2
The Structure of Perception and Action

> I have devoted so much time to the discussion of the problem of syntax, not only because language is one of the most important products of human cerebral action, but also because the problems raised by the organization of language seem to me to be characteristic of almost all other cerebral activity. There is a series of hierarchies of organization; the order of vocal movements in pronouncing the words, the order of words in the sentence, the order of sentences in the paragraph, the rational order of paragraphs in a discourse.
>
> (Lashley, 1951, pp. 121-122)

In general approach, the present book reverses the traditional strategy discussed in Chapter 1 of treating perception and action separately, because I begin with the evidence for shared perception-production units, which play a role in both perception and action. By taking up action next, the book also reverses the traditional strategy of giving priority to perception. I attempt first to specify a detailed set of theoretical processes for sequencing and timing the production of speech and other skilled behaviors involving shared perception-production components. I then examine how these shared perception-production components give rise to perception, and I develop a theory with applications to classical perceptual problems such as categorical perception, perceptual invariance, the nature of perceptual errors, perception of the distal stimulus, perception of sequential inputs, and the problem of ambiguity in perception.

My ultimate goal is a unified and general theory, unified in the sense of dealing with all aspects of dynamic or on-line perception-production, and general in the sense of dealing with these dynamic aspects at all levels, including, in the case of speech, the muscle movement, phonological, and setential levels. Along the way, I review a wide range of empirical findings from various domains of inquiry (mainly cognitive psychology, neuropsychology, psycholinguistics, cybernetics, and motor control), but my main aim throughout is to develop the new theory in as detailed a manner as possible.

To facilitate exposition, I develop the theory in stages corresponding to issues raised in Chapter 1. What are the common components that perception and

The structure of perception and action. Ch 2 (pp. 14-38) in MacKay, D.G. (1987). *The organization of perception and action: A theory for language and other cognitive skills* **(1-254). Berlin: Springer-Verlag.**

production systems share? How do these common components function in a theory of sequencing and timing in speech production? What processes involving these common components give rise to perception? How can asymmetries between perception and action be explained in a theory incorporating shared perception–production components? What functions did common perception–production components evolve to serve? And what role does perceptual feedback play in ongoing action? The present chapter addresses the first of these issues, Lashley's (1951) hypothesis concerning shared units for perceiving and producing speech.

The Mental Node Hypothesis

The mental node hypothesis is the cornerstone of the node structure account of the relationship between perception and action. Under the mental node hypothesis, some of the nodes for perception and production are identical. These mental nodes or shared perception–production units represent neither sensory experience nor patterns of muscle movement but higher level cognitive components common to both perception and production (see also the "hidden units" of Rumelhart, McClelland, & the PDP Research Group, 1986). By definition, mental nodes are neither purely motor nor purely sensory but both, and they become active during perception, production, and cognition (e.g., internal speech). For example, mental nodes in the language modality represent phonological units, such as segments and syllables, and sentential units, such as words and phrases.

However, not all of the components for speech perception–production are shared. The basilar membrane and associated auditory pathways register speech inputs but play no role in speech production, for example. Nor do the muscles for the respiratory, laryngeal, velar, and articulatory organs contribute to speech perception. Here, then, are two separate systems that do not share both perceptual and production functions. One system contains sensory analysis nodes, which represent the patterns of auditory input. The other system contains muscle movement nodes, which represent the patterns of muscle movement for producing speech sounds.

The hypothesis at issue is whether a common set of nodes becomes primed when we perceive a word (or sentence) and when we produce it, either aloud, or within the imagination (internal speech). Although I focus on examples from speech here, this mental node hypothesis is intended to apply not just to speech, but to all systems for everyday action and perception. A common set of mental nodes is assumed to be involved, for example, when a chess player perceives and comprehends a sequence of chess moves or generates the same sequence of moves either on the board or within the imagination. The mental nodes for comprehending and generating chess moves are of course distinct from the sensory nodes that analyze the visual pattern of the chess board and from the motor nodes that generate the sequence of muscle contractions for moving the pieces.

For readers interested in other (nonspeech) perception–action systems, D. G. MacKay (1985) discusses the mental nodes involved in hammering a nail, shifting gears in a standard gear-shift automobile, and the generation of Morse code.

Figure 2.1 provides a general overview of the mental node hypothesis. The mental nodes send "top-down" outputs to the muscle movement nodes during production and receive "bottom-up" inputs from the sensory analysis nodes during perception. These sensory analysis nodes also analyze self-generated perceptual feedback, represented by the broken line in Figure 2.1. In what follows I first specify the types and structure of connections between mental nodes and review various sources of evidence for mental nodes. I then explore some implications of the mental node hypothesis for the nature of interactions between the perception and production of speech. Finally, I conclude the chapter with some limitations and possible extensions of the mental node hypothesis.

Types of Mental Nodes

Mental nodes fall into three functional classes based on their dynamic properties (discussed in Chapter 1) and on the structure of their connections with other nodes. *Content nodes* represent the form or content components of an action or perception; *sequence nodes* represent the order in which content nodes become activated; and *timing nodes* determine when to activate the sequence nodes, which in turn activate the content nodes. All three types of nodes normally play a role in both perception and production. However, I focus here on the structure

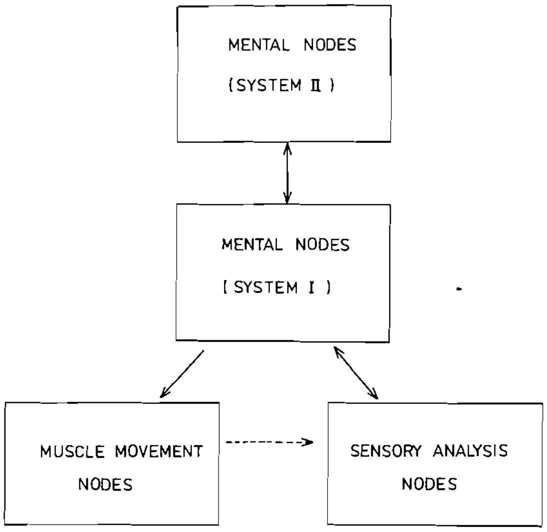

FIGURE 2.1. An overview of the mental node hypothesis. The solid arrows represent internal connections between mental nodes, muscle movement nodes, and sensory analysis nodes, while the broken arrow represents self-generated feedback.

of connections between content nodes. Indeed, when I use the term *node* in the remainder of this chapter, I refer to content nodes. I discuss sequence and timing nodes and how they interconnect with content nodes, in the subsequent chapters on processing.

The Structure of Connections Between Mental Nodes

The top-down connections between mental nodes can be described as "more-or-less hierarchic," rather than "strictly hierarchic." To illustrate this distinction, I will begin by analyzing a strict hierarchy and then discuss why, in general, top-down connections only form more-or-less hierarchies.

Top-down connections between the nodes representing the sentence "Theoretical predictions guide research" (Figure 2.2), provide an example of a strict

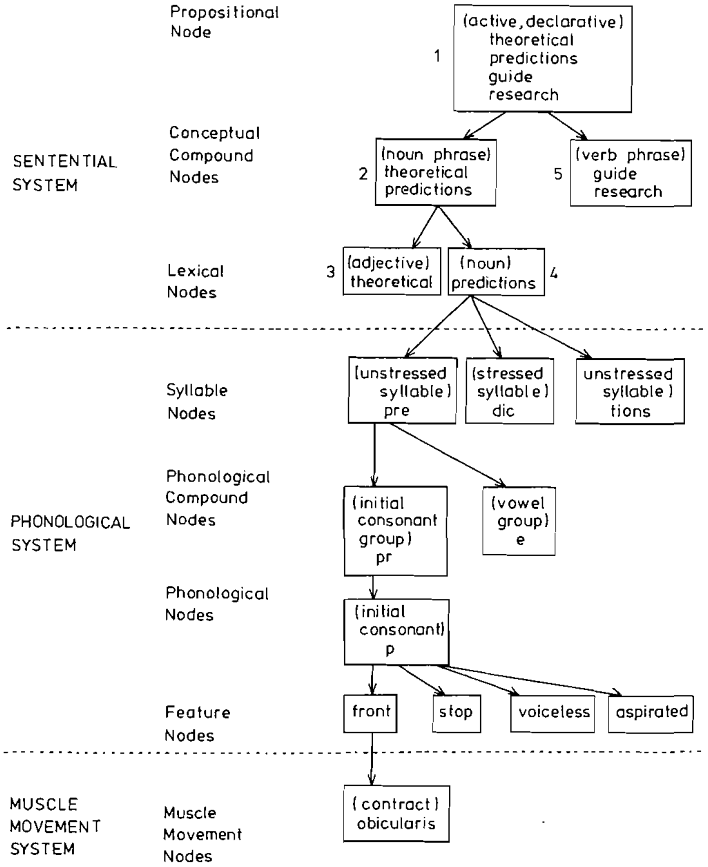

FIGURE 2.2. A sample of top-down connections for producing the preplanned sentence "Theoretical predictions guide research."

hierarchy. Following a notational convention developed in D. G. MacKay (1982), I refer to particular nodes by means of a two-component label: the content that the node represents appears in italics, followed immediately in parentheses by its sequential domain (explained later). The significance of this two-component label will become apparent when I discuss activating mechanisms in Chapter 3. Thus, the highest level node representing the entire thought underlying the sentence in Figure 2.2 has the content "Theoretical predictions guide research," occurs in the domain (active declarative), and is labeled *theoretical predictions guide research*(active declarative). This particular node is connected to two other nodes, labeled *theoretical predictions*(noun phrase) and *guide research*(verb phrase) (Figure 2.2). *Theoretical predictions*(noun phrase) is connected with two lexical nodes, *theoretical*(adjective) and *predictions*(noun). These lexical nodes are connected with specific phonological nodes, representing syllables (e.g., *pre*); phonological compounds (e.g., *pr*); segments (e.g., *p*); and features (e.g., the one representing the frontal place of articulation of *p*). Later in the chapter, I discuss some of the data supporting the particular units and connections illustrated in Figure 2.2, but the reader is referred to D. G. MacKay (1972; 1973b; 1978) and Treiman (1983) for details of the full range of supporting evidence. Numerals next to each node illustrate order of activation during production.

A more complex but otherwise similar hierarchy of nodes is assumed to underlie the control of muscle movements, but so little is known about the detailed nature and structure of connections within the muscle movement system for speech, or any other action system, that such a hierarchy cannot be represented here. Figure 2.2 illustrates nothing of this hierarchy of connections and indicates only one of the hundreds of muscle movement nodes that must become activated in producing the sentence "Theoretical predictions guide research." I simply do not know what all of the remaining muscle movement nodes are, let alone the structure of their interconnections; and even if I did, this information would be too complex to include in a form resembling Figure 2.2.

HANGING BRANCHES AND MORE-OR-LESS HIERARCHIES

Top-down hierarchies are in general only more-or-less hierarchies because some top-down connections in some node structures do not go all the way to the ground (the lowest level muscle movement nodes that give rise to behavior). These "hanging branches" are connections that exist but are not used for generating behavior in the current context. Because hanging branches do not cause their connected nodes to become activated, they represent a break in the hierarchic chain of command leading to behavior.

Hanging branches occur whenever context automatically determines the choice between two or more highly practiced response alternatives. Context-determined response specification is a very general phenomenon that occurs at all levels of a response hierarchy (D. G. MacKay, 1982; 1983), and because the mechanism is the same at all levels, I have chosen a higher level example from D. G. MacKay (1982) for purposes of illustration; the contextually determined specification of the definite versus indefinite article in English. Figure 2.3 shows the top-down connections for producing the noun phrase "the theory," in the sentence "The theory proved helpful." The node representing this noun phrase can be coded *a/the theory*(noun phrase) (see D. G. MacKay, 1982, for supporting arguments). That is, the information "definite versus indefinite determiner" isn't represented directly at the noun phrase level, but becomes specified at a lower level with the help of contextually available information. In the example under consideration, context specifies whether or not the theory in question has already been mentioned in the ongoing conversation and thereby determines the appropriate response alternative, *the*.

What is the mechanism underlying context-dependent response specification? In this particular example, the mechanism works as follows: *A/the theory*(noun phrase) is connected to *theory*(noun), and to both determiner nodes, *a*(determiner), and *the*(determiner). Each of these determiner nodes also receives a connection from another source. The other source for *a*(determiner) is a node representing the concept "new or never previously mentioned," whereas the other source for *the*(determiner) is a node representing the concept "old or

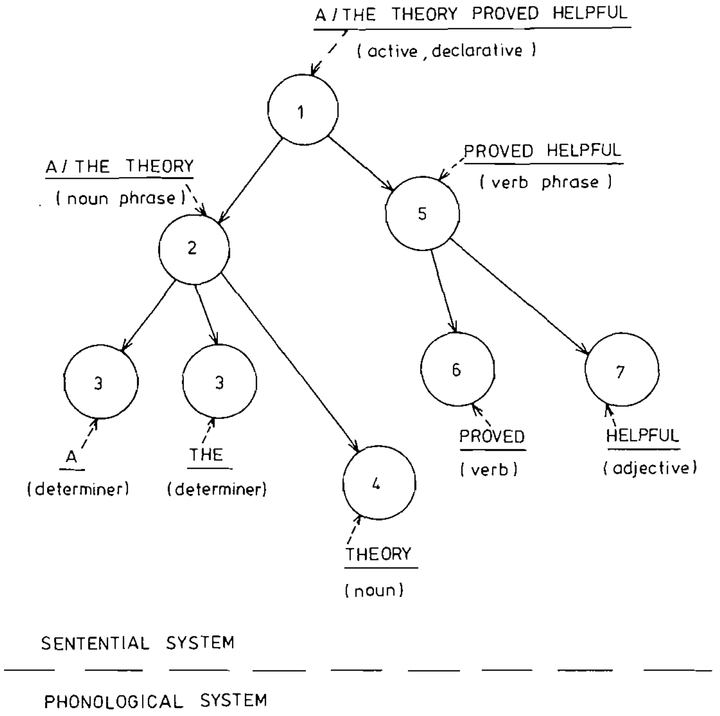

FIGURE 2.3. The top-down connections for producing the noun phrases "the theory," and "a theory," in the sentences "The theory proved helpful" and "A theory proved helpful."

previously mentioned." Because *a*(determiner) does not receive priming from its contextual source, whereas *the*(determiner) does, *a*(determiner) receives less priming than *the*(determiner), and so cannot become activated under the most-primed-wins principle when the activating mechanism is applied to the domain of determiner nodes. The connection to *a*(determiner) therefore represents a hanging branch, because it exists but is not used for generating behavior in this particular context. In general then, the most-primed-wins principle acts as an either–or gating mechanism so that when nodes in the same domain receive simultaneous priming, only the node receiving most priming from whichever (e.g., contextual) sources will become activated.

The extensiveness of context-dependent response specifications remains to be determined. For example, a similar contextual priming process could in principle select between the nouns *solely* versus *totally* in a context such as "He was solely/totally responsible for that." The process of context-dependent response specification could also help resolve the longstanding debate over the coexistence of syllable and morphological units in speech production. The debate revolves around the fact that morphemes and syllables are non-isomorphic at the surface level. Two different morphemes can map onto the same syllable, and two different syllables can map onto the same morpheme. For example, in the words *incapable* and *imprudent*, two different syllables, /in/ and /im/, represent the same negative prefix. This non-isomorphism between syllables and morphemes has led some to argue that either morphemes are a unit, or syllables, but not both. However, contextual specification via the most-primed-wins principle enables the hierarchic organization of units that are non-isomorphic at the surface level. It is perfectly possible for both syllables and morphemes to be units within the node structure theory. Both morpheme and syllable nodes represent abstract concepts, rather than surface elements per se and can connect to several different phonological nodes in the same domain. Lower level contextual sources of priming then determine which of these same-domain alternatives becomes activated. In the example under consideration, the contextual source of priming that determines whether /im-/ versus /in-/ gets produced is the place of articulation of the subsequent consonant. Moreover, many other contextually determined phonological modifications, alternative plural forms (/s/ versus /z/ versus /ez/, as in *lips*, *lids*, and *lunches*) and past tense forms (/t/ versus /d/ versus /ed/ as in *chipped*, *proved*, and *cheated*) (see Heffner, 1964, for other examples) could be determined in the same way (D. G. MacKay, 1983). For example, the lexical node for the word *proved* could connect to a syllable node, *prove*(stressed syllable), and a node representing the archiphoneme /D/, which represents the past tense abstractly by connecting with both nodes in the (voicing) domain, +*voice*(voicing), and −*voice*(voicing), as well as the other phonological feature nodes for producing /t/ versus /d/. Thus, contextual priming from the preceding consonant determines whether the −voice of /t/ or the +voice of /d/ gets activated under the most-primed-wins principle.

Action Hierarchies

An action hierarchy consists of all of the nodes that become activated in producing a preplanned behavior, including the full set of activated muscle movement nodes. Figures 2.2 and 2.3 illustrate aspects of what is, and is not, included in an action hierarchy. Figure 2.2 includes only (but not all) aspects of the action hierarchy for producing the preplanned sentence "Theoretical predictions guide research." However, Figure 2.3 includes more than just (aspects of) the action hierarchy for producing the sentence "The theory proved helpful." Nodes that receive first-order priming but do not become activated are not part of an action hierarchy, and because the hanging branch, *a*(determiner), does not become activated, it is not part of the action hierarchy for producing this particular sentence. Action hierarchies are therefore real or strict hierarchies, and not more-or-less hierarchies, and in general fail to represent the full structure of top-down connections between any given pair of nodes in the network.

Can we expect to find anatomical or neurophysiological structures in the brain that resemble action hierarchies such as the one in Figure 2.2? The likelihood of finding such structures using current technology is extremely remote. Action hierarchies are defined not by structure alone but by the occurrence of a process (activation), and we currently lack physiological definitions of either activation or priming, which would allow us to physiologically distinguish an action hierarchy from its hanging branches. Other structures must also be distinguished: the sequence nodes for activating the content nodes, the timing nodes, and other content nodes contributing connections, sometimes from other modalities. As illustrated later in the chapter, a single lexical content node typically receives connections not just from within the language modality but from many other visual, sensory, and conceptual modalities as well. And even if we could distinguish these other connections from the action hierarchy itself, anatomical action hierarchies will not be as neatly laid out as Figure 2.2. The brain lacks the systematic spatial arrangement that has been built into Figure 2.2 for ease of presentation, with the left-to-right dimension representing the order in which nodes become activated, and the up–down dimension representing the direction of priming.

Perceptual Hierarchies

Perceptual hierarchies are the input analogues of action hierarchies. They include all and only the nodes that become activated in perceiving a unitary input sequence. Nodes that only become primed, but not activated, are not part of a perceptual hierarchy. Figure 2.4 shows a typical perceptual hierarchy.

You will note that Figure 2.4 contains no sensory analysis nodes. I can say very little about how sensory analysis nodes for auditory inputs are connected to one another. Like top-down hierarchies of muscle movement nodes, bottom-up hierarchies of sensory analysis nodes are extremely complex and diverse, and the

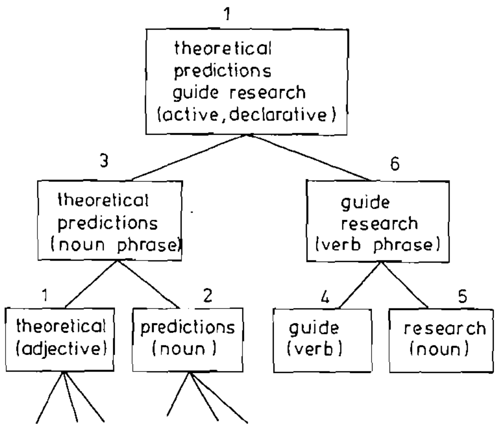

FIGURE 2.4. The perceptual hierarchy for normal comprehension of the sentence "Theoretical predictions guide research." Numerals next to each node illustrate the most likely order of activation during comprehension.

structure of their interconnections is currently unknown. For example, Lisker (1978) was able to catalogue 16 acoustic differences that could serve to distinguish a single phonological feature (the voicing of /p/ versus /b/) in a single phonological context (the words *rabid* versus *rapid*). Which sensory analysis nodes represent these acoustic differences? And what is the structure of interconnections between these nodes? All we can currently say is that acoustic analysis nodes deliver priming to phonological nodes.

This gap in our knowledge is unfortunate, but may not be especially important for an analysis of perceptual hierarchies. As we will see when I discuss perceptual processes in Chapter 4, perceptual hierarchies are quite flexible and only sometimes include sensory analysis nodes. In general, sensory analysis nodes only become primed, not activated, during everyday sentence perception. This means that sensory analysis nodes are not part of the perceptual hierarchies for normal sentence comprehension, because perceptual hierarchies only contain nodes that become activated and not just primed. Indeed, I will argue that even phonological nodes do not become activated during everyday sentence comprehension, so that Figure 2.4 represents the complete perceptual hierarchy for normal comprehension of the sentence "Theoretical predictions guide research." Numerals next to each node illustrate the most likely order of activation during comprehension.

Any given perceptual hierarchy represents only a small part of the network of bottom-up connections that become primed during perception of a unitary input sequence such as a sentence. Unlike action hierarchies, which *sometimes* represent more-or-less hierarchies, perceptual hierarchies *always* represent more-or-less rather than strict hierarchies. Hanging branches are not part of the perceptual hierarchy currently undergoing activation and constitute a universal characteristic of perception. Every activated node in a perceptual hierarchy primes many connected nodes that do not become activated. Bottom-up connections within the phonological system can be used to illustrate these hanging branches. Consider, for example, the syllable *pre* in the word *predictions*, illustrated in Figure 2.4. The syllable node, *pre*(unstressed syllable), has bottom-up connections not just with *predictions*, but with lexical nodes representing every word containing the syllable *pre*: for example, predominant, preformed, prepare, prehistoric . . . hanging branches all.

The Network in Overview

What can we say about the overall network of mental nodes for language perception–production? The existence of hanging branches means that the flow of information in the theory is neither strictly hierarchical, nor strictly heterarchical in nature. The overall network is structurally heterarchic, but functionally hierarchic. Structurally, everything can be said to connect with everything else via some relatively small number of connections in the node structure theory. The flow of priming automatically follows these existing connections and is therefore multidirectional or heterarchic in nature during both perception and action.

Functionally, however, the network is hierarchic. The activation process transforms the heterarchical connections of the overall network into local hierarchies that represent the functionally essential structures for perception and action. The next chapter discusses in detail how this activation process works: The present chapter on structure only shows that these hierarchies are there for potential use.

In the remaining chapters of the book, I argue that the heterarchic characteristics of the theory overcome the disadvantages of strictly hierarchic theories, which postulate a unidirectional flow of information but fail to explain the functional plasticity of behavior. I also argue that the hierarchic characteristics of the theory overcome the disadvantages of strictly heterarchic theories, which postulate a multidirectional flow of information but are too flexible to enable sequentially ordered action (see also Kelso & Tuller, 1981).

Systems of Nodes for Perception and Action

Functionally, nodes are organized into systems, but again only with respect to the process of activation. By definition, nodes organized into one system can be activated independently of the nodes organized into another system, and the next two chapters discuss the activation mechanisms that determine this functional organization of nodes into systems.

Figure 2.2 illustrates nodes within three different systems for perceiving and producing speech: the (speech) muscle movement system, the phonological system, and the sentential system. Activating sentential system nodes without activating nodes in the other two systems results in sequentially organized thought. Activating nodes in both the sentential and the phonological system, without activating nodes in the muscle movement system, results in internal speech (D. G. MacKay, 1981). Activating nodes in all three systems at once results in fully articulated speech. Readers interested in analogous systems of nodes for producing everyday actions such as shifting gears in a car and carrying out a preplanned shopping trip are referred to D. G. MacKay (1985).

MODALITIES FOR PERCEPTION AND ACTION

Functionally, systems of nodes are organized into modalities, once again, via the process of activation. Nodes organized into one modality can be activated independently of nodes organized into another modality. For example, the language modality includes the language comprehension systems (including connections from the basilar membrane), and the language production systems (including connections to the lungs, larynx, velum, and articulatory organs for speaking). Modalities can also contain modalities. In the case of someone who knows two languages, for example English and Ameslan (American sign language), the language modality can be said to contain an English modality and an Ameslan modality.

Systems can participate in more than one modality, and the traditional sensory organs and pathways for vision, touch, hearing, smell, and taste all participate in several modalities. For example, the basilar membrane participates in one modality when listening to speech and in another modality when comprehending complex auditory concepts such as a police siren or a familiar musical stanza. Similarly, the retina participates in one visual modality when we comprehend a printed page and in another visual modality when we comprehend complex visual concepts such as houses and trees.

Motor end organs also participate in many different modalities, as when the tongue is used for speaking versus chewing, for example. Systems of mental nodes can likewise participate in several modalities. When producing Ameslan, for example, virtually the same sentential nodes as for English become engaged, but systems for producing hand and body movements become engaged instead of the phonological and speech muscle movement systems (D. G. MacKay, 1982).

Because different modalities interconnect extensively, nodes in one modality regularly prime connected nodes in other modalities. What makes a modality modular is that its nodes can be *activated* independently from nodes in other modalities (D. G. MacKay et al., 1987).

The McGurk effect can be used to illustrate how the modalities for speech and vision can interact via priming but are independently activated. McGurk and MacDonald (1976) had subjects listen to and observe a video recording of a person saying simple syllables, their task being to identify the syllables. The auditory syllables were dubbed in synchrony with the speaker's lip movements, but the auditory syllables sometimes differed from the lip movements. The subjects' task was to say what syllable they *heard*, and the results showed that visual features such as lip closure exerted a strong effect on what phoneme the subjects reported hearing. With a conflict between visual /pa/ and auditory /ta/, for example, subjects usually reported hearing /pa/ rather than /ta/. Apparently the visual modality nodes representing facial gestures such as lip closure are connected to and prime their corresponding phonological nodes in the language modality, and thereby influence which segment node receives most priming and becomes activated under the most-primed-wins principle. However, the visual modality nodes do not themselves become activated and give rise to perception; the subjects were unaware that visual events contributed anything whatsoever to their perception (McGurk & MacDonald, 1976).

Evidence for Mental Nodes

Many findings can be seen to support the mental node hypothesis previously discussed. Here I briefly mention four very general classes of phenomena, leaving more detailed evidence and predictions for later in the book.

Parallel Empirical Effects

As expected under the mental node hypothesis, many variables have parallel effects on perception and production. Practice is one of these variables. Repetition facilitates both production (D. G. MacKay, 1982) and perception; even recognition and discrimination thresholds for sensory qualities improve as a function of practice (Woodworth, 1938).

Complexity is another variable with parallel effects on both perception and production. By way of illustration, consider the time to perceive and produce simple (one-syllable) versus complex (two-syllable) words. On the perception side, two-syllable words are harder to identify than one-syllable words with the same frequency of occurrence, the same length in letters, and the same initial segment(s). Spoer and Smith (1973) tachistoscopically presented one- versus two-syllable words and found that subjects took longer to identify the two-syllable words (e.g., *paper*) than the one-syllable words (e.g., *paint*).

On the output side, Klapp, Anderson, and Berrian (1973) likewise presented subjects with one- versus two-syllable words controlled for initial segment(s) and length in letters, but this time the subjects' task was simply to read the words aloud as quickly as possible. The dependent variable was production onset time, the time from visual presentation of the words until acoustic onset of the subject's output. The two-syllable words required slightly (15 ms) but significantly longer onset times. To rule out a perceptual interpretation of this complexity effect, Klapp et al. had subjects produce the same words in a picture-naming task, and again, production onset time was longer for two-syllable than one-syllable words. This control finding implicates a production effect, rather than a purely perceptual effect, because number of syllables is only relevant to saying the words in this condition; the input involved pictures, which do not have syllables.

In summary, complexity has parallel effects on the input and output side, and these parallel effects are readily explained under the mental node hypothesis, where two-syllable words involve more underlying nodes than one-syllable words both in perception and in production. By way of illustration, Figure 2.5 compares the mental nodes for producing *court* and *color*, words that have identical initial segments and identical length in letters but differ in number of syllables. However, more mental nodes become involved in perceiving and producing the two-syllable word *color* than the one-syllable word *court* (see Figure 2.5). Needless to say, the parallel effects of complexity on perception and production could have arisen independently in separate rather than shared node structures, but this view requires a separate explanation for the independent emergence of these parallel structures.

The mental node hypothesis also generates some new and more refined predictions concerning the relation between production onset time and the structure of words and syllables. Two factors contribute to production onset lags under the node structure theory. One concerns the set of content nodes that must become activated before the first muscle movements for producing a word or action can begin. Because activation takes time, the more underlying nodes that must be activated, the longer will be the lag that precedes production onset. This factor

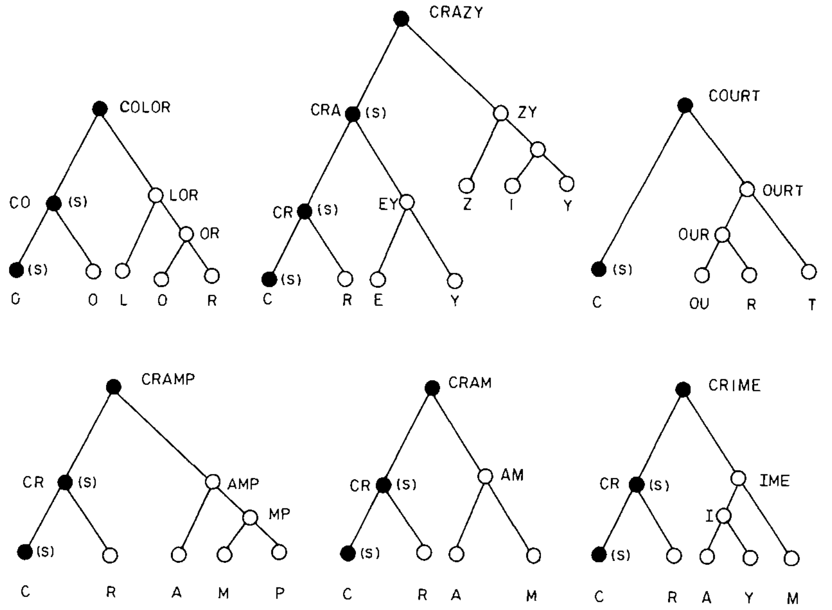

FIGURE 2.5. The structure of mental nodes for producing the words *color*, *crazy*, *court*, *cramp*, *cram*, and *crime*. Note that *crime* and *court* have equivalent length in letters, phonemes, and syllables but that *crime* and *color* have different lengths in syllables, while *cram* and *cramp* have different lengths in phonemes. (S) indicates the occurrence of a sequential decision, and filled circles indicate nodes that must be activated prior to activating the first segment node of these words.

by itself is sufficient to explain why production onset times are longer for two-syllable words than for one-syllable words that begin with the same initial segments. More mental nodes must become activated before the first segment node of a two-syllable word can become activated.

The other factor contributing to production onset lags concerns the number of sequential decisions that are required before the first muscle movements for producing the word can begin. As discussed in Chapter 3, a sequential decision is required whenever two or more nodes in different domains receive first-order priming beginning at exactly the same time. These sequential decisions take time, and the more sequential decisions that must be made, the longer the production onset time.

A detailed examination of these theoretical factors influencing production onset time leads to some new and counterintuitive predictions. The theory predicts *different* onset times for some word pairs that are equivalent in length and predicts *equivalent* onset times for other word pairs that differ in length, as measured in either syllables or segments. By way of illustration, Figure 2.5 compares the sequential decisions and mental nodes for producing the words *color*, *court*, *crime*, *crazy*, *cram*, and *cramp*. The letter *S* in parentheses (S) indicates a sequential decision, and shading indicates which nodes must be activated before the first segment node of these words can become activated. Note that *crime* and *court* have equivalent length, whether measured in letters, phonemes, or syllables. However, more sequential decisions must be made and nodes activated prior to activating the first segment node of *crime* than of *court* (Figure 2.5). The theory therefore predicts longer production onset times for words such as *crime*, which have an initial consonant cluster, than for otherwise similar words, such as *court*, which do not.

Now consider *cram* and *cramp*. These words differ in number of letters and phonemes, but they do not differ in the number of mental nodes that must be activated before the first segment node. *Cramp* only becomes more complex than *cram* after *c*(initial consonant) has become activated (Figure 2.5). In short, the theory predicts equivalent production onset times when differences between node structures arise *after* activation of the first segment node, all other factors being equal. *Crazy* and *cramp* illustrate a case where not all other factors *are* equal. As a two-syllable word, *crazy* requires more sequential decisions than *cramp* before activation of *c*(initial consonant).

Finally, consider *crime* and *color*. These words differ in number of syllables but *not* in sequential decisions and number of mental nodes prior to activation of the first segment node (Figure 2.5). The theory predicts *identical* production onset times when equivalent node structures and sequential decisions precede activation of the first segment node in one- versus two-syllable words, such as *crime* and *color*.

Interactions Between Perception and Production

The mental node hypothesis predicts interactions between perception and production involving the same mental nodes. An example is the phenomenon of perceptual–motor adaptation, which was first reported by Cooper and Nager (1975). The subjects first listened to a synthesized acoustical stimulus resembling either /pi/ or /ti/, which was repeated continuously for about a minute over earphones. To completely eliminate muscle movement factors, the subjects held a bite board firmly in their teeth and were instructed not to mouth the sounds. After this "perceptual adaptation" phase, the subjects produced aloud the syllable /pi/ or /ti/. The dependent variable during this "test" phase was voice onset time, which was determined to be the time between the release burst of the plosive and the onset of laryngeal pulsing. The results indicated that voice onset time during *production* of /pi/ and /ti/ became systematically shorter following repeated *perception* of either /pi/ or /ti/.

The mental node hypothesis provides a simple explanation of this finding. The perception and production of segments is mediated by common components (feature nodes) that become satiated as a function of repeated activation and respond less strongly to priming. Satiation of the feature node −*voice*(voicing) during perception therefore makes it more likely that +*voice*(voicing) will become activated in error under the most-primed-wins principle during production.

W. E. Cooper, Blumstein, and Nigro (1975) obtained an effect of production on perception of approximately equal magnitude, which further strengthens this conclusion. Their subjects repeatedly produced a sequence of syllables, each beginning with a labial place of articulation: /ba ma va/. There were two conditions of articulation. Under one condition the subjects repeated the syllables aloud with normal auditory feedback, and under the other they whispered the syllables while white noise masked their auditory feedback. After repeating the syllables for a minute, the subjects identified a set of auditorily presented acoustic stimuli, which varied along the place of articulation dimension: /ba/, /da/, and /ga/. Some of the subjects showed systematic adaptation effects (with or without white noise masking their auditory feedback), and these same subjects showed an equal degree of adaptation in the (standard) perceptual adaptation task, where they *listened* to repeating speech sounds and then identified other speech sounds with varying degrees of similarity to the adaptation stimuli.

The results for these subjects indicate that speech production can influence speech perception, and this finding adds further support to the hypothesis that phonological nodes provide a common substratum underlying both perception and production. It should be noted, however, that some subjects showed no effects of either perceptual or motor adaptation on the identification of speech sounds (W. E. Cooper et al., 1975). This variability seems attributable to individual differences (some subjects appear to be especially susceptible to adaptation), standard measurement error, and the small magnitude of effects in this paradigm (even statistically reliable differences only amounted to 3 ms in some cases). Needless to say, the theoretical importance of a statistically significant and independently replicated effect is not proportional to the absolute magnitude of the effect.

It should also be noted that subsequent studies reviewed in W. E. Cooper (1979) have shown that sensory analysis nodes are also subject to adaptation.

Adaptation effects diminish by anywhere from 5% to 50% when adaptation stimuli are presented to one ear and test stimuli are presented to the other ear, indicating satiation at both binaural (mental node) and monaural (sensory analysis) sites.

Shadowing Latencies

The latencies observed in the shadowing of speech inputs further illustrate the close relationship between perception and production (Kozhevnikov & Chistovich, 1965; Porter & Lubker, 1980). In shadowing experiments, subjects hear a word or sentence, which they produce aloud with as little lag as possible. The surprising result in these studies is that some subjects can shadow with lag times as short as 100 ms between acoustic onset of input and output, even with nonsense syllables as stimuli. These shadowing latencies are faster than the fastest auditory reaction times to a pure tone stimulus (about 150 ms), using either a single-alternative key press or a single-syllable response. These short shadowing times are all the more remarkable because shadowing involves a much larger set of response alternatives, a factor normally associated with *increased* reaction time. There apparently exists an intimate relationship or direct connection between mechanisms for perceiving and producing speech (see also McLeod & Posner, 1983), and this intimate relationship is directly explained under the hypothesis that the phonological nodes for perceiving and producing speech are identical.

Speed–Accuracy Trade-off in Perception

The node structure theory was originally designed to explain the trade-off relationship between time and accuracy in motor and mental skills (D. G. MacKay, 1982), and mental nodes readily capture speed–accuracy trade-offs in perceptual recognition. To recognize an object (word), the highest level node representing the object (word) must receive greater priming than any other extraneous node in its domain when the activating mechanism is applied. Whereas the priming for extraneous nodes is unpredictable, approximating a Gausian distribution with resting level as mean, priming for the appropriate or primed-from-below node summates systematically over time and must eventually exceed the priming of every other node in its domain if the stimulus duration is sufficient. But shortening the stimulus duration increases the likelihood of error, that is, the probability that some other node will be receiving more priming than the appropriate node at the time when the activating mechanism is applied.

Specific Units for Perceiving and Producing Speech

So far we have examined evidence for the hypothesis that perception and production involve identical units above the sensory analysis and muscle movement levels. This mental node hypothesis is to some extent independent of exactly what these units are and does not rise or fall on the basis of evidence for, or against, some particular unit such as, say, the syllable. The mental node hypothesis predicts only that units playing a role in perception will also play a role in production and vice versa. With this prediction in mind, let's look briefly at the evidence for specific units, first in perception and then in production.

Studies of speech perception over the past 50 years (for a review, see H. H. Clark & Clark, 1977) call for a hierarchy of abstract units including distinctive features (e.g., unvoiced); segments (e.g., /p/); syllables (e.g., *pre*); words (e.g., *predictions*); and larger sentential constituents, such as noun phrases (e.g., *theoretical predictions*); and verb phrases (e.g., *guide research*).

Available data are consistent with the hypothesis that above the sensory analysis and muscle movement levels, perception and production involve identical units. For example, recent studies of speech errors (Fromkin, 1973) indicate that the preceding perceptual units also play a role in production. Indeed, the error data for speech production go beyond the perceptual data. Many recently discovered production units have yet to be examined in studies of auditory speech perception. Within the structure of words, these recently discovered production units include word stems, stem compounds, prefixes, and suffixes, and all of these units are specific; that is, they interact only with units of the same domain or type. For example, adverbial suffixes constitute a different type of unit from past-tense suffixes, because adverbial suffixes do not substitute in error with past-tense suffixes and vice versa (D. G. MacKay, 1979).

Within the structure of syllables, the recently discovered production units include the initial consonant group, or onset (the consonant or consonant cluster preceding the vowel); the vowel group, or rhyme (the vowel and subsequent consonants within the syllable); the final consonant group, or coda (the consonants following the vowel); the vowel nucleus (a simple vowel plus a glide and/or liquid); and the diphthong (simple vowel plus glide) (D. G. MacKay, 1979).

The mental node hypothesis predicts that all of these recently discovered production units will play a role in perception and more generally that each new abstract unit discovered in studies of production will have a counterpart in perception and vice versa. Needless to say, a great deal of additional research is needed to test this general prediction. The perceptual units that are yet to be conclusively demonstrated include word stems, stem compounds, prefixes, and suffixes, initial consonant clusters, final consonant clusters, the vowel group, and in some respects, the syllable itself.

A great deal more work also remains to be done in order to apply the mental node hypothesis to the detailed nature of phonological features. For example, several findings suggest that the phonological representations of voicing and place of articulation may differ. Both Cooper, Billings, and Cole (1976) and Meyer and Gordon (1983) observed interactions between perceiving versus producing the voicing feature, but Gordon and Meyer (1984) found no such interactions between perceiving versus producing the place of articulation feature. W. E. Cooper et al. (1976) likewise experienced difficulty using the selective adaptation technique to demonstrate interactions between perceiving versus producing place of articulation. Perhaps the sensory analysis and muscle movement nodes that represent what we now call place of articulation are connected directly with segment nodes, so that no intervening feature nodes represent place of articulation per se. Although this would explain the missing interaction, it seems too early, given our current state of knowledge, to commit a general theory on this issue.

Experimentally Induced Speech Errors

The newly discovered production units discussed above have received support from three sources: studies of naturally occurring errors (e.g., D. G. MacKay, 1972), studies of the relative ease of perceiving and producing "secret languages" resembling Pig Latin and Double Dutch (Treiman, 1983; D. G. MacKay, 1973b), and studies of experimentally induced speech errors (e.g., Baars, Motley, & MacKay, 1975). To illustrate this latter source of data, I discuss in detail the study of D. G. MacKay (1978) on experimentally induced speech errors, which not only provides data on production units such as the diphthong, but suggests an interesting means of testing the prediction that identical units play a role in perception and production.

The task was as follows: Subjects heard a series of tape-recorded syllables, presented at a rate of one every 20 s, and listened for the presence of a critical segment, either /p/ or /b/, which might or might not occur in the syllable. If the syllable contained a /p/, the subjects had to change it mentally to /b/ and produce the resulting syllable as quickly as possible. Conversely, if the syllable contained a /b/, the subjects had to change it mentally to /p/ and produce the resulting syllable as quickly as possible. For example, if the subjects heard the syllable *ban*, they said *pan* as quickly as possible, and if they heard the syllable *nip*, they said *nib* as quickly as possible. As a check for possible misperceptions, after each trial the subjects wrote down the syllable that they thought had been presented.

The original purpose of the experiment was to investigate the nature of phonological rules and to test the hypothesis that the distinctive feature *voicing* constitutes an independently controllable unit within the speech production system. Interesting evidence on both of these issues was obtained. Of interest here, however, is the fact that the subjects made hundreds of errors in both perception and production. Moreover, perception and production errors were similar and had a systematic bearing on the internal organization of phonological units within the syllable. Perceptual errors were determined from what the subjects wrote down as the syllable they perceived from the tape recording. Production errors included only the incorrect responses that occurred when the stimulus was perceived correctly and were therefore operationally independent from perceptual errors. I will first examine two classes of production errors and then discuss the implications of the perceptual errors for future research.

DIPHTHONG SIMPLIFICATIONS

Diphthong simplifications were studied as a class of speech errors involving complex vowels (D. G. MacKay, 1978). Under one descriptive system, that of

Chomsky and Halle (1968), complex vowels are indivisible units that might be represented E, A, I, O, U. Under another descriptive system, that of Gleason (1961), complex vowels consist of two units that might be represented /iy, ey, aey, ow, uw/. Diphthong simplifications occurred when subjects simplified a diphthong in the transformed syllable by dropping its glide, either /w/ or /y/ in Gleason's (1961) description. A typical example involved the correctly perceived stimulus /piyk/. Subjects should have said /biyk/, but frequently said /bik/, inadvertently dropping the glide, /y/. Production errors such as these suggest that somewhere in the phonological system, diphthongs consist of a simple vowel plus a glide, as in /iy, ey, aey, aw, ow, uw/. However, this is not to say that Chomsky and Halle (1968) were wrong and that diphthongs are not also indivisible units higher up in the phonological system. *Both* descriptions are correct under the node structure theory. At one level, diphthongs constitute an indivisible unit represented by a single superordinate "diphthong node." At another level, diphthongs constitute two units, represented by a vowel node and a glide node, which send bottom-up connections to and receive top-down connections from the diphthong node.

The existence of higher level phonological units, such as diphthong nodes, suggests an interesting solution to a number of unresolved controversies concerning underlying phonological representations. Examples are the debates over the divisible versus indivisible nature of affricates such as /ch/, or rhotacized vowels such as /er/, and of velar nasals such as /-ng/. The node structure theory suggests that both sides of these debates are correct. Just as words are indivisible units at one level, but not at another, affricates, rhotacized vowels, and velar nasals are indivisible units at one level but consist of separate subcomponents at another level.

COMPLEX VOWEL SUBSTITUTIONS

Complex vowels participated in another class of errors called complex vowel substitutions in D. G. MacKay (1978). Subjects making these errors inadvertently substituted a complex vowel for a simple vowel plus a liquid, either /l/ or /r/. An example is the substitution of *powk* for *pork*, where a glide, /w/, has replaced the liquid, /r/. Although few in number and complex in nature, errors resembling these complex vowel substitutions also occur in everyday speech production. Examples, borrowed from Fromkin (1973), are the misproduction of *soup* as *serp* and *goal* as *girl*.

Taken together, complex vowel substitutions and diphthong simplification errors suggest a new domain of vowel nucleus units, which are expressed in the surface output as a simple vowel plus either a liquid or a glide. In short, nodes in the domain (vowel nucleus) each connect with two subordinate nodes: a vowel node and one other node representing either a liquid, as in the case of rhotacized vowels, or a glide, as in the case of diphthongs.

Experimentally Induced Perceptual Errors

The mental node hypothesis predicts that production units, such as the vowel nucleus discussed above, will play a role in perception and vice versa, and the procedures of D. G. MacKay (1978) suggest an interesting technique for testing this hypothesis. Consider the misperceptions of "noncritical" consonants, that is, any consonant in the syllables except for the ones the subjects were instructed to look for, /p/ and /b/. An example is the misperception of *nip* as *mip*. When noncritical consonants were misperceived, the misperceived consonant usually differed from the actual consonant by a single distinctive feature, most often in place of articulation ($p = .90$) rather than any other feature or feature cluster ($p = .10$). More importantly, place of articulation misperceptions were highly systematic; the place of articulation of the substituting consonant was usually more frontal than the actual place of articulation of the misperceived or substituted consonant. The misperception of *napt* as *mapt* provides an example. The substituting consonant, /m/, is more frontal than the substituted consonant, /n/. This bias toward perceiving a more frontal place of articulation was highly reliable and reflected the fact that subjects were instructed to listen for /b/ and /p/, which are consonants with a frontal place of articulation. When other (control) subjects simply listened to tapes of the same syllables, and wrote down what they heard, the bias toward frontal misperceptions disappeared.

These place-of-articulation misperceptions were therefore experimentally induced, and it should be possible to induce other types of misperceptions in the same way. For example, this induction technique could be adapted to test the hypothesis that vowel nucleus units play a role in perception. If subjects are instructed to press a key as rapidly as possible to indicate the occurrence of, say, a liquid (i.e., /r/ or /l/), then, just as in production, perceptual substitution errors of liquid for glide should be common occurrences, especially when the glide is part of a diphthong, as in the substitution errors *pork* for *powk*, *serp* for *soup*, and *girl* for *goal*.

Evidence for Symmetric Connections

Having discussed some general classes of phenomena that are consistent with the mental node hypothesis, I now examine one of the implications of mental nodes, namely that some of the connections between mental nodes must be symmetric or parallel. By symmetric I mean that the bottom-up connection between two nodes has a corresponding top-down connection and vice versa. Except for the lowest level mental nodes, bottom-up and top-down connections must be symmetric whenever identical nodes are involved in perception and production. By way of illustration, the connections between corresponding nodes in Figures 2.2 and 2.3 are symmetric. The bottom-up connections in Figure 2.3 parallel the top-down connections in Figure 2.2. Symmetric connections such as these help to make sense of the otherwise puzzling production phenomena and parallels between perception and production, which follow.

Bottom-up Effects in Speech Production

Symmetric connections readily explain recent evidence for bottom-up effects in speech production. As an example of one of these effects, consider the speech errors known as blends, which occur when a speaker inadvertently combines two (or more) simultaneously appropriate words (D. G. MacKay, 1972). An example is the error *sotally*, a combination of the words *solely* and *totally* in the context "He was sotally (solely/totally) responsible for that."

The main determinants of blends are syntactic and semantic similarity. As in the above example, words that become blended belong to the same syntactic class and are virtually interchangeable in meaning within their particular context of occurrence. The seemingly straightforward explanation is that the lexical content nodes for two (or more) semantically similar words in the same domain receive precisely equal priming and become activated simultaneously under the most-primed-wins principle.

However, D. G. MacKay (1973b) and Dell (1980) showed that this top-down explanation fails to account for an additional *bottom-up* effect. Specifically, D. G. MacKay (1973b) found that words involved in blends were phonologically as well as semantically similar with greater than chance probability, and Dell (1980) showed that phonological and semantic similarity independently and reliably influence these errors. Dell (1980) also reported a parallel phenomenon for word substitutions. Like blends, substituted words are usually syntactically and semantically similar (e.g., *table* and *chair*), but some (e.g., the substitution of *pressure* for *present*) (Fromkin, 1973) are phonologically similar as well. Dell (1980) also demonstrated that this phonological similarity effect exceeded chance expectation even for syntactically and semantically similar substitutions. (See Dell, 1985a; Dell & Reich, 1980; Harley, 1984; Stemberger, 1985 for other bottom-up effects taking place during speech production.)

These findings indicate that lower level processes (the phonological representation) can influence higher level processes (the selection and misselection of which word gets produced), and such phenomena are problematic for theories postulating separate perception and production components, with strictly top-down processes for production (see also Harley, 1984). However, bottom-up effects during production are readily explained in the node structure theory (as well as similar theories such as Dell, 1985a). Because the mental nodes for perception and production are identical, the bottom-up connections required for perception automatically prime the lexical content nodes for phonologically similar words, which can then become activated in error under the most-primed-wins principle. However, these phonologically similar errors will be very rare under the node structure theory, mainly because bottom-up priming is a weak (second-order) effect. Indeed, these errors seem most likely to occur when speakers have rehearsed internally what they want to say just prior to saying it, so that higher level nodes become activated twice, in two passes as it were, first during internal speech, and subsequently during overt speech. On the initial, internal speech pass, the mental node for the correct lexical item becomes activated and then self-inhibited, providing the basis for the error that occurs on the second, overt pass. Bottom-up priming arising from the first, internal speech activation will make the lexical nodes for phonologically similar words most primed if the lexical node for the correct word is still undergoing self-inhibition. However, if, as normally occurs, overt speech is produced in a single pass, without prior internal speech, the node for the correct word will become more primed than any of the nodes receiving second-order, bottom-up priming in the same domain, greatly reducing the probability of phonologically similar errors.

Perceptually Based Production Errors

Irrelevant but simultaneously ongoing perceptual input sometimes causes errors in production, and this phenomenon is difficult to explain in theories postulating separate components for production versus perception. Meringer and Mayer (1895) and Norman (1981) compiled several naturally occurring speech errors of this type, but the Stroop effect represents a well-known experimental demonstration of the same phenomenon (Norman, 1981). Subjects in Stroop studies are presented with color names printed in several different colors of ink, and the task is to ignore the word and name the color of the ink as quickly as possible. Errors are especially frequent when the color name differs from the name of the ink (e.g., the word *green* printed in red ink), and the most common error is "data driven": the printed name (*green*) substitutes the required name describing the color of the ink (*red*).

The Stroop effect is readily explained under the node structure theory, and other similar theories, where the same mental nodes are involved in perception and production and the most primed node in a domain becomes activated automatically, regardless of its source of priming. A high-frequency word such as *green* will prime *green*(color adjective) faster and more strongly than will the visually presented color green. Because the naming of a color is a relatively rare activity, color nodes will have relatively weak (i.e., slowly transmitting) connections with their corresponding word nodes. This does not mean that Stroop interference is completely describable in "race model" terms, because priming does not automatically cause activation in the theory. However, it *does* mean that color naming will either take more time, or exhibit more errors in Stroop experiments, because in order to become activated and give rise to perception, the lexical node representing the color must achieve more priming than the lexical node representing the color name.

Top-Down Effects in Perception

Symmetric connections readily explain top-down effects in both speech and visual perception. To illustrate one such effect, consider Leeper's (1936) study, in which subjects were presented with an ambiguous figure such as Jastrow's rabbit-duck, and then answered the question, "Can you see the duck?" The subject perceived the duck and not the rabbit because the question primed (top-down) the nodes representing the visuoconceptual components of ducks. With the added bottom-up priming from the figure itself, these "duck-nodes" received the most priming and became activated under the most-primed-wins principle, thereby causing perception of the duck. The "rabbit nodes," on the other hand, only received bottom-up priming, and being less primed, did not become activated, so that the rabbit was unperceived.

Extensions of the Mental Node Hypothesis

Having outlined some general sources of evidence for the mental node hypothesis, I now argue that the hypothesis as developed so far is too simple and requires extensions along the following lines.

Semisymmetric Connections

Semisymmetric connections are one of the main reasons why the evidence discussed for symmetric connections is needed. Top-down and bottom-up connections do not *always* run in parallel, even for mental nodes. Some mental nodes have some connections that are asymmetric, for example, those that contribute a bottom-up connection but receive no corresponding top-down connection. (See Grossberg, 1982, for the contrasting claim that strictly symmetric connections are in general essential for stable cognitive coding, and see Rumelhart, McClelland, & the PDP Research Group, 1986, for some models that only incorporate symmetric connections.) By way of illustrating these asymmetric connections, consider again the McGurk effect, the fact that seeing someone produce a speech sound can influence how the auditorily presented sound is perceived. Visual features such as lip closure exert a strong effect on what phoneme subjects report hearing when they see the lip movements for one syllable while hearing the sound of a different syllable (McGurk & MacDonald, 1976). Presented with a visual /pa/ and an auditory /ta/, for example, subjects usually report hearing the /pa/. Nodes representing visual lip movements apparently connect bottom-up with phonological nodes, so as to influence which phoneme node receives most priming and becomes activated.

However, there are neither logical nor empirical grounds for postulating a symmetric top-down connection between phonological nodes and the visual nodes representing lip movements. For example, hearing a speech sound over the telephone doesn't normally cause or even enable one to visualize how its production might *look*. This suggests an asymmetry. Visual nodes representing lip movements send bottom-up connections to phonological nodes but receive no top-down connections in return.

The lowest level mental nodes in an action hierarchy always have semisymmetric connections. By way of illustration, phonological feature nodes have semisymmetric connections. Connections with higher level phonological nodes are symmetric, but connections with lower level (muscle movement and sensory analysis nodes) are asymmetric. For example, phonological feature nodes send top-down connections to the muscle movement nodes for articulating speech but receive no corresponding bottom-up connections in return.

Other, higher level nodes may also have semisymmetric connections. Some mental nodes that are necessary for producing a behavioral sequence may lack corresponding bottom-up connections for reasons of structural economy and speed of processing. Consider monosyllabic words such as *desk*, for example. Producing this word requires both a lexical node, *desk*(noun), and a syllable node, *desk*(stressed syllable). Without the syllable node for monosyllabic words, speakers would be unable to produce the rhythmic timing characteristics of English (Chapter 5). However, the syllable node, *desk*(stressed syllable), may be unnecessary in perception, and may even slow down the perceptual process. That is, in monosyllabic words, phonological units such as *d*(initial consonant) and *esk*(vowel group) may connect with their lexical node directly rather than indirectly via a syllable node such as *desk*(stressed syllable). This would speed up perceptual processing but would introduce asymmetric connections at relatively high levels in the network. Testing for such high-level asymmetries is an important area for further research.

Kinesthetic and Muscle Spindle Inputs

So far I have discussed mental and muscle movement nodes as if they formed an either–or dichotomy. I represented muscle movement nodes as having no sensory or perceptual functions whatsoever. This representation is only partly correct. Although the distinction between mental and muscle movement nodes is functionally important, an analysis of kinesthetic and muscle spindle inputs suggests that this sensory–motor dichotomy is too simple. Even the very lowest level muscle movement nodes, which connect with the muscles themselves, receive direct connections from *some* sensory nodes. Specifically, sensory fibers located in spindles within the muscles connect with the lowest level alpha motorneurons, which move the muscles.

Kinesthetic feedback returns to muscle movement nodes at an only slightly higher level, perhaps still in the spinal cord. Kinesthetic input is anatomically specific and cannot be considered to connect directly with even the lowest level mental nodes representing, in the case of speech, distinctive features or phonemes. Rather, kinesthetic inputs must connect with and prime higher level muscle movement nodes.

This analysis suggests that muscle movement nodes make up a modality consisting of several hierarchically organized systems. Moreover, these muscle movement systems must themselves consist of subsystems that can be independently activated. In speech production, for example, we can activate the supralaryngeal subsystem independently from the laryngeal subsystem, which enables us to whisper, producing the same articulatory gestures but devoicing all of our speech sounds. Or we can activate the supralaryngeal articulatory subsystem independently from all other subsystems, producing lip, tongue, and jaw movements without any sound, the so-called mouthing of speech sounds.

Nature and Degree of Sensory–Perceptual Connectivity

If, as the preceding discussion suggests, the number and nature of inputs from the sensory–perceptual–cognitive systems provide the primary basis for distinguishing between mental versus muscle movement nodes, different systems of mental nodes can be distinguished in the same way. By way of illustration, compare the connections to phonological versus sentential system nodes. In particular, compare the connectivity of *p*(initial consonant) versus *pear*(noun). *P*(initial consonant) receives two possible sources of relatively direct input: from acoustic analysis nodes representing the phoneme and from visual nodes representing the lip movements. The sentential node *pear*(noun) on the other hand receives five possible sources of relatively direct input: from phonological nodes representing the word, including *p*(initial consonant); from visual concept nodes representing the visual form of a pear; from orthographic nodes representing the word *pear*; and finally, from olfactory and gustatory representations, because pears can be recognized and named from their smell and from their taste. This example further illustrates the nested nature of modalities. The visual modality contains at least three other modalities for representing visual lip movements, orthography, and visual form. Note, however, that different lexical concept nodes will receive different types of sensory–perceptual–cognitive input. For example, *dog*(noun) must receive an additional source of input from the auditory concept system, because dogs can be recognized from the sound of their bark. In general, then, nodes in higher level systems receive many connections from a variety of high- and low-level systems, whereas nodes in lower level systems receive fewer connections and mostly from low-level rather than high-level systems.