5
Temporal Organization of Perception and Action

> The production of rhythm in general is made possible by the preparation of movements and their coordination. . . . Without preparation, performance is a series of discrete reactions to external events. The development of motor skill can be traced as the progress from reactive movement to movement fluency, coupled with a flexibility in tailoring action to the details of an infinite variety of contingencies.
>
> (Shaffer, 1982, p. 110)

Theories of perception and action must account for three basic aspects of skilled behavior, and so far, the book has examined only the first two: What components underlie the organization of everyday action and perception, and what mechanisms activate these components in proper sequence?

I turn now to the third and most frequently neglected theoretical problem, that is, timing in perception and action. How do we produce behaviors with components of different durations, and how do we produce these behaviors at different rates and with different rhythms or patterns of durations? I begin with some general phenomena or constraints that any theory of timing must address, and then construct a node structure theory of timing that incorporates these general constraints and makes predictions for future test.

Requirements for a Viable Theory of Timing

The preceding chapter already has noted two important requirements for theories of timing: independence of mechanisms for timing and sequencing and a closer relationship between mechanisms for timing and sequencing than between mechanisms for timing and representing the form or content of behavior. In addition, theories of timing must address six other fundamental issues discussed in this chapter: Where in the specification of output components is rate and timing determined? What mechanisms underlie the production of rhythmic outputs? How and why is periodicity or near-miss periodicity achieved in complex skills such as typing? What is the relationship between the timing mechanisms for

The temporal organization of perceptual and action. Ch 5 (pp. 90-110) in MacKay, D.G. (1987). ***The organization of perception and action: A theory for language and other cognitive skills*** **(1-254). Berlin: Springer-Verlag.**

perception and action? What accounts for our ability to flexibly adjust the rate and timing of behavior? Why and how do actions with different timing characteristics interact with one another?

The Distributed Nature of Timing and Sequencing

The most fundamental constraint on theories of timing is that timing is an "everywhere" or distributed characteristic. Each and every component, from the lowest level nodes controlling muscle movements to the highest level nodes representing, say, sentential concepts, must be activated at some rate and for some duration. Rhythm, rate, and timing permeate the entire process of producing well-practiced behaviors, and cannot be tacked on as an independent stage at some point in the theoretical specification of the output.

To see why timing must be distributed in this sense, it is only necessary to examine existing "stage of processing" proposals (including one of my own), where rhythm and timing are treated as an "afterthought," a late stage of processing introduced just before or during the programming of muscle movements. Consider, for one example, my stage-of-processing theory (D. G. MacKay, 1971) of timing in speech production (for other examples, see Fromkin, 1973; Shaffer, 1982). Under my proposal (D. G. MacKay, 1971), the entire syntax, semantics, and phonology for producing a sentence are first constructed and then recorded within a simultaneous store or nontemporal spatial display called the phonological buffer. Only then are timing and rhythm added as part of the output specifications in the simultaneously displayed phonological string.

Stage-of-processing proposals, such as this one, face many unsolved problems. One is the complexity and reduplication of information that is required for the simultaneous store (phonological buffer). Rhythm and timing depend on information associated with units at every level (sentences, phrases, words, syllables, and segments), and the proposed spatial display must incorporate all of this information before timing specifications can be added. Because these same specifications are also needed in order to construct the sentence in the first place, adding timing after rather than during construction duplicates the specification process.

An even more serious problem for theories that represent timing as a stage of processing is speed–accuracy trade-off, one of the most pervasive phenomena in the study of skilled behavior. For all known skills, increased speed leads to increased errors in activating components, whether lower level muscle movement components, or higher level mental components. The muscle movement level has been the source of most demonstrations and theories of speed–accuracy trade-off (Meyer, Smith, & Wright, 1982), but speed–accuracy trade-off has also been demonstrated for higher level mental components. For example, consider speech errors such as the substitution of "coat-thrutting" for "throat-cutting." Here components within the phonological system have become interchanged (Dell, 1980), and in a study of experimentally induced speech relatively constant. For example, speakers of English produce the phrases "ran lickety split" and "the black horse" in about the same amount of time in the sentence "The black horse ran lickety split across the field," despite the two extra syllables in *lickety* (H. H. Clark & E. V. Clark, 1977; Lehiste, 1970). Considering all sentences of English, this regularity is only approximate (Fowler & Tassinary, 1981), but *producing* stress-timed rhythm could in principle facilitate *perception*, making it possible for listeners to adopt a strategy of listening most carefully and effectively at times when stressed syllables occur. And this strategy would pay off, because stressed syllables generally contain more phonetic and phonological information than unstressed syllables (Huttenlocker & Zue, 1983).

Several sources of evidence indicate that people do, in fact, expect stressed syllables at particular points in the rhythm of a sentence. Huggins (1972) is one example. By shortening one segment in a word and prolonging another, Huggins created sentences containing minor timing irregularities which subjects attempted to detect in a task requiring discrimination between these and other sentences without timing irregularities. The results showed that irregularities were more detectable when they altered the interval between stressed syllables in a sentence than when they left this interval unchanged. Changing the time between stressed syllables apparently violates the timing expectations that listeners impose on an input.

A phoneme monitoring experiment by Shields, McHugh, and Martin (1974) further illustrates the timing expectations associated with stressed syllables. Subjects pressed a key as quickly as possible after hearing the phoneme /b/ in sentences such as "You will have to curtail any morning sightseeing plans, as the plane to BENkik leaves at noon." The target sound, /b/, began a two-syllable nonsense word that received stress on either the first or the second syllable (BENkik versus benKIK) and was detected faster in the stressed syllable. However, the critical factor was the relation of the stressed syllable to the ongoing rhythm of the sentence. No difference between stressed versus unstressed syllables was found when the same words were presented in a list of nonsense syllables without sentential rhythm. People apparently listen for stress-timed rhythm only when processing real sentences.

The importance of stressed syllables in perception of English sentences appears yet again in an experiment by Allen (1972). Subjects repeatedly listened to a recording of a spoken sentence and were then asked to tap a key in time with its rhythm. That subjects were able to respond at all to this instruction is remarkable, and that they responded consistently by tapping in time with the stressed syllables is even more remarkable. It is as if stressed syllables provided the beat for an internal clock that controlled both the finger taps and perception of sentential rhythm.

DISRUPTIVE INTERACTIONS BETWEEN PERCEPTION AND ACTION

Wing (1980) demonstrated a disruptive interaction between perceptual and motoric timing that also calls for an explanation in terms of a common timing mechanism for perception and action. Wing's subjects first learned to generate a motoric rhythm, using the key-press continuation paradigm discussed previously. In this study, however, a brief tone sounded during each press, and unknown to the subjects, the tone generator was programmed to delay or advance the tone by about 10 to 50 ms relative to the normal time when a randomly selected tap was to be made. For delayed tones, subjects inadvertently tended to delay the next one or two taps in the same direction as the tone, but not as much as would be expected if feedback determined the timing of subsequent responses (Keele, 1986), and advanced tones had no significant effect whatsoever. Auditory feedback can therefore influence the internal clock that generates finger-tapping responses but not in the manner of feedback control theory (see also Chapter 10).

The question is whether Wing's (1980) effect is related to feedback per se, or whether *any* auditory input would have a similar effect. This question was answered by Pokorny (1985), who demonstrated a similar interaction between perceptual and motoric timing by simply presenting a tone, unrelated to ongoing feedback, during the time that subjects attempted to generate a series of equal-interval finger taps. Occurrence of the tone tended to increase the intertap interval in which it occurred, even when subjects were instructed to ignore the tones. Pokorny (1985) also noted that the tone interacted with the timing mechanism, rather than with the mechanism for producing any particular response, because subsequent responses occurred at the appropriate times, rather than compensating for the response that had been mistimed. It was as if the timer were a resettable rather than persistent pacemaker. Occurrence of the tone simply reset the timing mechanism, which then continued to generate its preprogrammed equal intervals (see also Keele, 1987).

TEMPORAL COUPLINGS BETWEEN PERCEPTION AND ACTION

Lashley (1951) reported an interesting interaction between input and output rhythms that likewise suggests that perception and action may share some of the same timing mechanisms. Lashley observed temporal couplings of a perceptual rhythm (listening to a band) with ongoing motoric activities such as walking, breathing, and speaking. Specifically, when someone is listening to a salient rhythm such as a marching band, the perceptual rhythm tends to cause the listener to fall in step, gesture, breathe, and even speak in time with the band. Such an interaction suggests that the same timing mechanisms may govern both the perceptual processes underlying listening to the band and the motoric processes underlying walking, breathing, and speaking (see also Keele, 1987).

Learned and Unlearned Aspects of Timing

Some aspects of timing are unlearned, and other aspects are learned, and both must be explained in theories of timing. For example, unlearned peripheral characteristics of the speech organs clearly influence timing. Because the jaw must move farther to produce an open vowel, open vowels have systematically longer durations than closed vowels (Lehiste, 1970). Differences between languages, on the other hand, illustrate a learned aspect of timing. An example is the more varied use of durational information in so-called "syllable-timed" languages such as French as compared with "stress-timed" languages such as English (Cutler, Mehler, Norris, & Segui, 1983).

Constant Relative Timing

As the overall time to produce a behavioral sequence changes due, say, to an increase in rate, the proportion of time required to produce some segment of the sequence remains constant within wide limits. This phenomenon, known as *constant relative timing*, has been observed for a large number of behaviors (e.g., walking, running, typing, handwriting, speech, lever rotation) and can be considered a general law of behavior. As a single example of this general law, consider the work of Shapiro, Zernicke, Gregor, and Diestel (1981). Subjects walked at various speeds on a treadmill, and the proportion of time required to execute the four basic phases of a step (lift, stride, heel contact, and support) was found to remain relatively invariant at the different speeds. If lift phase of a step took up 20% of step duration at a slow rate, it required about 20% at a faster rate. Timing only remained constant within limits, however. When the treadmill was accelerated beyond a certain point, for example, subjects broke into a jog, and the temporal configuration of their behavior suddenly and dramatically changed. Walking and running clearly have different temporal characteristics and are controlled by different underlying mechanisms, which nevertheless *both* conform to the law of constant relative timing. Handwriting and transcription typing also exhibit constant relative timing. Here, changes in overall output rate have been found to scale the duration of response components in almost perfect proportion, as would occur with a change in rate of a low-level internal clock (Shaffer, 1978).

Interestingly, constant relative timing has also been observed for the *involuntary* changes in the rate of behavior that occur as a result of practice. Components in a behavior sequence automatically speed up as a result of practice, and these changes in relative duration sometimes exhibit constant relative timing as well. For example, D. G. MacKay and Bowman (1969) had subjects practice producing a sentence as quickly as possible, and found that they produced the sentence faster after 12 trials of practice than after only one. More importantly, different components of the sentence speeded up proportionally; the relative durations of words and of syllables remained constant at the fastest speed. If a word took up 10% of total sentence duration at the slower, less practiced rate, then it took up about 10% of sentence duration at the faster, more practiced rate (see also D. G. MacKay, 1974). The constant relative timing that occurs when behavior speeds up, either voluntarily or involuntarily (as a result of practice), places fundamental constraints on theories of timing.

Theories postulating an on-line process of calculation for altering durations of behavioral components produced at different rates (e.g., Shapiro et al., 1981) have difficulty explaining constant relative timing, because the phenomenon appears in the behavior of insects and crustaceans, where such calculations are unlikely. Constant relative timing also appears immediately after subjects change their rate, without the lag times that seem necessary for temporal calculations. Rate-dependent changes in timing call for a more automatic mechanism that does not recompute movement time on the basis of rate.

DEVIATIONS FROM CONSTANT RELATIVE TIMING

Constant relative timing cannot be expected for all response components. In particular, not all changes in speech rate can be expected to scale proportionally over the durations of vowels versus consonants. With voluntary changes in speaking rate, vowels exhibit much more "elasticity" than do consonants. Vowels can be prolonged almost indefinitely to slow down the rate of speech, but a greatly prolonged stop consonant no longer resembles speech.

Such observations suggest that two different timing mechanisms may control production of speech sounds, one for consonants, the other for vowels. Consistent with this hypothesis, Tuller, Kelso, and Harris (1982) found that the lag with which a consonant is initiated following a vowel remained constant at different rates of speech, but only when compared to the interval between vowels. Neither consonant nor vowel duration per se remained constant relative to overall utterance duration.

Temporal Interactions

Theories of timing must explain why concurrent actions with different timing characteristics tend to interfere with one another. When unpracticed subjects attempt to produce several actions at once, they experience considerable difficulty when concurrent movements conflict in timing, but little difficulty when concurrent movements are temporally compatible or occur at harmonically related times. A study by Klapp (1979) provides a clear demonstration of this effect of temporal compatibility on the ability to time concurrent activities. Subjects pressed telegraph keys, one for each hand, in time with tones presented periodically to the two ears via headphones. The goal was to maximize temporal overlap of the key press and the tone to the corresponding ear.

Tones arriving at the two ears were either temporally compatible or incompatible. In the temporal compatibility condition, rhythms to the two ears were synchronous and harmonically related. One series proceeded at twice the rate of the other. In the temporally incompatible condition, rhythms to the two ears were equally fast on the average but were desynchronized, that is, they occurred at harmonically unrelated times. The results were straightforward; average temporal overlap of tone and key press was greater when the two rhythms were temporally compatible than when they were temporally incompatible (Klapp, 1979).

Klapp (1981) extended this finding by showing that temporal incompatibility disrupts concurrent actions with two quite different motor systems: speech and finger movement. The subjects simultaneously tapped telegraph keys and prouced syllables in time with corresponding perceptual rhythms that were either temporally compatible or temporally incompatible with one another. As before, the greatest disruption occurred with temporally incompatible inputs, as if the same internal clock were being used for timing both speech and hand movements.

Effects of Practice

Effects of practice are everywhere apparent in the timing literature and must be explained in theories of timing. For example, in skills that do not require rhythmic timing, such as typing, near-miss periodicity only becomes apparent when the skill has been highly practiced. Genest (1956) found that intervals between typestrokes came closer and closer to perfect periodicity as typists became progressively more proficient, but observed no periodicity whatsoever during early stages of learing to type. Periodicity apparently emerges as a function of practice.

The periodic timing of keystrokes becomes especially obvious when highly skilled typists transcribe experimentally constructed materials known as "alternation passages." These passages contain phrases, such as "authentic divisors," where normal typing conventions require alternate hands for each stroke, thereby minimizing interactions between keystrokes with fingers of the same hand. In typing these passages, the interkey intervals of expert typists become nearly equal, and subsequent strokes tend to compensate for deviations from perfect periodicity. That is, an especially fast stroke tends to follow, and make up for, an especially slow one and vice versa. As Shaffer (1980, p. 116) points out, this zig-zag, or negative serial covariance, sometimes approaches the theoretical limit that could be expected for a perfectly periodic internal clock.

Practice also plays a role in effects of temporal incompatibility (discussed previously). The original effects of temporal incompatibility (Kelso, Southard, & Goodman, 1979; Klapp, 1979; 1981) were obtained with unpracticed subjects, and it is clear that skilled performers can learn to produce temporally incompatible activities. For example, Lashley (1951) observed that expert pianists can readily produce temporally incompatible rhythms, such as a 3/4 rhythm with one hand (3 beats per measure) and a 4/4 rhythm with the other. Shaffer (1980) likewise observed that concert pianists could produce temporally incompatible finger movements by shifting one hand off the beat maintained by the other, for example. It is as if practice enables each hand to become controlled by its own internal clock, or by independent pulses from a single internal clock with a very rapid pulse rate, so that every Nth pulse triggers the program for one hand, and every Mth pulse triggers the program for the other.

The speedups in behavior that themselves result from practice also exhibit practice effects (D. G. MacKay & Bowman, 1969). If components of a behavior sequence have received relatively equal or asymptotic degrees of prior practice, as was the case for the phonemes that subjects in D. G. MacKay and Bowman (1969) produced, then proportional or constant relative speedups in these components can be observed as a result of practice. However, when different components of a behavior sequence have had unequal and nonasymptotic levels of prior practice, unpracticed components speed up faster than practiced components, so that relative timing changes. Consider for example the findings of Seymour (1959), von Treba and Smith (1952), and Wehrkamp and Smith (1952). Subjects practiced making a three-component response in the following sequence: (1) grasp a small object, (2) move it over to a box, and (3) drop it into the box as quickly as possible. With errors held constant, the time required to move the object improved faster with practice than did the time required for grasping it and dropping it into the box. The reason is related to prior practice. The actions of grasping and releasing small objects have had extensive prior practice, and so benefit little from further practice. In contrast, the action of moving the arm with a particular load, in a particular direction, and over a particular distance has received relatively little practice, and so improves quickly with additional practice.

The Node Structure Theory of Timing in Perception and Action

As developed so far, the node structure theory postulates a hierarchy of content nodes for representing the form of perception and action, while an independently stored set of (sequence) nodes codes serial-order rules and determines the sequence in which content nodes become activated. I turn now to timing nodes: their structural organization, rationale, and functioning; how they interact with one another; when their connections become established; and how they generate near-miss periodicity. I then focus on applications of the theory to phenomena such as constant relative timing. I conclude with evidence for a central thesis of the node structure theory, that one and the same timing node can play a role in both perception and action.

The node structure theory of timing was explicitly intended to meet the constraints on theories of timing discussed previously. Timing in the node structure theory is an inherent part of all output specifications, from the lowest level nodes controlling muscle movements to the highest level nodes representing sentential concepts, and is therefore a distributed or everywhere characteristic. Rhythm, rate, and time permeate the entire process of producing well-practiced behaviors and are not tacked on as an independent stage at some point in the specification of output components. Rather, each and every content node in an output hierarchy is activated at some rate, with some periodicity, and with some duration or is not activated at all. Finally, mechanisms for the timing and sequencing of behavior in the node structure theory are independent but closely related. In other words, separate node systems in the theory represent the form, sequence, and timing of behavior, but timing nodes are more closely related to the sequence nodes than they are to content nodes for representing the form of behavior. This relationship occurs because timing nodes have direct or first-order connections with sequence nodes but not with content nodes.

Structural Characteristics of Timing Nodes

A timing node is connected to and activates a set of sequence nodes that can be seen to fit the definition of a domain. For example, the sentential timing node is connected to the domain of sentential sequences nodes, and the phonological timing node is connected to the domain of phonological sequence nodes. Because all nodes are part of some domain, domains can therefore be considered a distributed rather than a local characteristic of node structures.

Timing nodes are connected with and activate sequence nodes in the same way that sequence nodes are connected with and activate content nodes, that is, via the most-primed-wins principle. When a timing node becomes activated, it multiplies the priming of the domain of sequence nodes connected to it until the most primed sequence node reaches threshold and becomes activated. By determining when the sequence nodes become activated, timing nodes therefore determine the temporal organization of the output.

Timing nodes play an essential role in the organization of content and sequence nodes into systems, and indeed, can be considered a defining characteristic of automatized systems. An automatized system is a set of nodes that becomes activated by means of a unique set of timing nodes. Timing, sequence, and content nodes bear a hierarchic relationship to one another in the activation of skilled behavior. Each timing node is connected with a domain of sequence nodes, and each sequence node is connected with a domain of content nodes. A single timing node can therefore be said to activate an entire system of sequence and content nodes at some particular rate. As already noted, for example, at least three timing nodes are required for producing speech at any given rate, the *sentence time* node, the *phonological time* node, and the *muscle time* node. The sentence time node is connected to the dozens of sequence nodes representing sequential rules for English sentences. The phonological time node is connected to the dozens of sequence nodes representing sequential rules for English phonology. And the muscle time node is connected to the dozens of sequence nodes for sequencing muscle movements in producing English speech sounds.

The relation between timing nodes and content nodes is always indirect, always one-to-many, and sometimes also many-to-one. Timing nodes always have a one-to-many relationship to content nodes, because each timing node is connected to many sequence nodes, each of which serves (indirectly) to activate many content nodes. However, some timing nodes also have a one-to-many relation to content nodes, because several timing nodes can serve (indirectly) to activate the same content node. Consider the muscle movement nodes that move the arm, for example. One set of timing nodes controls the arm during reaching, whereas a different set of timing nodes controls the coordination of the arms (and legs) during walking.

The Rate of Perception and Action

Timing nodes determine the rate at which perception and action can occur. Different timing nodes generate different periodicities or average rates of activation. For example, the sentential timing node exhibits a slower periodicity than the timing node for the speech muscles, because muscle flexions and extensions are produced orders of magnitude faster than words and other sentential components. The periodicity of a timing node determines the overall rate of perception or action. Consider speech rate, for example. To determine a desired rate of speech, speakers need only adjust (voluntarily) the overall periodicity or pulse rate of the relevant timing nodes (e.g., fast, normal, or slow).

As in production, the actual rate setting of the timing nodes in perception is partly individual specific, and partly situation specific, determined by the perceived rate of input, for example. This enables the input, and perception of the input, to take place at different rates within wide limits (see also Chapter 4). For example, rate settings for the timing nodes of speaker and listener need not match for speech perception to occur in the node structure theory. The only constraint is that the perceiver's rate setting not be so slow that priming has decayed to below commitment threshold by the time that the next pulse from the timing node arrives, and not so fast that so little priming has built up that the probability of activating the wrong node exceeds the error criterion (D. G. MacKay, 1982).

THE ENGAGEMENT AND DISENGAGEMENT OF TIMING NODES

The way that timing nodes are connected to sequence nodes enables simple higher level decisions regarding timing nodes to selectively engage or disengage whole systems of content nodes. Internal rather than overt speech is produced, for example, by engaging the timing nodes for the sentential and phonological systems and disengaging those for the muscle movement system. As a consequence, phonological components for a sentence become activated in proper serial order and prime their corresponding muscle movement nodes (D. G. MacKay, 1981), but full-fledged movement of the speech musculature does not ensue. No content nodes for muscle movement can become activated, because no timing and sequence nodes within the muscle movement system have been activated.

A high-level decision is required to determine which timing nodes are to become engaged and when. How fast an action unfolds, for example, is determined by altering the periodicity of a timing node or by selectively engaging timing nodes with different periodicities. For example, a given system may have three timing nodes, one each for fast, medium, and slow rates. In the case of hierarchically organized systems—say, the sentential, phonological, and muscle movement systems for speech—the timing nodes corresponding to a given rate are coupled, fast with fast, medium with medium, and slow with slow. As a result, only a simple decision (fast, medium, or slow) must be transmitted to one of the systems, in order to set the rates for the remaining systems. Beyond being a single and simple decision, however, little is known about how such decisions are represented and how they are executed. We do know that some perceptual events such as the unexpected occurrence of an auditory tone during an equal-interval finger-tapping task can cause resetting or reengagement of a timing node, so that the intervals generated begin again from the tone (Pokorny, 1985).

In this sense, timing nodes are flexible or resettable pacemakers rather than persistent pacemakers (Keele, 1986). Once activated, timing nodes don't simply continue activating sequence nodes indefinitely into the future, uninfluenced by external events.

Selective engagement of different sets of timing nodes can also determine the mode or modality of output. Consider, for example, two common modes of speech production: overt versus whispered speech. Overt speech engages all of the timing nodes for the sentential, phonological, and muscle movement systems. Whispering engages exactly these same timing nodes, except for the muscle movement timing nodes that control voicing within the larnyx. Selectively neglecting to activate this one timing node devoices the output, as during whispering (for further details, see D. G. MacKay & MacDonald, 1984).

The ability of bilinguals to rapidly alternate between languages illustrates another switch in output modality that could be achieved by selectively engaging timing nodes. Moderately proficient bilinguals require only about 0.2 s to switch between languages (Kolers, 1968), and fluent bilinguals, for example, French Canadian broadcasters, can switch languages with virtually no lags whatsoever. They produce the sentence "In one corner of the room stood three young men" about as fast as the mixed, French-English sentence "Dans une corner of the salle stood trois jeunes hommes."

How can the phonological systems for different languages become engaged so quickly? Content nodes representing such a sentence within the sentential system are identical for both languages (D. G. MacKay & Bowman, 1969), but phonological nodes for the two languages differ. This means that proficient bilinguals can automatically produce words in first one language and then the other by engaging the phonological timing nodes for the two languages in alternation. Thus, activating a lexical content node primes phonological nodes in both language systems, and alternate engagement of the phonological timing node for English and then French in rapid succession automatically causes production of words in first one language and then the other.

Of course, switching between different output modes only becomes necessary when both modes require simultaneous and incompatible use of identical muscles. The different muscles required for compatible output modes *can* be activated simultaneously. For example, English and American sign language (Ameslan) are compatible output modes for many sentences, involving identical sentential nodes but different muscle movement nodes. Fluent English-Ameslan bilinguals can simultaneously produce many sentences in both languages without interference–in theory, by simultaneously engaging the timing nodes for both output modes.

Practice, Automaticity, and Periodicity

Timing nodes represent the last stage in the establishment of automaticity in the node structure theory (D. G. MacKay, 1987). Timing nodes cannot be used to control behavior during early stages of skill acquisition when new connections between content and sequence nodes are being formed. The theory therefore predicts that timing periodicities will only become apparent when a skill has been highly practiced and is automatic. Genest's (1956) observations on periodicity in typing are a case in point. Periodicity was observed during the later stages of learning to type but not during the earlier stages.

OTHER EFFECTS OF PRACTICE

The original temporal compatibility effects of Kelso et al. (1979) and Klapp (1979; 1981) were obtained with unpracticed subjects, and the node structure theory predicts that skilled performers can learn to produce temporally incompatible activities with prolonged practice. Practice enables the expert to develop independent timing nodes for controlling temporally incompatible activities. As a result, concurrent activities are no longer constrained to begin and end at the same time, as would be the case if the activities shared a single timing node. This explains Lashley's (1951) observation that expert pianists can produce without interference a 3/4 rhythm with one hand (3 beats per measure), and a 4/4 rhythm with the other. It also explains Shaffer's (1980) observation that concert pianists can shift one hand off the beat maintained by the other. It is as if independent internal clocks have been developed to control each hand. Time-sharing is another possibility. For example, a single internal clock may be emitting a very rapid pulse rate, so that, with the help of a counter, every Nth pulse executes the program (i.e., activates the most primed nodes) for one hand, and every Mth pulse executes the program for the other. As D. G. MacKay (1985) points out, counters are theoretically necessary for explaining other aspects of timing behavior, and counters enable a pacemaker timing node to meter out intervals of (almost) any length, thereby acting as an interval timer in the sense of Keele (1986; 1987).

The Coupling of Timing Nodes

Unlike sequence nodes, different timing nodes can be coupled so as to emit pulses in unison. These couplings enable simultaneous control and synchronization of a large number of systems for perception and action. In this way, a single timing node, hierarchically coupled to several subordinate timing nodes, can indirectly activate nodes at regular time intervals, within either a perceptual system, an action system, or several perception–action systems that are simultaneously generating coordinated movements. As a result, evidence for periodicity in a perceptual system, an action system, or several simultaneously coordinated perception–action systems can be construed as evidence for timing nodes.

Coupled timing nodes provide a mechanism whereby the activity of different output systems can become coordinated. The precise and naturally occurring synchronizations of speech and gesture that occur when people simultaneously refer to and point toward an object (Levelt, Richardson, & Heij, 1985) illustrate how coupled timing nodes can coordinate different output modalities. The coordinated activity of the three systems for producing overt speech illustrate the hierarchic coupling of timing nodes within the same modality. The sentence, phonological, and muscle movement timing nodes have different average pulse rates. The *phonological time* node generates more pulses per second than the *sentence time* node, because phonemes are produced faster than words (by a factor of about 5 on the average). The muscle time node generates even more pulses per second, because muscle movements are produced faster than phonemes and words. However, the three timing nodes must operate in conjunction. If the *sentence time* node is speeded up, the *phonological time* and *muscle time* nodes must be speeded up proportionally. If a higher level timing node becomes decoupled and generates pulses too quickly, relative to timing nodes at lower levels, behavior will break down, and large numbers of errors will occur. With the possible exception of "cluttering," a close relative of pathological stuttering, catastrophic malfunctions of this sort seem to be rare, and one wonders what mechanisms have evolved to prevent them. Perhaps a built-in coupling connection prevents the timing nodes for different naturally coordinated systems from becoming desynchronized in this way. However, not all couplings between timing nodes are built-in. To illustrate an acquired (and unfortunately rather complex) coupling between timing nodes, I examine the phenomenon of (approximately) stress-timed rhythm in greater detail in the following section.

STRESS-TIMED RHYTHM AND COUPLED TIMING NODES

Stress-timed rhythm means that for a given rate of speech, the interval between stressed syllables tends to remain relatively constant. We require about the same amount of time to produce the words *summertime* and *mealtime* in the sentence "Mealtime differs in the summertime," despite the extra syllables and segments in *summertime*. The node structure theory tolerates two different accounts of the production of stress-timed rhythm. Both accounts require coupled timing nodes for activating stressed and unstressed syllables, and a counter, which is a mechanism that can both count the number of pulses that a timing node generates and engage or disengage the timing node. (For empirical and theoretical arguments in support of counters, see D. G. MacKay, 1985.)

The more complex account of stress-timed rhythm is this. The sequence nodes for the domain (stressed syllable) are connected to a stressed-syllable timing node, which causes stressed syllables to become activated at regular intervals. The timing of unstressed syllables introduces the complexity. Sequence nodes for the domain of unstressed syllable nodes are connected with the counter, plus a set of at least four timing nodes. These four unstressed-syllable timing nodes are coupled with the stressed-syllable timing node, but have a periodicity that is 2, 3, 4, or 5 times as fast. Once a stressed syllable becomes activated, the counter registers how many unstressed syllable nodes have been primed or readied for activation prior to the next stressed syllable. The counter then engages the appropriate unstressed-syllable timing node for generating either 1, 2, 3, or 4 pulses, so that the interval between stressed syllables remains equal.

The simpler explanation of stressed-timed rhythm is that a single clock and counter controls both stressed and unstressed syllables. The counter counts the pulses and causes activation of a stressed syllable on every sixth beat, say, while causing activation of unstressed syllables on the five intervening beats. However, the stressed syllable becomes lengthened in the surface output, depending on the number of unstressed syllables primed to be activated during the intervening interval.

Deviations From Periodicity

Because timing nodes for skilled behavior have their own endogenous rhythm and become active at regular intervals, the generation of periodicity is a general characteristic of the node structure theory, and the occurrence of perfect, or nearly perfect, periodicity in behavior and in electromyographic activity that precedes behavior (reviewed previously) is therefore to be expected. What require special attention under the theory are *deviations* from periodicity in skilled behavior. As discussed below, the node structure theory predicts deviations from periodicity under three conditions: when the skill involves multiple levels of timing control; when low-level timing factors within, say, the muscle movement system mask the periodicities of a higher level timing node; and when the skill depends on the processing of external feedback that is unpredictable in time.

MULTIPLE LEVELS OF TIMING CONTROL

Deviations from stress-timed rhythm in the production of English sentences illustrate how aperiodic events at one level can mask the periodic events at higher or lower levels. Measurements by Lehiste (1977), for example, show that the interval between stressed syllables may deviate by a factor of five for different sentences of English. Several factors contribute to these deviations from stress-timed rhythm under the theory. One is the time required for activating phonological and muscle movement nodes below the syllable level. For example, with lexical stress equated, more time is required to produce a seven-segment syllable, such as *splints*, than a two-segment one, such as *in*. Adding consonants to a syllable compresses the duration of each segment to some extent, but not so much as to cause invariance. The context in which a surface segment occurs can also influence its duration, because vowels become greatly lengthened when the following consonant is voiced rather than unvoiced (e.g., Lehiste, 1970). Finally, as discussed earlier in the chapter, some segments are more "compressable" than others. For example, vowel duration can change more as a function of lexical stress and overall speaking rate than can consonant duration.

How are these deviations from stress-timed rhythm to be accounted for in theories of English sentence production? Under the node structure theory, a single timing node activates nodes for stressed syllables at regular intervals, but low-level factors, such as number of segments per syllable, sequential interactions between the segments, and the temporal inflexibility of consonants, conspire to alter the periodicity of stressed syllables in the surface output. Such factors can distort a higher level (or underlying) periodicity, because an activated content node only primes its subordinate content nodes and cannot completely determine when they become activated.

Factors above the level of the syllable can also influence timing in the surface output and can distort a periodicity occurring among syllable nodes. An example is the constituent-final pause, which occurs in most, if not all languages. Speakers almost invariably insert pauses after major syntactic constituents, as in "When Mary leaves, [pause] Sam will be upset." Without this pause, listeners could readily mishear and misunderstand this sentence as, "When Mary leaves Sam, [pause] we'll be upset" (W. E. Cooper & Paccia-Cooper, 1980). Constituent-final pauses are necessary to prevent such misunderstandings and automatically override the timing periodicities of lower level units, such as syllables. As a result, demonstrating these lower level periodicities at the surface level will be difficult.

Mechanisms underlying production of constituent-final pauses raise interesting subsidiary questions. W. E. Cooper and Paccia-Cooper (1980, see also Gee & Grossjean, 1983) argue that constituent-final pauses arise from syntactic or sequential rather than semantic mechanisms, and the syntactic nature of this mechanism is readily explained under the node structure theory. Constituent-final pauses reflect activation of the sequence node SENTENTIAL PAUSE, which has effects resembling those of the sequence node PAUSE for spacing Morse code letters (discussed in D. G. MacKay, 1985). SENTENTIAL PAUSE participates in the sequential rules for sequencing phrases and sentences, and at a much higher level than the Morse code PAUSE. As would be expected under the node structure theory, SENTENTIAL PAUSE plays as much of a role in perception as in production. For example, people hear pauses at the major constituent boundaries of experimentally constructed sentences, even when no pause is actually present (Lieberman, 1963).

LOW-LEVEL MASKING OF PERIODICITY

Typing illustrates how low-level factors can mask a higher level periodicity. Indeed, low-level masking of periodicity is so common in typing that deviations from periodicity have been less of a research issue than the existence of periodicity itself (Gentner, 1983; Norman & Rumelhart, 1983). Recall that interkey intervals tend to approach equality only when highly skilled typists transcribe specially constructed materials, "alternation passages," where the hands alternate for each stroke. What is the basis for this periodicity, and why does it only become manifest with alternating keystrokes and after extended practice? The main determinant under the theory is a timing node controlling the domain of sequence nodes for activating keystrokes. The periodicity of the keystrokes reflects the periodicity of the timing node. However, this periodicity only becomes evident in proficient typists, because timing nodes only become connected during the final stages of automatization when, for example, visual input is no longer needed to control the keystrokes. Alternation passages also help to reveal the periodicity because between-hand strokes eliminate two major muscle movement factors that normally mask higher level periodicities. One factor is preparation time. When fingers of the same hand must strike two keys in succession, positioning the hand and relevant fingers requires a preparation time that varies with location of the immediately prior key. Interactions between fingers of the same hand moving toward about-to-be-typed keys are the other factor. Whenever successive keys require same-hand fingers to move in opposite directions, interstroke intervals lengthen.

Similar sources of "low-level noise" suggest an explanation for the deviations from periodicity observed by Wing (1980) in finger-tapping movements and in electromyographic activity that precedes onset of these tapping movements. Timing in typing, speech, and finger tapping are therefore similar. In all three behaviors, lower level factors introduce deviations from periodicity imposed by higher level timing nodes.

Higher level factors also play a role in masking lower level periodicities in these and other skills, and it is possible to manipulate these masking effects. In typing, for example, one can expect keystroke intervals to become aperiodic when skilled typists type freestyle, or from dictation, rather than transcribing from printed texts, as in most studies to date. Dictation and freestyle typing introduce higher level timing factors, which mask the periodicity of the timing node for activating keystrokes in a manner resembling constituent-final pauses in speech.

Applications of the Theory

To illustrate how the theory works in further detail, I now reexamine phenomena such as constant relative timing and the effects of temporal compatibility in perception and action.

Constant Relative Timing

As already noted, the phenomenon of constant relative timing calls for an automatic timing mechanism (without internal calculations for determining the temporal durations of output components produced at different rates), and the node structure theory provides such a mechanism. Constant relative timing is an automatic and emergent property in the node structure theory (D. G. MacKay, 1983). By way of simple hypothetical example, compare a behavior being produced at normal rate, with the same behavior being produced at about half that rate. To generate output at the normal rate, the timing node controlling the lowest level components of the behavior activates a sequence node and its most primed content node at regular intervals, say every n ms. Following a higher level decision to cut the output rate in half, a timing node with half the number of cycles per second becomes engaged. As a result, the same content nodes will become activated every 2n ms, and duration of each component will double. In the absence of factors such as those discussed previously that mask or distort this periodicity, the system will automatically exhibit perfect constancy, because changes in overall rate bring corresponding and proportional changes in the onset time for each component.

Data from handwriting and from transcription typing fit this surprisingly simple account remarkably well. In both skills, changes in overall rate of output have been found to scale the duration of response components in almost perfect proportion, as would occur with a change in rate of a low-level timing node. As already noted, however, changes in overall rate of speech are not scaled proportionally over the durations of all segments. In particular, Tuller et al. (1982) found that consonant durations remained constant relative to vowel durations over substantial changes in speech rate, but only when compared to the interval between vowel onsets. Neither consonant duration nor vowel duration per se remained constant relative to overall utterance duration. The reason is that vowels are much more "elastic" than consonants. Vowels can be prolonged until breath runs out, whereas stop consonants are difficult to prolong, and in any case, they no longer sound like speech when greatly prolonged. Vowels and consonants apparently engage two separate but coupled timing nodes, and although both are muscle movement timing nodes, the one for consonants has a narrower range of rate settings than the one for vowels to which it is coupled.

Consider now the way that components speed up as a result of practice. The theory predicts that behaviors will exhibit constant relative timing in the speed-ups that result from practice, but only if all segments of the behavior sequence have had roughly equal prior practice, as would be the case for the words and syllables of the sentences practiced in D. G. MacKay and Bowman (1969) (see also D. G. MacKay, 1982). However, not all behaviors will exhibit constant relative timing in the speed-ups that result from practice, because less practiced nodes benefit more from further practice than do more practiced nodes (D. G. MacKay, 1982). As a consequence, when segments of a behavior sequence have had unequal prior practice, unpracticed segments will speed up faster than practiced segments, so that relative timing will change. This explains Seymour (1959), von Treba and Smith (1952), and Wehrkamp and Smith's (1952) observations that the time required to move a small object improves faster with practice than does the time required for grasping it and dropping it into the box. Nodes for grasping and releasing small objects have had extensive prior practice and so benefit little from further practice, whereas nodes for moving the arm when carrying a particular load, in a particular direction, and over a particular distance have received relatively little practice, and so speed up a great deal with additional practice.

Temporal Compatibility in Perception and Action

To further illustrate the difficulty of generating temporally incompatible responses, consider the findings of Kelso et al. (1979). Subjects placed their index fingers on starting keys directly in front of them, and following a "go" signal, simultaneously moved both hands as quickly as possible to hit targets to the left and right of starting keys. The target for one hand was much larger than the target for the other, and according to Fitts' law (Keele, 1981), the hand for the larger target should begin to move sooner and reach its destination faster than the hand for the smaller target. However, detailed analysis of the movements showed that both hands began to move at the same time following the go signal, proceeded at the same speed, and contacted their targets simultaneously. A single timing node was apparently coordinating the simultaneous motion of the hands, so that the phases of both movements assumed the same temporal pattern as the more difficult of the two actions.

Kelso (in a related experiment cited in Keele, 1981) made another important discovery that bears on the relationship between sequencing and timing, and should apply to both skilled and unskilled performers under the node structure theory. Using the same apparatus and procedures discussed above, Kelso instructed subjects to terminate one movement before the other, with as small a difference in termination time as they could generate. These instructions had two unexpected effects. First, subjects greatly increased the time to begin movement, far beyond the generated difference in termination time. Second, contrary to instructions, subjects often desynchronized the *initiation*, as well as the termination, of their movements, starting *and* stopping one movement before the other.

Both findings are readily explained under the node structure theory. Instructions to desynchronize simultaneous movements require the formation of a serial-order rule, instantiated via inhibitory connections between sequence nodes, so as to terminate movement of, say, the right hand first, then the left. These sequence nodes must become connected to the proper content nodes; if connected in error to the content nodes controlling the entire movement, rather than just those controlling termination, then movement initiation will also become desynchronized, as was found. Moreover, activating these sequence nodes takes time, and must precede activation of the content nodes, which explains why more time was required to *begin* movement with either hand in this task.

COGNITIVE CLOCKS AND TEMPORAL COMPATIBILITY

A central thesis of the node structure theory is that higher level timing nodes for perception and action are shared. Timing nodes controlling systems of mental nodes are both perceptual and motoric. Timing nodes represent cognitive clocks, which play a role in both perception and action. As a result, the node structure theory predicts effects of temporal compatibility in perceiving input strings and, more importantly, interference between simultaneous perception and production of such strings. Some of the evidence supporting these predictions is discussed in the following section (see also Keele, 1987).

Temporal Compatibility and the Perception of Rhythm

Klapp, Hill, and Tyler (1983) were the first to demonstrate effects of temporal compatibility in perception. Subjects listened to a series of tones while watching a series of flashes arriving at times that were either temporally compatible or temporally incompatible with the tones. The task was to press a single switch as quickly as possible to indicate that either the tones or the flashes had stopped. Reaction time was faster when the rhythms were temporally compatible (harmonically related) than temporally incompatible. Because the task was to monitor rather than to produce the rhythms, this effect of temporal compatibility cannot be attributed to the production of a motor response. Rather, rhythmic compatibility must play a role in perception as well as production of rhythms. For unpracticed subjects, a single timing node apparently governs the timing expectations for both visual and auditory perception (see also Keele, 1987).