hand corner of the screen and the word *visual* or *auditory*, indicating the modality of the first stimulus, on the left of the screen (the presentation was similar for the second stimulus, on the right). After a pause, the word *first* was presented on the screen, and the first stimulus was heard or seen. If the stimulus was visual, a colored square was presented on screen, surrounding the cardboard box in which the LED was suspended, in order to focus attention on the stimulus.

After the ISI, the word *first* disappeared from the left-hand side of the screen, the word *second* was displayed on the right hand side of the screen, and the second stimulus started. If the first stimulus was visual, the square surrounding the LED disappeared. If the second stimulus was visual, it was replaced with a square of a different color.

When the second stimulus was finished, if it had been visual the surrounding square disappeared, and a notice to respond was displayed, indicating which key was *same* and which was *different*. The two keys were the “z” and the “/” key, consistently mapped to *same* or *different* for each subject, but randomized between subjects. After each response, the next trial was immediately initiated. After every 72 trials, the subjects were allowed to take a break.

## EXPERIMENT 1

### Effects of Base Rate

The first experiment was designed to study the effects of modality and base rate, as was discussed above.

### Method

**Subjects.** Twenty-one subjects were obtained from the University of Illinois subject pool.

**Stimuli.** All of the rhythm patterns consisted of seven intervals, created out of four basic ratio components (1, 2, 3, and 4), although not all rhythms had each of these ratios. Stimulus details are given in Table 1. Because of these simple ratios, all the rhythms had a musical flavor to them.

### Results

The results are displayed in Figure 1. The abscissa represents the base rate, whereas the four modality conditions provide the graph parameter.

The results were analyzed according to the signal detection model, with *different* trials treated as signal trials and *same* trials as noise trials. There were a great many 0 cells, however, so that rather than use  $d'$  as the dependent variable, the following measure, henceforth referred to as *accuracy*, was used:

$$\frac{\sin^{-1} \sqrt{p(\text{hits})} - \sin^{-1} \sqrt{p(\text{FA})}}{\frac{\pi}{2}}$$

This formula is like percentage correct, in that it ranges from 0 to 1 (for positive  $d'$ s) and has no problems with 0 cells, but it is less sensitive than percentage correct to changes in bias under the signal detection model (Table 2).

The experimental design consisted of the three levels of base rate and the four levels of modality, but the  $F$  tests implied that this design was too global to capture the most interesting features of the data, so tests with fewer degrees of freedom were used to answer the questions of substantive interest, using .05 as the critical  $p$  value.

First, the AA and VV conditions were compared with a 2 (modality)  $\times$  3 (base rates) analysis of variance (ANOVA). The AA condition was better than VV, replicating the expected temporal advantage of auditory stimuli [ $F(1,16) = 80.7, p < .001$ ]. There was an effect of base rate [ $F(2,32) = 3.9, p < .05$ ] because of a general decline in accuracy as the tempo slowed down, but this was mainly due to a decrement in the AA condition toward the VV condition; the interaction of base rate and modality was significant [ $F(2,32) = 3.4, p < .05$ ].

The two mixed modality conditions were extracted and subjected to a 2 (modality)  $\times$  3 (base rates) ANOVA. As was predicted, the AV condition was more accurate than VA [ $F(1,16) = 6.6, p < .05$ ]. In contrast to the single-modality conditions, there was an improvement as the tempo slowed down [ $F(2,32) = 3.9, p < .05$ ], but the interaction of modality and base rate was not significant [ $F(2,32) = 2.2, p = .13$ ].

Finally, we turn to the main test, comparing the cross-code and the single-code hypotheses. This hypothesis was

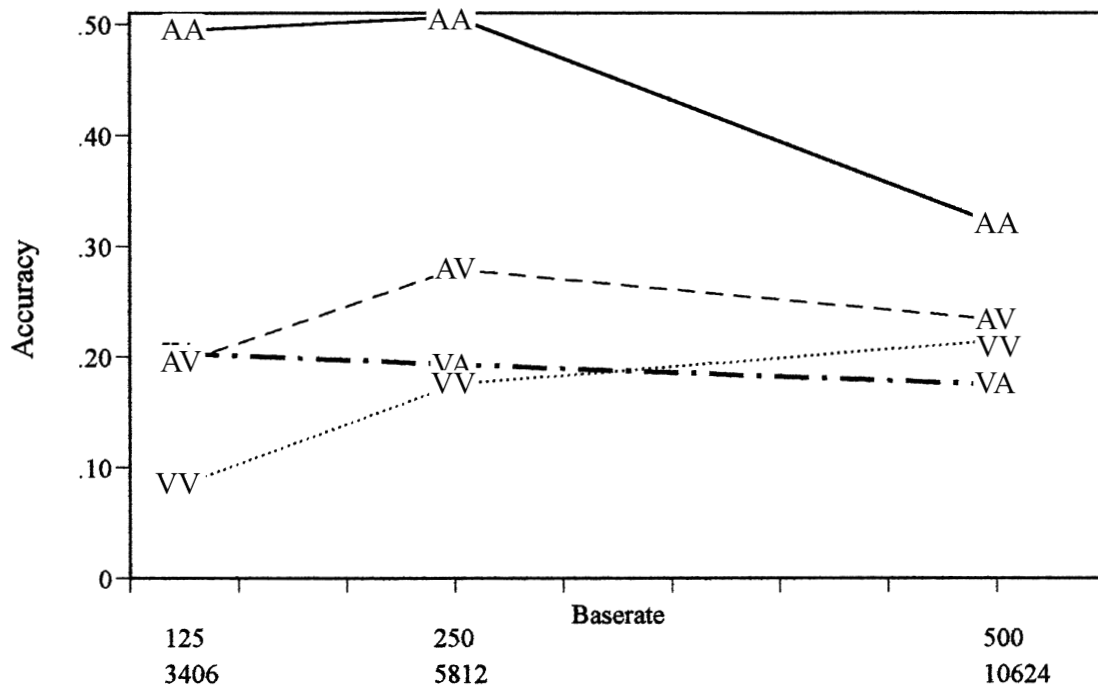


Figure 1. The results of Experiment 1. See the text for a description of the independent variable. The abscissa is labeled with the durations in milliseconds of the base rate (shortest interstimulus interval) and, below that, the duration of an average trial, including both rhythms to be compared and the interval between them.

tested separately for each base rate by parameterizing the four means at each rate as a  $2 \times 2$  ANOVA, with modality of the first and the second stimuli as the two factors. We expected a main effect of both positions, so that auditory stimuli always were better in either position. Of primary interest was the interaction. If comparison of stimuli in different modalities invoked a cross-code decrement, there would be an interaction such that the mixed modality conditions were performed worse than the single-modality conditions. Conversely, if there was no interaction, this would yield evidence favoring the single-code hypothesis.

The results of these tests differed by base rate. At the 125-msec base rate, there was a significant interaction, in-

dicating that the mixed modality conditions led to worse performance than would be expected on the basis of the components [ $F(1,16) = 38.4, p < .001$ ]. At the base rate of 250 msec, the effect began to weaken [ $F(1,16) = 4.6, p < .05$ ], and it disappeared at 500 msec [ $F(1,16) = 0.4, p = .56$ ]. In general, there appeared to be a convergence of the different modality conditions as the tempo slowed down. One is led to hypothesize that further slowing down of the base rate would have led to the ultimate convergence of the modality conditions, with performance declining to near chance levels.

**Discussion**

The results can be integrated into a coherent picture of the processing of rhythmic stimuli in working memory. Auditory rhythms are well processed at fast presentation rates, but this advantage decays with time in pretty much the same way that it does for nonrhythmic stimuli. Much of the auditory advantage seems to have been lost by the time the total trial length reaches about 10 sec.

At fast base rates, the mixed modality conditions are more difficult than the purely visual condition, but as presentation rate slows down, performance actually improves and lies between those for the purely auditory and the purely visual conditions. Thus, there appears to be a code-crossing overhead for the fast stimuli, but not for the slow ones. This is consistent with the hypothesis that as tempo slows down, the proprietary auditory code has more time to decay, and subjects begin to rely more on some kind of

**Table 2**  
**A Comparison of the Percentage of Hits Minus the Percentage of False Alarms (a Transform of Percentage Correct) With the Dependent Measure Used in This Research, Under the Signal Detection Model**

$d'$	Range: $p(\text{hits}) - p(\text{fa})$	Range: Arcsine Square Root Transform
0.5	.19	.10
1.5	.48	.23
2.5	.63	.33
3.5	.76	.48

Note—For each  $d'$ , percentage of hits and percentage of false alarms were calculated for a wide range of response biases. As the table indicates, our dependent measure had a smaller range than  $p(\text{hits}) - p(\text{false alarms})$ , indicating less sensitivity to response bias.

common code. This code might be generic (i.e., amodal), or it might be due to a conversion of everything into the auditory modality (i.e., humming in one's head).

Finally, it should be noted that the general advantage for the auditory stimuli could be explained by assuming that the particular beeps used were clearer than the visual flashes, but other facts, such as the difficulty with the mixed modality rhythms, as well as the experiments of Glenberg et. al. (1989), would not be so readily explained by such a hypothesis.

### EXPERIMENT 2 Chunking

In Experiments 2–4, we attempted to weaken or destroy the auditory advantage by manipulating the structure of the rhythms. The hypothesis being tested was that audition's advantage was due to its superiority at musical encoding, which we presumed to involve some sort of postcategorical grouping scheme (see, e.g., Essens & Povel, 1985; Martin, 1972), availing itself of a presumed auditory superiority with temporal order (Glenberg et al., 1987; Glenberg & Fernandez, 1988; Glenberg & Swanson, 1986). If audition's advantage arises from a postcategorical musical-encoding scheme, our manipulations should weaken this advantage, whereas this would not be the case if audition's advantage relies on a precategorical echoic memory.

In Experiment 2, each member of the *different* pair had the same number of rhythmic groups or chunks, as well as the same number of elements per group. An example of how this was done is illustrated in the stimulus pair (1112)(113) and (1113)(112). This made the two stimuli

making up a pair impossible to differentiate by counting the number of elements, groups, or elements per group.

#### Method

The experimental design was the same as that in Experiment 1, except for the substitution of the new set of stimuli. Each rhythm consisted of six intervals. Each rhythm was created out of the ratios 1, 2, and 3, with four of the pulses being 1s, and the remaining two being a 2 or a 3. The *different* trials were created by taking one of the rhythms and changing one or both of the 2/3s to a 3/2 in one of the rhythms, thus ensuring that both stimuli had the same number of chunks and elements per chunk.

#### Results

Results are graphed in Figure 2. Since the *different* pairs were designed to be more difficult to discriminate, it is not surprising to see a decline in overall performance. However, contrary to the hypothesis that giving the stimuli the same groupings would eliminate the auditory advantage, the purely auditory condition (AA) still had a robust advantage over VV [ $F(1,20) = 45.8, p < .001$ ]. Qualitatively, there appear to be similarities with the first experiment: Condition AV appears to be better than VA, and the mixed modality conditions appear to improve with slower base rates. However, none of these effects was significant. The difference between AV and VA was not significant [ $F(1,20) = 3.0, p = .10$ ]. The simple main effects of base rate were not significant for any condition except AA, and a test for the effect of modality that included the conditions VV, AV, and VA was not significant [ $F = 2.65, p = .08$ ]. To an extent, floor effects were at play here;  $t$  tests revealed that conditions VA125, VA250, VA500, and AV500 did not differ significantly from the chance level of 0.

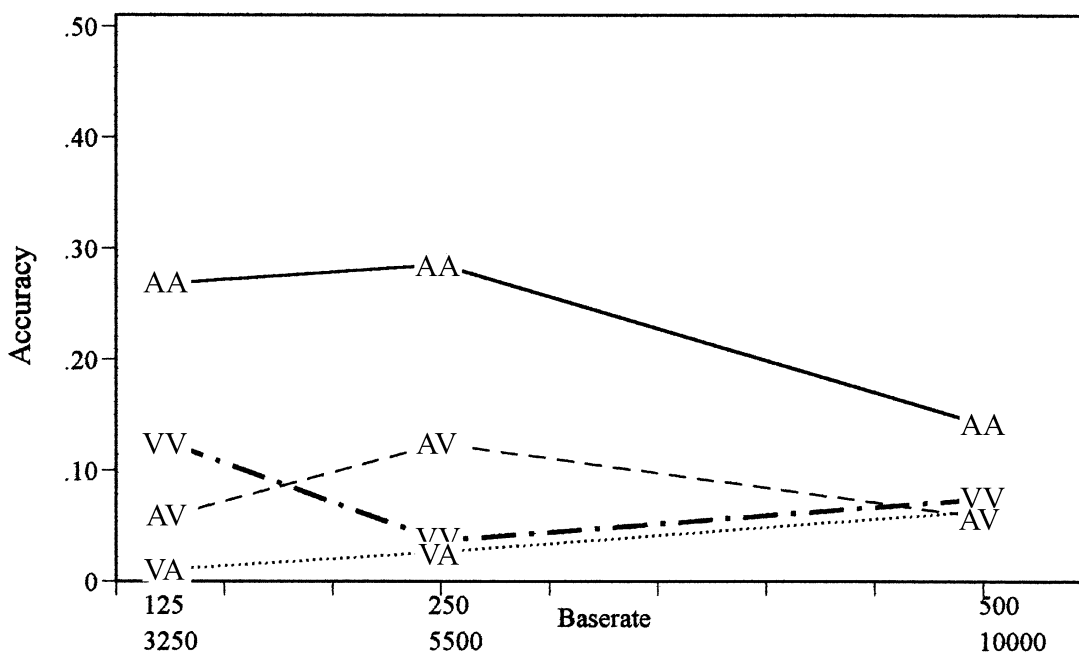


Figure 2. The results of Experiment 2.

However, there were significant interactions between the first and the second stimulus at 125 msec [ $F(1,20) = 43.5, p < .001$ ] and at 250 msec [ $F(1,20) = 7.1, p < .05$ ], but not at 500 msec [ $F(1,20) = 2.3, p = .14$ ]. This replicates the result in the first experiment, showing the difficulty of cross-modality comparisons at fast, but not at slow, presentation rates.

There also were simple main effects of modality at 125 msec [ $F(3,60) = 17.9, p < .001$ ], and at 250 msec [ $F(3,60) = 15.4, p < .001$ ], but not at 500 msec [ $F(3,60) = 1.6, p = .19$ ]. So here again, we see a tendency of convergence, owing to the worsening performance of the AA condition as the base rate slows down.

### EXPERIMENT 3 Complex Ratios I

Our failure to eliminate the auditory advantage in Experiment 2 may have occurred merely because the manipulation was musically naive in assuming that musical grouping relies on temporal proximity just as visual grouping relies on spatial proximity. In contrast, a binary hierarchical representation (e.g., Essens & Povel, 1985; Martin, 1972) would group things differently. Consider the pattern 1 2 1 1 2. The chunking strategy would use the longer intervals as chunk boundaries, thus parsing the pattern as (1 2)(1 1 1 2). In contrast, a musical strategy would parse it into equal length groups of four beats—that is (1 2 1)(1 1 2).

In Experiment 3, we attempted to more radically undermine the possibility of musical encoding by using complex

ratios among the stimulus elements, which are known to be difficult to process musically (Collier & Wright, 1995; Essens & Povel, 1985). Using this strategy, Glenberg and Jona (1991) were able to reduce the auditory advantage.

#### Method

The experimental design was the same as those in the prior experiments, except that the patterns were created out of the following set of ratios: 1, 1.33, 1.6, 2.1, 2.25, 2.5, 2.6, and 3. Each pattern contained between four and seven distinct durations, typically containing five distinct durations. Four of the 12 different pairs differed by one element, 4 by two elements, and 4 by three elements.

#### Results

The results are seen in Figure 3. The auditory advantage was destroyed at 250 msec, but not at 125 msec. The simple main effect of modality was significant at 125 msec [ $F(3,72) = 21.1, p < .001$ ], but not at 250 msec [ $F(3,72) = 1.1, p = .35$ ]. The effect of modality at 500 msec again became significant [ $F(3,72) = 3.4, p < .05$ ], apparently because the advantage of the AA condition was not decaying toward chance as rapidly as the remaining three conditions.

At 500 msec,  $t$  tests indicated that there was no significant difference from chance for condition VV [ $t(24) = 0.92, p = .37$ ], condition VA [ $t(24) = 0.45, p = .66$ ], and condition AV [ $t(24) = 0.22, p = .83$ ] but that condition AA was significant [ $t(24) = 3.61, p < .001$ ]. (All the other conditions differed significantly from chance as well.)

An ANOVA that included all the conditions but AA showed an effect of base rate [ $F(2,48) = 8.88, p = .001$ ], but not of modality [ $F(2,48) = 2.10, p = .13$ ].

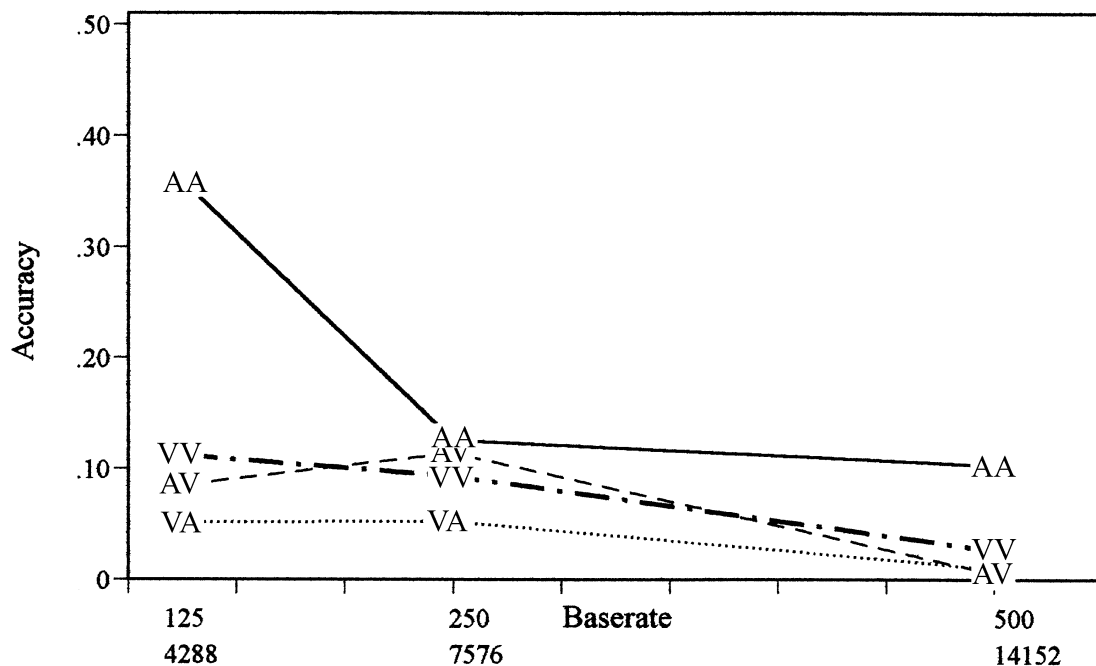


Figure 3. The results of Experiment 3.

**EXPERIMENT 4  
Complex Ratios II**

Experiment 4 compared two alternative explanations for Experiment 3's failure to eliminate the auditory advantage, in contrast to Glenberg and Jona (1991). Noting that the middle base rate in Experiment 3 showed no auditory advantage and that the use of complex ratios resulted in stimuli that averaged 1,038 msec longer than those in the first experiment, it seemed possible that, perhaps, the middle base rate was farther out on the decay curve than in Experiment 1. Alternatively, it could be that, at the fastest rate, the subjects were able to categorically encode the durations into a set of long-short durations and thus apply musical-encoding strategies. The first explanation is consistent with the view that the auditory rhythmic advantage is precategorical, whereas the latter is consistent with the view that it is postcategorical.

In order to address the first hypothesis, Experiment 4 used complex ratios but shortened the stimulus durations. In order to address the second hypothesis, we constructed a set of stimuli, each with only two different durations, and another set with many different durations, as in Experiment 3. We hypothesized that the former would be easier to categorically encode than the latter and, thus, should show better performance if the auditory advantage is due to categorical encoding.

**Method**

For the *many types* subset of stimuli, there were six intervals, concatenated out of the ratios 1, 1.33, 1.43, 1.5, 1.6, 1.75, 2.1, 2.25, 2.5, and 2.75. No ratio was used more than twice in each rhythm, and the rhythms had no more

than two pairs of duplicate ratios. For the *different* rhythm pairs, two differed by one duration, two by two durations, and two by three durations.

For the *two-ratios* subset of stimuli, there were eight rather than six intervals, in order to keep the total rhythm length approximately equal. The ratios used were 1, 1.43, 1.5, 1.75, 2.33, 2.5, and 2.75. However, each rhythm was concatenated out of 1s and one other of these ratios; thus, there were only two different durations used for each rhythm. For the *different* trials, both rhythms in each pair used the same two rhythm ratios, differing only in the placement of one duration.

**Results**

The results are depicted in Figures 4 (many different ratios per stimulus) and 5 (only two unique ratios per stimulus). The advantage has reappeared at the 250-msec base rate for both stimulus sets. Conditions AA do not differ between the two stimulus sets at 125 msec [ $t(29) = 1.31, p = .20$ ], although they do differ at 250 msec [ $t(58) = 2.94, p < .01$ ], owing to a mysterious improvement at this rate for the *two distinct ratios* stimulus set. This difference could be due to superior categorical encoding of this condition, but this would not explain the lack of difference at the fastest rate. In sum, the data are most consistent with the simpler explanation, that the auditory advantage requires shorter stimuli.

**EXPERIMENT 5  
Varying the Interstimulus Interval**

In Experiments 1-4, we manipulated trial duration by manipulating presentation rate (base rate), thus effectively

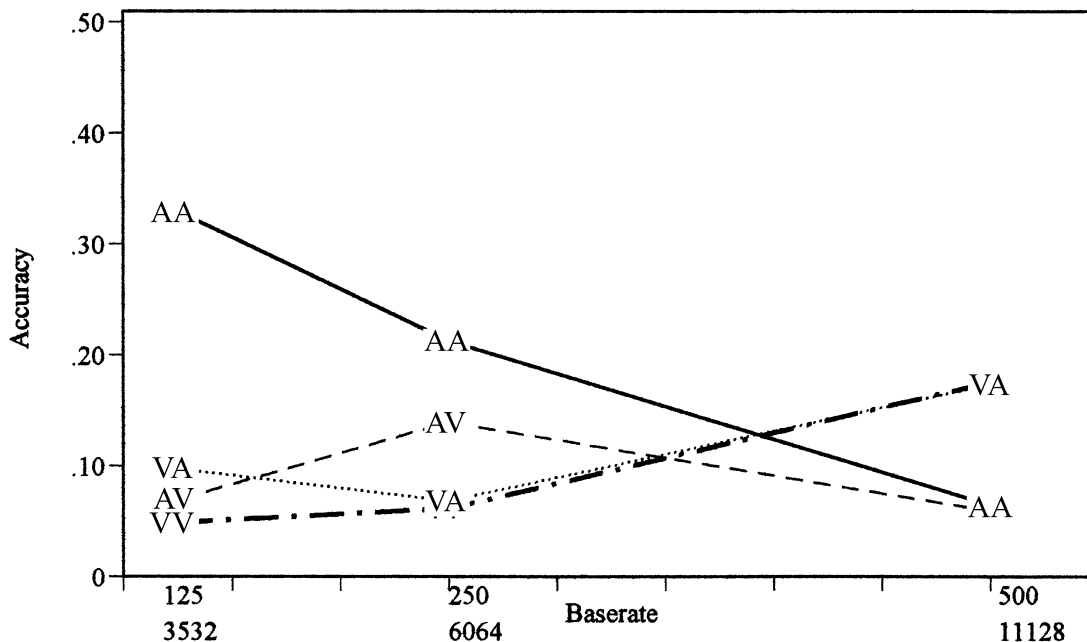


Figure 4. The results of Experiment 4, many different durational types.

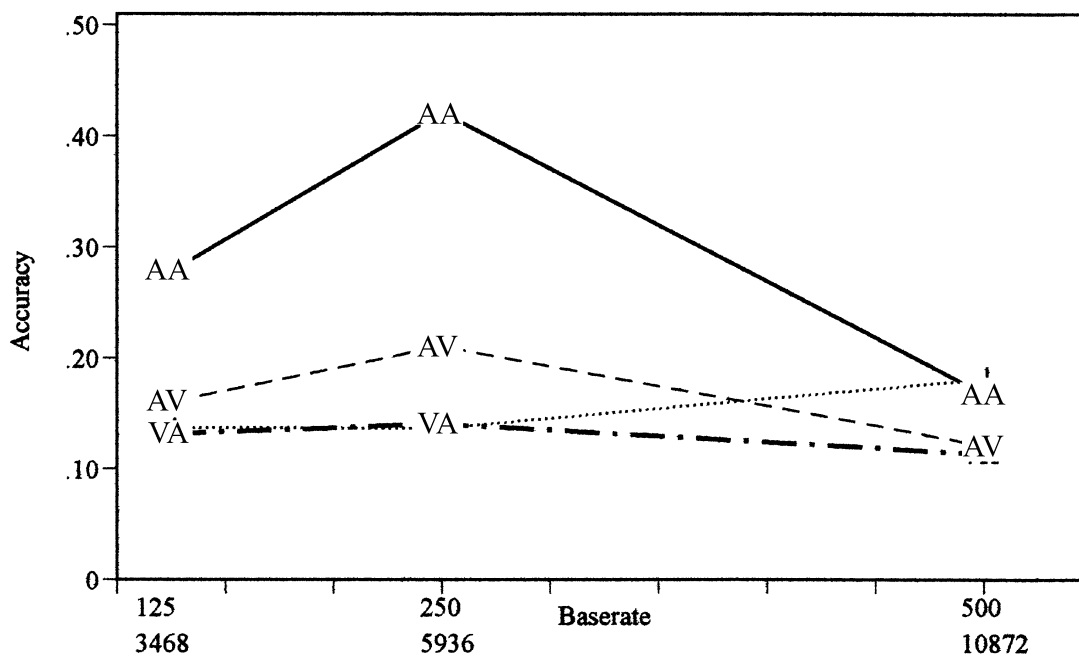


Figure 5. The results of Experiment 4, two different durational types.

confounding total trial duration and rate. The discussion thus far has assumed that total duration is the critical factor, but it could be the case that there are intrinsic differences between processing slow and fast stimuli, independent of the total stimulus durations. Perhaps faster auditory stimuli are easier to perceive as musical gestalts. To contrast this postcategorical hypothesis with the hypothesis of a simple precategorical echoic store, total trial duration was manipulated by varying the ISI rather than the presentation rate.

#### Method

The experimental design and stimuli were identical to those of Experiment 1, except that all the stimuli were presented at a base rate of 125 msec but the ISIs were 1,000, 2,500, and 7,250 msec, so that total trial lengths of both experiments were virtually identical.

#### Results

Overall, performance (Figure 6) appears worse than in the first experiment. But this is true even at an ISI of 1,000 msec, the condition identical with that observed at the fastest base rate of Experiment 1, so perhaps the difference was due to a chance sampling of less capable subjects or to the influence of the other conditions on subjects' strategy choices.

The effects of the manipulation of ISI were similar to the manipulation of base rate in the prior experiments. Again, AA was better than VV [ $F(1,27) = 33.3, p < .001$ ]. The accuracy of AA declined with increasing ISI [simple main effect of ISI on AA;  $F(2,54) = 23.4, p < .001$ ]. The modalities differed at the ISIs of 1,000 msec [ $F(3,81) = 18.2, p < .001$ ] and 2,500 msec [ $F(3,81) = 9.4, p < .001$ ],

but not at 7,250 msec [ $F(3,81) = .407, p = .41$ ]. Condition VV was unaffected by ISI [ $F(2,54) = 2.89, p = .75$ ].

As in Experiment 1, condition AV appeared to be better than VA, although this time the difference was not significant. In contrast to Experiment 1, however, the two modalities appeared to decline with increasing ISI, rather than to improve. This decline is significant for condition AV [ $F(2,54) = 5.15, p < .01$ ], but not for VA [ $F(2,54) = .415, p = .66$ ].

Finally, the two-way crossover interaction of first and second modalities is significant at the 1,000-msec ISI [ $F(1,27) = 11.4, p < .01$ ], but not at the other two ISIs. This is again consistent with the notion that there is a code-crossing handicap only when the trial durations are short.

In sum, the results of Experiments 1 and 5 were sufficiently alike to strengthen the belief that they were due to classical stimulus decay effects, rather than to special properties of fast rhythms over slower ones.

#### MUSICAL BACKGROUND

We examined whether musical background had an effect on the results. Mean accuracy across all the conditions was calculated for each subject in all the experiments, which was then correlated with information on the questionnaire on musical background. Correlations were negligible with objective indices, such as total number years of active involvement with music, number of years of lessons, and years since last actively involved with music. However, there was a correlation of .44 ( $n = 119, p < .001$ ) with a subjective rating of degree of interest in



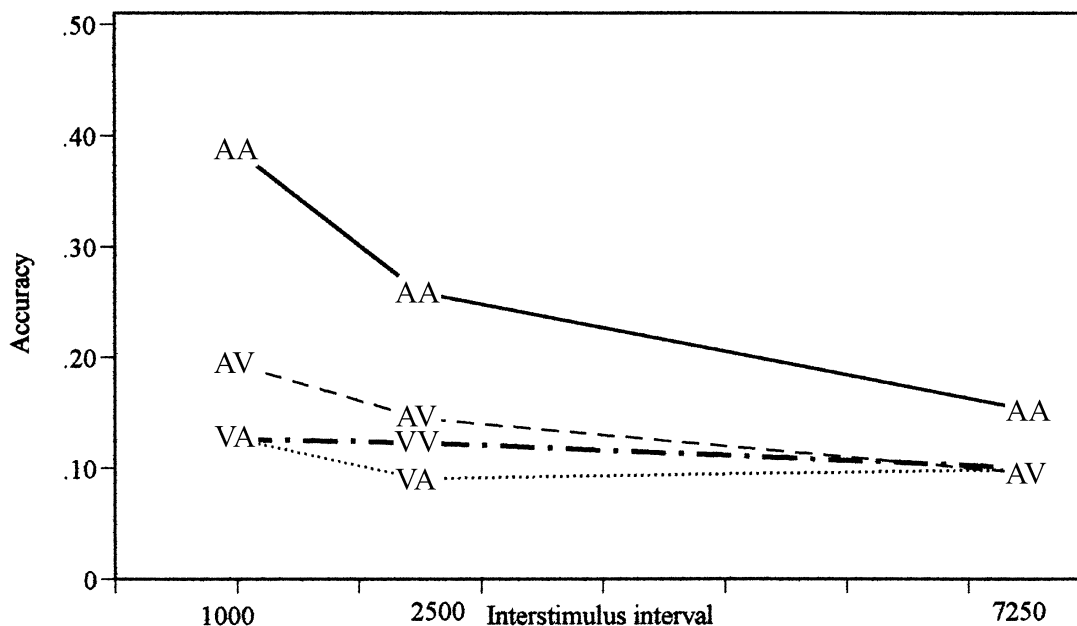


Figure 6. The results of Experiment 5.

music on a 7-point scale. Subsequently, we averaged the results separately for a low and a high musical interest group for each experiment but found no consistent interpretable difference between the groups. In particular, condition AA was superior to the other conditions, regardless of musical experience.

### GENERAL DISCUSSION

The five experiments demonstrated that same-different judgments of auditory rhythms are more accurate than those of visual rhythms. In most cases, audition showed a classic decay curve effect with either slower presentation rates or longer ISIs so that accuracy approached that of visual rhythms at slower rates.

At fast presentation rates, mixed modality rhythm pairs were as difficult as or more difficult than the visual pairs. At slower presentation rates, the mixed modality rhythms were in between the wholly visual and auditory conditions. Our interpretation is that the auditory advantage relies on some sort of proprietary code, which causes a cross-code decrement, but that during the decay interval the patterns were converted into a shared code.

Note that when Balch and Muscattelli (1986) studied melodic and spatial contour judgments, using the same four modality conditions that we did, they found a visual rather than an auditory advantage. This is consistent with the view that vision is fundamentally spatial, whereas audition is temporal.

Unlike our predecessors, we were unable to eliminate the auditory advantage at the fastest presentation rate. This advantage did not require auditory subvocalization

(unlike Watkins et al., 1992), nor did making the rhythms more complex eliminate the advantage (unlike Glenberg & Jona, 1991). The critical difference between these experiments and ours is that our fastest presentation rates resulted in substantially shorter stimuli than did those of our predecessors. Glenberg and Jona's complex ratio experiment was similar to our *two-types* stimulus condition of Experiment 4 (Figure 5), at our slowest presentation rate. In this condition, we did not observe an auditory advantage [VV vs. AA at 500 msec;  $t(29) = -0.897, p = .38$ ]. In contrast, at the slowest base rate in our Experiment 1, in which simple rhythms were used, there was a mild auditory advantage [VV vs. AA at 500 msec;  $t(16) = -2.573, p = .02$ ]. In sum, at these slower presentation rates the use of complex ratios reduced the auditory advantage, so that our results are consistent with those of Glenberg and Jona.

Turning next to the results of Watkins et al. (1992), they obtained no auditory advantage without concurrent silent mouthing, but then neither did we at comparably slow rates. In our view, concurrent mouthing reinstates the auditory advantage normally lost at these slower rates by interfering with the visual-to-auditory recoding process of the visual rhythms.

In addition, Watkins et al. (1992) eliminated the auditory advantage when the elements making up each stimulus were dissimilar. Specifically, a mixture of different environmental sounds eliminated the auditory advantage. We believe that this was due to the difficulty of forming a good auditory stream out of mixed environmental sounds (Bregman, 1990). This is consistent with the various results of Penney and associates (Penney, 1980, 1989; Penney & Butt, 1986), who have shown that recall of multimodal

streams of digits is biased toward recall of a single modality. All of this is also consistent with the cross-code decrement seen in our experiments at fast base rates; multimodal patterns do not form good rhythms because they are not heard as single streams.

Putting our results together with those of our predecessors, very short stimuli yield a robust auditory advantage, whereas long stimuli yield no auditory advantage. For stimuli of medium duration, there is an auditory advantage that can be weakened by various manipulations. We believe that it would be necessary to experiment with distracting stimuli of both modalities during the ISI to eliminate the auditory advantage for the fast stimuli.

In any case, the generality of theories of working memory and auditory short-term store cannot be tested by relying solely on verbal stimuli. Rhythmic stimuli are non-linguistic, multimodal, and parametrically manipulable, and they allow for complex encoding schemes. Thus, they provide an excellent alternative.

#### REFERENCES

- BALCH, W. R., & MUSCATELLI, D. L. (1986). The interaction of modality condition and presentation rate in short-term contour recognition. *Perception & Psychophysics*, **40**, 351-358.
- BREGMAN, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- COLLIER, G. L., & WRIGHT, C. E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 602-627.
- CROWDER, R. G., & GREENE, R. L. (1987). On the remembrance of times past: The irregular list technique. *Journal of Experimental Psychology: General*, **116**, 265-278.
- ESSENS, P. J., & POVEL, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, **37**, 1-7.
- GLENBERG, A. M., EBERHARDT, K. A., & BELDEN, T. M. (1987). The role of visual interference in producing the long-term modality effect. *Memory & Cognition*, **15**, 504-510.
- GLENBERG, A. M., & FERNANDEZ, A. (1988). Evidence for auditory temporal distinctiveness: Modality effects in order and frequency judgments. *Journal of Experimental Psychology*, **14**, 728-739.
- GLENBERG, A. M., & JONA, M. (1991). Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition*, **19**, 514-522.
- GLENBERG, A. M., MANN, S., ALTMAN, L., FORMAN, T., & PROCISE, S. (1989). Modality effects in the coding and reproduction of rhythms. *Memory & Cognition*, **17**, 373-383.
- GLENBERG, A. M., & SWANSON, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 3-15.
- MARTIN, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, **79**, 487-509.
- PENNEY, C. G. (1980). Order of report in bisensory verbal short-term memory. *Canadian Journal of Psychology*, **34**, 190-195.
- PENNEY, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, **17**, 398-422.
- PENNEY, C. G., & BUTT, A. K. (1986). Within- and between-modality associations in probed recall: A test of the separate-streams hypothesis. *Canadian Journal of Psychology*, **40**, 1-11.
- SCHAB, F. R., & CROWDER, R. G. (1989). Accuracy of temporal coding: Auditory-visual comparisons. *Memory & Cognition*, **17**, 384-397.
- WATKINS, M. J., LECOMPTE, M. N., & FISH, S. B. (1992). Short-term memory for the timing of auditory and visual signals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 931-937.

(Manuscript received September 21, 1998;  
revision accepted for publication May 17, 1999.)