18 Constraints on theories of sequencing and timing in language perception and production

Donald G. MacKay

Abstract

This chapter reviews some fundamental phenomena which theories of sequencing and timing in language perception and production must address. I begin with the problem of sequencing: how do speakers order the words in sentences, and how do they order the morphemes, syllables and segments that make up the words? I address five basic constraints on theories of sequencing.

Preparation for sequencing. Speech errors, reaction time experiments and neurolinguistic disorders indicate that a priming or preparation stage is necessary for the sequential activation of speech production units.

Separate mechanisms for sequence versus content. Evidence from speech errors, word games and simultaneous translation indicate that the units representing words, syllables and speech sounds must be separate from the mechanisms for sequencing these units.

A special relationship between sequencing and the initiation of behavior. Theories of sequencing must explain why it takes longer to initiate a preprogrammed output such as a word when the output consists of a sequence of subcomponents, e.g., syllables, than when it consists of a single subcomponent, all other factors being equal.

The sequential error regularity. Theories of sequencing must explain why substituted and substituting components in speech errors usually belong to the same sequential class; e.g., nouns substitute with other nouns, and not with adverbs; vowels substitute with other vowels, and not with consonants.

Different mechanisms for sequencing and timing. The mechanisms for timing speech production units must be different from the mechanisms for sequencing these units: the same mechanism cannot both time and sequence behavior.

LANGUAGE PERCEPTION AND PRODUCTION
ISBN 0-12-052750-2

I next outline some phenomena which must be explained in theories of sequential perception, but which violate an assumption which has become part of virtually every theory of perception and memory published to date. Under this 'sequential isomorphism assumption', perceptual sequences mirror the sequence of external events which has occurred in the real world. Example violations include click detection, phonological fusions, phonemic restorations, and the fact that subjects can respond to higher level units such as words and syllables before they can respond to lower level units such as segments.

Finally, I outline some general constraints on theories of timing in language production and perception: the distributed nature of timing; the occurrence of periodicity, and its relation to skill or practice; interactions between the timing mechanisms for perception and production, and between the timing mechanisms for different output systems such as speech and finger movement; and phenomena such as constant relative timing which characterize many different skills.

Introduction

Theories of language perception and production must deal with three basic questions: What content units or components represent language perception and production, and how are these units organized? How are the content units activated in proper sequence during everyday language perception and production? And what mechanisms are responsible for timing, or determining when and how rapidly these units become activated?

So far, the chapters of the present book have been dealing in various ways with the first of these problems, the nature and structure of the components, and especially the relation between the components for perceiving and producing speech, whether spoken or spelled. I too have examined this issue in other reports (see MacKay, 1985a, b, 1987), and have concluded that exactly the same units play a role in both perception and production above the distinctive feature level (see also Meyer and Gordon, 1983) and that an entire hierarchy of units is required. Included within the hierarchy are units representing distinctive features, segments, initial consonant groups (syllable onsets), vowel groups (rhymes), final consonant groups, syllables, morphemes, words, phrases and sentences (see MacKay, 1985a, b, 1987).

In this chapter, I examine the two remaining issues: how do shared perception–production components become activated in proper sequence and at the proper time and rate? My goal is to develop a list of fundamental phenomena or constraints that viable theories of sequencing and timing must explain.

I will begin with constraints on theories of sequencing, first in production, and then in perception. I will then examine constraints on theories of timing in both perception and production.

Sequencing in language production

How do we execute sequences of behavior in proper serial order when we do, and in improper order when we make errors? Language production has provided the most extensively studied example of the problem of sequencing. Although other cerebral activities raise similar issues, and may even make use of similar mechanisms (see Lashley, 1951; MacKay, 1985a; Mateer, 1985), sequencing is especially complex and interesting in the case of language. One reason is that the issue of sequencing arises at many different levels at once in language production. How do we produce sentences one after the other in logical order? How do we order the words within sentences? How do we order the morphemes, syllables and segments that make up the words? And finally, how do we order the muscle movements that give rise to the sequence of articulatory gestures making up a segment? Our everyday capacity to organize and to produce such a hierarchy of simultaneous, nested sequences is probably fundamental to our uniquely human ability to use spoken language. Our ability to perceive and produce written language is derivative of this more basic ability: the left-to-right and top-to-bottom spatial arrangement of orthography simply mirrors one or more of the levels of sequencing in spoken language.

Constraints on theories of sequencing

The list of fundamental questions which theories of language sequencing must address is relatively small: Is there a nonsequential or preparatory stage which precedes the sequential activation of language behaviour? How is the sequencing mechanism related to the output units for language production? Can sequencing be accomplished by the mechanisms responsible for timing? What is the relationship between the mechanisms for sequencing and timing in language production? I elaborate on these issues below.

Preparation for sequencing

Lashley (1951) was the first to recognize that a priming or preparation stage is necessary for sequencing: According to Lashley, a set of output units must be primed or simultaneously readied for activation before an independently stored sequencing mechanism can activate and impose order on them. Lashley (1951) outlined three sources of support for his idea that simultaneous priming precedes sequential activation. One was anticipations, where a unit occurs before its time, the most frequently occurring type of speech error. In anticipatory errors, an upcoming or soon-to-be-produced word or speech sound becomes produced before its time, as in, 'We have a laboratory in our . . .' instead of, 'We have a computer in our laboratory.' Anticipations indicate that, prior to actual activation, soon-to-be-produced units are simultaneously pre-excited, primed, or readied for activation. Otherwise, why would an upcoming or about-to-be-produced unit be so much more likely to intrude than any other unit in the speaker's vocabulary?

Another argument for a (simultaneous) preparatory stage prior to (sequential) activation is that, 'a general facilitation, a rise in the dynamic level' seems necessary for the performance of many sequential activities (Lashley, 1951, p. 187). For example, when sufficiently aroused, brain-damaged patients can execute sequences of behavior that under normal circumstances they cannot. For example, an aphasic who is unable to produce the word 'watch' in a laboratory test, may exclaim, 'Give me my watch!', when the experimenter pretends to make off with his watch (Teuber, 1965, personal communication). Such examples suggest that an output sequence cannot become activated unless its units have received sufficient priming: Of course, motivational factors helped to provide the priming in this particular neurolinguistic example, whereas factors associated with the specific word or action being produced normally provide the primary source of priming.

Lashley (1951, p. 189) also noted evidence from studies of reaction time and of word association indicating that a preparatory stage preceding activation can facilitate specific patterns of action. 'Reaction time, in general, is reduced by preliminary warning or by instructions which allow the subject to prepare for the specific act required. In controlled association experiments, the subject is instructed to respond to the stimulus word by a word having a certain type of relation to it, such as the opposite or a part of which the stimulus is the whole: black–white, apple–seed. The result is an attitude or set which causes the particular category to dominate the associative reaction.' It is as if controlled association instructions simultaneously prime or ready-for-activation, a large number of specific responses, thereby short-circuiting the first stage of the prime-then-activate process, so that the response can be produced soon after presentation of the stimulus.

Lashley's third basis for assuming that priming precedes sequential activation during production is that perception exhibits a similar process. To demonstrate perceptual priming, Lashley auditorily presented to his audience the garden path sentence, 'Rapid righting (writing) with his uninjured hand saved from loss the contents of the capsized canoe.' As might be expected, a sudden reinterpretation of the word 'writing' ('righting') took place once the audience heard the last two words of the sentence. On the basis of this demonstration, Lashley argued that the units for comprehending the word 'righting' ('writing') could not become activated until the phrase 'capsized canoe' had occurred, and so must have been held in a state of readiness or partial activation 'for at least 3 to 5 seconds after hearing the word' (p. 193). Thus, priming or readying-for-activation precedes actual activation during comprehension, and by analogy, during production as well, because 'the processes of comprehension and production of speech have too much in common to depend on wholly different mechanisms' (Lashley, 1951, p. 186).

Lashley's distinction between the processes of priming and activation is of course recognized in at least *some* recent theories of language production (e.g., MacKay, 1982, 1985a, b), and can be seen to provide a solution to the interaction-encapsulation issue, one of the main outstanding problems with the concept of modularity. The problem is that modules such as speech perception–production seem to exist as relatively autonomous processing systems but nevertheless interact extensively with one another, so that if modularity requires 'encapsulation of processing' (see Fodor, 1983), there are no modules. However, modules can be both interactive and encapsulated if *priming* is distinguished from *activation* as in MacKay (1982, 1985a, b), because priming is automatic and unencapsulated within modules, whereas activation requires a module-specific activation mechanism (sequence and timing nodes) and is therefore encapsulated or confined within particular modules.

Independence of sequence and content

The mechanism for sequencing behavior must be separate from the units which represent the content or form of the behavioral sequence. And in particular, the basic units making up a sequence of language units must be independent of the mechanism which sequences these units. To see why this is so, consider a set of theories incorporating non-independent sequencing and content mechanisms; chain association theories, the first class of theories discussed in Gordon and Meyer (Chap 20). There is no independent sequencing mechanism in chain association theories: unidirectional links between the units representing the content of behavior provide the representation of sequence. Activating the first content node directly primes, and indirectly causes activation of the second (connected) content node, and so on, until the entire sequence has been produced.

Many variants of this unidirectional bond assumption have been proposed, and the bonds are usually assumed to be excitatory in nature. But not always. For example, Estes (1972) proposed a chain association theory where the bonds are inhibitory rather than excitatory. The first unit inhibits the remaining units, the second inhibits all but the first, the third inhibits all but the first two, and so on. For example, in producing a simple word such as *act*, a superordinate node representing the entire word becomes activated, and primes its three subordinate nodes representing the segments, /a/, /c/, and /t/. Now under the unidirectional bond assumption, the first element, representing /a/, inhibits the other two, and the second element, representing /c/, inhibits the third, representing /t/. Thus, the first element, not being inhibited by any of the others, achieves the greatest degree of priming, and becomes activated under a most-primed-wins principle (see MacKay, 1982). The second, no longer being inhibited by the first, now has the greatest priming, and becomes activated, releasing the third from inhibition, and so on.

Lashley (1951) anticipated the basic problem with this, and other recently proposed chain association theories. The problem is that links between the basic output components will interfere with one another. For example, inhibitory links between the content nodes for the word *act* will interfere with the production of *cat* and *tack*, or any other words containing the same components in a different order. Extrapolating to a normal 50 000 word vocabulary, the conflicting inhibitory connections between the phonological components in this theory would simply prevent speech production altogether.

Theories confounding the sequencing mechanism with the production units therefore fail to explain the production of language sequences *per se*. These theories also predict sequential errors which do not occur, and have difficulty explaining the ones which do occur (see MacKay, 1970). Because they postulate non-independent mechanisms for sequence and content, chain association theories also have difficulty explaining the flexibility observed in sequential behavior. Children's word games, such as Pig Latin, illustrate the nature of this flexibility (see MacKay, 1972, and Treiman, 1983). When playing Pig Latin, children quickly and easily impose a new order on the segments of both never-previously-encountered nonsense syllables, e.g., *snark*, and frequently used words, e.g., *pig*. When children produce the word *pig* as *igpay*, for example, no painful process of unlearning the old habitual sequence is required, as might be expected if the old sequence were built into the output units themselves by means of unidirectional bonds. Instead, the sequencing mechanism appears to operate on the basis of rules which apply to an indefinitely large number of behavioral units, and which can be easily altered so as to produce never-previously-encountered forms such as *arksnay* (see MacKay, 1972).

Lashley noted one final set of phenomena calling for independence of the sequencing mechanism and the content units themselves; the ability to translate freely from one language to another using different word orders. An experienced translator does not have to proceed word by word, but quickly and easily alters the order of the components making up the original idea when translating into a target language with different word order. Such flexibility suggests that sequence is not part of a lexical concept or idea *per se*, but is imposed on the idea by language-specific rules or sequencing mechanisms.

Bilingual sequencing errors likewise suggest that the sequencing mechanism is independent of the words and ideas being sequenced. Sometimes bilinguals inadvertently impose the *wrong* order on words: For example, a native speaker of German may unconsciously adopt aspects of German syntax when attempting to speak rapidly in English, postponing the verb to the end of a frequently encountered English expression. Such errors simply could not occur if the sequencing mechanism consisted of links between the units representing language-specific words.

Sequencing and the initiation of behavior

Theories of sequencing must explain a special and repeatedly demonstrated relationship between sequencing and the initiation of behavior. A large number of recent studies have shown that it takes less time to initiate a preplanned behavior which consists of a single component than one which consists of a sequence of components. This relationship between sequencing and the initiation of behavior is an embarrassment to chain association or horizontal link theories, even ones augmented with vertical links such as Estes (1972) and Wickelgren (1979). It also presents problems for theories incorporating a scanning mechanism, including a long-abandoned theory of my own (MacKay, 1969). In 'scanning' theories, a behavioral sequence is loaded into a memory buffer in preparation for sequencing, and behavior becomes initiated by a scanner which sweeps over the buffer from e.g., left to right. Thus, a subject who is prepared to say the word *paper*, for example, has already loaded the word into the output buffer; following a go signal, the word can then be produced by sweeping the scanner over the buffer, causing activation of the initial /p/, followed by the remaining segments of the word in proper order. This process is of course independent of word length, so that the scanner should trigger the initial /p/ of a one-syllable word such as *paint* no faster than the initial /p/ of a two-syllable word such as *paper*.

Available data do not support this prediction, however, For example, Klapp, Anderson and Berrian (1973) investigated the time required after a go signal to begin to say a large number of one- versus two-syllable words such as *paint* and *paper*. All of the words were five letters long, and began with the same segment. The results showed that response time was significantly longer for two-syllable than one-syllable words, a finding replicated in other studies and for other language behaviors besides speech, e.g., Morse code (Klapp and Wyatt, 1976), and typing (Sternberg, Monsell, Knoll, and Wright, 1978).

Errors in sequencing

Theories of sequencing must of course explain how sequential errors occur. Not just the fact that sequential errors occur, but the detailed nature of the regularities that have been observed in these errors. An example is the sequential class phenomenon, one of the most general regularities observed to date. The phenomenon is this: when a speaker inadvertently substitutes one linguistic component for another, the substituted and substituting components almost invariably belong to the same sequential class. Cohen (1966) originally observed this regularity in errors involving interchanged words. An example is the error, 'We have a laboratory in our own computer', where one noun (*laboratory*) interchanges with another (*computer*). As in this example, nouns generally interchange with other nouns, verbs with other verbs and not with, say, nouns or adjectives (Cohen, 1967). Even 'Freudian slips' such as, 'He found her crotch, I mean, watch', adhere to this sequential class rule. Because both *watch* and *crotch* are nouns, this (invented) error obeys the sequential class regularity, even though, as Fromkin (1973) points out, semantic (Freudian) factors may simultaneously contribute to such errors.

The sequential class regularity has also been observed for errors involving (1) morphological components: prefixes interchange with other prefixes, suffixes with other suffixes, and never prefixes with suffixes (MacKay, 1979), (2) syllabic components: initial consonant clusters interchange with other initial clusters, and final with final, but never initial with final (MacKay, 1972), and (3) segmental components: vowels interchange with vowels, consonants with consonants, and never vowels with consonants (MacKay, 1972). In short, the sequential class regularity holds for all levels of speech production, and a viable theory of sequencing must explain this fact.

Exceptions to the sequential class rule. Even though exceptions to the sequential class rule are rare, they must also be explained in theories of sequencing because they display interesting regularities of their own. Consider the following examples from Fromkin (1973): 'She was waiting her husband for' (instead of, 'waiting for her husband'), and 'I don't want to part this book with' (instead of, 'to part with this book'). These regularities pose three questions: Why do these errors violate the sequential class rule (both errors involve a noun phrase changing places with a verb particle)? Why are these errors so rare? And why do these errors result in a sequence (Verb + Noun Phrase + Verb Particle) which is appropriate for other expressions such as 'She called the man up'?

Different mechanisms for sequencing and timing

Another general constraint on theories of sequencing is that the same mechanism cannot both time and sequence language behavior: sequencing cannot be achieved by a timing mechanism, and timing cannot be achieved by a sequencing mechanism (see also Keele, Chap 21). To see why timing and sequencing require different mechanisms, let us examine the two hypothetical alternatives. Consider first the possibility that a timing mechanism is by itself responsible for both sequencing and timing in speech production. This hypothetical timing mechanism is able to generate the sequence of phonemes in a word by specifying their time of production, and sequencing errors arise because phonemes have been improperly timed. The word *cat*, for example, might be misproduced as *act* because the *a* has been produced relatively early, and the *c* produced relatively late. Likewise, at a higher level, the phrase 'in the car', might be misproduced as 'in car the', because the noun is produced relatively early, and the article relatively late.

Unfortunately for this hypothetical account, no such errors occur: proficient speakers never simply misorder components in time. As discussed here, substituted components in actually occurring speech errors do not just exchange places in time, but virtually always belong to the same sequential class. For example, in the error, 'cake the ring of teas' instead of 'take the ring of keys,' the segments /*t*/ and /*k*/ exchange temporal positions, but they also belong to the same domain or sequential class, *initial consonant group* (see MacKay, 1972). This sequential regularity would not be expected if a timing mechanism determines sequencing.

Consider now the opposite possibility, that a sequencing mechanism determines both sequencing and timing, an idea proposed by Norman and Rumelhart (1983). Norman and Rumelhart's theory of typing incorporates a sequencing mechanism, but no timing mechanism, and timing of a keystroke in their theory depends on how long it takes to sequence a set of preprogramed keystrokes. Under this view, errors in the timing and sequencing of typestrokes are one and the same: when typestrokes occur out of sequence, one component is being activated especially early, and the other is being activated especially late. No one has proposed a similar hypothesis for speech production, for reasons which should be obvious from an examination of the speech error illustrated above.

However, it is important to stress that the Norman–Rumelhart hypothesis also encounters difficulties in explaining typing. Consider the findings of Grudin (1981) on the timing of keystrokes in transposition errors, e.g., *the* mistyped as *hte*. Grudin's data showed no tendency for one key to come especially early, and the other especially late in a large number of transposition errors produced by skilled typists. Rather, the keys exchanged places both in sequence, and in time, just as in speech errors. For example, assume that a skilled typist normally types the word *the* correctly with about 140 ms. between hitting *space* and *t*, and 75 ms. between hitting *t* and *h*. Grudin found that if this typist produced the transposition error *hte*, timing remained the same; about 140 ms between space and *h*, and 75 ms between *h* and *t*. The wrong components occurred at the right time. This finding indicates that timing is independent of the behavior being timed, and this independence could only occur with separate mechanisms for determining the content, sequencing and timing of behavior.

Grudin's findings also indicate that timing is being 'programmed' in proficient typing, and this is an especially important fact for theories of sequencing and timing, because typing is a skill which does not demand consistent or accurate timing, unlike say, music, Morse code, or speech. Apparently a timing mechanism plays a role in language production even when precise timing is unnecessary.

Constraints on theories of sequencing in perception

I turn now to sequencing in perception, the problem of how we perceive input sequences in proper serial order when we do, and improper order when we make errors. This problem places as many constraints on psychological theories as Lashley's problem of serial order in behavior, but has been largely ignored in psychology: studies of perception over the past 150 years have concentrated mainly on static visual displays, and have devoted relatively little attention to the perception of input sequences.

To illustrate the problem of sequential perception, I begin with the most frequently overlooked constraint on theories of sequential perception: effects of practice. Warren and Warren (1970) noted that we can perceive the serial order of sounds in familiar words such as *sand* at rates of 20 ms per segment, but require over 200 ms per sound for perceiving the order of unfamiliar sound sequences such as a hiss, a vowel, a buzz and a tone (when recycled via a tape loop). One interpretation of these findings attributes this difference to practice or familiarity: sequences of speech sounds are much more familiar than non-speech sequences such as *hiss–vowel–buzz–tone*. Another interpretation focuses on acoustic differences between speech vs. nonspeech sequences (see Bregman and Campbell, 1971). However, this second interpretation will not do for Warren's (1974) experimental demonstration of how practice facilitates the recognition of nonspeech sequences. Subjects in Warren (1974) repeatedly listened to non-speech sounds in sequences which were initially unrecognizable, e.g., *hiss–vowel–buzz–tone*, and after about 800 trials of practice, the subjects became able to identify the order of these sounds with durations of less than 20 ms. per sound. This order-of-magnitude effect of practice on the recognition of auditory sequences is all the more interesting because similar perceptual reversals are observed in the speech perception of children, but not adults. Children often reverse adjacent segments in perceiving a word, misperceiving 'spaghetti' as 'psghetti' or 'snow' as 'nows', for example (Allen, 1981), but adults virtually never make such errors. This developmental difference suggests that effects of practice on the recognition of sequence represent a general constraint on theories of sequential perception.

Violations of sequential isomorphism

Why is the problem of sequencing in perception often considered trivial and uninteresting? One reason seems to lie in an implicit, but fundamental assumption which has become built into virtually every theory of perception and memory published to date. Under this '*sequential isomorphism postulate*', perceptual sequences invariably mirror the external sequence of events in the real world. If correct, this 'first-in-first-perceived' postulate indeed renders the problem of serial order in perception trivial and uninteresting. In fact, however, whole classes of striking and well-documented phenomena, discussed below, violate sequential isomorphism (see also Cutler, Chap 2; and Huttenlocher and Goodman, Chap 19) and provide strong constraints on theories of sequential perception. Needless to say, however, the fact that sequential isomorphism appears to predominate most of the time in the remainder of our lives provides an additional constraint on theories of sequential perception.

The perceptual precedence of higher level units

Theories of perception must explain why units which end later in an input sequence are sometimes perceived more quickly than units which end sooner. The recognition of segments vs. syllables provides an example: subjects require more time to identify a segment than a syllable within a sequence of nonsense syllables, even though the segment ends sooner than the syllable in the acoustic stimulus. The original experiment by Savin and

Bever (1970) can be used for purposes of illustration because many subsequent studies have replicated their basic findings, and come to the same conclusion (see Massaro, 1979).

Savin and Bever (1970) had subjects listen to a sequence of nonsense syllables with the aim of detecting a target unit as quickly as possible. There were three types of targets: an entire syllable, e.g., *splay*, the vowel within the syllable, i.e., *ay*, and the initial consonant of the syllable, i.e., *s*. The subjects were instructed to press a key as soon as they detected their target, and the surprising result was that reaction times were faster when the target was the entire syllable rather than either the initial consonant or the vowel in the syllable. Theories of perception must therefore explain why higher level units, and in particular, a syllable or word, can be detected before the phonemes making up the syllable or word.

Sequential illusions

Sequential illusions occur whenever the surface units of an input sequence are perceived as coming sooner or later than they actually occur in the real world. I discuss two examples below.

Phonological fusions. Phonological fusions occur when a subject wearing earphones is presented with an acoustic stimulus such as *banket* in one ear, and *lanket* in the other ear: even with a sizeable (e.g., 200 ms) onset lag or temporal asynchrony between the stimuli, subjects often report hearing *blanket*, a fusion of the two inputs (Day, 1968; and Cutting and Day, 1975). If perception accurately represented the input sequence, subjects would perceive the *l* followed by the *b*, because the order of arrival at the acoustic level is *l* followed without overlap by *b*. Some subjects in fact do perceive the input sequence veridically, but there are large individual differences, and most subjects do not: instead they fuse the inputs, and report that the *b* preceded the *l* (see Day, 1968).

As their name suggests, phonological fusions depend on a phonological rather than on an acoustic representation of the input: phonological factors readily influence the probability of fusion, whereas lower level factors within the acoustic analysis system do not. One of these phonological factors is wordhood: fusions sometimes occur when both inputs are words, but they occur much more frequently when both inputs are non-words, such as *banket* and *lanket*. Words are also the most common type of fusion response, regardless of whether the stimuli are words or nonwords (Day, 1968).

Another phonological factor is sequential permissability: fusions always result in phonological sequences which are permissable or actually occurring within the listener's language. Percepts which violate phonological rules (e.g., *lbanket*) never occur, even when nonoccurring sequences represent the only possible fusions. For example, simultaneous presentation of *bad* and *dad* never results in fusions such as *bdad* and *dbad*, because initial *bd* and *db* do not occur in English.

By way of contrast, acoustic factors have little or no effect on the likelihood of fusion: Cutting and Day (1975) found that number of fusions remained constant when the fusion stimuli differed in intensity, in fundamental frequency, and in allophonic characteristics, as when one stimulus contained a trilled /r/ and the other an untrilled /r/. Explaining the detailed nature of phonological fusions, and other sequential illusions, provides a fundamental challenge for theories of perception.

Phonemic restorations. Phonemic restorations represent another sequential illusion. When subjects listen to a sentence containing a word such as *legi*lature*, where the *s* has been masked by a cough (*), they hear the word intact, and are unable to accurately locate the cough within the sequence of phonemes, or tell which phoneme is missing when informed that the cough has physically replaced a single speech sound (Warren and Warren, 1970). This inability to locate the cough in the sequence of phonemes violates sequential isomorphism and must be explained in theories of sequential perception. Click localization studies (see Fodor, Bever, and Garrett, 1974, for a review) provide a similar example.

Constraints on theories of timing in language perception and production

I turn now to the third basic problem, timing in languge perception and production. How do we produce language units of different durations? And how do we produce these units at different rates, and with different rhythms or patterns of durations?

What fundamental phenomena or constraints must theories of timing address? The section on sequencing has already discussed one of these constraints: independence of the mechanisms for timing and sequencing the basic language units. In addition, theories of timing must address six other fundamental issues discussed below: Where in the specification of output components is rate and timing determined? What mechanisms underlie the production of rhythmic outputs? How is periodicity or near-miss periodicity achieved in language skills such as typing, handwriting and speech? What is the relationship between the timing mechanisms for language perception and production? What accounts for the ability to flexibly adjust the rate and timing of behavior? Why do activities 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 in speech production, from the lowest level components controlling muscle movements, to the highest level components representing sentential concepts, must be activated at some rate, and for some duration. Rhythm, rate and timing permeate the entire process of language production, and cannot cannot be tacked on as an independent stage at some point in the theoretical specification of output processes.

To see why timing must be a distributed characteristic, it is only necessary to examine existing 'stage of processing' proposals, including one of my own (MacKay, 1969). These proposals treat rhythm and timing as an 'afterthought', a late stage of processing introduced just before or during the programming of muscle movements. For example, in my (1969) stage-of-processing proposal for timing the producing of speech, the entire syntax, semantics and phonology for producing a sentence are first constructed, and then stored within a simultaneous or nontemporal spatial display. Only following this construction and storage stage are timing characteristics such as speech rate specified as part of the output.

Stage of processing proposals such as this one face many unsolved problems. One is the complexity and reduplication of information which is required for the simultaneous display. 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 specifications are also required for constructing the sentence in the first place, adding timing at one particular level in the construction of a sentence complicates the process.

An even more serious problem for stage-of-processing theories of timing 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 the activation of components, whether low-level muscle movement components, or high-level mental components. As an example of speed-accuracy trade-off in the activation of high-level mental components, consider phonological speech errors such as the substitution of 'coat-thrutting' for 'throat-cutting'. Here components within the phonological system have become interchanged, and in a study of experimentally induced speech errors, MacKay (1971) demonstrated that such errors increase as a function of speech rate (see also Dell, 1985). These findings cannot be explained if timing is determined *after* the specification and mispecification of phonology: for rate to influence phonological errors, the phonological components and their rate of output must become specified at the same time. And because rate also influences errors occurring at other levels, both above and below the phonological level, rate and timing must be specified throughout the entire hierarchy of units for producing a sentence, from the lowest level muscle movements to the highest level phrase and lexical concepts (see MacKay, 1987 for details). In short, rate must be a distributed characteristic, specified everywhere, rather than at just one point in the hierarchy of output units.

Monitoring, rate and errors. A possible counterargument in favor of stage-of-processing theories of timing is that errors increase with rate not because rate is a distributed characteristic, but because various output monitoring devices become suspended at faster rates, thereby allowing more errors. To be taken seriously, this explanation of speed-accuracy trade-off requires a great deal more theoretical specification and empirical support. There currently exists no empirical evidence for monitoring devices which are independent of the output mechanisms themselves, and no evidence that hypothetical monitoring devices of this sort are 'suspendable' (see MacKay, 1987). The concept of perceptual monitoring as a final stage in production also has difficulty with the time characteristics of how errors are detected and corrected: error detection and correction is so rapid as to sometimes precede the full-blown appearance of an error in the surface output (see Levelt, 1984). And even if monitoring and production are viewed as parallel rather than serial processes, the monitoring counter-argument has difficulty with the fact that perception can proceed much faster than production at maximal rate (see MacKay, 1985b, 1987): suspending perceptual monitors could not facilitate output rate if perception proceeds in parallel with and faster than production.

The generation of periodicity

Proficient performance of language skills such as speech production, Morse code, typing and handwriting has been shown to exhibit perfect or nearly perfect periodicity in various ways which must be explained in theories of timing. For example, Wing (1978) demonstrated 'near-miss' periodicity in handwriting for the time between successive downstrokes and upstrokes of subjects producing the letters m, n, v and w. The temporal deviations from perfect periodicity were quite small, and tended to alternate with one another, temporal undershoot on one stroke followed by temporal overshoot on the next, and vice-versa. Similar zig–zag alternations in the time between adjacent components have also been observed for speech (in the durations of successive syllables in an utterance; Kozhevnikov and Chistovich, 1965), and for skilled typing (Shaffer, 1978, Wing, 1980). Typestroke periodicity becomes especially obvious when highly skilled typists transcribe specially constructed materials (Shaffer, 1980). These 'alternation passages' contain phrases, such as 'authentic divisors', where normal typing conventions require a different hand on each stroke, so that interactions between successive movements with the same hand cannot occur. In typing these passages, the inter-key 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 'negative serial covariance' sometimes approaches the theoretical limit that could be expected for a perfectly periodic internal clock. Shaffer's conclusion is especially interesting because neither rhythm nor precise timing is necessary for executing typestrokes. Apparently people not only *can*, but normally *do* generate near-miss periodicity, even when neither rhythm nor precisely timed output is required.

Effects of practice on timing

Effects of practice are everywhere apparent in the timing literature, and must be explained in theories of timing. For example, language skills such as typing only exhibit near-miss periodicity following extensive practice. Genest (1956) found that the interval between typestrokes came closer and closer to perfect periodicity as typists became progressively more proficient, but observed no periodicity whatsoever during the early stages of learning to type (see also Shaffer, 1978, 1980, discussed above).

Interactions between timing mechanisms for perception and production

On-line interactions between perceptual events and the timing of ongoing speech and action have frequently been observed, and suggest that systems of perception and production may share some of the same timing mechanisms (see also Keele, Chap 21; MacKay (1987); and Keele, Pokorny, Corcos, and Ivry, 1985). Lashley (1951) was the first to note such an interaction between a perceptual rhythm (listening to a marching band) and ongoing motoric activities, including walking, breathing and speaking: 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 interactions suggest that identical timing mechanisms govern perceptual processes such as listening to music, and motoric processes such as walking, breathing and speaking.

Prosodic flexibility

Timing in syllable production is flexible rather than built in, and theories of timing must explain our ability to learn and to produce different types of language rhythm. An example is the more varied use of durational information in 'stress-timed' languages such as English as compared to 'syllable-timed' languages such as French (see Cutler, Mehler, Norris and Segui, 1983).

Constant relative timing

As the overall time to produce a behavioral sequence changes due to a voluntary decision to increase rate, the proportion of time required to produce some segments of the sequence often remains constant. This phenomenon, known as constant relative timing, has been observed within limits for many 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, Shapiro, Zernick, Gregor and Diestal (1981) had subjects walk at various speeds on a treadmill, and found that the proportion of time required to execute the four basic phases of a step (lift, stride, heel contact and support) remained virtually invariant at the different speeds. If the lift phase required 20% of the duration of a step cycle at a slow rate, it required about 20% at a faster rate. Relative timing only remained constant within a limited range of rates, however: When the treadmill was accelerated beyond a certain point, subjects broke into a jog, and the temporal configuration of the components of their strike changed dramatically. Walking and running clearly have different temporal characteristics, and are controlled by different underlying mechanisms, which both conform to the law of constant relative timing. Handwriting and transcription typing also exhibit constant relative timing. Here 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 internal clock (Shaffer, 1978).

Interestingly, constant relative timing has also been observed for *involuntary* changes in the rate of language production. Components in a sentence speed up involuntarily as a result of practice, and these changes in relative duration sometimes exhibit constant relative timing as well. For example, MacKay and Bowman (1969) had subjects practice producing a sentence as quickly as possible over 12 trials of practice, and found that the maximal speech rate increased systematically with practice (see also MacKay 1974). More importantly, different components of the sentence speeded up proportionally: the relative duration of words and syllables remained constant at the faster speed. The constant relative timing that occurs when behavior speeds up, due either to voluntary rate changes, or to involuntarily effects of practice, places fundamental constraints on theories of timing.

Theories postulating a computational process for calculating the durations of behavioral components have difficulty explaining constant relative timing because the phenomenon appears in the behavior of insects and crustaceans (see Shapiro, Zernick, Gregor, and Diestal 1981), where such computations are unlikely. Constant relative timing also appears immediately after a voluntary decision to change rate, without the lag times which seem necessary for computing the new temporal values. The mechanism for adjusting timing on the basis of rate must instead be automatic and noncomputational in nature.

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 if stop consonants are greatly prolonged, they no longer resemble speech sounds. Such observations suggest that different timing mechanisms may control the production of consonants versus vowels, and consistent with this hypothesis, Tuller, Kelso and Harris (1982) found that at different rates of nonsense syllable production, the durations of consonants remained constant relative to vowels, but only when compared to the interval between vowel onsets. Neither consonant nor vowel duration *per se* remained constant relative to overall utterance duration.

Temporal interactions between different activities

Theories of timing must explain why concurrent activities with different timing characteristics tend to interfere with one another. For example, speech exhibits temporal interactions with finger movement: when subjects attempt to produce a sequence of syllables and a sequence of finger taps at the same time, they experience considerable difficulty if the movements conflict in timing, but little difficulty if the concurrent movements are temporally compatible, or occur at identical or harmonically related times.

Klapp (1981) had subjects press a telegraph key in time with tones presented periodically via headphones to one ear, and simultaneously produce a syllable in time with tones presented to the other ear. The goal was to maximize the temporal overlap of keypresses and syllables with the tone to the corresponding ear. The tones arriving at the two ears were either temporally compatible, or temporally incompatible. In the temporal compatibility condition, the rhythms to the two ears were harmonically related: One series proceeded at twice the rate of the other. In the temporally incompatible condition, the rhythms to the two ears were equally fast on the average, but were desynchronized, or occurred at harmonically unrelated times. The results were straightforward; the average temporal overlap of tone and behavior was greater in the temporally compatible condition than in the temporally incompatible condition, as if the same internal clock was needed for timing both speech and hand movements.

Summary and conclusions

This chapter has reviewed a number of fundamental constraints or characteristics which can be used to evaluate current theories of sequencing and timing in language perception and production. Indeed, I have already used some of these characteristics to illustrate problems with four general classes of theories: stage of processing theories of timing, chain association and scanning theories of sequencing, and monitoring theories of the processing of perceptual feedback. However, a general evaluation of extant theories on the basis of these characteristics remains to be carried out.

The list of characteristics developed here is short, and undoubtably incomplete, and will surely grow as the field progresses. Contained within the list, however, are the *sine qua non* requirements for a viable theory: Theories of sequencing and timing which lack one or more of these characteristics can be considered incomplete or inadaquate. And in addition to providing a standard against which to evaluate current theories, the criteria of this chapter can be used to develop new and more adequate theories. Indeed, I myself hope to use these criteria in developing such a theory (see MacKay, 1987).

What sort of theory will be needed for explaining the fundamental phenomena of sequencing and timing in language perception and production? The theory must postulate a hierarchy of units, including units representing sentences, phrases, words, morphemes, syllables, segments and distinctive features. These units must be separated from the mechanisms for activating them in proper sequence, and the sequential activation of these units must be preceded by a priming or preparation stage in the theory. Moreover, the sequential activating mechanism must operate on classes of units, so as to explain why substituted and susbtituting components in speech errors usually belong to the same sequential class; e.g., verbs substitute with other verbs, and not with pronouns; syllable initial consonants substitute with other syllable initial consonants, and not with final consonants.

On the perceptual side, the theory must somehow allow violations of sequential isomorphism, so that the perceived sequence can differ from the actual sequence for relatively unfamiliar, rapidly presented external or real-world events. And somehow, the theory must also give precedence to higher level units, so that more rapid responses can be generated for higher level units such as words and syllables than for lower level units such as segments.

Timing must employ different mechanisms from sequencing in the theory. And the timing mechanisms must act like an internal clock which is basically periodic in nature and can be sped up to introduce proportional changes in the duration of different output components. The timing mechanisms must also develop with practice, and must operate throughout the entire hierarchy of production units, so as to capture the distributed nature of timing. Finally, language perception and production must have identical or closely interacting timing mechanisms; and so must different output systems such as speech and finger movement.

How general are these characteristics? Do they apply to other skills besides language perception–production? As Keele (Chap 21) points out, timing in language skills may reflect a more fundamental ability which characterizes many other skilled behaviors. And sequencing likewise reflects a fundamental ability with basically similar manifestations in speech and other everyday activities (see Lashley, 1951). The present chapter may therefore provide a rough outline of what a general theory of sequencing and timing in behavior must eventually explain.

Acknowledgements

This paper was prepared in part while the author was a member of the Research Group on Perception and Action at the Center for Interdisciplinary Research (Z.i.F.), University of Bielefeld, and was presented to the conference entitled, ‘Common Processes in Speaking, Listening, Reading, and Writing’, which was held at the Z.i.F., Bielefeld, July 1–5, 1985. A preliminary version of this chapter appeared as Perception and Action, Report 84, 1985, Zentrum fuer interdiciplinaere Forschung, Universitaet Bielefeld, FRG. The author thanks M. Studdert-Kennedy and an anonymous reviewer for comments on an earlier version of this chapter.

References

Allen, G. D. (1981). Suprasegmental constraints on segmental representation: Research involving speech production. In J. Myers, J. Laver, and J. Anderson (Eds) *The cognitive representation of speech*. Amsterdam: North-Holland.

Bregman, A. S. and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Quarterly Journal of Experimental Psychology*, **25**, 22–40.

Cohen, A. (1967). Errors of speech and their implication for understanding the strategy of language users. *Zeitschrift fuer Phonetik*, **21**, 177–81.

Cutting, J. E., and Day, R. S. (1975). The perception of stop-liquid clusters in phonological fusion. *Journal of Phonetics*, **3**, 99–113.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1983). A language-specific comprehension strategy. *Nature*, **304**, 159–60.

Day, R. S. (1968). *Fusion in dichotic listening*. Unpublished doctoral dissertation, Stanford University.

Dell, G. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science*, **9**, 3–23.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton and E. Martin (Eds). *Coding processes in human memory*. Washington, DC: Winston.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. A., Bever, T. G., and Garrett, M. F. (1974). *The psychology of language*. New York: McGraw-Hill.

Fromkin, V. A. (1973) (Ed). *Speech errors as lingustic evidence*. Paris. Mouton.

Genest, M. (1956). L'analyse temporelle du travail dactylographique. *Bulletin du Centre d'Etudes et Recherches Psychotechniques*, **5**, 183–91.

Grudin, J. (1981). *The organization of serial order in typing*. Unpublished doctoral dissertation, University of California, San Diego.

Keele, S. W., Pokorny, R. A., Corcos, D. M. and Ivry, R. (1985). Do perception and motor production share common timing mechanisms?: A correlational analysis. *Acta Psychologica*, in press.

Klapp, S. T. (1979). Doing two things at once: The role of temporal compatibility. *Memory & Cognition*, **7**, 375–81.

Klapp, S. T. (1981). Temporal compatibility in dual motor tasks II: Simultaneous articulation and hand movements. *Memory & Cognition*, **9**, 398–401.

Klapp, S. T., Anderson, W. G., and Berrian, R. W. (1973). Implicit speech in reading reconsidered. *Journal of Experimental Psychology*, **100**, 368–74.

Klapp, S. T., and Wyatt, E. P. (1976). Motor programming within a sequence of responses. *Journal of Motor Behavior*, **8**, 19–26.

Kozhevnikov, V. A., and Chistovich, L. A. (1965). *Speech articulation and perception*. Washington, DC: Joint Publications Research Service.

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

Levelt, W. J. M. (1984). Spontaneous self-repairs in speech: processes and representations. In M. P. R. van den Broecke and A. Cohen (Eds) *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dortrecht: Foris, pp. 105–11.

MacKay, D. G. (1969). Forward and backward masking in motor systems. *Kybernetik*, **6**, 57–64.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, **8**, 323–50.
MacKay, D. G. (1971). Stress pre-entry in motor systems. *American Journal of Psychology*, **84**, 35–51.
MacKay, D. G. (1972). The structure of words and syllables: evidence from errors in speech. *Cognitive Psychology*, **3**, 201–27.
MacKay, D. G. (1979). Lexical insertion, inflection and derivation: creative processes in word production. *Journal of Psycholinguistic Research*, **8**, 477–98.
MacKay, D. G. (1982). The problem of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, **89**, 483–506.
MacKay, D. G. (1985a). A theory of the representation, organization, and timing of action with implications for sequencing disorders. In E. A. Roy (Ed.) *Neuropsychological studies of apraxia and related disorders*. New York: North-Holland, pp. 267–308.
MacKay, D. G. (1985b). Asymmetries between the perception and production of speech. The Research Group on Perception and Action, Report # 8, ZiF, University of Bielefeld, FRG.
MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer.
MacKay, D. G., and Bowman, R. W. (1969). On producing the meaning in sentences. *American Journal of Psychology*, **82**, 23–39.
Mateer, K. (1985). Common sites for oral motor sequencing and phoneme identification: Evidence from electrical stimulation of the language cortex. Paper presented to the conference on 'Common Processes in Listening, Speaking, Reading, and Writing.' held at the ZiF, University of Bielefeld, July 1–5.
Massaro, D. W. (1979). Reading and listening. In P. A. Kolers, M. E. Wrolstad, and H. Bouma (Eds). *Processing of visible language*. New York: Plenum, pp. 331–540.
Meyer, D. E., and Gordon, P. C. (1983). Dependencies between rapid speech production and perception: Evidence for a shared sensory-motor voicing mechanism. In H. Bouma, and D. G. Bouwhuis (Eds) *Attention and Performance X*. Hillsdale, NJ: Erlbaum.
Norman, D. A. and Rumelhart, D. E. (1983). Studies of typing from the LNR research group. In W. E. Cooper (Ed.) *Cognitive aspects of skilled typewriting*. New York: Springer, pp. 45–66.
Savin, H. B., and Bever, T. E. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, **9**, 295–302.
Shaffer, L. H. (1978). Timing in the motor programming of typing. *Quarterly Journal of Experimental Psychology*, **30**, 333–45.
Shaffer, L. H. (1980). Analysing piano performance. In G. E. Stelmach, and J. Requin (Eds) *Tutorials in motor behavior*. Amsterdam: North-Holland.
Shapiro, D. C., Zernicke, R. F., Gregory, R. J., and Diestel, J. D. (1981). Evidence for generalized motor programs using gait pattern analysis. *Journal of Motor Behavior*, **13**, 33–47.
Sternberg, S., Monsell, S., Knoll, R. L., and Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.) *Information processing in motor control and learning*. New York: Academic.
Tuller, B., Kelso, J. A. S., and Harris, K. S. (1982). Interarticulator phasing as an

index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 460–72.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, **15**, 49–74.

Warren, R. M. (1974). Auditory temporal discrimination by trained listeners. *Cognitive Psychology*, **6**, 237–56.

Warren, R. M., and Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American*, **223**, 30–6.

Wickelgren, W. (1979). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice Hall.

Wing, A. M. (1978). Response timing in handwriting. In G. E. Stelmach (Ed.) *Information processing in motor control and learning*. New York: Academic.

Wing, A. M. (1980). The long and short of timing in response sequences. In G. E. Stelmach, and J. Requin (Eds) *Tutorials in motor behavior*. Amsterdam: North-Holland.