4
Perceptual Sequencing and Higher Level Activation

> The purpose of perception is not to produce an end-product (such as a percept), but to constrain actions in such a way as to continuously reveal useful aspects of the environment.
>
> (Michaels & Carello, 1981, p. 95).

> The much-worked claim that "illusions" and "failures of perception" are instances of failed inference . . . has about as much intellectual force as a cough in the night.
>
> (Turvey, Shaw, Reed, & Mace, 1980, p. 275).

Having examined the problem of sequencing in action, I turn now to the problem of perception, especially the problem of how we perceive input sequences in proper serial order when we do and improper order when we make errors. As in the previous chapter, I will begin with some general constraints that apply to any theory of perception and then construct a node structure theory of perception that incorporates these general constraints and makes predictions for future test.

General Constraints on Theories of Perception

Differences between sensation and perception, the problem of sequencing in perception, and basic phenomena such as perceptual constancies impose fundamental constraints on theories of perception. Needless to say, this chapter cannot comprehensively summarize all of the research related to these constraints. After all, perception has been the subject of many thousands of studies, and whole books have been written on categorical perception alone (Harnad, 1986), one aspect of the phenomenon of perceptual constancy. However, I do intend to touch on the basics and to put forward some strong claims as to their underlying basis.

Basic Phenomena

At least three basic phenomena must be explained in theories of perception: perceptual constancies, the category precedence effect, and effects of context on perception.

Perceptual sequencing and higher level activation. Ch 4 (pp. 62-89) in MacKay, D.G. (1987). *The organization of perception and action: A theory for language and other cognitive skills* (1-254). Berlin: Springer-Verlag.

Perceptual Constancies

As Fodor (1983) points out, constancies of form, size, color, and phonology function as follows:

> . . . to engender perceptual similarity in the face of the variability of proximal stimulation. Proximal variation is very often misleading; the world is, in general, considerably more stable than are its projections onto the surfaces of transducers. Constancies correct for this, so that in general percepts correspond to distal layouts *better than* proximal stimuli do. . . . (Fodor, 1983, p. 60)

Constancies of phonology concern the fact that phonemes do not have an invariant acoustic representation in the speech signal. People hear different allophones or acoustic variants as the same phoneme. The output side exhibits an analogous problem. The actual movements associated with producing a phoneme vary with the contexts in which the phoneme is produced.

The mechanisms responsible for phonological constancy are also responsible for categorical perception, the fact that speech perception fails to follow a continuously varied stimulus but is categorical or discontinuous in nature. For example, when voice onset time for a stimulus resembling either a /*d*/ or a /*t*/ is varied along a continuum, the resulting stimuli are not perceived as continuously varying, but as belonging to one phoneme category or the other.

Quite diverse perceptual systems have been found to exhibit categorical perception: human music perception (plucked-string versus bowed-string violin notes; Cutting & Rosner, 1974); human color perception (Lane, 1965); and the recognition of speech stimuli by primates and chinchillas (Kuhl & Miller, 1975). Moreover, humans do not *necessarily* perceive phonemes categorically. Massaro and Cohen (1976; 1983a) showed that subjects can, if appropriately instructed, use acoustic features to discriminate between test stimuli that fall within a phonemic category (see also McClelland & Elman, 1986). Another interesting exception is the fact that normal length vowels do not exhibit categorical perception (e.g., Pisoni, 1975).

Phonological constancies also exhibit trading relations. Whenever two or more cues contribute directly to a phonological distinction, say, between a voiced stop versus an unvoiced stop, one cue can be traded against the other (within limits). If one cue in a synthesized syllable is changed to favor one alternative and the other cue is changed to favor the other alternative, the effects become integrated, and perception remains constant; the change in one dimension offsets the change in the other (e.g., Massaro, 1981; Massaro & Cohen, 1976; Summerfield & Haggard, 1977).

The Category Precedence Effect

The category precedence effect concerns the fact that subjects can sometimes perceive an entire category of objects or words (e.g., letters versus digits) faster than a particular member of the category. We can use the original experiment of Brand (1971) for purposes of illustration because more recent category search

experiments (e.g., Prinz, 1985; Prinz, Meinecke, & Heilscher, 1985; Prinz & Nattkemper, 1985) corroborate the original results. Subjects in Brand (1971) were required to detect either (a) a particular digit embedded in a list of other digits or (b) *any* digit embedded within a list of letters. The results showed that response times were faster in condition (b) than in condition (a). *Any* digit among letters was identified faster than a particular digit among other digits.

How can perceiving that a character is a digit proceed faster than perceiving which particular digit it is? Is there some abstract and as yet unknown feature that distinguishes the *category* of letters from the *category* of digits? A follow-up experiment by Jonides and Gleitman (1976) ruled out this stimulus-based interpretation. The subjects were asked to detect either the *digit O* or the *letter O* as quickly as possible, with the *O* embedded either in a list of digits or in a list of letters. The *stimulus* for the letter *O* versus the digit *O* was therefore identical, but the results were as before. Subjects instructed to look for the *digit O* responded faster when the *O* was embedded within a list of letters than when it was within a list of digits. Conversely, subjects instructed to look for the *letter O* responded faster when the *O* was embedded within a list of digits than when it was within a list of letters. Category information (letter versus digit) can apparently facilitate perception independently of any possible surface feature for distinguishing one category from the other.

Context and the Part–Whole Paradox

As Fodor (1983) and others point out, the everyday fact that both prior and subsequent context facilitates the detection of letters, words, and objects is part of a theoretical paradox. Perception of a whole word, object, or scene seems to require perception of its parts, but at the same time is known to *influence* perception of its parts. Object contexts facilitate feature or line detection, and scene contexts facilitate object detection. For example, tachistoscopically presented objects are easier to identify when they form part of a real world scene than when they form part of a jumbled version of the same scene (Biederman, 1972). Similarly, a letter is easier to perceive within a familiar word than within an unfamiliar string of letters. The *D* in *WORD* is easier to perceive than the *D* in *WROD,* for example (e.g., McClelland & Rumelhart, 1981). Even acronyms induce a "word superiority effect." A letter is easier to perceive in a familiar letter string such as LSD or YMCA than in an unfamiliar letter string such as LSF or YPMC (Henderson, 1974). This finding cannot be explained in terms of bottom-up orthographic or phonological factors, and is paradoxical if identification of letters (parts) must precede identification of words (wholes). Resolving this part–whole paradox provides a basic challenge for theories of perception.

Differences Between Sensation and Perception

Differences between sensation and perception, such as those discussed in the following sections, impose additional constraints on theories of perception.

Sensations not Represented in Perception

As a general rule, at least some ongoing sensations are not represented in perception. We normally do not perceive the proximal stimulus, or full-blown pattern of sensory stimulation, but rather the distal stimulus or higher level conceptual aspects of an input. In speech perception, for example, we perceive and comprehend words but not low-level (e.g., phonemic and allophonic) characteristics of speech inputs. Similarly in visual perception, we perceive an object such as a lamp at some distance from ourselves, but we fail to perceive the disparity between the images in our two retinas that can provide the sole sensory basis for our distance judgment. Likewise in audition, we hear the sound of a car's horn as coherent and localized in space, but we fail to perceive the sensory events underlying this perception, for example, differences in arrival time of the sound to the two ears (e.g., Warren, 1982). Explaining why such sensations are not represented in perception, or more generally, why we perceive the distal, rather than proximal stimulus, imposes fundamental constraints on theories of perception.

Illusions: Perceptions not Represented in Sensation

As instances where perception fails to correspond to the input, illusions provide another challenge for theories of perception. The phonemic restoration phenomenon illustrates a typical illusion where an element missing in the input is nevertheless perceived. When an extraneous noise such as a cough or a pure tone is spliced into a magnetic recording so as to acoustically obliterate a speech sound in a word, the word sounds completely normal, and subjects are unable to tell which speech sound has been obliterated (Warren, 1970; 1982). For example, when subjects listen to a sentence containing the word *legi∗lature*, where a cough (∗) has been spliced in place of the /s/, the word sounds intact, and the missing /s/ sounds as real and as clear as the remaining acoustically present phonemes (Warren, 1970; 1982). The subjects somehow synthesize the missing /s/, and when informed that the cough replaced a single speech sound, they are unable to identify which sound is missing.

Effects of Unperceived Sensation on Behavior

A large number of findings indicate that sensations can influence behavior but at the same time fail to reach awareness. An example is the effect of allophonic variation on reaction time. Allophones are the set of perceptually indistinguishable acoustic variants of a phoneme. Subjects perceive all members of an allophone set as the same phoneme. Nevertheless, reaction time measures indicate that the acoustic differences between allophones undergo unconscious processing. When subjects make same-different judgments between acoustically presented pairs of phonemes, they are unaware of allophonic differences, but nevertheless respond "same" faster for two *identical* allophones than for two *acoustically different* allophones of the same phoneme (Pisoni & Tash, 1974).

As Fodor (1983) observed, findings of this sort are legion in studies of perceptual constancies, and these findings support the hypothesis that constancies of form, size, color, and phonology correct for the often misleading variability of proximal stimulation. Fodor also noted that "the work of the constancies would be undone unless the central systems which run behavior were required largely to ignore the representations which encode *uncorrected* proximal information" (1983, p. 60). What remains to be explained is how "the central systems which run behavior" ignore lower level variability when it comes to perception, but nevertheless respond to lower level variability when it comes to reaction time.

Sequencing in Perception

The problem of sequencing in perception is this: What mechanisms enable us to perceive and to represent input sequences in proper serial order when we do and in improper order when we make errors? Sequential perception presents as much of a challenge for psychological theories as does sequential behavior but has been relatively neglected. Studies of perception over the past century have concentrated mainly on static visual displays and have devoted relatively little attention to the perception of input sequences. In the following sections I discuss three general classes of phenomena that illustrate the problem of serial order in perception and impose constraints on theories of sequential perception.

Sequential Illusions

Sequential illusions occur whenever units coming later in an input sequence are perceived as coming sooner. Phonological fusions are one example. Phonological fusions occur, for example, when a subject wearing earphones is presented with an acoustic stimulus such as *lanket* in one ear followed by *banket* in the other ear. Even with a sizable (e.g., 200 ms) onset lag or temporal asynchrony between the stimuli, subjects often report hearing *blanket*, a fusion of the two inputs (Cutting & Day, 1975; Day, 1968). If perception accurately represented the input sequence, subjects would perceive the *l* followed by the *b*, because the input order 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* (Day, 1968).

As their name suggests, phonological fusions depend on a phonological rather than an acoustic representation of the input. Whereas phonological factors readily influence the probability of fusion, lower level factors within the acoustic analysis system do not. One of these phonological factors is "wordhood"; fusions are relatively rare when both inputs are words, but relatively common when both inputs are nonwords, 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 permissibility. Fusions always result in phonological sequences that are permissible or actually occurring within the listener's language. Percepts that violate phonological rules (e.g., *lbanket*) never occur, even when nonoccurring sequences represent the only possible fusions. For example, *bad* and *dad* never fuse, because nonoccurring sequences (*bdad* and *dbad*) are the only possible fusions.

By way of contrast, Cutting and Day (1975) found that acoustic factors have little or no effect on the likelihood of fusion. The probability of fusion remained constant when they changed the intensity and fundamental frequency of one of the fusion stimuli, or altered its allophonic characteristics by trilling an *r*.

The Perceptual Precedence of Higher Level Units

Theories of perception must explain why we sometimes perceive units that come later in an input sequence more quickly than units that come sooner. The time it takes to recognize segments versus syllables provides a good example. Subjects require less time to identify an entire syllable than its syllable-initial segment, even though the segment ends sooner than the syllable in the acoustic stimulus. The original experiment (Savin & Bever, 1970) can again be used for purposes of illustration, because many subsequent studies have replicated its basic findings and come to the same conclusion (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 such as *splay*; the vowel within the syllable, that is, *ay*; and the initial consonant of the syllable, that is, *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. Why do higher level units often take precedence in perceptual processing, and how in particular can a syllable or word be perceived before the phonemes making up the syllable or word?

Effects of Practice

Effects of practice represent a frequently overlooked constraint on theories of sequential perception. 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 speech sound, but we 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 sequences of nonspeech sounds such as *hiss-vowel-buzz-tone*. Another interpretation focuses on acoustic differences between speech versus nonspeech sequences (Bregman & Campbell, 1971). However, this second interpretation will not do for Warren's (1974) demonstration of how practice facilitates the recognition of nonspeech sequences. Subjects in Warren (1974)

repeatedly listened to initially unrecognizable sequences of nonspeech sounds such as *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. Theories of sequential perception must explain this order-of-magnitude effect of practice on sequence perception.

The Node Structure Theory of Perception

In my development of the theory so far, I argued that some of the nodes for perception and action are identical, and I illustrated some of the interconnections between these mental nodes for perceiving and producing speech. I then examined how mental nodes become activated in proper sequence during production. I turn now to the issue of perception: Which priming and activation processes involving mental nodes give rise to perception, and can the node structure theory handle the constraints on theories of perception previously discussed?

Priming is necessary for activation, and activation is necessary for perception, and I begin by discussing how activation takes place in perception. I then examine a general principle in the node structure theory whereby many of the nodes in a perceptual hierarchy only become primed rather than activated, and therefore never give rise to perceptual awareness. Various sources of evidence for this "principle of higher level activation" are discussed. Finally, I apply the node structure theory to the constraints imposed by the problem of serial order in perception.

The Most-Primed-Wins Principle in Perception

As noted in Chapter 1, the dynamic properties and activating mechanisms for mental nodes are identical in perception and production. During perception, activation within a system is sequential, requires a special activation mechanism (sequence node), and takes place at a rate specified by a timing node. However, the main sources of priming arrive bottom-up during perception, rather than top-down, as during production. Consider, for example, how the node *frequent*(adjective) becomes activated following presentation of the word *frequent* in perception. Sensory analysis and phonological nodes converge (many-to-one) to provide strong bottom-up priming to their connected nodes, and this priming summates on *frequent*(adjective), which then transmits second-order priming to ADJECTIVE, just as in production. With the next pulse from the timing node, this second-order priming enables ADJECTIVE to become activated as the most primed sequence node. Once activated, ADJECTIVE then multiplies the priming of all nodes in the *adjective* domain, but only the most primed one, normally *frequent*(adjective) in the present example, reaches threshold and becomes activated.

The mechanisms underlying the most-primed-wins principle therefore apply in the same way to activate nodes in both perception and production. This

most-primed-wins principle is especially important for explaining temporal context effects, where perceiving an ongoing input both influences and is influenced by what comes in earlier and/or later (e.g., McClelland & Elman, 1986; Salasoo & Pisoni, 1985; Warren & Sherman, 1974). When an activating mechanism is applied to some domain in the system, the most primed node always becomes activated, regardless of whether its priming arrived before, during, or after the current surface input; but because the activating mechanism for perceiving a unit is normally applied long after the unit has come and gone in the surface input (see the following discussion), both left-to-right and right-to-left context effects are to be expected under the theory.

In summary, the record or trace of an input in the node structure theory is duplex in nature (rather than unitary but malleable as in McClelland & Elman, 1986). There are two records, a priming record and an activation record. The activation record is all-or-none and self-sustaining (for a set period of time), has relatively permanent effects, and gives rise to perception. The priming record is graded, malleable, and temporary, does not self-sustain, and does not necessarily give rise to perceptual awareness. Priming decays over time when the activity of its connected (e.g., contextual) sources stops; it summates over time as long as its connected (e.g., contextual) sources of input remain active; and it becomes erased following activation of a node by its activating mechanism.

Activation of a node of course leaves intact the (rapidly decaying) priming record of the large number of other nodes that happened to have less than most-primed status at the time when the activating mechanism was applied. As we will see in Chapter 7 (and in D. G. MacKay, 1987), the priming–activation distinction provides a natural account of "subliminal effects" in perception, such as those seen in studies of ambiguity. This "dual-trace" aspect of the node structure theory contrasts sharply with McClelland and Elman's (1986) TRACE theory, where a single, unfolding record gets "crunched" (destructively altered) at regular intervals, say every 25 ms, on the basis of available "right and left" context. Interestingly, Lashley (1951) pointed to the source of evidence that may eventually distinguish between these two accounts: garden path sentences such as "Rapid writing/righting with his uninjured hand saved from loss the contents of the capsized canoe" (discussed in Chapter 2). The node structure theory predicts that the perceptual switch from "writing" to "righting," which occurs at the end of this sentence, will be very rapid, because by then the node representing "righting" will have switched from less-than-most-primed to most-primed status, and can be activated immediately. Under other theories, however, the switch from "writing" to "righting" requires a time-consuming reanalysis of the entire sentence.

Perceptual Invariance and Categorical Perception

Why does phoneme perception tend to remain invariant across a variety of acoustic signals, so that we normally hear different allophones or acoustic variants of a phoneme as identical? The reason is that segment nodes receive bottom-up connections from a large set of acoustic analysis nodes, subsets of which characterize

different allophones, or context-dependent acoustic variants of the segment (Figure 4.1).

Now, one allophone can be considered prototypical (see also Massaro, 1981) and provides a better (e.g., less error-prone) acoustic stimulus for perceiving the phoneme, because it contributes more bottom-up priming than any other allophonic variant. However, differences between allophones in the bottom-up priming of segment nodes are normally never perceived in the theory because perception requires all-or-none activation. Under normal (error-free) conditions, the same segment node will invariably become most primed in its domain and activated (perceived), regardless of which of its allophonic variants is present in the acoustic input.

As we will see, the acoustic and phonological feature information underlying phoneme identification normally becomes primed, but not activated (the principle of higher level activation discussed later in this chapter), and certainly never (consciously) perceived. Most other theories also characterize the feature information underlying phoneme identification as inaccessible (see the motor theory of Liberman et al., 1962) and/or rapidly lost (see the dual code theories of Massaro, 1981; Pisoni, 1975). What the node structure theory does is provide a much more general basis for both the rapid loss (decay of priming) and the inaccessibility (nonactivation under the principle of higher level activation, described later) of feature information that these other theories simply assume.

The same principles of the theory explain an analogous phenomenon on the output side: the fact that movements associated with producing a phoneme vary with the contexts in which the phoneme is produced. Although a single node represents a given segment in both perception and production, any given segment node is connected to many different muscle movement nodes representing acoustic variants of the phoneme. Context-dependent priming arising within the muscle movement system then determines which of these muscle movement

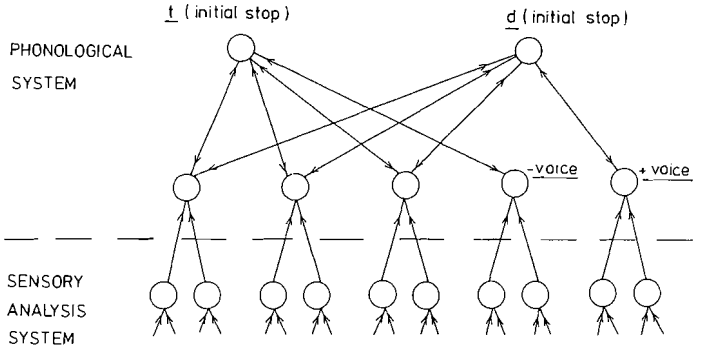

FIGURE 4.1. Connections from sensory analysis and phonological feature nodes to two segment nodes, representing syllable-initial /t/ and /d/.

muscle movement system then determines which of these muscle movement nodes becomes activated (D. G. MacKay, 1982), thereby introducing context-dependent motoric variation.

The same principles also explain why perception of speech sounds is generally categorical or discontinuous in nature. Stimuli varied along a perceptual continuum, such as voice onset time, are perceived discontinuously as belonging to one of two categories, such as /*b*/ or /*p*/, because at some point along the continuum, most-primed status will suddenly switch from one node to another within the relevant domain. For example, if an acoustic input resembles either a /*b*/ or a /*p*/, the /*b*/ node will receive more priming than the /*p*/ node when voice onset time is short, giving rise to perception of /*b*/ when the most-primed-wins principle is applied to the *stop consonant* domain. But when voice onset time is lengthened to the point where the /*p*/ node receives more priming than the /*b*/ node, a sudden discontinuity or categorical shift to perception of /*p*/ will occur. (See also McClelland & Elman's 1986 account, which is similar in some respects and goes into other aspects of categorical perception.)

Because most-primed-wins is a universal activating principle, applying at every level in every system, the phenomena of categorical perception and perceptual invariance should be universal as well. Although I have mainly used examples from speech perception to illustrate these phenomena here, the node structure theory predicts similar phenomena in other areas of perception, such as vision, touch, and music perception. The fact that categorical perception can be shown for human color perception, or for any other perceptual modality, is compatible with the theory. So is the fact that primates and chinchillas can perceive speech stimuli categorically, although it seems likely that their categories are acoustic rather than phonetic or phonological, and it remains to be explained why their acoustic nodes exhibit categorical sensitivity to voice onset times characteristic of English consonants.

Of course, the fact that synthesized vowels do *not* show categorical perception at first sight seems embarrassing for a general principle such as most primed wins. However, vowels, unlike consonants, can be described as musical chords, and if subjects are treating the vowels in categorical perception experiments as music rather than as speech (i.e., activating nodes representing patterns of pure tones), it is not surprising that these subjects fail to show boundary effects corresponding to English vowels; the categorical boundaries for acoustical tones are much narrower than those for speech. This analysis predicts categorical effects within a much narrower frequency range for subjects instructed to treat vowels as music, a prediction not shared by other theories (e.g., Pisoni, 1975) that assume inherently more persistent priming or short-term memory for vowel features than for consonant features.

Trading Relations

Under the node structure theory, trading relations illustrate how priming from lower level nodes summates at higher level nodes. An increase in bottom-up

priming arriving from one set of nodes that is sufficient to offset a decrease in priming arriving from another set of nodes will leave summated priming unchanged, so that the same segment node will be activated under the most-primed-wins principle, and perception will remain the same. Thus, if one cue in a synthesized syllable is changed to favor a voiced stop, and the other cue is changed to favor the corresponding unvoiced stop, summated priming remains the same as if no changes had been made; the change in one dimension offsets the change in the other. Needless to say, there are many other accounts of the trading-relations phenomenon (McClelland & Elman, 1986), and some have provided more detailed fits to the empirical data (e.g., Massaro, 1981). However, the node structure account differs from other accounts in several fundamental respects, such as the distinction between priming versus activation, and these differences must eventually be subjected to empirical test.

The Specialness of Speech

Like the motor theory of speech perception (Liberman et al., 1962), the node structure theory recognizes the specialness of speech among systems for perception and action. Speech systems are activated independently from other perception and action systems in the node structure theory. For example, one and the same auditory stimulus can be analyzed as a speech event, by activating the sequence nodes for the phonological system, or as a nonspeech event, by activating the sequence nodes for the auditory concept system. The staggering degree of practice that speech normally receives (D. G. MacKay, 1981, 1982) also makes speech special in the node structure theory, as does the self-inhibitory mechanism that content nodes for speech require to deal with self-produced feedback (Chapter 8). Whereas speech stimuli can be self-produced, people cannot self-produce visual stimuli such as the external world—except marginally in drawing, typing, writing or moving the eyes—and this means that not all systems representing the visual world require self-inhibitory mechanisms.

However, speech is not *fundamentally special* in the node structure theory. Similar node structures and degrees of practice can be achieved in principle, if not in practice, within other perceptual and motor systems. Moreover, although different speech and nonspeech systems and modalities differ in nodes, and perhaps also node structures or patterns of connections (D. G. MacKay, 1987), they do not differ in fundamental principles of activation.

The Category Precedence Effect

Why are categories of words and objects sometimes perceived faster than particular members of the category? In Jonides and Gleitman (1976), for example, how did subjects identify *any* digit among letters faster than a particular digit among other digits? This finding is paradoxical if it is assumed that identifying a particular exemplar is necessary for identifying its category. Dick (1971) deepened the paradox by showing that subjects can name a visually presented character from

100 ms to 200 ms faster than they can discriminate that the character is a digit versus a letter. Deepening the paradox further, Nickerson (1973) showed that identifying a character as a digit versus a letter and identifying what the character is required the same quality of information. For stimuli presented in noise, subjects failed to distinguish between a digit versus a letter unless they could also decipher which digit or letter it was (see also Prinz & Nattkemper, 1985).

Can the node structure theory explain all these seemingly contradictory findings? The basic category precedence effect follows directly from the speed–accuracy trade-off postulate of the theory (Chapter 2; and D. G. MacKay, 1982). Subjects activate the most primed node in the digit domain when instructed to look for a digit and the most primed node in the letter domain when instructed to look for a letter. Thus, with the *O* embedded in digits, errors are more likely when subjects are instructed to look for the *digit O* than for the *letter O*, because activating the most primed node in the letter domain will not suffer interference (an increase in the probability of errors) when irrelevant (extraneous) nodes in the number domain become primed, but activating the most primed node in the number domain *will* suffer interference. Because speed trades off with errors, this means that, with errors held constant, detecting the digit *O* among letters will take longer than detecting the letter *O* among letters.

However, the quality of information required to detect the *O* versus to classify the *O* as a letter or as a digit will be identical in the theory. Priming transmitted from a content node to its sequence node provides the basis for classification, and also provides the basis for identification, which can only occur when the target content node has greater priming than any other node in its domain. Moreover, the process of naming a letter is different and more direct than the process of generating a proposition such as "*O* is a letter," and it is not at all surprising that the naming process is faster.

More generally, between-category searches in the node structure theory can make use of existing or preformed connections between sequence and content nodes. To successfully detect a digit among letters, for example, the activating mechanism for the domain of digit nodes can simply be applied over and over on each trial. This general strategy has two consequences. One is that a node in the digit domain will become activated soon after it becomes primed, enabling the already discussed rapid recognition response. The other consequence is a high probability of "false alarms" to nontargets in the same category as the target. Repeated application of a sequence node will automatically activate whatever content node has greatest priming in the domain or category, regardless of whether the priming arises from a target or a nontarget. This explains an interesting finding of Gleitman and Jonides (1976) that subjects searching for a particular digit among letters respond incorrectly with very high probability to "catch trials" with a nontarget digit. For example, subjects instructed to search for a *3* among letters often respond "present" to a catch trial digit such as *6*. The reason is that the content node representing *6* will automatically become activated as the most primed digit node if the sequence node for digits is repeatedly applied.

Unlike between-category search tasks, within-category search tasks in the node structure theory cannot make use of existing connections. Successful detection of a *particular* digit embedded among other digits requires formation of new connections. A new domain, consisting in this case of a single node representing the target digit, must become established with its own special activating mechanism. Considerable practice is needed to strengthen the connections between sequence and content nodes within this new domain so as to achieve rapid reaction times for detecting the target digit among other digits. Once this practice has taken place, however, the node structure theory predicts disappearance of the category precedence effect. That is, with extensive practice, a particular digit among other digits can become as easy to detect as the same digit among letters. A similar process is required for explaining why consistent mapping conditions are superior to varied mapping conditions in the visual search tasks of Shiffrin and Schneider (1977).

The Principle of Higher Level Activation

Activation is *necessary* for perceptual awareness in the node structure theory but not *sufficient*. An additional mechanism, discussed in detail elsewhere (D. G. MacKay, 1987), is required in the theory for achieving perceptual awareness. This "consciousness mechanism" only complicates the present discussion, however, and in order to simplify exposition, I will pretend that activation is synonymous with perceptual awareness in the pages to follow.

As noted in Chapter 2, not all nodes in a bottom-up hierarchy, such as the one in Figure 2.4, become activated during perception, the way they do in a top-down hierarchy during production. Only higher level nodes normally become activated and give rise to everyday perception. In particular, I will argue that only nodes above the phonological system become activated in perceiving everyday speech. This "principle of higher level activation" is extremely general and will be illustrated with examples from visual and auditory perception, as well as from speech perception.

I begin with the logical basis for the principle of higher level activation, the fact that activating lower level nodes is *unnecessary* in perception. I then show why activating lower level nodes is *undesirable*, and I discuss the optimal level for activation to begin during everyday speech perception. Finally, I discuss various phenomena whose explanation seems to require a principle of higher level activation.

Why Lower Level Activation is Necessary in Production

To understand why activation at lower levels is unnecessary in perception, it helps to examine the reasons why activation is necessary *at all levels* in *production*. Two reasons stand out. One is that even lower level components such as segments must be produced in sequence, so that nodes *must* become activated

in production, rather than just primed, because activation is sequential whereas priming is not. The corollary fact that segment units activated during production never reach awareness (except when an error occurs; see Chapter 9) further illustrates the need for a theoretical distinction between activation and awareness.

The commitment threshold is the second reason for activating nodes at all levels in production. Recall from Chapter 1 that a minimum level of priming, designated the commitment threshold, is required to activate a node. To become activated, content nodes must not only be most primed in their domain; their priming must reach commitment level, so that multiplication of priming via the sequence node can reach activation threshold and activate the node.

In production, then, higher level nodes must pass on sufficient priming to reach the commitment threshold of the lowest level nodes in an action hierarchy in order for behavior to occur. However, because top-down connections are one-to-many, that is, they diverge rather than converge, priming from highest to lowest level nodes cannot summate during production. This means that, without activation along the way, priming transmitted to the lowest level muscle movement nodes would fall below commitment threshold.

Why Lower Level Activation is Unnecessary in Perception

Transmission of priming is quite different in perception than in production. For temporal and structural reasons discussed in the following paragraphs, lower level nodes, without themselves becoming activated, *can* pass on sufficient bottom-up priming during perception to enable higher level nodes to accumulate enough priming to reach commitment threshold and (indirectly) become activated.

Why is *bottom-up* priming passed on so efficiently? Three fundamental factors play a role. One is the fact that bottom-up connections are convergent or many-to-one (Figure 4.2). Because convergent connections allow priming to summate, lower level nodes, without themselves becoming activated, can combine or converge to pass on sufficient bottom-up priming to reach the commitment threshold of their connected nodes.

Linkage strength also contributes to the efficiency of lower level bottom-up priming during perception. Lower level connections have greater linkage strength than higher level connections (D. G. MacKay, 1982), so that priming can be efficiently transmitted via these strong bottom-up connections between lower level nodes, without the help of activation along the way.

Temporal parameters are the third factor contributing to the efficient summation of lower level bottom-up priming. Convergent priming from lower level connections is simultaneous, or nearly simultaneous. Two or more lower level nodes prime a connected node at identical or nearly identical times during speech perception. For example, all four feature nodes illustrated in Figure 4.2 prime their connected segment node at the same time. In contrast, higher level nodes generally prime a connected node at different and nonoverlapping times.

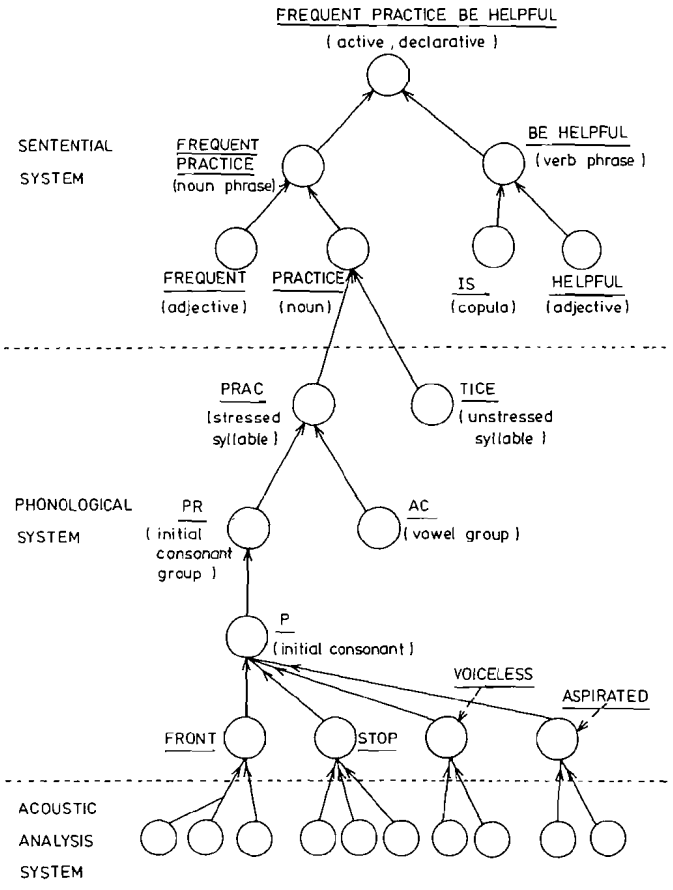

FIGURE 4.2. The bottom-up hierarchy of nodes for perceiving the sentence "Frequent practice is helpful."

Consider the arrival of convergent priming to the proposition node in Figure 4.2, for example. Priming from the verb phrase node will normally begin about 2 or 3 s after priming from the noun phrase node, and by that time, priming from the noun phrase node will already have begun to decay. In general, time lags between priming inputs will increase with the level of a node in the hierarchy, so that decay of priming will present more of a problem for higher than lower level nodes. As a result, lower level nodes generally receive greater temporal summation of priming than higher level nodes.

To summarize, higher level nodes must become activated during perception in order to transmit enough priming for connected nodes to reach commitment threshold and, indirectly, to become activated under the most-primed-wins principle. However, even when they do not become activated, lower level nodes pass on enough (second-order) bottom-up priming for their connected (higher level) nodes to reach commitment threshold and, indirectly, to become activated.

Activating these lower level nodes is therefore unnecessary, and this constitutes a preliminary or logical basis for the principle of higher level activation.

To illustrate this logical basis in greater detail, consider the bottom-up connections to the word *practice* in Figure 4.2. To facilitate exposition, assume that the sensory analysis nodes representing the acoustic input provide the equivalent of first-order priming to phonological feature nodes. Without becoming activated, each feature node therefore passes on somewhat weaker (second-order) priming to its connected segment nodes. However, because each segment node receives bottom-up connections from at least four feature nodes, the second-order priming from all four feature nodes may summate to at least the level of first-order priming from a single node. The segment nodes transmit this summated priming to their connected phonological compound and syllable nodes, and again, because of convergent summation, temporal overlap, and high linkage strength, the combined degree of second-order priming may remain comparable to that of first-order priming from a single *activated* node. Because first-order priming invariably suffices to meet the commitment threshold and, in fact, constitutes the normal basis for activation during production, activating lower level nodes is unnecessary in perception for transmitting sufficient priming to higher level nodes.

The efficient transmission of lower level bottom-up priming is only one reason why activating lower level nodes is unnecessary in perception. Another reason is that sequential perception is unnecessary for highly practiced, lower level sequences of components. Consider perception of the phonemes in the word *legislature*, for example. As long as the higher level node, *legislature*(noun), can become activated, it is always possible to reconstruct (top-down) what the lower level sequence of phonemes "must have been." I will illustrate the details of this reconstruction process later in the chapter when I discuss the phonemic restoration phenomenon, where a phoneme can be absent from an input sequence, but nevertheless perceived in sequence by top-down reconstruction, that is, by priming resulting from activation of its lexical content node.

Why Unnecessary Activation is Undesirable

Up to now, I have been arguing that activation of lower level nodes is *unnecessary* in perception: priming suffices. I now carry the argument a step further by noting that unnecessary activation is undesirable and should not occur. The main reason is that activation is more time consuming than priming and should be avoided, if possible, in order to speed up the rate of perceptual processing.

Why is activation so much slower than priming? Two reasons stand out. One is the temporal bottleneck caused by the self-inhibitory process that follows activation but not priming (Chapter 8). The other is the sequential rather than parallel nature of activation. An activating mechanism (sequence node) must first receive a buildup of priming and then become activated via a pulse from its timing node. Only then can it activate its most primed content node via multiplication of priming. Activating more than one node at a time is virtually impossible,

and rate of activation must not be so fast as to induce errors in either perception or production (D. G. MacKay, 1982). This further reduces the rate of activation, relative to priming. Activation may also require more energy or effort than priming, and this may make unnecessary activation even more undesirable.

Errors are another reason for not activating lower level nodes in perception. Perceptual inputs are much more ambiguous at lower levels than at higher levels, where ambiguity is defined in terms of the relative degree of priming of "intended" versus extraneous nodes within the same domain (see also Chapter 7). So defined, ambiguity is a major cause of errors at all levels of perception. For example, the phonological feature nodes representing + versus − consonantal will both receive some degree of priming at the point when the /s/ in the word *legislature* occurs, because acoustic cues for consonants and vowels overlap in the acoustic signal (see McClelland & Elman, 1986, among others). Input at the phonological feature level can therefore be considered relatively ambiguous, and activation of phonological feature nodes could easily result in error, that is, activation of the wrong feature node, − consonantal in this example. The resulting misperception and awareness of error would disrupt perceptual processing, making activation of phonological nodes undesirable. However, the probability of error drops sharply at higher (e.g., lexical) levels, because unlike acoustic cues for phonological features, cues for different words rarely overlap (McClelland & Elman, 1986). Higher level activation also contributes "noise resistance." An extraneous sound could completely mask the /s/ in *legislature*, for example, without changing the most primed status of *legislature*(noun), because no other node in the *noun* domain could receive comparable bottom-up priming and become activated in error (see following discussion).

Another reason for not activating lower level nodes in perception concerns one of the most fundamental purposes of perceiving: to generate adaptive action. I argue below that perception (i.e., activation) of low-level components can interfere with, rather than promote, adaptive action. I begin with the observation that actions based entirely on low-level components are neither necessary nor desirable in everyday human behavior. For example, consider the phonological nodes for producing and perceiving speech in a normal, turn-taking conversation. Activating phonological nodes during perception primes muscle movement nodes, in effect preparing the muscle movement system for *producing* the just-perceived sequence of phonemes. However, immediately repeating a just-heard phonological sequence is by and large neither necessary nor desirable in everyday conversations. What is normally required is a new and adaptive response, rather than a repetitive one, and activing phonological nodes representing the previous input could only slow down or interfere with such a response. On the other hand, activating higher level (e.g., lexical content) nodes provides the primary basis for forming new connections (D. G. MacKay, 1987), not just within the sentential system, but within other systems representing visual cognition, for example, and generating adaptive rather than repetitive responses generally requires the formation of new connections.

The Optimal Level for Activation

To summarize, activation incurs costs and benefits. Although activation costs time, and perhaps also effort, it enables perceptual awareness, which is desirable at the highest possible levels to ensure adaptive action. It follows from this analysis that activation will become cost-effective at some optimal level in the hierarchy. Below the optimal level, costs of activation (time, errors, and effort) outweigh benefits, and above the optimal level, benefits of activation (perceptual awareness and adaptive action) outweigh costs.

What is the optimal level? At what level should activation begin during everyday speech perception for example? I will argue that lexical content nodes represent this optimal level, at least for adults perceiving familiar words under favorable acoustic conditions. Consider first the degree of priming arriving at lexical content nodes, relative to higher level phrase and proposition nodes. Bottom-up linkage strength, temporal summation, and convergent summation from phonological nodes is so great that second-order priming alone can be considered sufficient to meet the commitment threshold of a lexical content node. For example, most lexical nodes have undergone thousands, and sometimes many millions, of prior activations over the course of a lifetime (D. G. MacKay, 1982), so that the strong bottom-up connections to these lexical nodes will transmit sufficient second-order priming to reach commitment threshold and permit activation.

In contrast, phrase nodes normally receive insufficient bottom-up priming unless their connected lexical nodes become activated and contribute additional, first-order priming. Linkage strength of bottom-up connections to phrase nodes is relatively weak, because phrases, like propositions, are by and large new and receive much less practice than lexical units. Most phrase nodes have undergone very few prior activations, and many have undergone none whatsoever (D. G. MacKay, 1982). As a result, second-order priming will normally fall below commitment threshold of phrase nodes. Activating lexical content nodes therefore becomes necessary for passing on sufficient priming to enable phrase nodes to become activated.

Another reason why lexical nodes are the first to require activation in perception is that words represent the first level where sequence cannot be stored in advance. As Chomsky (1957) pointed out, it is reasonable to suppose a memory representation for the sequence of phonological components making up a word, but it is unreasonable to suppose a similar representation for the sequence of words making up most sentences; there are just too many possible sentences to store them all. Thus, because activation is required for sequencing in the node structure theory, lexical content nodes must become activated in order to represent and, if necessary, retrieve the sequence of words in a sentence.

A final reason for activating lexical nodes first in perception is that lexical nodes are the first units in a bottom-up hierarchy that interconnect with mental nodes outside the language modality. In order to comprehend the word "apple,"

for example, *apple*(noun) must be connected to the visual concept nodes representing apples. Similarly, in order to name a visually perceived apple, visual concept nodes representing the apple must send a return connection to *apple*(noun). Thus, activating lexical nodes enables nodes in other mental systems to become strongly primed and activated so as to generate adaptive rather than "repetitive" responses to verbal inputs.

Flexibility of Higher Level Activation

Higher level activation is a relative rather than an absolute principle. As discussed in greater detail elsewhere (Chapter 5, and D. G. MacKay, 1987), there exists a mechanism for activating lower level systems of nodes, and this mechanism is called into play whenever an input is novel, degraded, or requires selective attention. These situations can be said to cause a downward shift in the cost-effective level for activation. Activating lower level nodes incurs costs such as reduced rates of processing, and perhaps also greater effort, but paying these costs is necessary in these situations in order to provide sufficient bottom-up priming to reach the commitment threshold of higher level nodes.

In any given experimental situation, some subjects may be more willing than others to pay the cost of lower level activation, and this may explain why studies using degraded or unfamiliar stimuli often exhibit large individual differences. These individual differences are sometimes the subject of unnecessary controversy. An example is the controversy over level of processing in studies of perceptual-motor adaptation (Repp, 1982). Some studies such as W. E. Cooper, Blumstein, and Nigro (1975, discussed in Chapter 2) obtained small but positive effects of perceptual-motor adaptation, and concluded that higher level units (mental nodes) common to perception and production were responsible for their results. Other studies failed to show perceptual-motor adaptation, and concluded that adaptation effects occur exclusively at an early stage in auditory processing, prior to phonological analysis.

Conflicting conclusions are also to be expected for less-than-optimal stimuli such as synthetically constructed nonsense syllables. If most of the subjects in one set of studies are analyzing the stimuli (in this case, activating nodes) at a sensory analysis level, whereas most of the subjects in another set of studies are analyzing the stimuli at the phonological level, conflicting results are inevitable. Individual differences in the level of activation may also be responsible for recent controversies over categorical perception (Massaro, 1981). Under the node structure theory, phonemes will be perceived categorically if phonological nodes alone become activated, but not if sensory analysis nodes also become activated. It is therefore not the case that subjects can *only* respond to speech stimuli in terms of absolute phonological categories. Subjects can apply the most-primed-wins principle *below* the segment level, even though they don't normally do this, and this unusual strategy enables perception of acoustic features for discriminating between test stimuli that fall within a phonological category, a phenomenon reported in Massaro and Cohen (1976) and elsewhere. Conflicting results

associated with phonological fusions, discussed later in the chapter, also seem attributable to individual differences in level of activation.

Evidence for the Principle of Higher Level Activation

The principle of higher level activation does not apply during *production* in the node structure theory (unlike in Dell's, 1985b, theory). In order to *produce* the phonemes of a word in proper sequence, phonological nodes must invariably become activated. The principle of higher level activation is a *perceptual* principle, and several lines of empirical evidence can be shown to support the hypothesis that phonological nodes normally become primed but not activated during *perception*.

The Recognition of Segments Versus Syllables

The fact that it takes more time to identify segment targets than syllable targets (Savin & Bever, 1970) provides strong support for the principle of higher level activation. Such findings cannot be explained if all nodes in an input hierarchy must become activated or if activation of higher level nodes always requires activation of lower level nodes. Something like the principle of higher level activation is required. That is, the subjects initially must have activated only higher level (syllable) nodes, enabling rapid perceptual recognition of syllable targets. Perception of segment targets required an extra step, activation of segment nodes via multiplication of priming from the appropriate sequence node.

Reaction Times for Allophones

The principle of higher level activation explains the large body of findings indicating that lower level information can influence behavior (via priming), but nevertheless fail to reach awareness (which requires activation). An example is the effect of allophonic variation on reaction time. Even though all members of an allophone set are perceived as the same phoneme, unconsciously processed acoustic differences between allophones nevertheless influence same–different reaction times (Pisoni & Tash, 1974). Under the principle of higher level activation, only higher level (in this case, phonological, but not sensory analysis) nodes become activated, giving rise to perceptual awareness of phonemes, but not allophones. Some allophonic variants nevertheless prime their phoneme node more strongly than others, thereby enabling it to become activated more quickly (with error criterion held constant). However, sensory analysis nodes representing different allophones do not themselves become activated and give rise to perceptual awareness, so that allophonic variants influence reaction time, but not perception.

Perception of the Distal Stimulus

As expected under the principle of higher level activation, we normally perceive the distal stimulus, or higher level conceptual aspects of an input, and not the

proximal stimulus, or pattern of sensory stimulation. In visual perception, for example, we perceive how far away an object is, but fail to perceive retinal disparities, the sufficient sensory basis for that perception. Similarly in audition, we hear the sound of a car's horn as coherent and localized in space, but we fail to perceive the sufficient sensory basis for localization, differences in arrival time of the sound at the two ears (Warren, 1982). The reason is that priming from the sensory analysis nodes representing sensory events is passed on so automatically, and so effectively, that full-fledged activation and perceptual awareness normally never occur at the sensory analysis level.

Noisy Input and the Phonemic Restoration Effect

Speech perception is remarkably insensitive to everyday noise and other input degradations (e.g., McClelland & Elman, 1986), and this efficiency is readily explained under the principle of higher level activation. For example, when an extraneous (nonspeech) noise such as a cough or pure tone acoustically obliterates a speech sound in a word, the word sounds completely normal, and subjects are unable to tell which speech sound has been obliterated (Warren, 1970). Subjects somehow synthesize the missing sound, and when informed that the cough has replaced a single speech sound, are unable to identify which sound is missing.

Phonemic restorations cannot be attributed to allophonic or coarticulatory cues in segments adjacent to the replacement sound because subjects restore the missing phoneme of a contextually appropriate word even when the extraneous sound has replaced an "incorrect" or deliberately mispronounced phoneme in a word. If allophonic cues were responsible for restorations, subjects should have perceived the *mispronounced* version instead of restoring the appropriate word (Warren, 1982). Moreover, the same missing segment can be perceived as many different phonemes, depending on the sentential context that precedes or follows the replacement sound (Warren & Sherman, 1974). Finally, similar restorations occur in other perceptual modules, where explanation in terms of coarticulatory cues is out of the question. For example, when an extraneous noise replaces a note in a familiar melody, the missing note undergoes perceptual restoration in the same way as the missing phoneme in a word (Warren, 1982).

Phonemic restorations are readily explained under the principle of higher level activation. For example, consider the sentence "The state governors met with their respective legi*latures convening in the capital city" (from Warren, 1970). Lexical content nodes become activated first under the principle of higher level activation, and for the input *legi*lature*, *legislature*(noun) will acquire greatest priming because of both bottom-up and top-down (right and left contextual) priming. Even though the cough (*) has obliterated the *s* in the acoustic waveform, no other node in the (noun) domain is likely to acquire as much priming. *Legislature*(noun) therefore becomes activated under the most-primed-wins principle, and contributes top-down priming to its connected nodes, including, *is*(vowel group), and *s*(final consonant group). By applying the most-primed-wins principle to the (final consonant group) domain, *s*(final consonant group) can therefore become activated, causing clear perception of the obliterated *s*.

This is not to say that bottom-up priming arising from the replacement sound cannot influence the restorability of a speech sound. As Samuel (1981) points out, acoustic similarity between original and replacement sound plays a role in how readily the original sound is restored. With fricativelike white noise as the replacement sound, fricatives are better restored than vowels, but the opposite is true with a pure tone replacement. However, although acoustic similarity can influence restorability, it is not necessary for the occurrence of restoration under the node structure theory. Restorations should still occur for replacement sounds that are completely unlike any speech sound whatsoever.

Finally, I stress again that the principle of higher level activation only represents the *normal* processing strategy for perceiving and comprehending speech. Not all tasks elicit this normal strategy. In the task of phoneme monitoring, for example, subjects search for a particular speech sound in an incoming sentence and respond as quickly as possible after perceiving it. Here, activating the target phoneme on the basis of bottom-up priming represents a superior strategy to higher level activation, which would slow the subjects down. This observation is consistent with Foss and Blank (1980) and Foss and Gernsbacher's (1983) evidence that phoneme detection is basically a bottom-up process in the phoneme monitoring task. (See McClelland and Elman, 1986, for a related account that goes into greater detail on the issue of when top-down effects are and are not observed in speech perception.)

Context and the Part-Whole Paradox

Many findings can be used to illustrate the facilitative effects of context, including the just-discussed phonemic restoration phenomenon. Restorations of acoustically obliterated speech sounds occur when word and sentential context enables a lexical content node to become most primed and activated. Activation of the lexical content node in turn causes top-down priming of phonological nodes, enabling the node for the missing speech sound to receive greatest priming in its domain and become activated under the most-primed-wins principle. The result is clear perception of the missing speech sound replaced or acoustically obliterated by the cough.

More generally, two related aspects of the theory are required to explain facilitative effects of context on the detection of words and objects: (1) top-down priming arising from the identity of nodes for perception and production and (2) the most-primed-wins principle, which ensures that the most primed node becomes activated, regardless of whether it receives its priming from above or from below.

Consider now the part–whole paradox, which is the fact that perception of a whole word, scene, or object seems to require perception of its parts, but at the same time, can influence perception of its parts, as in the "word superiority" effect. Effects of the whole on perception of its parts are readily explained under the principle of higher level activation. Normally, only the higher level node representing the whole word or object becomes activated; lower level nodes representing parts only become primed. Moreover, perception of the whole only

requires priming from *some* of its parts and never *requires* activation (i.e., perception) of any of its parts. Effects of the whole on subsequent perception of its parts are therefore unremarkable, because perception of the whole primes all of its parts top-down.

The part-whole paradox bears a theoretical relationship to the category precedence effect under the node structure theory. To illustrate this relationship, consider Bruner's (1957) demonstration that one and the same character can be perceived as a *B* among a sequence of letters but as a *13* among a sequence of numbers. How does context (numbers versus letters) bring about these differing perceptions? As in the category precedence phenomenon, such context effects are abstract in nature; the perceiver expects either numbers or letters *in general*, not a *specific* number or a *specific* letter. How can the abstract category of an unidentified stimulus precede and determine how the stimulus is perceived? Again the node structure theory provides a simple account of this and other examples of categorical context effects illustrated in Neisser (1976). The preceding (contextual) characters determine whether the activating mechanism (sequence node) for numbers or for letters becomes engaged, which in turn determines whether the most primed content node in the domain of the letters (B) or numbers (13) becomes activated, leading to perception of a letter versus a number.

Serial Order in Perception

What are the mechanisms whereby we perceive and represent input sequences in proper serial order when we do, and improper order when we make errors? The problem of perceptual sequencing has been virtually ignored in psychology, and is often considered trivial and uninteresting. The reason seems to lie in an implicit but fundamental assumption that has become built into virtually every theory of perception and memory published to date. Under this "*sequential isomorphism assumption*," perceptual sequences invariably mirror the external sequence of events in the real world: "first in" is "first perceived."

Sequential isomorphism usually holds in perception, but not always. Any theory of perception must explain why we usually perceive events in the order in which they occur, but there exist whole classes of striking and well-documented exceptions to this general rule, and we have already encountered several in the present chapter. I discuss the significance of these nonisomorphisms first, and then develop the node structure theory of sequential perception, show how it handles the nonisomorphisms, and examine some of its predictions for future test.

Violations of Sequential Isomorphism

Phonological fusions represent a clear violation of sequential isomorphisms. When presented with the acoustic stimulus *lanket* to one ear, followed 200 ms later with *banket* to the other ear, subjects should perceive the *l* followed by the *b*, given sequential isomorphism, because the order of arrival of the input is

acoustic *l* followed without overlap by acoustic *b*. The fact that most subjects do not perceive the input this way, but instead fuse the inputs, and report that the *b* preceded the *l*, therefore violates sequential isomorphism (Day, 1968).

Phonemic restorations also violate sequential isomorphism. When subjects hear a sentence containing a speech sound masked by a cough (*), they 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. The detection of clicks in sentences provides another paradigmatic violation of sequential isomorphism (see Fodor et al., 1974, for a general review). Finally, the fact that subjects can recognize syllables before syllable-initial segments (Savin & Bever, 1970) violates sequential isomorphism, because syllable onsets precede syllable offsets and so should be perceived first, given sequential isomorphism.

The Node Structure Theory of Sequential Perception

Perception of sequence depends on the sequence in which nodes become *activated* under the node structure theory and not on the sequence in which they become *primed*. Sequence in perception and in the external world can therefore exhibit nonisomorphisms. Although priming necessarily mirrors the detailed sequence of events in the real world, node activation does not, and node activation determines sequential perception. Because only higher level nodes normally become activated and give rise to perception (the principle of higher level activation), the priming of lower level nodes representing the sensory sequence doesn't necessarily determine the sequence perceived. For example, only the lexical content node for a familiar word normally becomes activated in perception, so that segments making up the word become primed but not activated and perceived in sequence.

Novel Sequences

Why is perception of rapidly presented and unfamiliar sequences so difficult? The main reason under the node structure theory is that perception of sequence requires activation. Priming is fundamentally simultaneous rather than sequential, at least for content nodes. A content node receives simultaneous priming from any number of other content nodes, with no indications as to sequence (D. G. MacKay, 1982). Reconstructing a sequence requires that sequence nodes become engaged, so that activation can occur, and this activation process takes time (see preceding discussion, and Chapter 7).

Why is the added time required for activation especially problematic for unfamiliar or novel sequences? One reason is that perceiving novel sequences requires activation of lower level nodes. The fact that a sequence is novel means that there are no nodes for representing the higher level components of the sequence. This means that sequential activation and perception must occur especially rapidly for novel sequences because inputs arrive more rapidly at lower levels than at higher levels. By way of illustration, compare the relative rates of

activation required for word nodes versus segment nodes. If lexical nodes for words must be activated every 500 ms, on the average, phonological nodes for segments would have to be activated about five times as rapidly, say, one every 100 ms on the average. Without top-down help, then, perceiving a phoneme sequence requires activation at five times the rate that perceiving a word sequence requires. However, activation takes time and has a maximal rate, so that lower level sequential perception will break down at some rate of input where higher level sequential perception can still occur. Perceiving novel sequences therefore requires relatively slow rates of input because novel sequences require activation of lower level nodes (see preceding discussion).

As we will see, familiar sequences also enjoy the advantage of enabling listeners to reconstruct what the lower level sequence "must have been," and this reconstruction process is not possible for unfamiliar or novel sequences. Perceiving a novel sequence requires formation of an appropriate hierarchy of connections between content nodes, and each of these content nodes must become connected to a sequence node. Because forming these new connections requires additional time, and perhaps also extensive practice (D. G. MacKay, 1987), novel sequences can only be perceived at relatively slow rates of presentation or following extensive practice. The reason we perceive sequences of speech sounds so quickly and so effortlessly is that we have already had so much practice at doing so (D. G. MacKay, 1981; 1982).

Perceptual Lags

When the appropriate nodes and connections for representing a familiar input sequence such as a word have been formed and strengthened, the time course of perception is no longer locked into that of the input under the theory. The time to perceive becomes flexible, so that input and perception can proceed at different and variable rates, within limits. Indeed, because of the problem of ambiguity, discussed previously in this chapter and in Chapter 6, perception not only *can* but *should* lag behind the input by a considerable period.

How long a lag can be tolerated between input and perception? Limits to the lag are set by the degree of priming and its rate of decay for the nodes in question. Lags cannot be so long that priming decays below the commitment threshold of higher level nodes. The dichotic listening task illustrates the general nature of this limit. When subjects in dichotic listening experiments shadow one channel, or activate nodes representing what has been said on that channel, they can subsequently perceive what has been said simultaneously on the other channel up to several seconds earlier (D. G. MacKay, 1973a; 1987; Norman, 1969). This lag between input and perception is only possible because priming takes several seconds to decay. Sufficient priming remains after a few seconds so that the nodes representing information arriving on the other (unattended) channel can still be activated and give rise to perception. Of course, with delays longer than a few seconds, so much of the priming for an unactivated node will have decayed that activation and perception can no longer occur.

Applications of the Theory

I now reexamine the violations of sequential isomorphism, discussed earlier, in order to illustrate how the node structure theory handles the constraints these violations impose.

Phonological Fusions

How are phonological fusions explained under the node structure theory? Simultaneous presentation of forms such as *banket* and *lanket* automatically prime higher level nodes representing phonological compounds, syllables, and words, and using the principle of higher level activation (the normal strategy for everyday perception), the fusion, *blanket*, is the only possible perception. Because there are no lexical content nodes for *banket* or *lanket*, *blanket*(noun) will receive more priming than any other lexical node and automatically become activated. Even though *l* precedes *b* in the input, a fusion such as *lbanket* is impossible, because speakers of English don't have a node for representing, say, the syllable *lban*. A 200-ms temporal asynchrony between *lanket* and *banket* of course *can* be perceived at the sensory analysis level, but only by abandoning the principle of higher level activation and by adopting the unusual perceptual strategy of activating sensory analysis nodes.

The fact that appropriate sentential contexts increase the probability of fusing two simultaneous inputs is explained under the theory in the same way as other context effects. For example, the probability of fusing *pay* and *lay* to give perception of *play* increases in the context "The trumpeter will pay/lay for a while" (Cutting & Day, 1975). The reason is that the sentence context primes (top-down) the lexical node *play*(verb), which therefore acquires more priming than *pay*(verb) and *lay*(verb), so as to become activated under the most-primed-wins principle.

The node structure theory predicts that fusion responses will have greater frequency of prior occurrence than their input stimuli at the lexical, syllable, and phonological compound levels (all other factors being equal). The reason is that the lexical content node for a familiar word has connections with high linkage strength and accumulates more priming than the node for the potential fusion, which cannot therefore become activated under the most-primed-wins principle. This explains why common words tend not to fuse. Consider, for example, the dichotically presented stimuli *pin* and *sin* and their only possible fusion response, the lower frequency word, *spin*. When lexical content nodes for *pin*, *sin*, and *spin* become primed in the phonological fusion task, the high-frequency alternatives, *pin* and *sin*, because of the greater linkage strength of their connections, accumulate more priming than the lower frequency alternative, *spin*. As a result, the stimulus words, *pin* and *sin*, become perceived under the most-primed-wins principle, and not the potential fusion response, *spin*. Word stimuli fuse less often than nonword stimuli for a similar reason: Nonword stimuli, such as *lanket* and *banket* have no lexical content nodes that compete for priming with the fusion response (*blanket*).

Consider now the individual differences contributing to fusion versus nonfusion (Day, 1968). One hypothesis attributes these individual differences to the level at which nodes are becoming activated. Fusers are activating nodes at higher (lexical and phonological) levels, using the principle of higher level activation, whereas nonfusers are activating nodes at lower (sensory analysis) levels, that is, below the phonological level where fusions can take place. By activating nodes below the normal "level of processing" for everyday speech, nonfusers therefore achieve more accurate perception of the actual acoustic sequence.

Nonfusers could of course apply this strategy more generally to other input modalities, which may explain Keele and Lyon's (1982) demonstration that nonfusers for speech tend to be nonfusers for tones, accurately perceiving the order of tones presented simultaneously with a slight onset lag. Fusers for speech likewise tend to be fusers for tones; they experience difficulty determining which tone came first, perhaps because they are only activating higher level nodes in both systems. Under this "level of processing" explanation of these results, practice, feedback, and instructions such as, "pay attention to the sounds themselves" should suffice to transform these fusers into nonfusers.

An alternate hypothesis, suggested by Keele and Lyon (1982), holds that fusers have difficulty discriminating the order of onset for all stimuli, whether speech or nonspeech, because their timing nodes innately are less finely tuned. This being the case, the node structure theory predicts that fusers will display a timing deficit in both production and perception, because the same timing nodes control both (Chapter 5). Moreover, neither practice, feedback, nor instructions will eradicate the deficit.

Phonemic Restorations

I have already discussed the node structure account of why a phoneme that has been masked by an extraneous sound such as a cough (*) sounds as real and as clear as the remaining acoustically present phonemes, even when subjects have been informed that the cough has physically replaced a single speech sound. I turn now to the sequential issue: Why aren't subjects able to accurately locate the cough within the sequence of phonemes? Why isn't the cough perceived in its true (isomorphic) position in the sequence? Why does the cough sound sequentially independent of the phonemes of the word, as if coexisting in a separate perceptual space (Warren & Warren, 1970)?

In the node structure theory, the cough (*) is represented by content nodes that are unconnected to the speech perception nodes—there are no content nodes and serial-order rules for representing the vowel group, *i**, syllable, *gi**, word, *legi*lature*, or lexical concept, *legi*lature*. Even though speech and nonspeech noises share the same basilar membrane, perceiving nonspeech noises involves separate content and sequence nodes in an independent perceptual system, analogous in some ways to a separate sensory system. This explains why the cough (*) is poorly localized with respect to the speech sounds, and why (in part) the cough seems to coexist in the separate perceptual space from the sentence. It also explains why speech sounds replaced by silence do not become restored. Silence

is an acoustic feature of stops and becomes perceived as a speech sound in sequence with other speech sounds.

Effects of Practice

How can we perceive the serial order of sounds in words such as *sand* at rates of 20 ms per speech sound, whereas we require over 200 ms per sound for determining the order of unfamiliar sound sequences, such as *hiss-vowel-buzz-tone* (when recycled via tape loop)? This order-of-magnitude difference reflects an effect of practice under the node structure theory. Perceiving the sequence of speech sounds in *sand* depends on prior establishment of underlying nodes for the concept *sand*; the word *sand*; the stressed syllable *sand*; the initial consonant group *s*; the vowel group *and*; and the final consonant group *nd*—all of which constrain the perception of *sand* and conspire against perception of, say, *nsad* (for which there are no existing initial consonant group and syllable nodes corresponding to *ns* and *nsad*). By contrast, no underlying nodes for subsequences such as *hiss-vowel* or *buzz-tone* have been formed to constrain perception of a never previously encountered sequence, such as *hiss-vowel-buzz-tone*.

This view also captures Warren's (1974) demonstration that practice enables sequential identification of previously unfamiliar nonspeech sequences such as *hiss-vowel-buzz-tone* with durations of less than 20 ms per sound. The reason under the node structure theory is that practice enables formation of hierarchically organized higher level nodes, each representing what Warren (1974, p. 253) terms a "temporal compound," an aggregate or cluster of auditory items that is distinct from all other clusters. More specifically, nodes representing the sequence (*hiss-vowel-buzz-tone*) become connected to superordinate nodes representing, for example, a *hiss-vowel*, and a *buzz-tone*, which in turn become connected to a *hiss-vowel-buzz-tone* node. Once such mental nodes have been formed and extensively practiced, they can be activated by their corresponding sequence nodes, even with brief and recycling stimuli, because rate of priming and activation increases as a function of practice (D. G. MacKay, 1982).

Other Sequential Effects

The present chapter has only touched on some of the sequential effects that have been reported in the speech perception literature. There are others, none of which are in conflict with the node structure theory. One example is the effect of phonotactic rules or sequential constraints on phoneme identification (Massaro & Cohen, 1983b; McClelland & Elman, 1986). Another example is the fact that segment changes in "experimentally mispronounced" words are easier to detect at the beginning of a word than at the end (Marslen-Wilson & Welsh, 1978). This finding reflects the fact that prior to applying a lexical activating mechanism, priming from word-initial segments has more time to summate than priming from word-final segments. Word-initial "mispronunciations" will therefore contribute more to the total priming summated from all sources that a lexical node receives and so will play a bigger role in determining which lexical node receives most priming and gets activated when the activating mechanism is applied.