Psychological Review
1972, Vol. 79, No. 6, 487–509

RHYTHMIC (HIERARCHICAL) VERSUS SERIAL STRUCTURE IN SPEECH AND OTHER BEHAVIOR[1]

JAMES G. MARTIN[2]
University of Maryland

Lashley regarded rhythmic action as a central concept in the problem oi serial order in behavior, and implied that it was a natural link between the perception and production of connected speech. Rhythm means relative timing between adjacent and nonadjacent elements in a behavior sequence (hence nonsequential dependencies), as opposed to temporal ordering (concatenation) of elements, with implications for motor and perceptual behavior in real time. Two simple rules for generating rhythmic (hierarchically organized) behavior sequences are proposed and applied to speech, music, and other meaningful sound patterns. The temporal constraints on rhythmic sequences lead to several hypotheses concerning their efficient perception and production.

Just over 20 years ago, Lashley (1951) published his influential paper on the problem of serial order in behavior, emphasizing speech and discussing a number of topics he believed to be highly interrelated, including associative-chain versus hierachical conceptions of behavior, determining tendencies, the generality of syntax, central control of motor patterns, rhythmic action, and others. All but one of the topics mentioned have been widely discussed, the first in particular (e.g., Adams, 1968; Chomsky, 1959; Greenwald, 1970; Halwes & Jenkins, 1971; Keele, 1968; MacNeilage, 1970; Miller, Galanter, & Pribram, 1960; Milner, 1970; Neisser, 1967; Wickelgren, 1969). The last of these topics, Lashley's views on rhythmic action, has not been considered in detail by anyone with the exception of Lenneberg (1967). But Lashley (1951) strongly implied that rhythmic action and hierachical motor organization were highly related concepts and even went on to hint that rhythmic action might be the natural link between the perception and production of connected speech. This latter theme, which is potentially one of the most interesting in his entire paper, has attracted little attention. To be sure, Lashley was not quite clear on the point, but another reason for its neglect could be the several misconceptions that are commonly held about rhythmic action. One of these is that rhythm implies *only* periodic, repetitive behavior, like walking or breathing, and hence is too simple a concept to be of interest for speech behavior. Another quite different misconception is that rhythm necessarily entails *strict* temporal regularities, usually along with complex, aesthetically motivated elaborations, as in music, and hence is too esoteric to apply to natural speech sounds except by trivial analogy. Lashley's (1951) views were not limited in either of these ways. By making clear what rhythmic action is and by showing how it has consequences for theories of motor and perceptual behavior in real time, this paper suggests that rhythmic action is one of the central concepts in the schema Lashley presented.

Rhythm, which means temporal patterning, is a concept based on motor functioning. Natural movement sequences as produced by the motor system are determined by a variety of factors including temporal constraints,

[1] This article was supported by Grant MH 16726 from the National Institute of Mental Health, United States Public Health Service, and by a grant from the Bio-Medical Sciences Support Committee, University of Maryland, to the Center for Language and Cognition.

[2] The author expresses thanks to Charles N. Cofer and to an anonymous referee for extensive comments on an earlier version. Nancy Hurlbut, Cathy Klemic, Dianne Reamy, Judith Sprei, and Judith Zachai assisted in the data collection and analysis.

Reqests for reprints should be sent to James G. Martin, Department of Psychology, University of Maryland, College Park, Maryland 20742.

some of which are based on the dynamics of inertia, etc. These constraints necessarily are reflected in the organization of sounds produced by human movements, which include speech and music. In the case of speech these constraints influence the segmental details of the acoustic signal, but they should be expected also to affect at least some aspects of the morphology and syntax of any language at the level of syllable strings since they play a role in determining what can be naturally and easily spoken.

The constraint on speech sounds, or on any other real-time sequence of behavioral elements, that is directly implied by the concept of rhythm is *relative timing,* which means that the locus of each (sound) element along the time dimension is determined relative to the locus of *all* other elements in the sequence, adjacent *and* nonadjacent. This is to say that sequences of sounds, speech or otherwise, that are rhythmic will possess hierachical organization, that is, a coherent internal structure, at the *sound* level. The alternative to a rhythmic sequence, which is less restrictive, is that the loci of the elements in the sequence are only successive (concatenated) in time. Sound sequences like these cannot have a structured, internal organization, although entities that they might represent can, of course, have such an organization, albeit one that is relatively highly abstract.

These alternatives, relative timing (rhythmic patterning) as opposed to concatenation, are related to quite distinct theoretical points of view concerning real-time behavior. It is more obvious that speech and other temporally unfolding behavior sequences proceed left to right than that they might possess internal structure. Hence it is relatively easy to adopt a concatenative (associative) view about these sequences, which will be reinforced by the inevitable finding that there are dependencies between adjacent elements in the sequence. Local (adjacent) dependencies, however, rarely reveal the role that these elements play in the more wholistic, hierarchically organized structures of which they are a part. The alternative to the associative view is to begin by carving nature at the joints, so to speak, by considering first the wholistic unit which *has* the internal structure and then by analyzing the dependency relationships between all elements in terms of the unit. Witness Chomsky's (1957) revolution in linguistics, which depended in part on his choice of the sentence rather than some other unit as the object for study. Wholistic real-time sound units (i.e., rhythmic patterns) also have a hierachical internal structure, as is demonstrated later in the paper.

Relative timing as opposed to concatenation also has logically distinct implications for production and perception models. The production of rhythmic sounds entails that the temporal locus, hence duration, of each sound element is related to each other locus in the resulting pattern. This requirement places a heavier constraint on timing mechanisms than does the concatenative alternative, which requires only that the resulting sounds be produced in the correct temporal order. Furthermore, the temporal patterning relations between nonadjacent sounds, hence movements, which rhythm implies seem to call for present timing, hence a predominantly central rather than peripheral control of the timing mechanisms.

In the case of perception, rhythmic constraints on production entail that the sound inputs during perception will be temporally *patterned.* Patterned speech sounds could be redundant with respect to linguistic message elements to a far greater extent than sounds that are only concatenated. Furthermore, since rhythmically patterned sounds have a time trajectory that can be tracked without continuous monitoring, perception of initial elements in a pattern allows later elements to be *anticipated* in real time. Suppose that some elements in a sequence are more informative than others. If these informative elements are nonadjacent and temporally predictable, then certain efficient perceptual strategies (e.g., attention cycling between input and processing) might be facilitated. Perception of concatenated sounds, on the other hand, would seem to require continuous attention. More generally, perceptual requirements should be expected to differ for sounds that can be expected at times T_1, T_K, *and* T_N, as opposed to sounds expected at unspecified times. These and other logical implications of relative timing for production and perception should be translatable into empirical consequences.

There is, however, no grammar and no model of speech perception or production that incorporates the constraints imposed by relative timing to speech sequences extending beyond a few phonemes; in all present grammars and psychological models, temporal relationships existing between elements are represented as concatenations only.[3] These concatenations will serve most grammars adequately, and even many psychological models, for example, models of sentence comprehension. Since the relationships of interest in such models are abstract, and since subjects in experiments testing these models will be provided either with inputs spoken by humans or inputs they can read, problems of accounting for the processing of the speech sounds themselves are bypassed and questions concerning any relations between the abstract entities and the sounds do not arise. But if relative timing is the true situation in speech, it has consequences for any theory or model concerned with connected speech in real time. Up to the present time the only such theories or models have been those in the theoretically and practically important areas of automatic speech synthesis and automatic speech recognition. The current status of these areas rather accurately reflects present understanding of speech production and perception, and in each of them relative timing could be an important consideration. A machine could produce concatenated speech, which in the present view would be temporally distorted but which would be fairly intelligible to humans nevertheless—given the high redundancy in speech (Ladefoged, 1967b) and the perceptual strategies available to the native listener. However, a general-purpose machine routine for recognizing natural connected speech that ignored temporal relationships would be inefficient at best and might not work at all if much of the more critical, or more reliable, information in the acoustic signal were systematically related to the temporal dimension of the signal.

In the next section, the meaning of rhythmic action is made explicit by means of two formal descriptive rules (a grammar) that are proposed for describing the timing of elements in a natural temporal pattern. These rules generate element sequences having a hierarchical structure that are here called rhythmic patterns. In the section following, these rules for rhythmic structure are applied to the patterns of speech, and then some implications for the phonetic reality of these patterns are considered. If speech has the rhythmic structure that the rules describe, then constraints are placed on the possible mechanisms that might be postulated for efficient speech perception and production. Several hypotheses for a speaker–listener model are presented, together with bits of evidence that might be relevant to each. The hypotheses are related to the general notion that speaking and listening are dynamically coupled rhythmic activities, such that linguistic information is encoded rhythmically into the signal by the speaker and decoded out of it on that basis by the listener. Finally, some speculations concerning the implications of rhythm for other research areas are offered. The emphasis in this paper is upon speech, but music and other meaningful temporal patterns are considered also, on the theory that they have a common base and are rooted in the motor and perceptual capabilities of the human organism.

Rhythmic Patterns

Two rules for the formal description of simple natural rhythmic patterns are proposed here. They are intended to relate a variety of intuitions and other facts about rhythmic structure, some of which are mentioned during the discussion.

Rhythmic patterns are defined as event sequences in which some events (elements) are marked from others, for example, loud (or high) sound (or left foot) versus soft

[3] Nor is there a theory of auditory temporal pattern perception which uses relative timing. In fact, except for Garner's work (e.g., Garner, 1970; Garner & Gottwald, 1968; Preusser et al., 1971; Royer & Garner, 1966, 1970) there has been little research on auditory temporal pattern perception since about the time of Woodrow (1909).

(or low) sound (or right foot); call the marked elements "accents." The accents recur with some regularity, regardless of tempo (fast, slow) or tempo *changes* within the pattern (accelerate, retard). Together an accented and unaccented element constitute the simplest rhythmic pattern; more complex patterns consist of patterns within patterns; that is, they are hierarchically organized (Martin, 1969b, 1970b).

Natural rhythmic patterns can take a variety of surface forms, but nevertheless have a simple underlying structure that can be characterized by one or the other of two obligatory rules. These rules express the relation between (*a*) relative accent level on the elements in a sequence and (*b*) relative timing of the elements.

The first rule is called the accent rule and applies to more or less continuing pattern sequences, the simplest case being the two elements represented by filled dots in Figure 1a. The numerals *under* the tree refer to relative accent level on the two elements (level one highest), showing that the accent naturally falls on the first element of a duple. In the four-element sequence represented by the four equidistant dots in Figure 1b, the strongest accent is on the first element, but the second strongest is on the third element, indicating the generally alternating character of accent level.

Now notice that left branches are labeled zero, right branches are labeled one. These represent binary numbers that can be obtained by reading up or down the tree and are used to determine either the accent level of a given element in the sequence or its serial position (not labeled) in the sequence. To compute accent level, read up the tree, convert the binary number to decimal number, and add 1. Recall that binary number 11 equals decimal number 3, 10 equals decimal number 2, 01 equals decimal number 1, and so on. Hence, accent level on the last element in Figure 1b, for example, is 3 + 1 = 4. Similarly, to obtain serial position, read down the tree, convert binary to decimal, and add 1. This reciprocal relation between accent level and accent location (timing) in the sequence is the accent rule and is general for any number of nodes.

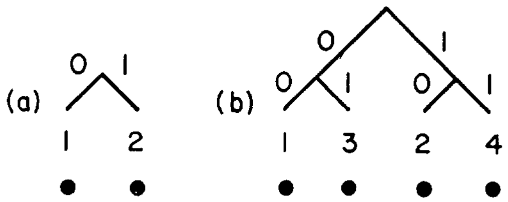

FIG. 1. Simple rhythmic patterns (see text for explanation).

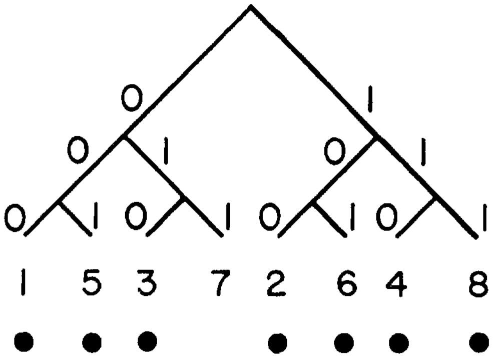

FIG. 2. Rhythmic tree with one null branch (see text for explanation).

The three-node-level tree in Figure 2 has a null branch, that is, the fourth element is missing, but with no effect on the remaining elements since accent level is determined not by serial position but by location along the time line, whether or not the element in question is actually realized. This applies not only to staccato elements, which are separated by silent intervals as in Figure 2 but also to more or less continuous sounds such as legato musical phrases or strings of syllables, in which case the temporal patterning would refer to the onset of each musical note or syllabic vowel.[4] Notice further in Figure 2 that the third element is necessarily lengthened if the sound sequence is continuous. Thus the duration of an element in continuous sounds depends upon rhythmic context and is rhythmically independent of its accent level, from the point of view of relative timing. Accent level correlates with duration, but for other reasons, as discussed later.

Notice in Figure 2 that the elements in first and fifth position are those most

[4] As is seen later, the stipulation that lower-level right branches can be null, that is, are somewhat optional, permits great flexibility in application of the rule.

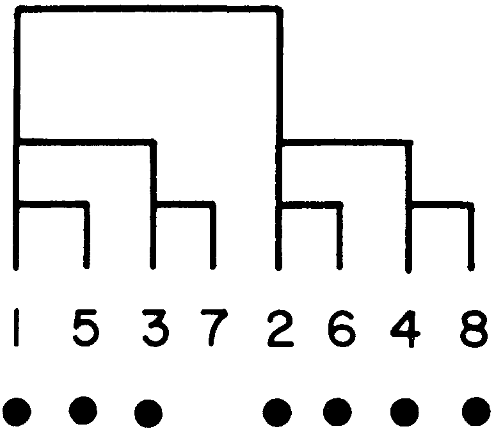

Fig. 3. Another way of drawing the rhythmic tree seen in Figure 2.

strongly accented. To demonstrate to yourself that these accent levels are the "natural" ones, tap an eight-element sequence on your desk, repeatedly, as fast as you can, accenting the first and the fifth elements. Then try the same thing but accent the first and the sixth elements. This pattern is more difficult than the first one, but not merely because accenting elements one and five degenerates the eight-pattern into two four-patterns. To demonstrate this, tap again, comparing the pattern having accents on elements one and six with another having accents on elements one, three, and five. The latter does not degenerate into equal-sized sequences, and it requires programming an additional accent, but it is generally easier nonetheless. Related to the view that the accent rule describes natural patterns is an observation by Woodrow (1951). He noted that listeners presented with continuous sequences of evenly spaced identical sounds often reported subjective grouping into fours, with stronger subjective accent on the first, and weaker subjective accent on the third, element of the group.

Figure 3 is exactly equivalent to Figure 2 and shows more clearly the relation between timing and accent levels. Time intervals are subdivided to the right via lower nodes but without affecting the temporal and accentual relations between elements springing from higher nodes. In this sense these elements are hierarchically rather than serially dependent. To the extent that right branches on lower nodes are optional, this tree maps onto a variety of terminal strings, but they will all have the same underlying structure; this is inherent in the meaning of rhythmic structure.

To illustrate its use, the accent rule in tree notation is applied below to some simple musical phrases, with appropriate musical notation added accordingly (pitch notation omitted). For the nonmusical reader it may be helpful to point out that nothing of relevance is missed if he prefers to read the tree only and to ignore the musical notation. The musical reader may appreciate that the musical notation is a shorthand for the tree diagram but, more important, that the tree diagram makes explicit the accent patterns usually implied by musical notation and tacitly known though never taught. Tree diagrams are as little used for learning music as for learning to read.[5]

In Figure 4, the vertical bars are boundaries between measures giving a two-measure phrase. Notes connected together by horizontal bars are subdivisions of one beat, giving two beats to a measure. The word *Old* falls on the first half of the first beat and receives strongest accent, the word *had* receives second strongest accent, and so forth.

In Figure 5, the first beat is further sub-

[5] That is, the rhythm rule as stated here makes explicit what the musician (or schoolchild) knows tacitly in the same sense in which a linguistic rule makes explicit what the speaker of the language knows tacitly. Rhythm appears to be taken so much for granted in music training that there is only one book on rhythmic theory although there are many on melody, harmony, and counterpoint (Cooper & Meyer, 1960). The latter are culturally determined to a far greater extent than is rhythm.

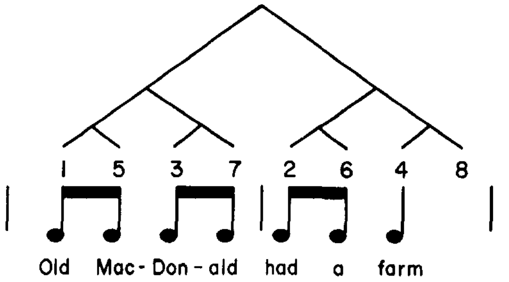

Fig. 4. The first phrase of a familiar tune. (Accent levels generated by the rhythmic tree can be seen above the notes.)

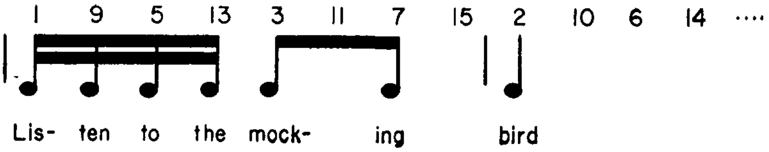

Fig. 5. Another familiar musical phrase. (Accent levels are those generated by a tree with four node levels.)

divided (each bar connecting notes halves their durations), requiring another node level in the tree but not a change in the relations between the three (higher level) notes on the beat. The tree has been omitted to save space, but it should be clear that the sequence of accent levels above the notes in Figure 5 is that generated by the accent rule for a four-node-level tree. The number of node levels is strictly determined by the element having shortest relative duration. As Figure 5 shows, adding a node level generally results in increased numbers of null branches.

Subdivisions of an interval can occur in other ways as well, as in the incomplete phrase in Figure 6. Examples like these show that the pattern of notes giving a tune its own rhythmic flavor can vary but nonetheless will map onto trees defined by the obligatory accent rule.[6]

Whereas the accent rule applies more or less to repeating patterns or to nonfinal musical phrases, the second rule, called the terminal rule, applies to nonrepeating or terminal sequences like musical cadences, as well as to a variety of speech units. Stated formally, it is: First, apply the accent rule to the tree to obtain the sequence of accents, then reverse the positions of accent levels one and two. The consequence of this rule, as is apparent later, is that it allows primary accent to appear not only on the initial element of the sequence but anywhere in the pattern. It is descriptively far more flexible than the accent rule.

The detailed mechanics for use of the terminal rule are shown later. For the moment it is convenient to consider an example of the results of the rule. Figure 7 shows the last two measures of the *Old MacDonald* song. Notice that the sequence of accent levels (shown above the notes) is exactly like the output of the accent rule for the three-node-level tree seen in Figure 4 except that accent levels one and two are reversed. In this context the effect is that the phrase then ends on a strong accent, which is typical of most simple musical cadences. In the next section, the terminal rule is shown to apply also to the rhythmic patterning of speech.

Rhythmic Structure of Speech

"Rhythm" here as in conventional usage refers to the pattern of accents/stresses on

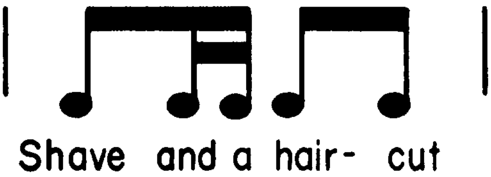

Fig. 6. Part of a familiar cadence.

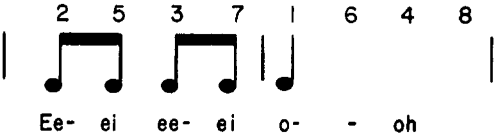

Fig. 7. The last phrase of a familiar tune (see text for explanation).

[6] Of course musical rhythms quickly depart from or elaborate upon simple rules such as these, but notice that syncopation and other interesting devices are effective only when played against the simpler underlying patterns, real or implied. More to the point here, complex rhythms are not required for ordinary speech and probably most poetry. Thus these rules have not been stated in music theory because they are too simple to be of interest. They have not, however, been stated in studies of linguistics or prosody either, to the best of the author's knowledge, although they are of theoretical importance in these areas. However, the importance of relative timing, etc., in verse has been known (or argued) for a long time (see, e.g., Lanier [1880]).

Incidentally, any reader still waiting for the resolution of the incomplete phrase in Fig. 6 is probably aware that the strong accent beginning the next measure is not actually sounded but only implied (and felt), a conventional rhythmic device.

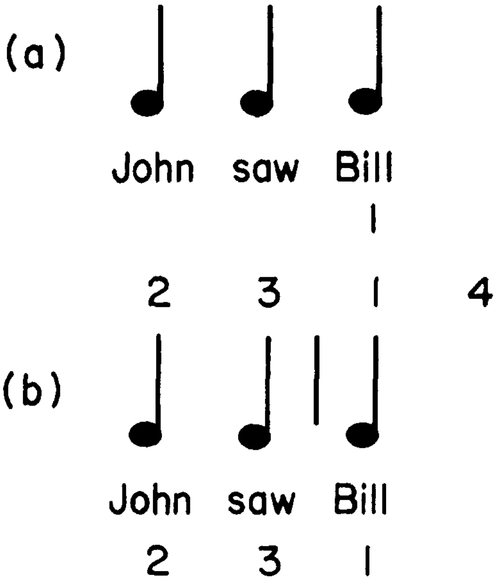

Fig. 8. Applying the terminal rule to a simple sentence (see text for explanation).

a string of syllables. The rhythmic pattern as defined here is assumed ordinarily to consist of up to seven syllables or so; it is thus a prosodic unit often corresponding to breath groups (Lieberman, 1967), syntagma (Kozhevnikov & Chistovich, 1965), tone groups (Boomer & Laver, 1968; Halliday, 1963), or phonemic clauses (Boomer, 1965; Trager & Smith, 1951). In its usage here, the term "accent" is equivalent to linguistic "stress" except that (formally) accent level is applied relationally to *all* syllables in a pattern. Hence it includes and orders not only all varieties of linguistic stress such as emphatic, primary, sentence, word stress, etc., but also unstressed and reduced vowels. How many of these levels are to be called accented is then an arbitrary choice. This relational approach is thus neutral with respect to the terminological question of "degrees of stress" to be allowed in English (cf. Chomsky & Halle, 1968; Halliday, 1963; Vanderslice, 1968, 1970) or in any other language.

Application of the Rhythm Rule to Speech

On the theory presented here, the terminal rule postulates the relations holding between relative accent level and relative timing for each syllable in the rhythmic pattern. Thus, for example, if accent levels are assigned to syllables in a string, then syllable durations are fixed accordingly. Alternatively, given the relative duration of each syllable in the string plus location of primary accent, the remaining accents are determined. More generally, given some minimum amount of information concerning accent levels and duration on syllables in a string, one could state (or predict) the remaining accent levels and/or syllable durations.

To illustrate, consider Figure 8a, which shows a string of three syllables (elements) of equal duration, with the last syllable marked for primary accent. (Notice that duration of the last syllable in any string is irrelevant since the rule applies not to syllable durations but to syllable loci, specifically, their vowel onsets.)

To place accent level on each syllable, the rhythm rules are applied as follows. (*a*) First determine the minimal tree required. Three evenly spaced elements require a two-node-level tree to yield four terminal branches, one of which will be null. (*b*) Apply the accent rule to the tree to obtain the sequence of accent levels 1324. (*c*) Reverse levels one and two by the terminal rule to obtain 2314. (*d*) Map this sequence onto the syllable string such that accent level one corresponds to the syllable marked for primary accent. The result of this maneuver, shown above the notes in Figure 8b, places relative accent level on each of the three syllables in the string. The last accent level generated by the rule is null, that is, the rightmost tree branch is not actualized. The resulting accent pattern for the elements actualized, 231, is recapitulated below the words in Figure 8b. A vertical bar marks the boundary between "measures" as in Figure 7; this bar will always precede primary accent. Since primary accent always falls on the first left branch in the right half of the rhythmic tree, any pattern can be represented as falling within two "measures." Thus the outer two vertical bars can be omitted, as they are in all remaining examples. Again, primary accent is not restricted to final position by the terminal rule but can occur anywhere in the pattern, as is seen later.

Now consider the three examples in Figure 9, in which the monosyllabic names *John* or *Bill* in Figure 8 have been replaced by the bisyllabic names *Henry, Mary,* or *Marie,* which require one interval subdivision. Put another way, one additional differential relative syllable duration must be considered. Hence one additional node level is required, which generates the accent-level sequence 15372648 by the accent rule, converted to 25371648 by the terminal rule. With primary accent located on the accented syllable of the last word, the sequence is then mapped onto each sentence to yield the accent levels as indicated above the notes. More generally, a sequence of X elements requires a tree with $2^N \geqslant X$ branches, N being the minimal number of node levels if elements are equal in duration. If any one element is one-half the duration of any other, $N + 1$ node levels are required. If any one element is one-fourth the duration of any other, $N + 2$ node levels are required, and so forth.[7]

Numerals below the notes in Figure 9 give accent levels for elements actualized, in rank order. Since three syllables here are arbitrarily counted as accented, the parentheses identify "unaccented" syllables. Thus the accent pattern for each of the three sentences is the same, 231. These examples illustrate the general principle, seen most obviously in languages like English, that syllable durations are compensatorily adjusted so that accented syllables will tend to fall at equidistant intervals. By contrast, on a concatenative view of syllable timing there would be no need for the duration of the word *saw* to change in the differing sentence contexts of Figures 9b and 9c.

It is important to note that these patterns are somewhat simplified. Accent level and syllable duration are positively correlated (Lehiste, 1970) so that, depending upon how the pattern was pronounced, an accented syllable, for example, *saw* in Figure 9c might be represented as somewhat longer and the following unaccented syllable as correspondingly shorter. At least one additional node level then would be required, in consequence placing lower relative accent on the unaccented syllable (vowel reduction?), but without changing the patterning between the three accented syllables. In principle at least, such details can be represented by the terminal rule whenever required. It is worth noting, incidentally, that each of the sentences in Figure 9 follows the accent pattern assigned by the transformational stress cycle (Chomsky & Halle, 1968).[8] The terminal rule, however, is assumed to be more general, as may be seen next.

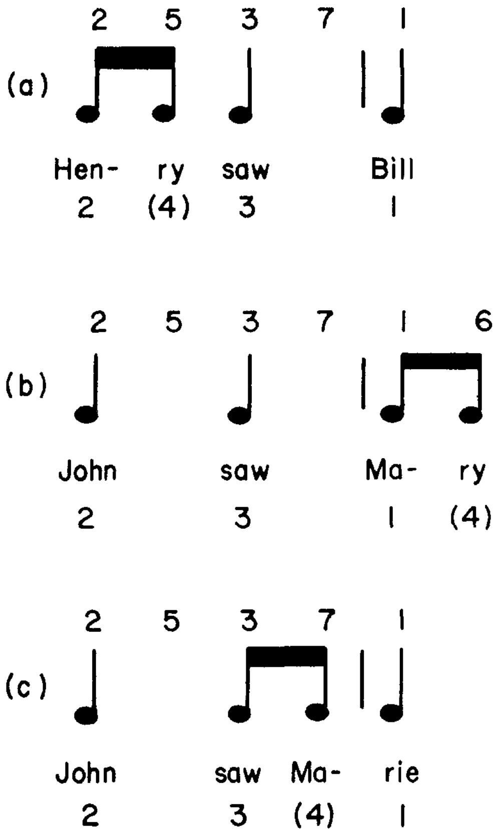

FIG. 9. Applying the terminal rule to three sentences (see text for explanation).

[7] One additional node level is also required if primary accent falls temporally within the first half of the sequence. In this case, the right half of the tree generated by the above criteria for node levels will not cover the elements in the sequence. One additional node level gives the required result, although now nearly all (sometimes, all) branches on the left half of the tree are null.

[8] The rhythm rule is easily shown to apply also to the idealized pronunications of words in isolation, although in most cases some intuition about syllable durations is required since phonology gives only notional time. Thus the relation between syntax and stress in English as expressed by the transformational cycle appears to be a special case of the rhythm rule.

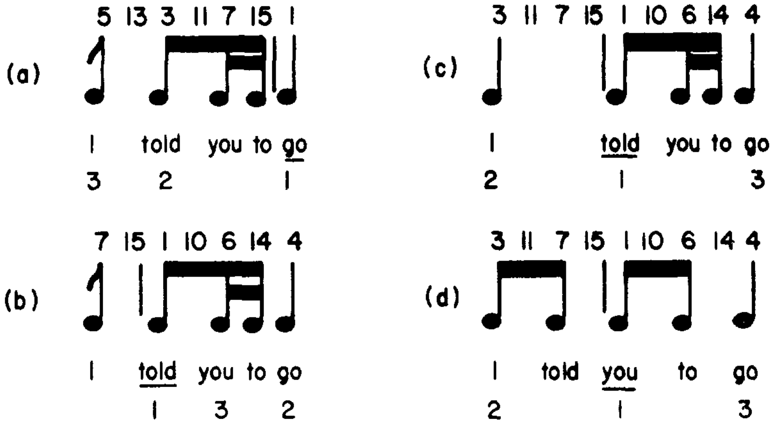

Fig. 10. Four rhythmic versions of the "same" sentence. (The underlined word has been emphasized).

A given sentence, phrase, etc., can be spoken in a variety of ways, being influenced by context, mispronunciation, "foreign accent," regional dialects, the use of emphasis or contrast, and so on. The effects of some of these factors on accent have been discussed by others (e.g., Allen, 1968; Bierwisch, 1968; Bolinger, 1965; Bresnan, 1971; Newman, 1946; Vanderslice, 1968). Consideration of such factors points to another general principle of speech rhythms illustrated here, which is that whenever these factors affect accent placement in an utterance, the temporal pattern of the utterance is affected also. In Figure 8b, emphatic accent on *saw* would give the accent pattern 213; on *John,* 132. To illustrate further, consider the effects of change of emphatic accent in Figure 10. Again the numerals above the notes represent accent level on each syllable by the terminal rule. Numerals below assign accent level in rank order to the three highest values. Although the time values (durations) of the syllables are the same in Figures 10a and 10b, the accent levels, hence rhythmic patterns, differ. Time values are different though accent levels are the same in Figures 10c and 10d. Primary accent is located on the same word in Figures 10b and 10c but the (somewhat unnatural) longer (earlier?) initial syllable in Figure 10c receives greater accent than in Figure 10b, thus changing the whole pattern.[9] Other variations are possible, but in every case a change of accent level or location requires a reorganization of the whole pattern. This principle is assumed to apply to mispronunciation, foreign accents, dialects, and all other factors affecting accent assignment in an utterance.

Notice, finally, that the principle of equal-interval accent holds in all of the examples above except Figure 10a. Given the temporal pattern pronounced exactly as indicated in Figure 10a, the rule requires the accent pattern indicated. If, however, the syllable *I* were shorter than indicated (which is not an unlikely prouncіation), the additional required node level would place lower relative accent on *I*, so that the third highest accent would then shift to *you,* giving equal-interval accent pattern 231. On the other hand, given the pronunciation as is, deciding *arbitrarily* in favor of four accent levels rather than three preserves the equal-interval accent principle since then the pattern is 3241. This is a terminological point; the main point is that the tree gives the temporal location of accent however defined (and of "vowel reduction" as well) as it is distributed in natural running speech.

In summary, the rhythmic patterns of speech are wholistic units that can be characterized in terms of the interrelated dimensions of relative accent level and relative timing. These patterns are idealizations; their point-by-point physical correlates will depend on many factors in any actual situation. Notions like these bear strongly on

[9] Put another way, placing stronger accent on the initial syllable of Figure 10c requires greater relative duration in this rhythmic context.

real-time theories of speech production and perception and are considered later in these terms. However, it is appropriate for the moment to consider first some implications for phonetic description.

Phonetic Considerations

A survey of phonetic research shows that interest in the prosodic characteristics and temporal organization of speech has grown rapidly over the past several years. Earlier the emphasis was predominantly on the segmental characteristics of speech, and questions concerning timing and prosodic features arose more rarely. Three of these questions are considered here. They are relevant because they concern speech timing, but also because they more or less directly illustrate the consequences of adopting one point of view or another about the relations holding between the elements in a string of speech sounds. A given point of view on this problem, whether explicit or not, implicates any number of assumptions about the "units" in speech, research questions to be asked, observations regarded as relevant, perceptual expectations, and so on. Thus the relative timing point of view, which assumes the rhythmic pattern as the unit of temporal organization, entails certain expectations concerning the phonetic manifestations of accent and timing, which differ from those that might be expected on a concatenative, or a neutral, point of view. These differences can be seen in each of the problems discussed below.

Acoustic correlates of accent. A conventional view concerning prosodic and segmental levels of speech, made explicit by Fry (1968), has been to regard segment strings (hence syllable strings) as in some sense the "vehicle" for prosodic phenomena. This view implies, at least by omission, a concatenative view of the temporal relationships between the elements in a string. From this it would seem to follow that when perceived accent occurs in an utterance, its physical correlates are to be located "on" the accented syllables themselves. Accordingly, much of the research on the question of the acoustic correlates of accent typically has been concerned with comparisons based on syllable pairs (e.g., *OBject, obJECT*) without running context, or at least without consideration of the effects of running context. Work in a variety of languages has shown that higher fundamental frequency, greater intensity, and greater duration are generally associated with accented, as opposed to unaccented, syllables, but that there are no simple correspondences between any one or a combination of these acoustic characteristics of a syllable and relative accent (Fry, 1968; Ladefoged, 1967a; Lehiste, 1970; see the latter for an extensive review of the research on prosodic phenomena).

By contrast, the key assumption of the present view is that the rhythmic organization of speech requires accented syllables to fall at roughly equal intervals in time. As shown earlier, this constraint entails that syllable durations, on accented and unaccented syllables alike, will vary according to utterance-specific context. But, in the present theory, utterance-specific context also affects the intensity and frequency characteristics of the syllables, just as it affects duration, one reason being that frequency and intensity characteristics are probably partly determined by the number and rate of articulatory transitions required between syllables in the course of producing a given utterance. Thus the relative values of these characteristics on accented syllables will vary with local context and should not be expected to correlate with perceived relative accent. Instead, the invariance holding between accented syllables is relative timing. Some additional minimal aspects of a rhythmic pattern are assumed to be acoustically marked in the syllable string; these often will include the end of the pattern, which is marked by pause or characteristic fundamental frequency contour, and primary accent (Cushing, 1969; Fonagy, 1966). This information is sufficient to "anchor" the pattern so that relative nonprimary accent on syllables can be "heard" in terms of the rhythm rule, that is, in terms of their location in the temporal pattern. Relative accent is thus a perceptual consequence of relative timing.

It seems clear that accurate reports of accent, in this view, would require some sort of perceptual orientation toward the whole extent of the temporal pattern. For instance, such an orientation is necessary in order to transcribe the temporal pattern of a speech or musical phrase into rhythmic notation once it is heard. But the conviction is strong that the acoustic manifestations of accent are "on" the accented syllables themselves, as attested to by the research devoted to the problem, and indeed it is deceptively easy to attempt comparative judgments based on pitch, loudness, etc. As Lieberman (1965) has shown, however, these judgments will often be wrong, even when they are made by sophisticated linguists. Judgments like these apparently are influenced by a variety of linguistic expectations, with conflicting expectations leading to conflicting judgments. In short, in the present view, relative timing is the proper basis for the perception of relative accent, but it cannot easily be separated from acoustic and other dimensions on which judgments of accent can be, and are, made. A statistical correlation between perceived relative accent and relative timing might emerge over a variety of listeners, listener expectations, and utterances, but would not be highly probable in any particular instance.

Some recent experiments, however, have separated timing from other aspects of running speech, and they show quite clearly that accent and timing are related. These experiments used connected speech materials in which temporal relationships had been artificially distorted. In one approach, judgments of accent on a syllable were shown to depend upon location of the syllable within the rhythmic context of an utterance when the effects of frequency, intensity, and duration of the "accented" syllable itself were ruled out. Listeners heard two versions of sentences and judged relative accent on the last two (monosyllabic) words. Tape splicing prior to the words placed one or the other of the words on the rhythmic "beat." Judged words were thus acoustically equivalent in the two versions. The word on the beat was judged more often as the accented word of the pair (Martin, 1970a).

In a very recent report, using extremely precise electronic splicing techniques, Huggins (1972a) artificially shortened or lengthened a single segment in naturally spoken sentences, for example, the first /m/ in *The hostel for paupers has two moody managers.* Listener reports indicated that as the segment shifted between shortened and lengthened, primary accent shifted between *two* and *managers.* These experiments clearly separate timing from other aspects of the stimulus and show that perceived accent does depend in part at least upon temporal context, and not solely upon the characteristics of the accented syllables themselves.

In another recent paper, Huggins (1972b) artificially varied adjacent segment durations in normally spoken sentences to study compensation effects between them and found that the temporal distortions were more likely to go unnoticed when the original timing relationships between the accented syllables in the sentences remained intact. Again, this experiment demonstrates the importance of relative timing in the perception of accent. Finally, it should be mentioned here that accent is seen as having other correlates than the acoustic characteristics already mentioned, for example, certain measures of physiological effort (Ladefoged, 1967b). Fonagy (1966) regards accent as based upon cues to effort that can be perceived by a listener, even without knowledge of the language. The present view that rhythm, hence accent, is based upon the constraints on natural movement appears to be somewhat compatible with these views.

"Stress-timed" versus "syllable-timed" languages. In the present view, natural constraints on movements entail that running speech must be rhythmic as defined here, so that accented syllables are generally separated from each other and will fall at roughly regular intervals. Obviously, this view also implies that equal-interval accent is a language universal. But the principle of equal-interval accent appears to be perceptually more convincing in the case of some languages than in the case of others, and in fact some languages have been classified as "stress timed" (equal-interval stress) in the former case as opposed to "syllable timed" (equal-length syllables) in the latter (Pike, 1945). Not many phoneticians seem to have accepted this classification, but it is nevertheless of

some interest to consider what might have motivated the choice of this particular dichotomy. It appears that the rhythm rule directly implies the perceptual phenomena upon which the classification is based. To illustrate, assume for the moment, as does the present theory, that all languages are stress timed, in the sense in which the term is used. When word stress in a language is fixed such that widely varying numbers of affixes and function words must be accommodated between stressed syllables, as is the case in English, so that the result is often the "crushing together" (Pike, 1945) of a large number of unstressed syllables, then the perceptual prominence of stressed syllables should be enhanced. This result is implied by the rhythm rule since subdivision of any interval in the pattern adds a node level to the tree, thus increasing the relative difference between the highest and lowest values of accent level placed on each syllable in the pattern. But when word stress in a language is relatively free, or is fixed on syllables at or near a word boundary (either one or the other boundary but not both, in a given language), then the number of syllables between stressed syllables will be fewer, on the average, than in a language like English, hence requiring less interval subdivision and resulting in the impression of evenly spaced syllables and less prominent stress at the same time. Thus English is called stress timed, while French and Spanish, which have late word stress (Delattre, 1965; Harris, 1969; Schane, 1968), have been called syllable timed (Abercrombie, 1967; Harris, 1969).[10] There are, of course, numerous other accent-related linguistic characteristics peculiar to these languages or to any others; but in the present theory these characteristics will sometimes enhance, sometimes weaken, the perceptual prominence of the underlying rhythmic structure that is always present. In short, stress timing is simply more obvious in some languages than others.

Some unpublished data from a study by the author support this view. Very briefly, the purpose (in part) was to show that some unlikely language candidates (viz., Spanish and French, as well as Italian) could be heard as stress timed under suitable conditions.

Tape-recorded excerpts of a monologue provided the stimulus materials in each language. In an indirect approach (*a*) musically trained listeners provided a rhythmic notation, hence temporal pattern, for each utterance. Then (*b*) fluent (mostly native) speakers of each language listened and reported accent patterns (three levels of relative accent on three specific syllables). Application of the terminal rule to the temporal patterns reported by the musicians then predicted accent location and level reported independently by the fluent speakers. The results were that when fluent speakers agreed with each other on which three syllables were accented, the syllables tended to be equidistant, as the theory predicts. Also noteworthy was that the patterns of relative accent on the three syllables followed the terminal rule significantly more often would be expected by chance.

Perceptual and physical time. The claim that English is stress timed has led to several attempts at physical measurement of interaccent intervals (see Allen, 1968a, 1968b; Bolinger, 1965), but these measurements have shown substantial variation between the intervals. In the present view, there are two reasons why a lack of correspondence between perceptual and physical measures of these intervals might be expected. The first is that tempos can change within a rhythmic pattern. When they do, these changes can usually be apprehended yet do not ordinarily affect the perceptual integrity of the pattern, that is, "beats" are still heard as beats; nor do they affect the listener's ability to "follow" or to anticipate later elements in time, since the rhythm is definite though not regular, in Lashley's (1951) words. These tempo changes would be revealed by systematic and cumulative changes in physical syllable durations, either increasing or decreasing, depending upon whether the change was positive or negative. But to detect these

[10] Note also that languages in which accent is located relative to word boundaries are unlikely to have a transformation cycle relating accent to syntax, since there is little of syntactic significance in a word boundary.

systematic changes and separate them from measurement error in any given utterance, it is necessary first to postulate a specific rhythmic pattern, that is, the pattern of syllable durations that would obtain in an even tempo, and then to compare departures from these values in the observed durations by curve-fitting or similar routine. For instance, relative time intervals from an utterance heard and transcribed in rhythmic notation might be plotted against cumulative measured vowel onsets. A straight line would indicate even tempo, a curved line would indicate an *accelerando* or *ritard.* Clearly, alternatives such as calculating "average" or "intrinsic" syllable durations without regard to specific pattern context would be of little value in this regard.

A second reason for expecting a lack of correspondence between physical and perceptual measures of the interval between accents involves interesting perceptual problems that have been little explored. One of these concerns the correlation between perceptual and physical measures of the dimensions of a rhythmic or any other auditory temporal pattern when the dimensions are considered independently as opposed to jointly. There is some evidence that the time dimension interacts with the accent-level dimension so that physical and perceptual measures of either dimension taken alone will not correspond one to one. Woodrow (1909), using nonspeech stimuli, showed that variations in physical relative intensity level on elements affect perceived relative timing. On the other hand, variations in physical relative timing affect perceived accent level (Martin, 1970a), as mentioned earlier.

Rhythmic Structure in Speech Production and Perception

Several Hypotheses about Speaking and Listening in Real Time

Speech utterances in the present view are temporally patterned, and many of their elements are temporally predictable. These characteristics of the auditory stimulus constrain the class of mechanisms that might be postulated for their efficient processing (Garner, 1970b). In this section, main interest centers on how linguistic information might be encoded into and decoded out of the speech signal at the rhythmic (prosodic) level, but some implications for the segmental level are considered also. As will be seen, there is reason to believe that the importance of prosodic as opposed to segmental factors has been underestimated in theories of the production and perception of everyday connected speech.

Hierarchical organization of the speech production program. Any evidence for rhythmic patterning in speech, that is, that the syllables in a string are timed relative to the whole pattern rather than concatenated, is also evidence for preset timing and hence leans toward the view favoring central rather than peripheral timing mechanisms during production.[11] A more specific hypothesis which follows is that since the accented elements dominate the temporal organization of an utterance, they must in some sense be planned first. Intervening, lower-level syllables then are planned subsequently in hierarchical fashion, by (metaphorically) reading the rhythm tree, level by level, from the top down. Planning here means at least selecting the time at which syllables will be actualized, although this decision must facilitate any number of other contingent decisions during production. In this view one might think of accented syllables as the main targets in the organization of the articulatory program. The hypothesis then predicts that decisions concerning accented syllables will precede those concerning unaccented syllables, hence that decisions concerning "content" words will precede those for "function" words, and so forth. Observed utterances do not often illuminate the mechanisms generating them; but errors, for example, slips of the tongue, sometimes do. Fromkin's (1971) interesting work on slips of the tongue revealed that whole words in an utterance could be transposed without affecting the overall accent pattern of the utterance, for example,

a laboratory in our own computer.
2 3 1

[11] Invertebrate physiologists appear virtually to take for granted that rhythmic behavior is predominantly under central control (see many papers in Wiersma [1967]).

Her interpretation of this result was that the syntax and the stress contour (accent pattern) of the utterance was generated prior to word selection. There are, however, only a few accent patterns, of which 231, the pattern in all but one of Fromkin's three-accent-level examples, is the most common. On the present theory, sufficient decisions about content words are made so that they can be ordered (or misordered), and the syllables marked for accent mapped onto the higher branches of the rhythmic tree. Function words, etc., lower down on the tree are organized subsequently. Thus reversals occur between elements at the same level, and reversals between a content and function word would not be expected (unless the function word was of the often-accented dangling participle type, e.g., *He called her UP*). Other data provide a bit more detailed support for the top-down generation hypothesis. Boomer and Laver's (1968) analysis of slips of the tongue in tape-recorded speech showed that reversals involved a pair of accented syllables most often, a pair of unaccented syllables less often, but rarely an accented and unaccented syllable. Alternatives to the top-down hierarchical generation hypothesis cannot be ruled out completely, but if they are to be viable they should be expected to take into account the priorities that accented syllables appear to have in production.

Finally, Goodglass, Fodor, and Schulhoff (1967) reported that certain aphasics had difficulty imitating unaccented words in three-word phrases when the unaccented words were phrase initial (*do BIRDS FLY*) but not when they were phrase medial (*DOGS can BARK*). They concluded that prosodic rather than grammatical characteristics of function words determined whether they were lost or retained in ungrammatical speech and further that these aphasics depended upon accented features of an intended utterance for initiating and maintaining speech. These data and conclusions, which imply "momentum" in speaking, are consistent with the present view that accented syllables are the articulatory targets of ballistic movements, and that intervening unaccented syllables are produced by "secondary" articulatory gestures "en route" to the target syllables.

Syntax and morphology in rhythmic patterns. Probably all languages provide for the alternation of accented and unaccented syllables in their morphological and syntactic structure, as in English in which "function" words and affixes intervene between the ordinarily accented words and syllables. To the extent that syntactic and other information is encoded at the rhythmic level, such information becomes available for perception independently of analysis at the segmental level. While it is too early to say whether highly abstract syntactic relations will be found to correlate with the prosodic contours of a rhythmic pattern, there are many perceptual decisions about syntax, morphology, and meaning that can be made on the basis of syllabic contour alone. For instance, the names *Mary* and *Marie* in Figures 9b and 9c could be distinguished from each other in the absence of clear segmental detail. Recognition that *John was hit by Bill* is a passive, not an active, sentence requires no decisions concerning the phonetic details in the function words. The syllabic contour of *John flew to WxYz,* in which *W* and *Y* are accented syllables, and *x* and *z* are unaccented syllables, rules out Boston and New York as possible destinations, as well as Minneapolis, assuming the latter is correctly pronounced, etc.

While these are simple examples, there is a good deal of evidence that rhythmic patterning carries a heavy information load in ordinary connected speech. This is particularly evident in cases where speech remains intelligible when segmental information has been distorted or destroyed. Most of the following studies did not refer to rhythmic patterning, and none gave explicit patterns, but the nature of the distortions in each of them was such as to leave the temporal, hence rhythmic, patterning of the signal intact. One obvious example of this from everyday experience is whispered speech, a distortion affecting mainly the frequency and amplitude but not time dimensions of the signal. The fact that whisperers apparently compensate by exaggerating stress (Fonagy, 1968) shows the effectiveness of rhythmic factors in communication.

Cherry and Wiley (1967) and Holloway (1970) gated "weakly-voiced segments"

(Holloway, 1970) out of running speech, replaced these destroyed segments with white noise, and found that the speech again became intelligible. Both papers proposed rhythmic continuity as one aspect of the signal left unaffected although, clearly, segmental information had been removed. While the result is hardly surprising since the missing segments are just those that would be masked in noisy environments, it does show the contribution of prosodic features to intelligibility.

Of the various distortions that nevertheless do not destroy speech intelligibility (Licklider & Miller, 1951), one of the most severe is infinite peak clipping. Two musicians in the author's laboratory transcribed in rhythmic notation a passage of running speech distorted in this fashion.[12] Their rhythmic patterns agreed strongly with each other and with a subsequent transcription of the corresponding normal speech made later by one of them. With unlimited time and repeated listening, they made few errors of word identification, but when they did the errors were usually rhythmically identical to the original words. Lindner (cited in Bondarko, Zagorujko, Kozhevnikov, Molchanov, & Chistovich, 1970) has reported that spectrally distorted speech also retains its rhythmic pattern.

Blesser (1969) transformed the speech signal by rotating spectral energy around a central frequency of 1,600 Hertz [Hz.] so that, for example, energy at 2,000 Hz. became energy at 1,200 Hz. While this transformation distorts segmental information, temporal patterning and fundamental pitch contour remain unchanged. Pairs of subjects in isolated rooms learned to converse with each other rather successfully, hearing only the transformed speech of the other subject. It is tempting to suppose that their communicative success depended upon learning the transformation, but this was not Blesser's conclusion. Among his findings were that intelligibility scores on isolated syllables, words, and sentences did not correlate with communicative success, and that misheard sentences were often syntactically (and rhythmically) nearly identical to the input sentences but otherwise completely incorrect (e.g., *Hoist the load to your left shoulder* was reported as *Turn the page to the next lesson*). Among the conclusions Blesser drew from his study was that syntax was encoded into prosodic features, but more generally that intensive analysis of phonetic details would not solve the problems of speech segmentation and recognition. Studies like these appear to show that there is extensive information available in speech at the rhythmic level that can be used by the listener. They do not show, of course, that he ordinarily does or must use the information under ordinary listening conditions, but it does not seem wise to conclude that the information is used only as a last resort. Future research will determine the levels of complexity and detail of the syntactic and other information that are or can be encoded into the rhythmic contour of the speech signal. Conceivably some languages may make more informative use of rhythmic constraints on the sound patterning of speech than others.

Rhythmic deployment of phonetic detail. The hypothesis that accented syllables are the main targets of ballistic movements during the course of production suggests one possible consequence that also has implications for perception, namely, that accented syllables are articulated at the expense of lower-order intervening syllables in fast speech, or perhaps even normally. The more syllables intervening (subdividing) between accented syllables then, the greater the extent that phonetic detail on these syllables will be blurred. In music, a situation analogous in some respects exists when the performer, encountering a passage too fast to articulate, takes pains to get the notes on the beat and simple multiples thereof, slurring over the unaccented intervening notes, or even omitting them completely. The melodic outline usually is preserved nonetheless, and the slurred or omitted notes may even escape notice. In the case of speech, it is known that formants of weakly stressed vowels depart from "target" values, approaching a neutral value (schwa) during fast speech (Lehiste, 1970; Lindblom, 1963). These

[12] Gratitude is expressed to John Boehm, who made the recording.

variants on pronunciation contribute to the well-known enormous variability of the acoustic correlates of a given phoneme (MacNeilage, 1970) in ordinary connected speech. But schwa is uninformative concerning the phoneme "intended" by the speaker, and presumably many other vowel pronuciations are also, the more so that they approach schwa. These variants on the acoustic correlates of phonemes complicate the perceptual problem for any mechanism that approaches the signal as though every interval in the signal potentially contains useful segmental information, and hence that every interval must be taken account of to determine whether useful information is present or not. There is no evidence, however, that speech perception proceeds this way. The alternative view proposed here is that the perceptual analysis is optimized by prior information as to the more informative stretches of the signal, for example, on-target vowels plus preceding consonants, which appear to be programmed with them (Tatham, 1970). On the present rhythmic hypothesis these informative stretches, namely, accented syllables, are systematically related to the temporal dimension of the signal. Analysis of the utterance would consist in determining the rhythmic pattern of the syllable string by means of a wholistic (Neisser, 1967) "preliminary analysis" (Halle & Stevens, 1964), then analyzing for segmental detail by reading down the rhythmic tree, so that the more veridical stretches of the signal are analyzed first, with decisions about the lower-level details made conditional upon the results of the higher-level analysis. Such a routine is made possible by the fact that the rhythmic patterning can be extracted from the signal independently of a segmental analysis.

This hypothesis is made plausible by evidence that speech is rhythmically patterned, but it would be strengthened by evidence that accented syllables (and segments therein) are easier to recognize than unaccented syllables (and segments therein) when these two kinds of syllables are removed from linguistic context. In a study using such an approach Lieberman (1963) placed experimental words in two sentences each such that the word was informative (presumably accented) in one sentence or redundant (presumably unaccented) in the other, and gated the words out of the running speech context. Finding that the excised informative words were recognized more easily than the redundant words in listening tests, he concluded that they were more clearly articulated than the redundant words.

Another way of excluding linguistic context from elements to be recognized is to use listeners ignorant of the language. In the language study mentioned earlier, the pair of musicians who transcribed the Italian speech in rhythmic notation were also required to first transcribe the sounds even though they did not know Italian. Analysis of the transcriptions showed more agreement on accented than on unaccented vowels, as they were recorded in English alphabetic notation. These listeners of course were locating the sounds in coarse categories within the perceptual space of their own language. Presumably, phoneticians would often report schwa in the unaccented syllables, hence agreeing more often on these syllables, whereas the musicians, who knew nothing of phonetics, drew their guesses from a pool of English vowels. The relevant point here is that differential agreement appears to be based upon information tied to the rhythmic patterning of the utterances. These results and Lieberman's (1963) seem to be consistent with the hypothesis that the deployment of usable segmental information in the speech signal is temporally systematic. If the hypothesis is correct, future research may determine if and how the temporal information is ordinarily used in perceiving the speech.

Dynamic coupling of speaking and listening in real time. The hypothesis of the preceding section, which suggests a two-stage analysis with the segmental stage depending upon the preceding prosodic stage, bears some resemblances to the analysis-by-synthesis model of speech perception (Halle & Stevens, 1964), as implied above. The analysis-by-synthesis model postulates a preliminary analysis of the input signal prior to an (active) analysis-by-synthesis matching routine. The input signal which is the object of the preliminary analysis has not usually been specified, but could be the rhythmic pattern as implied above. Of particular relevance here is that whatever the size of the input signal, or "decision unit" (Miller, 1962), a sequential stage model appears to imply that the last stage of the analysis, and possibly even the preceding one, could not proceed until the end of the unit had been reached, since some terminal cue would seem to be required as the sign to start analyzing. In running speech this routine thus appears to amount to a continuous, repeating sequence of (*a*) accept input acoustic signal unit, (*b*) apply (first stage) preliminary analysis, (*c*) apply (second stage) matching or other routine, (*a*′) accept next input acoustic signal unit, (*b*′) . . . , etc. Even if Stages *a* and *b* overlapped in time it would still be true that the smaller the decision unit (e.g., syllable) the more rapidly and often the analytic stages would be required to alternate, whereas the larger the unit (e.g., rhythmic pattern), the faster the last stage would be required to operate. An alternative possibility, which would permit a large unit yet more leisurely analysis, is to suggest that Stage *a* above proceeds independently of and prior to Stage *b*, and that Stages *b* and *c* operate sequentially with respect to each other but simultaneously with the *following* Stage *a*′. That is, the two-stage analyses of a preceding pattern are distributed over the time during which a following pattern is simultaneously being accepted into acoustic memory. This notion, however, seems counterintuitive since we appear to understand speech as we hear it.

The rhythmic point of view provides a somewhat different alternative, which may be explicated as follows. As in the hypothesis of the preceding section, a prosodic and segmental analysis is assumed. Inherent in the rhythmic concept is that the perception of early events in a sequence generates expectancies concerning later events, in real time. When the events are sounds produced by continuous movements, the perception of these includes cues as to the movement dynamics involved in their production. Hence it is not simply, or not only, that discrete arrival times of accented syllables are induced from earlier timing relationships but also that the total array of time-varying cues in the continuous flow of speech will project ahead the general outline of the remaining prosodic contour. These cues telegraph not only tempo changes but more generally the whole thrust of the pattern of sounds yet to come. It is on this basis that one might say not that the listener "follows" the speaker but rather that the listener, given initial cues, actively enters into the speaker's tempo. Whereas the analysis-by-synthesis model is a feedback model (Halle & Stevens, 1964), the rhythmic model presented here appears to have also a feed-*forward* aspect (MacKay, 1967), in the sense that it adjusts ongoing processing based upon information about future, as opposed to present or past, states of the signal.

The consequences for running speech perception in this view are that temporal redundancy at the prosodic level permits a prosodic analysis to begin prior to the end of the decision unit. This in turn might even permit an earlier start toward segmental analysis, since the latter depends upon the results of the prosodic analysis. The temporal redundancy also permits accented syllables to be anticipated in real time, so that "attention focusing" on accented syllables becomes a logical possibility, as does "attention cycling," such that previous inputs could be processed during the low-information intervals between accented syllables. Real-time mechanisms like these appear to be of the sort needed to account for the fact that speech processing is so rapid (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

There is at present very little evidence in the literature that would bear on these considerations, although two studies can be mentioned. These concern the role of accented syllables in perception. Allen (1967) reported that tapping variances were smaller to accented syllables than to unaccented syllables when the subjects were required to tap in time with a single specified syllable in utterances presented repeatedly by means of a tape loop. Since it would seem that the arrival times for both kinds of target syllables were equally predictable in the sense that the subject had heard the utterance many times and knew "where" the syllable was in the utterance, the results may have to do with the ease with which the listener can coordinate his activities with those of the speaker. In the present view, the listener taps more accurately when his tapping targets are the same as the speaker's speaking targets, that is, accented syllables. The report by Huggins (1972b) mentioned earlier, which showed that listeners were insensitive to certain artificial distortions in speech when the original timing relationships between accented syllables were preserved, is also consistent with the view presented here. Further research may determine the extent to which efficient perceptual processing involves actively locking into the temporal flow of the speech signal.

Applications to Automatic Speech Synthesis and Recognition

These research areas are concerned either with natural running speech as inputs or with natural sounding speech as outputs. Here some suggestions are offered as to ways in which rhythmic structure considerations might be relevant.

Speech synthesis by rule. At the present time, speech synthesis programs typically concatenate segments whose durations are determined by dictionary look-up, with adjustments in duration based upon immediate segment context (see, e.g., Flanagan, Coker, Rabiner, Schafer, & Umeda, 1970; Mattingly, 1971; see also Denes, 1970; Mattingly, 1966; Rabiner, 1969; Rabiner, Levitt, & Rosenberg, 1969; Teranishi & Umeda, 1968). In the present theory, the temporal patterning of utterances constructed this way can be expected to deviate from those of natural speech. Additionally, since dictionary look-up durations presumably will always be of sufficient length so as to permit some phonetically differentiating detail, these details will be distributed relatively evenly throughout the utterance so that the resulting utterance contains more phonetic detail, or at least more evenly distributed detail, than the same utterance in natural speech. The requirements of relative timing in natural speech, on the other hand, often will obliterate phonetic detail when unaccented syllables are crowded together in utterance-specific rhythmic contexts. Possibly such deviations from natural speech account in part for the reported trade-off between naturalness and intelligibility in synthesized speech (Mattingly, 1971).

In principle at least, these speculations could be rather directly tested by experiments on synthetic speech. The general approach would be to mark syllables to be accented in the syllable string and then apply the rhythm rule to the string. Relative syllable durations would thus be determined, and hence also the phonetic details affected by duration. Whether such rhythmic constraints result in synthesized speech which is easier to listen to over long periods of time or is of better judged quality is then an empirical question to be settled by listening tests of various kinds, including tests that would be sensitive to long-term fatigue effects, etc., hence to the "cost" of processing distorted inputs.

Automatic speech recognition. Automatically synthesized speech can be less than perfect yet remain intelligible because of the redundancy in speech and the perceiver's ability to compensate for the inadequacies of the signal. It is the latter that allows judgments to be made about an inadequate input to the listener, so that trial-and-error procedures can be used to evaluate progress toward better quality synthesis. Automatic speech recognition, on the other hand, is seen to be particularly difficult because speech recognition by humans depends so little upon information in the acoustic signal and so heavily upon knowledge of the language and of the world not possessed by computers (Pierce, 1969). Nevertheless, the rhythmic hypotheses considered earlier suggest an approach that in principle at least might make it possible to extract more information from the acoustic signal than is now the case, and additionally, bring the perceiver's abilities into the automatic recognition routine. This approach involves both the analysis of the signal and the resulting printout. The first stage of analysis would consist in determining the syllabic pattern, by locating vowels or vowellike segments.

The second stage would consist in extracting the more veridical segmental information by means of a top-down analysis as suggested earlier. The printout would reflect the two-stage nature of this analysis by providing a symbol sequence such that the syllabic pattern is recognizable regardless of the accuracy of identification of any of the segments therein. Identifiable segments (those exceeding some cutoff decision criterion) would be represented by appropriate phonetic symbols. Unidentifiable vowellike segments might be represented by schwa; unidentifiable consonantlike segments might be represented by some neutral symbol. Such an output provides a sequence in which the general syllabic pattern may be made out although many of the details are blurred. But such a sequence, like poor ditto copy or poor handwriting, might be more or less readable, since strategically located and correctly identified consonants might cue whole words. Thus by bringing the perceiver's (reader's) knowledge of the language and of the world to bear on evaluation of the printed output, trial-and-error procedures become available to the speech recognition worker that are quite analogous to those enjoyed by those working in speech synthesis. And, possibly, the computer recognizing speech might require little more knowledge of the language and of the world than that required by the computers assigned to synthesize speech.

FURTHER, FINAL SPECULATIONS CONCERNING OTHER RESEARCH AREAS

Rhythmic structure considerations are potentially relevant in a variety of research areas. In this concluding section, several of these areas are mentioned. In some cases the relevance of rhythmic considerations is fairly clear, but in others it can be only highly speculative.

Immediate Memory

Work in this research area has depended heavily upon the use of digits, letters, and other spoken stimulus inputs. Adams investigated rhythmic factors in immediate recall of digit strings as early as 1915, but there has been relatively little interest in the rhythmic aspects of the speech stimulus in memory experiments until quite recently (e.g., Neisser, 1967).

Most contemporary experiments on the role of temporal factors in immediate memory have tended to focus on the grouping aspect of rhythmic sequences, using stimulus inputs designed to provide grouping by the insertion of pauses into the sequences. The effect of pauses has been seen as that of providing perceptual groupings and of providing the opportunity for coding, rehearsal, etc., during the pause interval (e.g., Bower & Springston, 1970; Laughery & Spector, 1972). The present view would go further than this, to assume that speaking digits and other verbal input strings requires an organized articulatory program just as does normal speaking and hence produces a structured stimulus to which the listener responds in the manner discussed earlier. Many elements of this structure persist (it is assumed) even when the speaker tries to speak without intonation. Factors like these are often irrelevant to the purposes of the memory experiment, but they will nevertheless play a role in ongoing processing of the stimulus string. Sometimes these factors will be particularly important, for example, when group sizes vary such that the subject does not know in advance how many items he will hear in a group (e.g., Bower & Winzenz, 1969; Laughery & Spector, 1972), in which case prosodic cues on group-initial elements could project ahead the general prosodic contour of the group, which includes cues as to the number of items it contains. The contribution of perceptual factors like these to memory experiments has not been systematically assessed, but there is evidence that they exist, as the following experiment shows.

Neisser, Hoenig, and Goldstein (1969) examined the "prefix effect" from a point of view quite similar to that of the present paper. Earlier, Dallett (1964), following Conrad's (1958, 1960) work, had shown that a redundant stimulus prefix spoken at the beginning of a seven-digit string reduced recall of the digits, even though the subject was not required to repeat the redundant digit, as compared to the condition in which the spoken redundant digit did not precede the string. Neisser et al. (1969) showed that the prefix effect disappeared when the redundant digit was uttered by a different speaker than the one speaking the digits to be recalled, so that the prefix could not be a part of the same production program. It appears that the listener actively responded to the whole spoken pattern, which, as the experimenters pointed out, required him to handle an eight-digit string at input and suppress the prefix before speaking the recall digits back. Hence recall of prefixed seven-digit strings yielded about the same percentage recall as eight-digit strings. This experiment seems to show that intonation contour or other aspects of the speech stimulus can affect recall, but it does not make clear which details are important. Probably a systematic analysis will require comparisons between normally spoken speech strings and other strings with elements timed the same way but with articulatory continuity removed by tape splicing or synthetic methods. It may even be possible to demonstrate that natural articulatory grouping can, under certain conditions, reduce or even eliminate the conventional advantage of pauses within input strings to be recalled (Martin, 1969a).

Serial Pattern Learning

Restle (1972) investigated the learning of serial patterns which, in contrast to materials used in the section above, have an inherent structure as described by structural trees (Restle, 1970; Restle & Brown, 1970). Pauses introduced during presentation faciliated pattern learning, as is typically the case in serial learning, but only when the pauses were appropriate for the internal structure of the pattern; that is, they separated subunits. It is a reasonable guess that learning of these meaningful sequences might be further facilitated if the timing of elements in a sequence during input not only marked subunits but also was temporally organized around an accent structure following the accent rule. It should also be noted that Restle's (1970) theory has been shown to be related to aspects of music structure.

Auditory Temporal Pattern Perception

In the present theory, natural auditory temporal patterns, speech or otherwise, are sound sequences with internal structure as described by the rhythm rules and with a wholistic unity. Hearing early elements in a pattern generates expectancies concerning later elements, in particular, the temporal location of accents. For instance, accents on elements one and three and/or five in an eight-element evenly spaced pattern follow the accent rule, whereas accents on elements one and four do not follow the rule and lead to expectancies of triplets, which are appropriate for nine-element patterns but inappropriate for eight-element patterns. Expectancies that are confirmed should make any kind of perceptual processing easier, whereas expectancies that are violated should make processing harder.

Most of the extensive work by Garner and his associates (Garner, 1970a; Garner & Gottwald, 1968; Preusser, Garner, & Gottwald, 1970; Royer & Garner, 1966, 1970) on auditory temporal patterns does not bear directly on this hypothesis since their methods were directed toward determining preferred organizations. But one of their findings was suggestive for the present theory. In these experiments, after hearing several continuous repetitions of a pattern, the subjects, whose task was always to report the pattern correctly, generally preferred to describe or track the patterns by placing runs either at the beginning or end of the pattern, regardless of the starting point of the pattern as it was presented to them. These were their preferred organizations of the pattern. Of interest here, however, is the case where there were no runs of substantial length, so that no run preference could apply.

In Royer and Garner's (1966) experiment, there were no runs longer than two elements in patterns R, S, and T. Their data show that for each of these patterns, response delays were less, on the average, when the starting point presented to the subjects followed the accent rule (i.e., placed accents on elements one and three) than when they did not follow the rule (i.e., placed accents on elements one and four). It appeared that when the patterns did not have the characteristics that permitted preferred organizations (two of these three patterns were the most difficult in the experiment), then the effects of starting point made a difference in the time required for the subject to confidently begin tracking the pattern. This interpretation is of course highly speculative, and there may be other explanations, but it is supported by evidence from an experiment designed explicitly to test this interpretation. Sturges and Martin (1971) presented patterns continuously but only twice; the second pattern was either exactly the same (repeated) or changed slightly. The tasks were to judge same or different and then write the pattern(s). Repeating patterns that followed the accent rule were more easily recognized as repeating than patterns that did not follow this rule. Similar results held for written recall.

Language Acquisition

Prosodic features have been a traditional concern in studies of child language. For instance, children appear to learn intonation patterns before they learn segmental features (Weir, 1966). Surely the rhythmic characteristics of speech will make a difference in the way sentences are heard by children. Take a sentence like *The girl the boy saw went home.* This sentence sounds equally natural if *boy* is more strongly accented than *saw* or vice versa; yet the two pronunciations could lead the child to give different imitations or paraphrases, which in turn would lead the developmental psycholinguist to different conclusions concerning the comprehension of embedded sentences. More generally, it would not be surprising if a great deal of child language data that seem refractory on syntactic grounds would make sense if considered in terms of real-time phonetic considerations.

Meaning

Rhythmic patterning applies to sounds, but is there anything to be said about semantics? Kozhevnikov and Chistovich (1965) have pointed out that the syntagma, their articulatory unit, has been defined also as a unit of meaning. McNeill (1971), who, incidentally, also cited Lashley's (1951) paper on serial order, has advanced the view that speech units contained within certain intonation contours are underlying structure sentences. Perhaps future work will show that ideas and meanings are efficiently encoded into natural rhythms. There is an old tune, from the 1930s or so, called "It Don't MEAN a Thing if it Ain't Got That SWING." It refers, of course, to the rhythms of music, but there may be a sense in which it is true for the rhythms of language also.

REFERENCES

ABERCROMBIE, D. *Elements of general phonetics.* Chicago: Aldine, 1967.

ADAMS, H. F. A note on the effect of rhythm on memory. *Psychological Review,* 1915, 22, 289–298.

ADAMS, J. A. Response feedback and learning. *Psychological Bulletin,* 1968, **70**, 486–504.

ALLEN, G. D. Two behavioral experiments on the location of the syllable beat in spoken American English. *Studies in Language and Language Behavior,* 1967, **4**, 2–179.

ALLEN, G. D. The place of rhythm in a theory of language. (Working Papers in Phonetics, No. 11). Los Angeles: University of California, 1968. (a)

ALLEN, G. D. Towards a description of stress-timing in spoken English. In E. Zale (Ed.), *Proceedings of the Conference on Language and Language behavior.* Ann Arbor: University of Michigan, 1968. (b)

BIERWISCH, M. Two critical problems in accent rules. *Journal of Linguistics,* 1968, **4**, 173–176.

BLESSER, B. A. Perception of spectrally rotated speech. Unpublished doctoral thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, 1969.

BOLINGER, D. L. Pitch accent and sentence rhythm. In I. Abe & T. Kanekiyo (Eds.), *Forms of English.* Cambridge: Harvard Press, 1965.

BONDARKO, L. V., ZAGORUJKO, N. G., KOZHEVNIKOV, V. A., MOLCHANOV, A. P., & CHISTOVICH, L. A. A model of speech perception by humans. (Trans. by Ilse Lehiste) (Working Papers in Linguistics, No. 6) Columbus: Ohio State University, 1970.

BOOMER, D. S. Hesitation and grammatical encoding. *Language and Speech,* 1965, **8**, 148–158.

BOOMER, D. S., & LAVER, J. D. M. Slips of the tongue. *The British Journal of Disorders of Communication,* 1968, **3**, 2–12.

BOWER, G. H., & SPRINGSTON, F. Pauses as recoding points in letter series. *Journal of Experimental Psychology,* 1970, **83**, 421–430.

BOWER, G. H., & WINZENZ, D. Group structure, coding, and memory for digit series. *Journal of Experimental Psychology,* 1969, **80** (2, Pt. 2).

BRESNAN, J. W. Sentence stress and syntactic transformations. *Language,* 1971, **47**, 257–281.

CHERRY, C., & WILEY, R. Speech communication in very noisy environments. *Nature,* 1967, **214**, 1164.

CHOMSKY, N. *Syntactic structures.* The Hague: Mouton, 1957.

CHOMSKY, N. A review of B. F. Skinner's *Verbal behavior. Language,* 1959, **35**, 26–58.

CHOMSKY, N., & HALLE, M. *The sound pattern of English.* New York: Harper, 1968.

CONRAD, R. Accuracy of using keyset and telephone dial, and the effects of a prefix digit. *Journal of Applied Psychology,* 1958, **42**, 285–288.

CONRAD, R. Very brief delay of immediate recall. *Quarterly Journal of Experimental Psychology,* 1960, **12**, 45–47.

COOPER, G. W., & MEYER, L. B. *The rhythmic structure of music.* Chicago: University of Chicago, 1960.

CUSHING, S. English as a tone language: The acoustics of primary stress. (Quarterly Progress Rep. No. 92) Cambridge, Mass.: M.I.T. Research Laboratory for Electronics, 1969.

DALLETT, K. M. Effects of a supplementary prefix on recall. *Journal of Experimental Psychology,* 1964, **67**, 296–298.

DELATTRE, P. *Comparing the phonetic features of English, French, German and Spanish.* Heidelberg: Verlag, 1965.

DENES, P. B. Some experiments with computerized speech. *Behavioral Research Methods and Instrumentation,* 1970, **2**, 1–5.

FLANAGAN, J. L., COKER, C. H., RABINER, L. R., SHAFER, R. W., & UMEDA, N. Synthetic voices for computers. *IEEE Spectrum,* 1970, **7**, 22–45.

FONAGY, I. Electro-physiological and acoustic correlates of stress and stress perception. *Journal of Speech and Hearing Research,* 1966, **9**, 213–244.

FONAGY, I. Accent and intonation in whispered speech. *Phonetics,* 1968, **20**, 177–192.

FROMKIN, V. A. The non-anomalous nature of anomalous utterances. *Language,* 1971, **47**, 27–52.

FRY, D. Prosodic phenomena. In B. Malmberg (Ed.), *Manual of phonetics.* Amsterdam: North-Holland, 1968.

GARNER, W. R. Good patterns have few alternatives. *American Scientist,* 1970, **58**, 34–42. (a)

GARNER, W. R. The stimulus in information processing. *American Psychologist,* 1970, **25**, 350–358. (b)

GARNER, W. R., & GOTTWALD, R. L. The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology,* 1968, **20**, 97–109.

GOODGLASS, H., FODOR, I. G., & SCHULHOFF, C. Prosodic factors in grammar—Evidence from aphasia. *Journal of Speech and Hearing Research,* 1967, **10**, 5–20.

GREENWALD, A. G. Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review,* 1970, **77**, 73–99.

HALLE, M., & STEVENS, K. S. Speech recognition: A model and a program for research. In J. A. Fodor & J. J. Katz (Eds.), *The structure of language: Readings in the philosophy of language.* Englewood Cliffs, N. J.: Prentice-Hall, 1964.

HALLIDAY, M. A. K. The tones of English. *Archivum Linguisticum,* 1963, **15**, 1–28.

HALWES, T., & JENKINS, J. J. Problem of serial order in behavior is not resolved by context-sensitive associative memory models. *Psychological Review,* 1971, **78**, 122–129.

HARRIS, J. W. *Spanish phonology.* Cambridge: M.I.T. Press, 1969.

HOLLOWAY, C. M. Passing the strongly voiced components in noisy speech. *Nature,* 1970, **226**, 178–179.

HUGGINS, A. W. F. Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America,* 1972, **51**, 1270–1278. (a)

HUGGINS, A. W. F. On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America,* 1972, **51**, 1279–1290. (b)

KEELE, S. W. Movement control in skilled motor performance. *Psychological Bulletin,* 1968, **70**, 387–403.

KOZHEVNIKOV, V. A., & CHISTOVICH, L. A. *Speech: Articulation and perception.* Washington, D. C.: Joint Publications Research Service, 1965.

LADEFOGED, P. *Three areas of experimental phonetics.* London: Oxford University Press, 1967. (a)

LADEFOGED, P. *Linguistic phonetics.* (Working Papers in Phonetics, **6**, Phonetics Laboratory). Los Angeles: University of California, 1967. (b)

LANIER, S. *The science of English verse.* New York: Scribner, 1880.

LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951.

LAUGHERY, K. R., & SPECTOR, A. The roles of recoding and rhythm in memory organization. *Journal of Experimental Psychology,* 1972, **94**, 41–48.

LEHISTE, I. *Suprasegmentals.* Cambridge: M.I.T. Press, 1970.

LENNEBERG, E. H. *Biological foundations of language.* New York: Wiley, 1967.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review,* 1967, **74**, 431–461.

LICKLIDER, J. C. R., & MILLER, G. A. The perception of speech. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

LIEBERMAN, P. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech,* 1963, **6**, 172–179.

LIEBERMAN, P. On the acoustic basis of the perception of intonation by linguists. *Word,* 1965, **21,** 40–54.

LIEBERMAN, P. *Intonation, perception, and language.* Cambridge: M.I.T. Press, 1967.

LINDBLOM, B. E. F. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America,* 1963, **35,** 1773–1781.

MACKAY, D. M. Ways of looking at perception. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form.* Cambridge: M.I.T. Press, 1967.

MACNEILAGE, P. F. Motor control of serial ordering of speech. *Psychological Review,* 1970, **77,** 182–196.

MARTIN, J. G. Rhythm vs. grouping in memory for digits. Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1969. (a)

MARTIN, J. G. Temporal structure and the perception of speech. Paper presented at the meeting of the Midwestern Psychological Association Chicago, May 1969. (b)

MARTIN, J. G. Rhythm-induced judgments of word stress in sentences. *Journal of Verbal Learning and Verbal Behavior,* 1970, **9,** 627–633. (a)

MARTIN, J. G. The rhythmic structure of speech, music and other meaningful sounds. Paper presented at the meeting of the Psychonomic Society, San Antonio, November 1970. (b)

MATTINGLY, I. G. Synthesis by rule as a tool for phonological research. *Language and Speech,* 1971, **14,** 47–56.

MATTINGLY, I. G. Synthesis by rule of prosodic features. *Language and Speech,* 1966, **9,** 1.

MCNEILL, D. Sentences as biological processes. Paper presented to the CNRS Conference on Psycholinguistics, 1971.

MILLER, G. A. Decision units in the perception of speech. *IRE Transactions on Information Theory,* 1962, **8,** 81–83.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

MILNER, P. *Physiological psychology.* New York: Holt, Rinehart & Winston, 1970.

NEISSER, U. *Cognitive psychology.* New York: Century, 1967.

NEISSER, U., HOENIG, Y. J., & GOLDSTEIN, E. Perceptual organization in the prefix effect. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8,** 424–429.

NEWMAN, S. E. On the stress system of English. *Word,* 1946, **2,** 171–187.

PIERCE, J. R. Whither speech recognition? *Journal of the Acoustical Society of America,* 1969, **46,** 1049–1051.

PIKE, K. L. *Intonation of American English.* Ann Arbor: University of Michigan, 1945.

PREUSSER, D., GARNER, W. R., & GOTTWALD, R. L. Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *American Journal of Psychology,* 1970, **83,** 151–170.

RABINER, L. R. A model for synthesizing speech by rule. *IEEE Transactions on Audio and Electroacoustics,* 1969, **AU-17,** 7–13.

RABINER, L. R., LEVITT, H., & ROSENBERG, A. E. Investigation of stress patterns for speech synthesis by rule. *Journal of the Acoustical Society of America,* 1969, **45,** 92–101.

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

RESTLE, F. Serial patterns: The role of phrasing. *Journal of Experimental Psychology,* 1972, **92,** 385–390.

RESTLE, F., & BROWN, E. Organization of serial pattern learning. In G. H. Bower (Ed.), *Psychology of learning and motivation.* New York: Academic Press, 1970

ROYER, F. L., & GARNER, W. R. Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception and Psychophysics,* 1966, **1,** 41–47.

ROYER, F. L., & GARNER, W. R. Perceptual organization of nine-element auditory temporal patterns. *Perception and Psychophysics,* 1970, **7,** 115–120.

SCHANE, S. A. *French phonology and morphology.* Cambridge: M.I.T. Press, 1968.

STURGES, P. T., & MARTIN, J. G. Rhythmic structure in auditory temporal patterns. Paper presented at the meeting of the Psychonomic Society St. Louis, November 1971.

TATHAM, M. A. A. A speech production model for synthesis-by-rule. (Working Papers in Linguistics, No. 6) Columbus: Ohio State University, 1970.

TERANISHI, R., & UMEDA, N. Use of pronouncing dictionary in speech synthesis experiments. *Reports of the Sixth International Conference on Acoustics,* 1968, 2, 155–158.

TRAGER, G. L., & SMITH, H. L. *Outline of English structure.* (Studies in Linguistics. Occasional Paper No. 3) Norman, Okla.: Battenburg Press, 1951.

VANDERSLICE, R. Synthetic elocution. (Working Papers in Phonetics, No. 8) Los Angeles: University of California, 1968.

VANDERSLICE, R. Occam's razor and the so-called stress cycle. *Language Sciences,* 1970, **13,** 9–15.

WEIR, R. H. Some questions on the child's learning of phonology. In F. Smith & G. A. Miller (Eds.), *The genesis of language.* Cambridge: M.I.T. Press, 1966.

WICKELGREN, W. A. Context sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review,* 1969, **76,** 1–15.

WIERSMA, C. A. G. (Ed.) *Invertebrate nervous systems.* Chicago: University of Chicago Press, 1967.

WOODROW, H. S. A quantitative study of rhythm. *Archives of Psychology,* 1909, **18,** No. 1.

WOODROW, H. S. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

(Received March 23, 1972)