INSTITUTE FOR THE

STUDY OF

INTELLECTUAL

BEHAVIOR

Preferred Rhythmic Structures in the Perception of Syllable Sequences

Alan Bell

Linguistics Department

University of Colorado

Institute for the Study of Intellectual Behavior

University of Colorado

Boulder, Colorado 80309

Technical Report No. 87

October 1979

ABSTRACT

Subjects listened to recorded sequences of repeated sounds of ambiguous rhythmic structure, and indicated the perceived rhythmic grouping of the sounds by pressing a telegraph key during each rhythmic group. The sequences were composed either of tones or of the synthesized syllable /ba/. The rhythmic structure of the sequences was marked by a difference either of pitch, intensity, or duration on every third sound; or by a longer or shorter interval between every third sound.

The results for nonspeech stimuli generally confirm the effects of the acoustic dimension of prominence upon the preferred placement of prominence within a rhythmic group that have been reported by Handel 1974 and earlier investigators. The same preferences were found for the /ba/ sequences as well, with the exception of low pitch prominence. Further work is required to establish whether the latter difference can be ascribed to differences between perception of nonspeech and speech.

Preferred Rhythmic Structures in the Perception of Syllable Sequences

Alan Bell

University of Colorado

Rhythmic perception and the prominence structure of rhythmic groups.

It is well established that many nonspeech stimuli, such as sequences of buzzes or tones, are perceived rhythmically in terms of repeating groups of sounds with one or more of the sounds of each group most prominent. This phenomenon occurs at appropriate rates (roughly between one and five repetitions per second) even with such completely unstructured and ambiguous stimuli as sequences of evenly spaced and identical sounds (Woodrow, 1909; Fraisse, 1956).

Assuming that speech is rhythmically structured, which is supported, in fact, by evidence of many sorts (Martin, 1972; Allen, 1975), it is reasonable to inquire whether general perceptual principles might not underlie some aspects of the rhythmic structures of speech. The preferences for placement of prominence in perceived rhythmic groups, which is the focus of the present paper, is an appropriate candidate for such an inquiry, for several reasons.

First of all, the hypothesis that the placement of prominence both in accentual systems and in rhythmic perception are affected by common cognitive processes is suggested by several striking parallels between the two. The parallels are found between preferred structures of accentual systems that are observed in comparisons over the world's languages and the preferred structures in rhythmic perception. For example, the repetition rates of accentual units in speech are in the same range as the rates for which rhythmic perception occurs. Further, accentual systems in language are typically characterized by a single major prominent element in each accent group; nonspeech sequences with rhythmic structures of this sort are easier to perceive and learn. Finally, accent placement in accentual systems is either initially or finally based; correspondingly structured nonspeech sequences are most easily remembered and learned, and moreover, when an ambiguous stimulus permits the perception of rhythmic groups with prominence at any position in the group, it tends to be perceived either in terms of rhythmic groups with an initial prominent element or in terms of groups with a final prominent element.

A second reason to consider prominence position is that it is feasible to investigate the interaction between the prominence structure of accentual systems or of perceived rhythmic groups and the acoustic dimensions which signal the structure. In most languages, accent is signaled by pitch, intensity, duration, or some combination of these. (These dimensions also have other important functions in speech, of course.) For nonspeech signals, the perceived place of prominence in rhythmic groups varies systematically according to the sensory dimension used to mark the prominent element, i.e. the preferred percept differs according to whether prominence is marked by pitch, intensity, or duration, in ways that are spelled out immediately below.

It is obviously necessary to determine whether such perceptions of nonspeech are also affected by the nature of the listener's language. Jakobson, Fant, and Halle (1952) conjectured that they are. Thus speakers of languages with an accentual pattern of initial stress would perceive rhythmic groups on nonspeech sounds with initial prominences, etc. The conjecture was not confirmed by Bell (1977), who found no tendency for the different positions of accent in Bengali, English, French, Persian, and Polish to influence the perception of speakers of those languages. The rhythmic perception of speech is apparently no more than possibly weakly language-dependent.

Perception of ambiguous auditory rhythmic structures.

The effect of sensory dimension on the perception of rhythmic groups can be observed in the relative ease of identifying and/or learning repeated sound combinations. For the very simple structures of the type that appear to be most relevant to the rhythmic structure of a language, a more sensitive probe is required. Ambiguous stimuli have proven to be satisfactory. The visual analogue is the Necker cube -- the two-dimensional sketch of the edges of a cube -- which is perceived as a cube as seen from below or else as a cube as seen from above, and not usually as an interesting two-dimensional design. The simplest auditory example is a regular succession of identical sounds, whicn, at appropriate rates, will be perceived as a succession of groups of sounds, with one or more members of the group more prominent than the others.

Although some people do perceive such stimuli as a succession of identical sounds, most hear a grouping of dyads, triads, or tetrads. This is the phenomenon called subjective rhythm. Now if to such sequences we add a bit of structure -- say we make every third element louder, or longer, or different in some way -- the listener will be disposed to perceive groups of three. The structure of the triads has three possibilities: initial prominence -- BA ba ba (a dactyl); medial prominence -- ba BA ba (an amphibrach); or final prominence -- ba ba BA (an anapest). And as with the Necker cube, the listener will be disposed to hear the triads as structured in one of these ways.

Which one of these rhythmic structures is heard is not equally likely for all kinds of prominent elements. The effect of the sensory dimension of prominence, as it is currently understood for nonspeech, can be summarized in four statements. (See Handel, 1974, for a recent treatment and reference to earlier work.) First of all, there is a general preference for groups with initial prominence. Secondly, this preference for initially prominent rhythmic structures is increased when the prominent element of the signal's rhythmic groups is marked by greater amplitude. This is a very robust effect, obtained many times under different conditions (Woodrow, 1909; Fraisse, 1956; Handel, 1974). Thirdly, with duration marking the prominent element, on the other hand, the tendency to perceive groups beginning with the prominent element is greatly reduced. Indeed, if the prominent element is made long enough, the perception of finally prominent groups becomes favored. Finally, with a higher or lower pitched element marking the rhythmic grouping, conflicting results have been obtained. Under some experimental conditions, a preference for initially prominent groups has been observed (e.g. Woodrow, 1911), whereas under others, either no preference or else a preference for finally prominent groups has been observed (e.g. Handel, 1974).

There are also consistent preferences, already mentioned above, which are largely independent of the sensory dimensions of the stimuli, and which depend on the structure of the sequences. The most important one for our purposes is that perception of triads with a medial prominence is greatly disfavored, which is a special case of a more general preference for initially or finally marked rhythmic groups, roughly speaking (Handel, 1974; Fraisse, 1956).

Experimental design.

The primary goal of the experiment was to extend earlier results on prominence in nonspeech to the perception of speech-like sequences. The paradigm of perception of ambiguous auditory sequences was used. Half of the subjects heard sequences of tones and half heard sequences of the synthesized syllable "BA." A second goal was to clarify the effect of pitch prominence and the effect of rhythmic patterns marked by intervals. We will return to the question of interval patterns shortly.

Experimental stimuli.

Subjects in the experiments listened to recorded sequences of repeated sounds, these sequences being about 45 seconds long. The sequences were prepared in digital form at 10,000 samples per second, converted to an analogue waveform, and recorded on audio tape.

The sequences were composed either of tone signals or of synthesized /ba/ syllables. The tone signals were generated by the Haskins Laboratories synthesizer by setting the first formant to twice the fundamental frequency and setting the amplitude of all higher formants to zero. The /ba/ syllables were synthesized using the OVE III synthesizer at Haskins Laboratories. Synthetic rather than natural speech was used so that the pitch, amplitude, and duration of the syllables could be independently controlled.

Either tones or "BA" syllables were combined to construct five types of sequences: null sequences, pitch sequences, amplitude sequences, duration sequences, and interval sequences. The rhythmic structure of each type was marked in a different way. All but the pitch sequences are illustrated schematically in Figure 1.

(Fig. 1 about here)

Null sequences were composed of a sequence of identical signals, each 186 msec. long, separated by 100 msec. The rate of presentation was thus 3.5/sec, toward the slow end of the rate of conversational speech. This rate was held constant for all sequences. The signals in the null sequences were either all low-pitched or else all high-pitched. The low-pitched tone signals had a fundamental frequency of 168 Hz; for the high-pitched tone signals it was 190 Hz. The low-pitched /ba/ syllables began with F_0 at 104 Hz, falling to 77 Hz over approximately the last half to two-thirds of the syllable. The high-pitched "BA" syllables began at 120 Hz, falling to 89 Hz.

Pitch sequences either had a low-pitched prominence, with every third tone or /ba/ low-pitched, or else had a high-pitched prominence, with every third sound a high-pitched tone or /ba/.

Amplitude sequences were constructed from high-pitched elements, with every third sound in the sequence having a higher amplitude than the others.

In the duration sequences, every third element was relatively longer than the others. As Figure 1 shows, the long element is of course longer than the ones in the null or amplitude sequences, but in addition the other two elements are correspondingly shorter. This is necessary if we are to hold the presentation rate constant and also to preserve equal intervals between the signals. This is of course not the only possible choice, and indeed some earlier studies, including Bell (1977), used duration sequences in which the prominent element was lengthened at the expense of the following interval. This, however, produces a sequence in which two rhythmic structures are superimposed -- a duration pattern and an interval pattern.

Interval patterns can be schematized as in Figure 2. The null sequence A in the figure has equal intervals, so its interval pattern is represented by ... x x x x x x ... where the x's stand for the signal elements. The duration sequence B1 is constructed simply by lenthening every third element; of course this makes every third interval shorter, producing the interval pattern shown. The experiment used duration sequences of the type B2, with a neutral interval pattern.

The effect of interval patterns was treated separately using sequences of the type C2 in Figure 2: the durations of the sounds were kept constant, every third interval was either shortened or lengthened, and the other intervals proportionately lengthened or shortened, respectively, thus holding the presentation rate constant.

(Figure 2 about here)

Interval sequences offer three possible rhythmic triad structures just as the sequences with prominent elements do. If every third interval is longer, then the three structures are

```
... |X   X   X|X   X   X|X   X   X|X   X   X|X  ...
...  X | X   X X | X   X X | X   X X | X   X X| ...
...  X   X  |X X   X  |X X   X  |X X   X  |X X  ...
```

If every third interval is shorter, then the three structures are

```
... |X  X  X|X  X  X|X  X  X|X  X  X|X  ...
...  X | X  X X | X  X X | X  X X | X  X X | ...
...  X  X | X X  X | X X  X | X X  X | X X  ...
```

In order to make the sequences perfectly ambiguous, their structure should have no perceptible beginning or end. There are a number of ways to accomplish this more or less satisfactorily. The sequences used here began with a gradual increase in amplitude, reaching full strength in about 5 seconds, and terminated with a similar fade in amplitude. The method poses some problems, which are discussed below.

Experimental procedure.

Each subject heard 13 sequences, as follows: 2 null sequences, high-pitched or low-pitched elements; 2 pitch sequence, high-pitched prominences or low-pitched prominences; 3 amplitude sequences, prominences of 3 dB, 6 dB, or 9 dB; 3 duration sequences, prominences of 1.25, 1.75, or 2.25 greater duration; and 3 interval sequences, with every third interval 0.7 times, 1.5 times, or 2.2 times the length of the other intervals.

Subjects were counterbalanced over 11 quasi-random orders of presentation. Each of the 11 orders began with a different nonnull sequence, which was followed by one of the null sequences. In every order a different sequence occurred in third position after the null sequence. The second null sequence fell between the sixth and eleventh positions. Every order had a different nonnull sequence before this null sequence, and every order had a different nonnull sequence in final position. Order of presentation in the eleven conditions was otherwise random.

The mode of subject response has been perhaps the greatest weakness of past research on rhythmic perception. It has mainly relied on self-report. In fact, as far as I can tell from his reports, Woodrow, the pioneer investigator in this area, determined the percept for a given sequence, continuing until Woodrow decided that a stable percept had been obtained. It is also possible to obtain results by asking subjects to use some schematic formula for writing what they hear as they hear it, e.g. /⌣ ⌣ / ⌣ ⌣ , a procedure used by Handel (1974) and Bell (1975). My experience with this method suggests to me that the possibility of confusion in reports of this sort is great enough for concern. For the present experiment, therefore, I turned to a more indirect method of recording subjects' presses of a telegraph key.

I first thought, mistakenly, that simply asking subjects to tap the key in rhythm to what they heard would be the best method. Too many subjects simply found the rate of presentation too fast to follow, with the amount of training that I could feasibly provide. But more importantly, the results did not admit sufficiently consistent interpretation, even for the subjects who were able to follow the stimuli. For many subjects, one would find clear responses such as Al illustrated in Figure 3, with clear separation of groups and longer response corresponding to the prominent element of the stimulus sequence. For too many subjects, though, the inter-group separations were not consistently greater than the intra-group separations. And in some cases, there occurred responses for which the beat did not correspond to the rhythmic pattern of the stimulus, as illustrated by response A2 in Figure 3.

(Figure 3 about here)

The response procedure that produced the most satisfactory results in pilot trials, and which was adopted for the present experiment, was to request subjects to press the key during the whole group that they heard. That is, when they heard the sequence occurring in rhythmic groups, they were to press the key at the beginning of each group and to release it at the end of each group. An ideal response of this kind would look like response B1 of Figure 3; most actual responses that were interpretable ranged from the extremes represented by responses B2 and B3 in Figure 3. Subjects were instructed that most people heard the sounds in groups of three, but that different people heard different groups.

The subjects were volunteers, students in the School of Education of the University of Colorado during the summer of 1977, and members of faculty and staff of the University. Subjects were native speakers of English, had no previous history of hearing difficulties, and had not had extensive musical training. About 10 per cent of the subjects could not perform the task after the training provided, and had to be replaced. Twenty-two subjects heard sequences of tones, and 22 heard sequences of "BA" syllables.

<u>Results</u>.

The responses were categorized according to the position of the prominent element in the groups indicated by the subject. Where the response indicated that the perceived grouping had shifted during the sequence, the first grouping indicated was chosen for analysis. The results are shown in Table 1 and plotted in Figures 4 through 7.

(Table 1 and Figs. 4-7 about here)

An inspection of the data shows that in general there is little difference between the results for the /ba/ syllables and tone, with the notable exception of the sequences with low pitch prominence and those with duration prominence of 125%. This impression is confirmed by the χ^2 statistic, even though its application here can only provide an approximate test, since some of the observed frequencies are so close to 1 or zero. For the amplitude sequences, χ^2 (2) for the number of initially-prominent groups (dactyls) is 5.1. For the interval sequences, χ^2 (2) for the number of X XX groups for the short interval sequence and of XXX groups for the long interval sequences is 2.2. Both values are below the .05 significance level of 6.0.

Turning to the effects of prominence type, we note first of all that the frequency of perception of rhythmic groups with medial prominence (amphibrachs) is uniformly low for amplitude, duration, and pitch, again with a notable exception: low pitch prominence.

The proportion of dactyls is highest for the amplitude sequences, and increases as the degree of prominence increases. The amplitude sequences and duration sequences are very significantly different in this regard (at a level less than .001), even omitting the deviant results for the 125% duration sequence. The proportion of dactyl responses for high pitch prominence sequences falls in the lower part of the range for amplitude sequences.

The indicated rhythmic groups for interval sequences show the following preference patterns: 1) That if every third interval is shorter, the perceived grouping will not divide the two closest signals (no instances observed), and that there is no significant preference between the two remaining patterns, X XX and XX X. 2) If every third interval is longer, there is strong preference for the perceived rhythmic group to coincide with the long interval, with the preference significantly greater when the interval is lengthened 220% than when it is lengthened 150%.

Discussion.

The results in general confirm the effects of prominence on perceived rhythmic structure described earlier, notably the increased preference for dactyls as the amplitude of prominence is increased, and the reduced preference for dactyls when greater duration marks the prominence.

The responses to the /ba/ duration sequences follow the pattern reported by earlier investigators, with preference for dactyls decreasing with increasing duration of prominence. The results are also consistent with the reports by Fraisse (1956) that a preference for finally prominent groups (anapests) occurs as the duration of the prominent element reaches 1.5 to 2.0 times that of the other elements. Thus the responses for the tone sequence with 125% duration prominence, with 16 of 22 subjects indicating an anapest is anomalous, since this is <u>greater</u> than the number indicating an anapest for the longer prominence durations of 175% and 225%. I am thus inclined to attribute the result to an experimental artifact, the most likely candidate for which is a bias introduced by the onset of the sequence.

The remarkable difference between the /ba/ sequences and tone sequences with low pitch prominence must be considered as possibly indicating a difference between nonspeech and speech perception. Unlike the duration prominence sequences, the two pitch prominence sequences cannot be presumed to be of the same general nature. So the similar pattern of perception for the high pitch sequences of both /ba/'s and tones does not necessarily cast suspicion on the results for the low pitch sequences. But why should low pitch act differently from high pitch as a prominent element? We may speculate that LH and HL pairs form subunits of rhythmic groups and that LH is disfavored initially, leaving HLH and HHL as preferred groups for the sequences with low pitch prominences. (Since preferences of prominence position differed only for the /ba/ sequences, the dispreference for the rising LH pairs would be specific to speech stimuli. This is at least consistent with the general preference for falling tones as opposed to rising tones in the world's languages with lexical tone contrasts.) Recall, however, that the pitch of the /ba/ stimuli and tone stimuli differed in a possibly significant way. While the tones had a constant pitch, the /ba/ syllables were synthesized with a falling pitch, in order to make them more speechlike. It is plausible then that a LH sequence is dispreferred because the low element is a low <u>falling</u> element, and it is not just dispreferred for speech. This could be tested by comparing the perception of tones with falling pitches. An experiment to do this is in progress.

In Bell (1977) an unexpectedly large proportion of subjects (48%) perceived sequences with duration prominence in terms of amphibrachs ( ), and very few perceived anapests ( ), contrary to findings by earlier investigators. This was explained as a consequence of an interference between the duration rhythmic pattern and the interval rhythmic pattern that was introduced by the shortened interval that followed the longer elements, as in B2, Figure 2. The present results for the short interval sequence, whose elements were of identical duration, show that XX X and X XX interval patterns are equally preferred. Since the explanation of Bell (1977) assumed that X XX groups were preferred to XX X groups, it must accordingly be modified in the following way. Grouping in terms of anapests ( ) corresponds to the heavily dispreferred interval pattern of X X X (which splits the shortest interval). This pattern would be perceived only if subjects were able to ignore the interval pattern. If they attended to the interval pattern, then they would report either a dactyl (and XX X) or an amphibrach (and X XX). The results still contradict those of Fraisse (1956), who obtained preferences for anapests with sequences in which duration prominence was similarly confounded with an interval pattern. The patterns may have interacted differently at the slower presentation rates (about 2/sec) used by Fraisse.

Two characteristics of the stimuli used in the present experiment need to be kept in mind in evaluating the results. The first is the nature of the gradual onsets. The amplitude envelopes were brought up from inaudibility to full value over about five seconds. While this doubtless reduces any bias of an initial starting point, it may not remove it. Nor is there any means of evaluating the effect. The pertinence of an initial bias effect, moreover, is not just restricted to the experimental paradigm. If general principles of rhythmic perception are relevant to accentual structures of speech, then initial bias is likely to be involved in some way, for speech is not a continuous stream of elements but is rather a series of chunks marked by distinct intonation contours and/or pauses. In order to interpret the patterns of perception of ambiguous rhythmic sequences both for speech and nonspeech, it is necessary to evaluate the effect of the onset of sequences, whether gradual or not. Experiments to make the necessary comparisons are in progress.

The second characteristic concerns the /ba/ sequences. Just how speechlike are they? From debriefing subjects, it appears that most of them heard the sequences as speech syllables. A significant number, however, reported either verbal transformation effects -- hearing bla, bla, bla,... or da, da, da,..., etc. -- or a "nonverbal" transformation in which the /ba/'s dissolved into noise pulses. The monotony of the single syllable evidently impairs the speechlike quality of the /ba/ sequences. Thus while they are certainly more speechlike than the tone sequences, at least some of the sequences were probably processed as nonspeech by subjects. This could have obscured some differences in patterns of perception for the two types of sequences.

Conclusions.

The general patterns of prominence placement in auditory rhythmic perception that have been established for nonspeech stimuli hold for most of the sequences of /ba/ syllables considered. A difference between perception of nonspeech and speech stimuli is most likely to be found for rhythmic structures with pitch prominence. Further work is needed to determine: 1) the effect of sequence onset, 2) the effect of different types of pitch prominence, and 3) the effect of less monotonous syllable sequences (such as combinations of /ba/'s and /da/'s, for example).

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at the 5th International Congress of Applied Linguistics, Montreal, August 21, 1978. The research was supported by the Council on Research and Creative Work of the University of Colorado and by NICHD Contract No. N01-HD-1-2420 to the Haskins Laboratories. I wish to express my appreciation for the hospitality of the staff at Haskins Laboratories and for the assistance of Mary Coberly with running the subjects and analyzing the data, and for the cooperation of the School of Education of The University of Colorado.

REFERENCES

Allen, George D. 1975. Speech rhythm: its relation to performance universals and articulatory timing. J. Phon. 3. 75-86.

Bell, Alan. 1977. Accent placement and perception of prominence in rhythmic structures. In L. M. Hyman, Ed., Studies in Stress and Accent, 1-13. Los Angeles: University of Southern California Linguistics Department.

Fraisse, Paul. 1956. Les structures rhythmiques. Louvain: Publications Universitaires de Louvain.

Handel, Stephen. 1974. Perceiving melodic and rhythmic auditory patterns. Journal of Experimental Psychology, 103. 922-933.

Jakobson, Roman, Gunnar Fant, and Morris Halle. 1952. Preliminaries to speech analysis. Technical Report 13, M.I.T. Acoustics Laboratory, 5th Printing, M.I.T. Press, 1963.

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

Woodrow, Herbert. 1909. A quantitative study of rhythm. The effects of variations in intensity, rate, and duration. Archives of Psychology 14. 1-66.

Woodrow, Herbert. 1911. The role of pitch in rhythm. Psychological Review 18. 54-77.

Figure 1. Four types of stimulus sequences

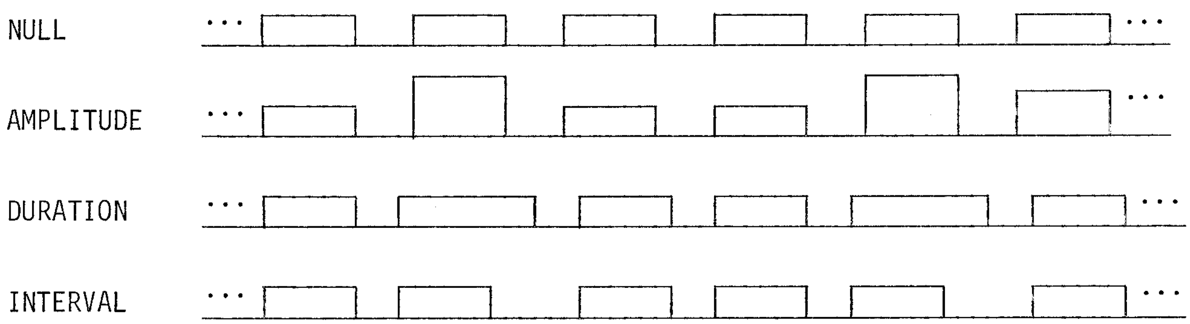
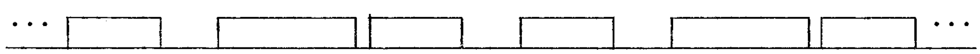


Figure 2. Duration and interval patterns

A. Null sequence

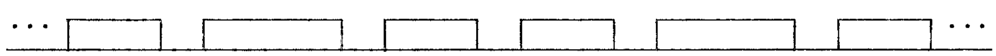
Interval pattern: ...X X X X X X X X...

B1. Duration sequence confounded with interval pattern



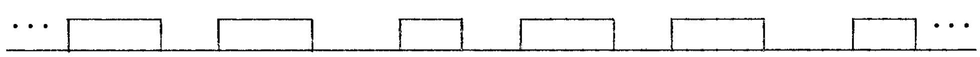
Interval pattern: ...X XX X XX X XX X...

B2. Duration sequence with equal intervals



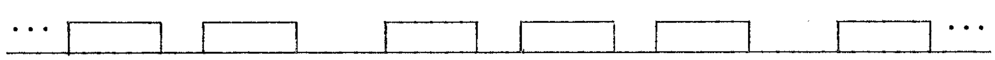
Interval pattern: ...X X X X X X X X...

C1. Interval sequence confounded with duration pattern



Interval pattern: ...XXX XXX XXX XXX...

C2. Interval sequence with equal durations



Interval pattern: ...XXX XXX XXX XXX...

Figure 3. Telegraph key response patterns

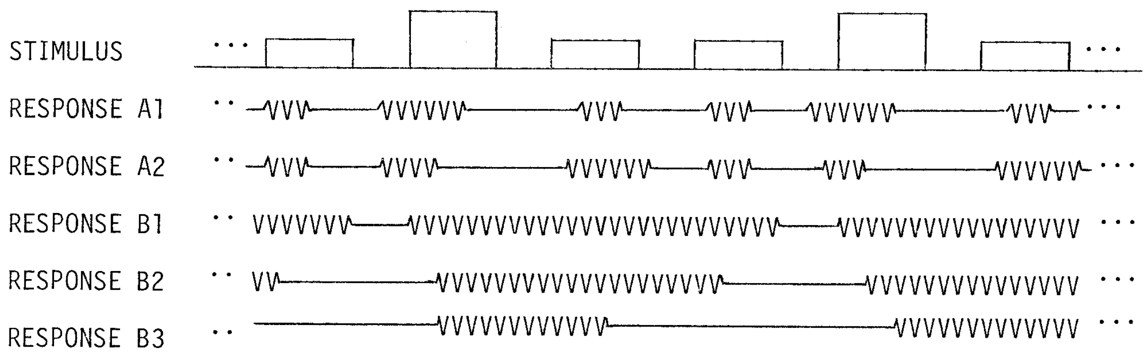
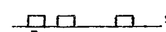
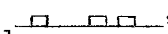
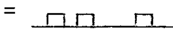


Table 1. Number of subjects by prominence structure of stimulus sequence and by prominence position of indicated rhythmic group. For the interval sequences, the "prominent" element of the rhythmic triad was taken to be the element before the different interval. Thus for the short (.70) interval sequence, initial = , medial = , and final = . For the long (1.50 and 2.20) interval sequences, initial = , medial = , and final = .

Prominence structure of sequence		Tone sequences: Prominence position of indicated rhythmic group				/ba/ sequences: Prominence position of indicated rhythmic group			
		Initial	Medial	Final	Not scorable	Initial	Medial	Final	Not scorable
Amplitude prominences	3 dB	14	1	4	3	9	3	7	3
	6 dB	19	2	1	0	16	1	5	0
	9 dB	19	0	3	0	21	0	1	0
Duration prominences	1.25	1	1	16	4	12	1	8	1
	1.75	7	3	11	1	6	2	14	0
	2.25	7	0	13	2	8	2	11	1
Pitch prominences	High	14	0	8	0	13	3	3	3
	Low	9	0	12	1	0	8	9	5
Interval structures	0.70	8	10	0	4	11	7	0	4
	1.50	4	1	15	2	1	5	15	1
	2.20	2	0	19	1	0	1	21	0

Figure 4. Number of subjects indicating dactyl rhythmic groups for sequences with pitch prominence.

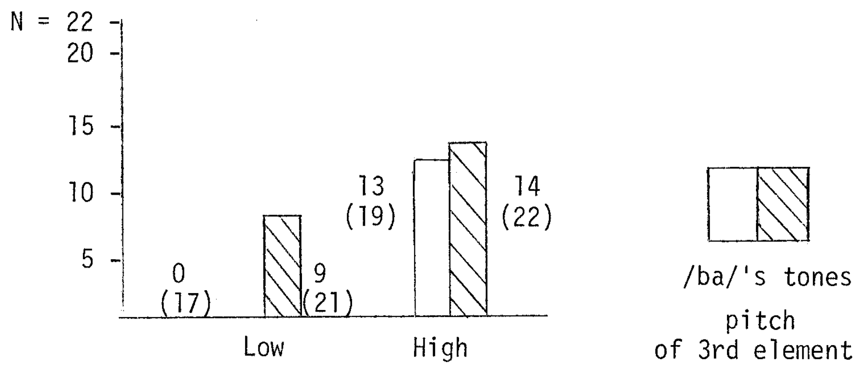


Figure 5. Number of subjects indicating dactyl rhythmic groups for sequences with amplitude prominence.

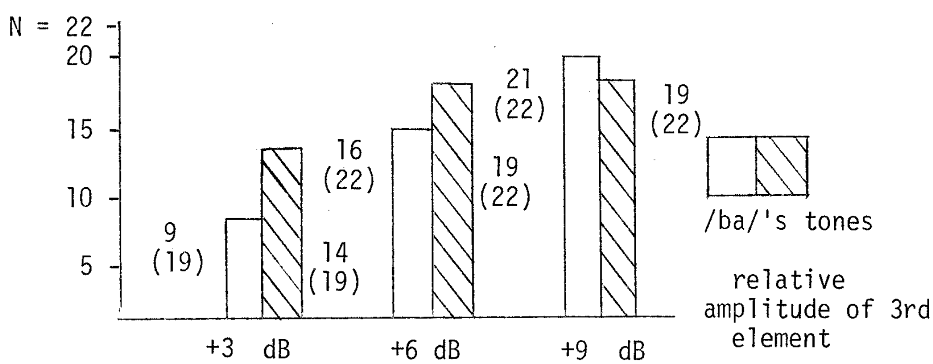


Figure 6. Number of subjects indicating dactyl rhythmic groups for sequences with duration prominence.

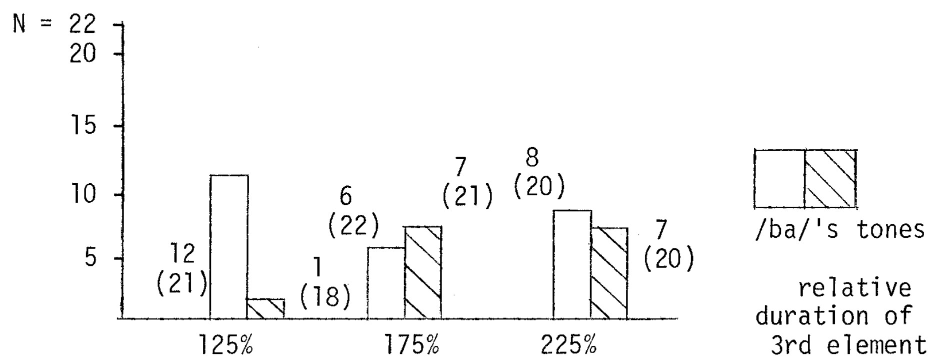


Figure 7. Number of subjects indicating specified rhythmic groups for sequences with interval patterns.