12 Asymmetries in the Relationship Between Speech Perception and Production

Donald G. MacKay
University of California, Los Angeles

What is the relation between perception and action? The present chapter describes a new theory of this relationship for the most proficient of human skills: the perception and production of speech. I develop the theory in two stages that reflect the classical dintinction between structure (hardware) and function (the real time processes that the hardware undergoes during production or perception). The structural issue concerns the relationship between the mechanisms for perceiving versus producing speech and represents a source of considerable controversy over the past several decades. Some, such as Lashley (1951), argue that perception and production share some of the same mechanisms because "the processes of comprehension and production of speech have too much in common to depend on wholly different mechanisms" (p. 186). Others have assumed separate rather than shared mechanisms for perception and production. For example, Wernicke used cases of aphasia to argue that production is localized in one area of the brain and perception in another, interconnected but separate area (see Straight, 1980). The motor theory (Liberman, Cooper, Harris, & MacNeilage, 1962; Studdert-Kennedy, Liberman, Harris, and Cooper, 1970) implicitly makes the same assumption because speech sounds such as stops are perceived with the help of the components that produce them in the motor theory, and this could only occur if the components representing stops differ for perception versus production.

MacKay (in preparation) examines in detail the available data bearing on this structural issue and argues that these and other theories that assume separate perception and production components are limited in detail and scope (e.g., none deal with the structural issue at the sentential level of the system) and are in a state of crisis (all have encountered fundamental phenomena that contradict their basic assumptions). The present chapter therefore attempts to develop a new theory wherein some of the components for perception and production are shared, and reviews various sources of evidence for these shared components (mental nodes representing higher level phonological units such as segments and syllables, and sentential units such as words and phrases).

The second, functional issue is in many ways more complex than the structural issue, even though structure places constraints on function. The issue is this: If there are shared perception-production components, how do they function in a theory of speech production? What processes involving these common components give rise to perception rather than production? And how do these processes involving shared components account for the basic facts of perception such as the regularities in perceptual errors?

The theory developed here addresses all these questions as well as some issues raised by a set of recently discovered asymmetries in the relationship between the perception and production of speech. For example, speech perception can proceed much more quickly than speech production: Computer-compressed speech remains perceptually intelligible at 5 to 7 times the rate that people can produce speech of equivalent intelligibility (Foulke & Sticht, 1969). I show here that this difference cannot be explained in terms of muscular or biomechanical factors but reflects a higher level processing difference between perception versus production. This raises the question of what these processing differences are that enable perception to proceed so much faster than production, especially if perception and production share identical higher level components. In reviewing this and many other processing asymmetries I discuss numerous empirical findings from various domains of inquiry (neuropsychology, psycholinguistics, cybernetics, motor control), but my main goal throughout is to develop the new theory of shared components in as detailed and general a manner as possible.

THE STRUCTURAL ISSUE: COMMON COMPONENTS

What are the common components underlying speech perception and production? I refer to these common components as nodes, i.e., processing units that share the same relatively simple structural characteristics and processing capabilities and respond to basic variables such as practice (repeated activation) in the same way (discussed later) during production and perception.

I begin by observing that not all of the nodes for perceiving and producing speech can be shared. The ear and associated auditory pathways register speech inputs but play no direct role in producing speech. Nor do the muscles for the respiratory, laryngeal, 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 that represent the patterns of auditory input. The other system contains nodes that represent the patterns of muscle movement for producing speech sounds.

The hypothesis at issue is whether there exists another system of nodes that

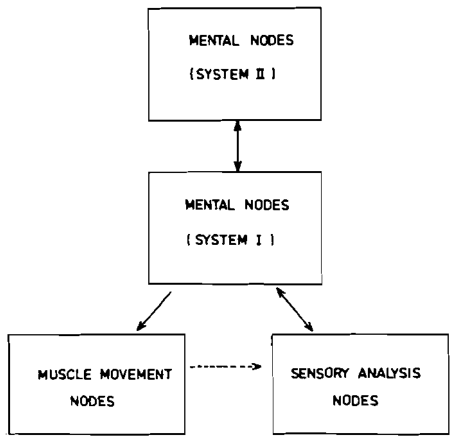

FIG. 12.1 An overview of the mental node hypothesis. The solid arrows represent internal connections between mental nodes, muscle movement nodes, and sensory analysis nodes whereas the broken arrow represents self-generated feedback.

represents neither sensory experience nor patterns of muscle movement but higher level cognitive components common to both perception and tion. Under this "mental node hypothesis," a common set of nodes is i when we perceive or imagine perceiving a segment, word, or sentence, a: we produce these components, either aloud or within the imagination speech). Although my examples here come from speech, this mental ı pothesis is not limited to speech but applies more broadly to all sys everyday action and perception. Under the mental node hypothesis, a s of mental nodes becomes involved when a chess player perceives a prehends a sequence of grandmaster chess moves or generates the same of moves either on the board or within the imagination. Such mental noc course distinct from the sensory nodes that analyze the visual patte chessboard and from the motor nodes that generate the sequence o contractions for moving the pieces.

Fig. 12.1 provides a general overview of the mental node hypoth mental nodes send "top-down" outputs to the muscle movement ı speech production and receive "bottom-up" inputs from the sensory nodes for speech perception. These sensory analysis nodes analyze pattern for the speech of others as well as self-generated feedback, re by the broken line in Fig. 12.1. Sensory analysis nodes are located in th cortex and associated sensory pathways whereas muscle movement located in the motor cortex and associated motor pathways. Muscle r and sensory analysis nodes have built-in connections at the lowest different types of highly specialized end organs and achieve much gre tice (use) than mental nodes, so that their processing can proceed at a m

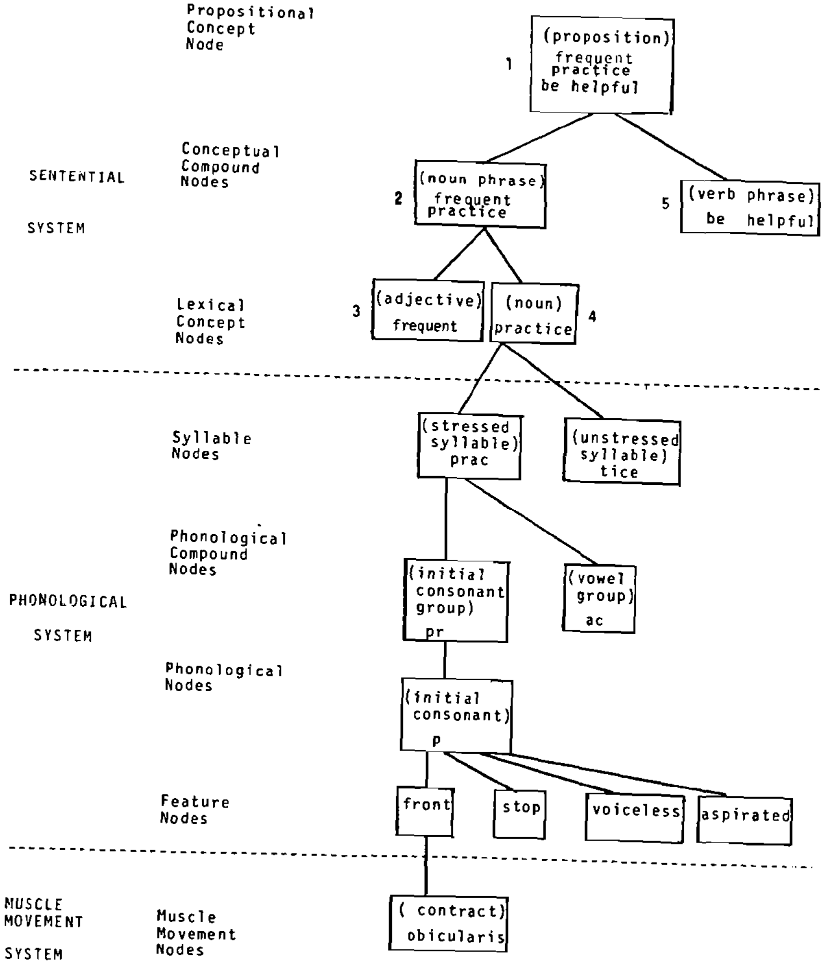

FIG. 12.2 Content nodes in the sentential, phonological, and muscle movement systems for components of the sentence "Frequent practice is helpful." The numbers adjacent the sentential components indicate order of activation during production (modified from MacKay, 1982). Lines between the phonological and sentential nodes represent both bottom-up and top-down connections. The connection to the muscle movement node (contract obicularis) is top-down only.

rate (see MacKay, 1982). However, mental, muscle movement, and sensory analysis nodes all speak the same "processing language" and communicate with one another in the same way (see following).

Mental nodes fall into three functional classes by virtue of how they connect with other nodes: *content nodes* for representing the form or content components of an action or perception; *sequence nodes* for activating and determining the

order of the components; and *timing nodes* for determining when to activate the components. All three types of nodes play a role in both perception and production. However, I concentrate here on the connections between content nodes, setting aside sequence and timing nodes and how they interconnect until the subsequent section on processing.

Content nodes must become activated in producing a sentence and are connected to one another in hierarchical fashion. I illustrate these hierarchical connections in Fig. 12.2 by means of an arbitrarily chosen sentence, "Frequent practice is helpful." Following a notational convention developed in MacKay (1982), I refer to particular nodes by means of a two-component label: The content that the node represents appears in italics followed by its sequential domain (explained later) in brackets. Thus, the highest level node representing the entire thought underlying this sentence has the content *frequent practice be helpful,* occurs in the sequential domain (active declarative) and is labeled *frequent practice be helpful*(active declarative). This node is connected to two other nodes labeled *frequent practice*(noun phrase) and *be helpful*(verb phrase) (see Fig. 12.2). *Frequent practice*(noun phrase) is connected with the lexical nodes *frequent*(adjective) and *practice*(noun). These lexical nodes are connected with specific phonological nodes, representing syllables (e.g., *prac*), phonological compounds (e.g., *pr*), segments (e.g., *p*), and features (e.g., the one representing the frontal place of articulation of *p*.)[1]

A more complex but otherwise similar hierarchy of nodes underlies 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, that such a hierarchy cannot be represented here. Bottom-up hierarchies of sensory analysis nodes are likewise extremely complex and diverse, and beyond current analysis. For example, although Lisker (1978) was able to catalogue 16 acoustic differences that could serve to distinguish a single phonological feature (the voicing of /p/ vs. /b/) in a single phonological context, both the nodes that represent such acoustic differences and the structure of their interconnections are currently unknown.

Evidence for Mental Nodes

The mental node hypothesis predicts and explains many classes of phenomena, both detailed and general. Here I briefly mention four general classes, leaving more detailed phenomena and predictions for later in the chapter.

[1]Nodes are dynamic and sequential rather than purely descriptive units (see MacKay, 1982). Thus, the word *practice* in this example requires syllable nodes in order to sequence its components, but not all words require syllable nodes. For example, a monosyllabic word such as *desk* is a sequential unit only at the lexical level, which means that its lexical node *desk*(noun) may connect directly with the sequential units *d*(initial consonant) and *esk*(vowel group) rather than a syllable node such as *desk*(stressed syllable). The reader is referred to MacKay (1972, 1973b, 1978) and Treiman (1983) for detailed evidence supporting the syllable structure implied by the connections discussed and illustrated here.

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: it facilitates both production (see MacKay, 1982) and perception (including recognition and discrimination thresholds; see Woodworth, 1938). Complexity is another. For example, Sphoer and Smith (1973) showed that the time to recognize tachistoscopically presented two-syllable words(e.g., *paper*) is longer than one-syllable words (e.g., *point*) equated for length in letters and frequency of occurrence, and, as predicted under the mental node hypothesis, Klapp, Anderson, and Berrian (1973) and others have demonstrated a parallel effect of syllabic complexity on the output side.

Interactions Between Perception and Production. The mental node hypothesis readily explains interactions between perception and production and vice versa, e.g., those demonstrated by Cooper and Nager (1975) and Cooper, Blumstein, and Nigro (1975) using adaptation techniques: As expected under the mental node hypothesis, repeated production of speech sounds (motoric adaptation) influences perception, and repeated perception (perceptual adaptation) influences production.

Shadowing Latencies. In shadowing experiments, a subject hears a word or sentence and simultaneously produces it 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 msec between acoustic onset of input and output, even with nonsence syllable stimuli; see Kozhevnikov and Chistovich (1965) and Porter and Lubker (1980). These shadowing latencies are faster than auditory reaction times for a single-alternative key press or for syllabic responses to a pure tone stimulus. These short shaddowing times are all the more remarkable because shadowing involves a very large set of response alternatives, a factor normally associated with increased reaction time. There apparently exists a highly compatible relationship or direct connection between the mechanisms for perceiving and producing speech and this compatibility is directly explained under the hypothesis that the phonological nodes for perceiving and producing speech are identical.

The Units for Perception and Production. A hundred years of research into speech perception have confirmed the need to postulate a hierarchy of abstract units, including distinctive features (e.g., unvoiced), segments e.g., $\boxed{p}$, syllables e.g., *prac*, words e.g., *practice*, and larger sentential constituents such as *frequent practice*(noun phrase) or *is helpful*(verb phrase) (See Clark & Clark (1977) for a review of relevent data). Recent studies of speech errors indicate that these same units play a role in speech production (see Fromkin, 1973), a finding consistent with the mental node hypothesis. However, the speech error data in fact go beyond the perception data, indicating additional units as yet unexamined in studies of auditory speech perception. Within the structure of words the additional production units include word stems, stem compounds, prefixes and suffixes, all specific as to type, e.g., adverbial suffixes constitute a basically different type of unit from past tense suffixes (see MacKay, 1979), and within the structure of syllables the additional production units include the initial consonant group (or onset, i.e., the consonant or consonant cluster preceding the vowel), the vowel group (or rhyme, i.e., the vowel and subsequent consonants within the syllable), the final consonant group (or coda, i.e., the consonants following the vowel), the vowel nucleus (a simple vowel plus a glide and/or liquid), and the dipthong (simple vowel plus glide) (See MacKay, 1972, 1978, 1979; and Treiman, 1983 for supporting data).

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 exhibit a counterpart in perception, and vice versa. Needless to say, a great deal of additional research is needed to test this hypothesis and its implications. One of these implications concerns phonological complexity as revealed by production onset time. As noted earlier, several investigators have reported that production onset times are usually longer for two- than one-syllable words, and the reason is that most two syllable words require the activation of more mental nodes before their first segment node can become activated. However, as MacKay (in preparation) points out, number of activations prior to production onset is not correferential with length in either syllables or segments, so that the mental node hypothesis generates some new and more refined predictions concerning the relation between production onset time and the structure of words and syllables. For example, the theory predicts production onset time differences for some word pairs with equivalent length (e.g., *crome* vs. *court*) and predicts *equivalent* production onset times for other word pairs with different lengths in either syllables (e.g., *crome* vs. *color*) or segments (e.g., *cram* vs. *cramp*). The reader is referred to MacKay (in preparation) for details underlying these predictions.

FUNCTIONAL ISSUE I: A THEORY OF OUTPUT PROCESSES INVOLVING MENTAL NODES

How do mental nodes function in the perception and production of speech? My first step in addressing this issue is to outline a theory of *production* incorporating mental nodes. The theory is an extended version of the node structure theory proposed by MacKay (1982) for explaining how practice makes behavior more fluent (faster, less prone to error) and more flexible (adapting readily to changed circumstances and transferring readily from one response mechanism to another). Minor modifications have been introduced to accommodate present purposes (to develop a unified theory of perception and production incorporating shared mental nodes), but readers familiar with the earlier theory may be inclined to skip this section.

Dynamic Properties of Mental Nodes

Mental nodes have four dynamic properties that are relevent to both perception and production: activation, priming, self-inhibition, and linkage strength.

Activation. Behavior occurs if and only if the bottom-most muscle movement nodes in a hierarchy such as the one illustrated in Fig. 12.2 become activated. Activation is always sequential and requires a special triggering mechanism to determine when and in what order the content nodes controlling the action become activated. By way of illustration, numbers adjacent to the sentential nodes in Fig. 12.2 represent order of activation.

Activation is all or none and is self-sustained, continuing for a specifiable period of time, independently of input from the sources that led originally to activation. During this period of self-sustained activation, a node simultaneously primes all nodes connected to it. A period of self-inhibition (discussed later) follows activation. Unlike other uses of the same term, activation in the node structure theory never spreads and never changes with "distance" or fatigue or the number of other nodes a node connects to.

Priming. Priming is required for activation and refers to a transmission across a connection that produces increased subthreshold activity in a connected node. The degree of priming varies with "distance" from the source: An activated node primes its connected nodes most strongly (first-order priming) whereas a node receiving first-order priming primes its connected nodes less strongly (second-order priming). Third-order priming from a single node is negligible and can be ignored in theories of production. Thus, priming spreads but only to a limited degree, and unlike other propositional network theories, activation always requires deliberate application of a special activating mechanism.

Priming summates across all simultaneously active connections and increases during the time that any given connection remains active. Consider for example how top-down priming summates during production for the numbered nodes in Fig. 12.2. Top-down connections are one-to-many, which introduces anticipatory effects into the theory (see MacKay, 1982). For example, node 1 becomes activated first, which simultaneously primes nodes 2 and 5 (see Fig. 12.2). However, node 5 cannot be activated until 2, 3, and 4 have been activated, so that priming of 5 represents "anticipatory priming", which continues to summate during the time that nodes 2, 3, and 4 are being activated. This anticipatory priming accumulates over time and facilitates the eventual activation of 5 and all other "right-branching" nodes in an output hierarchy. However, anticipatory priming also increases the probability of anticipatory errors, the most common

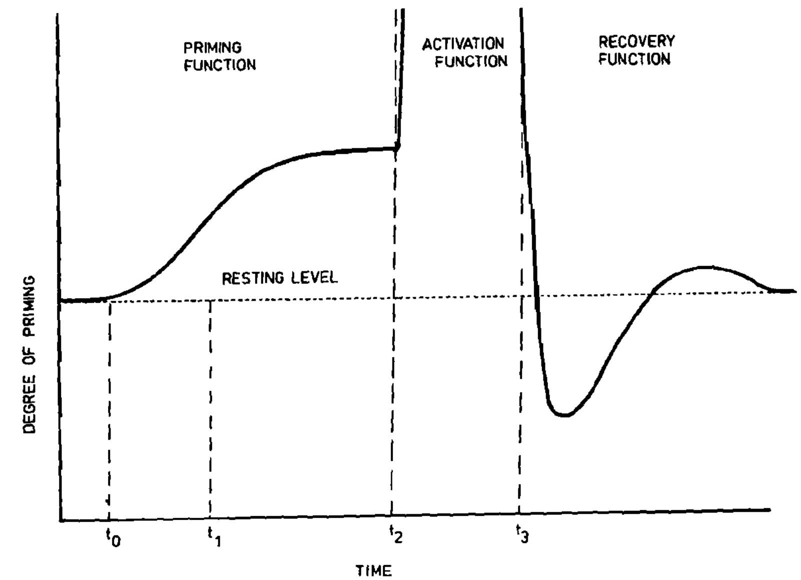

FIG. 12.3 The priming function (relating degree of priming and time following onset of priming, t_0), activation function (following application of the triggering mechanism at t_2) and recovery function (following onset of self-inhibition, t_3). See text for explanation (from MacKay and MacDonald, 1984).

class of error in speech production at either the phonological or sentential levels (see MacKay, 1982).

Unlike activation, priming is neither self-sustaining nor results in behavior when the bottom-most muscle movement nodes in an action hierarchy become primed: Priming between content nodes only summates to some subthreshold asymptotic level (see Fig. 12.3) and cannot directly cause a node to become activated. Also unlike activation, priming is order free or parallel in nature, requires no special triggering mechanism to determine when and in what order it occurs, and is not followed by a period of self-inhibition.

Self-Inhibition. After the nodes for producing components of skilled behavior become activated, they undergo a brief period of self-inhibition, during which their level of priming falls below resting level (see Fig. 12.3). Self-inhibited nodes then undergo a normal recovery cycle, which includes a period of hyperexcitability or postinhibitory rebound during which priming first rises above and then returns to resting level (see Fig. 12.3). Various sources of evidence for self-inhibition and the recovery cycle are discussed in MacKay (1986). With repeated activation of a node for prolonged periods of time (e.g., 5 minutes), fatigue sets in. During fatigue, the period of self-inhibition becomes extended, rebound from inhibition falls below normal resting level, and the node becomes generally less responsive to priming.

Linkage Strength. Linkage strength influences both the asymptotic level of priming and its rate of summation per unit time (represented by the initial slope of the priming function in Fig. 12.3). Linkage strength also determines how much and how rapidly priming becomes transmitted across a connection, and is itself determined by practice: the frequency with which a node has been activated via a particular connection in the past. As MacKay (1982) points out, linkage strength represents a long-term characteristic of a connection and explains a wide range of practice effects in the psychological literature.

The Sequential Activating Mechanism: Sequence Nodes

Mental nodes must be activated in proper sequence if an output is to be error free. Consider for example the mental nodes illustrated in Fig. 12.2. The highest level node *frequent practice be helpful*(active declarative) must be activated first. This simultaneously primes both *frequent practice*(noun phrase) and *be helpful*(verb phrase). However, only *frequent practice*(noun phrase) must become activated at this point, thereby simultaneously priming its connected nodes, *frequent*(adjective) and *practice*(noun), and so on down to the muscle movement nodes. The issue, then, is what mechanism causes ordered activation of simultaneously primed content nodes.

Sequence nodes represent that mechanism: They determine whether, when and in what order content nodes become activated. However, because each sequence node connects with and can activate any of the content nodes in its "domain," they are nonspecific in their effect (see MacKay (1982) for evidence in support of this nonspecificity). For example, the sequence node COLOR ADJECTIVE connects with and is responsible for activating all content nodes representing color adjectives (*red, green, blue, brown,* etc.), the set of nodes making up its domain. More generally, a sequential domain can be defined as set of response alternatives all of which share the same sequential privileges of occurence.

An activated sequence node multiplies the priming of every node connected with it by some large factor (e.g., 100) within a brief period of time. This multiplicative effect has no consequences for unprimed nodes but soon serves to activate (i.e., bring to threshold) the content node with the greatest degree of priming in its domain, normally the one that has just been primed from above via a connection from a superordinate content node. In producing the adjective *green,* for example, *green*(color adjective) must first become primed, either from above via a superordinate node such as *green apples*(noun phrase) or from below via say visual perception of either the color green or the printed word *green.*

Then COLOR ADJECTIVE must activate *green* (color adjective) as the most primed node in its domain. This "most-primed-wins" principle is extremely general and governs the activation of not just content nodes but sequence nodes as well (see MacKay, 1982).

A content node must of course attain some minimal degree of priming that can then be multiplied by the sequence node so as to achieve the threshold level required for activation. Below this minimal level, the degree of priming resulting from multiplication remains subthreshold, so that activation cannot occur.

Connections between sequence nodes represent serial order rules that determine order of activation for simultaneously primed sequence nodes. For the sentential system these serial order rules fall under the heading syntactic rules and for the phonological system, phonological rules. Thus, the sequence nodes ADJECTIVE and NOUN are connected in such a way as to represent the syntactic rule that adjectives precede nouns in English noun phrases.

An inhibitory connection is a simple means of achieving this order relation among sequence nodes. Under this proposal, ADJECTIVE inhibits NOUN and dominates in degree of priming when ADJECTIVE and NOUN receive simultaneous priming. However, once ADJECTIVE has been activated and returns to resting level, NOUN is released from inhibition and dominates in degree of priming, thereby determining the sequence (adjective + noun) for this and other noun phrases.

The Temporal Activating Mechanism: Timing Nodes

Timing nodes are the mechanism for the programming of timing in the node structure theory and are connected with sequence nodes in the same way that sequence nodes are connected with content nodes. Timing nodes also serve to organize the sequence and content nodes into systems. Thus, the sequence and content nodes in the aforementioned examples are part of the sentential system (see Fig. 12.2), and the sequence nodes for this system are connected with a sentential timing node. Sequence nodes within the phonological system are connected with a phonological timing node, and sequence nodes within the muscle movement system are connected with a muscle timing node.

Timing nodes become activated at specifiable points in time, priming the sequence nodes connected to them and activating the most primed one, following the most-primed-wins principle. By determining when the sequence nodes become activated, timing nodes therefore determine the temporal organization of the output. Different timing nodes have different periodicities or average rates of activation. For example, the sentential timing node exhibits a slower periodicity than the muscle timing node, because muscle flexions and extensions are produced faster than words and other sentential components.

Timing nodes also control speech rate. To determine the desired rate of speech, a speaker voluntarily adjusts the overall periodicity or pulse rate of the timing nodes (e.g., fast, normal, or slow). Control of the timing nodes also enables a speaker to selectively engage or disengage whole systems of nodes. During internal speech, the timing nodes for the sentential and phonological systems become active but not those for the muscle movement system. As a consequence, the phonological components for a sentence become activated in proper serial order and prime their corresponding muscle movement nodes (see MacKay, 1981) but no actual movement of the speech musculature ensues: Because none of the muscle sequence nodes have become activated, none of the content nodes within the muscle movement system can become activated.

A Simple Example

To illustrate how the timing and sequence nodes interact to determine whether, when and in what order the content nodes become activated in everyday speech production, consider the sequencing of the words *frequent* and *practice* in the sentence "Frequent practice is helpful." Fig. 12.4 illustrates the relevent content nodes (in rectangles), sequence nodes (in circles), and timing node (in triangle). Unbroken connections in the figure are excitatory, broken connections are inhibitory, and the dotted connection represents the inhibitory relationship between sequence nodes. Similar connections and processes are assumed for all sequentially organized mental nodes.

The node representing the sentential concept *frequent practice* (noun phrase) is activated first, simultaneously priming *frequent*(adjective)[2] and *practice*(noun), which immediately pass on second-order priming to their sequence nodes ADJECTIVE and NOUN. The inhibitory link between ADJECTIVE and NOUN temporarily reduces the priming level for NOUN so that ADJECTIVE becomes activated under the most-primed-wins principle following the first pulse from the timing node. ADJECTIVE therefore multiplies the priming of every content node in its (adjective) domain, and the one with the most priming in its domain, i.e., *frequent*(adjective), having recently been primed by *frequent practice*(noun phrase), reaches threshold soonest and becomes activated under the most-primed-wins principle.

Once a content node becomes activated, its sequence node must return quickly to resting level, because content nodes have a return connection to their sequence node that could cause reverberatory reactivation. Thus, once a content node becomes activated, it must quench or inhibit rather than further prime its

[2]Just as the domain of the words *green* and *red* is (color adjective), so the domain of *frequent* is (temporal adjective) rather than (adjective) as indicated here in order to simplify the illustration. Although all adjectives bear the same sequential relation to nouns in English, more specific domains such as (color adjective) and (temporal adjective) are needed in order to establish the appropriate sequencing among adjectives themselves. Thus, we say "frequent red lights" rather than "red frequent lights" following a sequential rule such as (temporal adjective + color adjective).

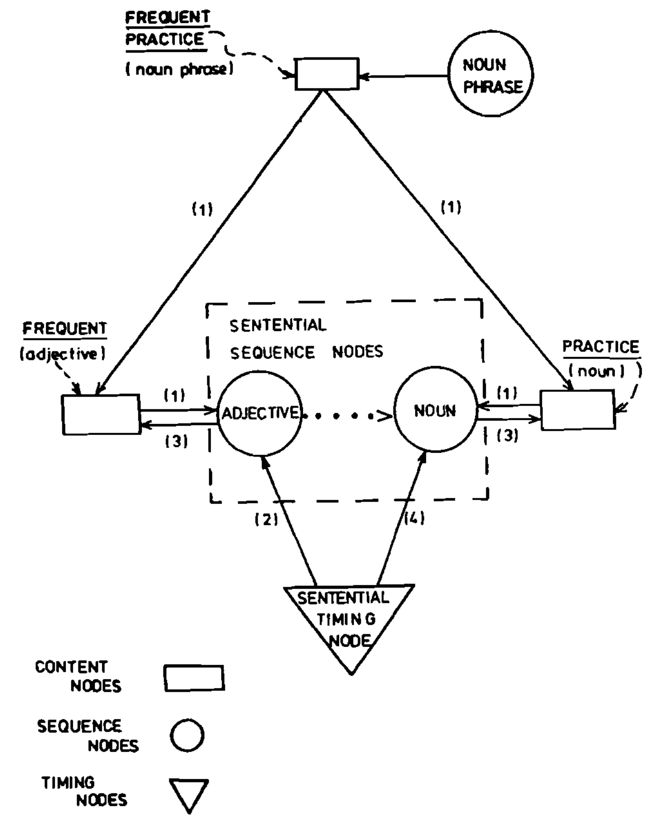

FIG. 12.4 The order of top-down processes (in brackets) underlying activation of content nodes (in rectangles), sequence nodes (in circles), and the concept timing node (in triangle) for producing the noun phrase *frequent practice* (from MacKay, 1982). The self-inhibitory and quenching mechanisms have been omitted for purposes of simplification.

corresponding sequence node so that only one content node becomes activated at any one time. This requires a special mechanism with a threshold, which if exceeded causes content nodes to inhibit rather than prime their sequence nodes.

Returning to the example in Fig. 12.4, quenching ADJECTIVE releases the inhibition on NOUN which now dominates in degree of priming and becomes activated under the most-primed-wins principle with the next pulse from the sentence timing node. NOUN therefore primes the entire domain of (noun) nodes, but having just been primed, *practice*(noun) has more priming than any other node in the domain, and becomes activated under the most-primed-wins principle.

STRUCTURAL ISSUE II: BOTTOM-UP CONNECTIONS FOR PERCEPTION

So far I have discussed some general phenomena which are consistent with the mental node hypothesis, and I have outlined a detailed theory of production involving activation of content nodes. Activation of a content node is also necessary for perception and proceeds via exactly the same structures (sequence and timing nodes) and dynamic properties (e.g., priming, and activation under the most-primed-wins principle) as production, a fact that must be kept in mind as we discuss Structural Issue II, the "bottom-up" corollary of the mental node hypothesis. Under this corollary, mental nodes are connected to one another via bottom-up connections that are necessary for perception. By way of illustration, bottom-up connections parallel the top-down connections in Fig. 12.2, so that each of the mental nodes illustrated there receive at least two bottom-up connections in addition to their single top-down connection.

To see why evidence for this corollary is needed, it is first necessary to note that top-down and bottom-up connections don't *always* run in parallel. Some nodes contribute bottom-up connections but receive no corresponding top-down connections in return. For example, there are neither logical nor empirical grounds for postulating top-down connections between phonological nodes and the visual nodes that represent facial movements such as lip closure. By way of illustration, hearing a speech sound over the telephone doesn't normally cause or enable one to visualize how its production might *look*.

However, there are good grounds for postulating bottom-up connections in the opposite direction. Consider the McGurk effect for example. McGurk and MacDonald (1976) had subjects observe a film of a person saying a simple syllable in synchrony with an auditory recording of a *different* syllable and found that visual features such as lip closure strongly influenced what phoneme the subjects reported *hearing*. Thus, with a conflict between visual [pa] and auditory [ta], subjects more often reported hearing [pa] rather than [ta]. This finding indicates that nodes representing visual events such as lip closure connect bottom-up with phonological nodes, thereby influencing which segment node receives most priming and becomes activated. And this means that top-down connections don't always parallel bottom-up ones.

Evidence for Bottom-up Connections

Parallel bottom-up and top-down connections help explain some otherwise puzzling speech production phenomena as well as some additional parallels between perception and production, discussed later.

Perceptually Based Production Errors. Irrelevant but simultaneously ongoing perceptual processes sometimes cause production errors, and this input-

output interaction 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 (see 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 for the ink (e.g., the word *green* printed in red ink): subjects erroneously substitute the printed name (*green*) for the required name representing the color of the ink *(red)*.

This Stroop effect is readily explained under the node structure theory, where the same mental nodes are involved in perception and production and the most primed node in a domain automatically becomes activated regardless of its source of priming. A high-frequency word such as *green* will prime its lexical content node faster and more strongly than will a visually presented color, because the naming of a color is a relatively rare activity. This does not mean that Stroop interference is completely describable in "race model" terms because priming doesn't automatically cause activation in the theory. However, it *does* mean that color naming will either take more time or exhibit more errors with than without the competing color name, 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. The mental node hypothesis also explains top-down effects in both speech and visual perception. To illustrate one such effect, consider Leeper's (1936) experiment where an ambiguous figure such as Jastrow's rabbit-duck is presented along with instructions such as "Can you see the duck?" The subject will perceive the duck but not the rabbit because the instructions prime the nodes representing the conceptual components of ducks. With the added bottom-up priming from the figure itself, these "duck nodes" receive the most priming and become activated under the most-primed-wins principle, thereby causing perception of the duck. The "rabbit nodes" on the other hand only receive bottom-up priming and, being less primed, do not become activated, so that the rabbit goes unperceived.

Speed-Accuracy Trade-Off in Perception. The node structure theory was originally designed to explain the trade-off relationship between time and accuracy (errors) in motor and mental skills (see MacKay, 1982), and bottom-up connections readily capture speed-accuracy trade-offs in perceptual recognition. To recognize an object (or word), the highest level node representing the object (or word) must receive greater priming than any other (extraneous) node in its domain when the triggering mechanism is applied. Although extraneous nodes receive unpredictable amounts of priming, with a distribution over time approxlisten to a sentence containing the word *legi* lature*, where the *s* has been masked by a cough*, the word sounds intact, and the missing *s* sounds as real and as clear as the remaining acoustically present phonemes (Warren, 1970). The subjects somehow synthesized the missing *s* and when informed that the cough replaced a single speech sound were unable to identify what sound is missing.[4]

The question of how this perceptual synthesis occurs is readily answered 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 word *legi*lature*, *legislature*(noun) will acquire greatest priming: even though the *s* has been obliterated in the acoustic waveform, no other node in the (noun) domain is likely to acquire as much top-down (contextual) and bottom-up 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, the *s* node can therefore become activated, causing clear perception of the obliterated *s*.

Consider now the cough and why it isn't perceived in its true (isomorphic) position in the sequence of speech sounds. The cough* is represented by nodes that are unconnected to the speech perception nodes—there are no nodes and serial order rules for representing the vowel group *i**, syllable *gi**, word *legi*lature* or lexical concept "legi*lature." This explains why the cough is poorly localized with respect to the speech sounds and why (in part) the cough seems to coexist in a separate perceptual space from the sentence (see Warren & Warren, 1970). Nonspeech noises are perceived via separate content and sequence nodes in a different perceptual system, analogous in some ways to a separate sensory system, even though speech and nonspeech noises share the same basilar membrane.

The Recognition of Segments Versus Syllables. The time it takes subjects to identify a segment versus a syllable within a sequence of nonsense syllables provides further support for the principle of higher level activation. The original experiment by Savin and Bever (1970) can be used for purposes of illustration because subsequent studies have replicated their basic findings and come to the same basic conclusion (see Massaro, 1979). Savin and Bever (1970) had subjects listen to a sequence of nonsense syllables with the aim of detecting a target unit

[4]Warren introduced an ingeneous control procedure to ensure that the missing sound was truly missing. In the control procedure, a word such as *legislature* was "mispronounced" as e.g., *legiklature*, and the mispronounced segment was replaced by the cough*. The results showed that again the subjects synthesized the missing *s*. Clearly then, this synthesis does not derive from coarticulatory cues present within the acoustic waveform because these cues differ for *s* versus *k*.

as quickly as possible. There were three types of targets; an entire syllable e.g., *splay*, the vowel within the syllable, i.e., *ay*, or the initial consonant of the syllable, i.e., *s*. The subjects pressed a key as soon as they detected their target and the surprising result was that reaction times were faster for syllable targets than for segment targets, either the initial consonant or the vowel in the syllable. Syllables apparently become perceived first, with perception of phonemes coming later.

These findings cannot be explained if all nodes in an input hierarchy must become activated, or if lower level activation is always necessary for higher level activation. However, the principle of higher level activation readily explains these findings: The subjects activated only the higher level (syllable) nodes on first encounter with the nonsense syllables, enabling immediate perceptual recognition of the syllable targets. Perception of the segment targets required an extra step: activation of segment nodes via multiplication of priming from the appropriate sequence node.

Aphasic Deficits. The principle of higher level activation predicts an asymmetry in the effects of brain damage on perception versus production (see MacKay, 1985) that receives support from recent studies of expressive and receptive deficits in aphasic, that is, the principle of higher level activation predicts that some lesions will impair only production whereas other lesions will impair both perception and production. Specifically, if a lesion selectively damages content nodes, then both production and perception will suffer under the theory. Selective damage to *sentential* sequence and/or timing nodes will likewise impair both production and comprehension, although the patient may still be able to produce and recognize *phonological* components. However, selective damage to *phonological* sequence and/or timing nodes will severely impair production, but will leave perception intact. The reason is that *phonological* sequence and timing nodes are unnecessary for the perception of common words, because phonological content nodes don't normally become activated (the principle of higher level activation). Moreover, for patients with intact comprehension, the theory predicts a specific type of production deficit involving the sequencing and timing of phonological components (see MacKay, 1985).

PERCEPTUAL ERRORS

As Freud (1901) and Meringer and Mayer (1895) pointed out, perceptual errors provide a means of (1) infering the otherwise hidden mechanisms of everyday speech perception, and (2) "testing" existing theories of perception, because theories that are incapable of explaining the errors that occur are incomplete or inadequate as accounts of the mechanisms underlying "veridical" perception. In this section I show how the node structure theory explains the regularities in imating a Gausian curve 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 is presented for long enough. But the shorter the stimulus duration, the greater the likelihood of error, i.e., the greater the probability that some other node receives more priming than the appropriate node at the time when the activating mechanism is applied.

FUNCTIONAL ISSUE II: THE PROCESSES UNDERLYING PERCEPTION

Having examined some implications of and evidence for bottom-up connections in a theory of production incorporating mental nodes, I turn now to processes that give rise to perception in the theory. Not only do mental nodes have the same dynamic properties in perception as in production, but the processes and mechanisms underlying the activation of a node are exactly the same in both. The only difference is that processes normally become initiated bottom-up rather than top-down in perception. Consider for example activation of the node *frequent*(adjective) following presentation of the word "frequent" in perception. Sensory analysis and phonological nodes provide strong and convergent (many-to-one) bottom-up priming that summates on *frequent*(adjective) and introduces second order priming to ADJECTIVE, just as in production. ADJECTIVE then becomes activated as the most primed sequence node following the next pulse from the timing node. Once activated, ADJECTIVE then multiplies the priming of all nodes in its (adjective) domain, but only the most primed one, normally *frequent*(adjective) reaches threshold and becomes activated.

As in production, the rate setting of the timing nodes in perception is partly individual specific and partly situation specific, determined by the perceived rate of input for example. In the node structure theory the input and perception of the input can proceed at different rates within wide limits: The rate setting for the timing nodes of speaker and listener need not match. The only requirements are that the perceiver's rate setting not be so slow that priming has largely decayed by the time the next pulse from the timer arrives, and not so fast that so little priming has built up that the probability of activating the wrong node exceeds the error criterion (see MacKay, 1982).

THE PRINCIPLE OF HIGHER LEVEL ACTIVATION

As already noted, activation is necessary for perceptual awareness. However, not all nodes in a hierarchy such as the one in Fig. 12.2 become activated during perception the way they do in production. Only higher level (e.g., sentential) nodes normally become activated and give rise to everyday perception. This principle of higher level activation is extremely general, applying to all types of perception. To illustrate the logical basis of this principle, I first show why activating lower level nodes is unnecessary in perception. I then show why activating lower level nodes is undesireable in perception and discuss the optimal level for activation to begin during everyday speech perception. Finally, I discuss various sources of evidence that support this principle of higher level activation.

Why Lower Level Activation is Unnecessary

Lower level nodes need not become activated during perception because of the efficiency with which they pass on priming to higher level nodes. This efficiency is attributable to two fundamental structural characteristics of bottom-up connections. One is the fact that bottom-up connections are many-to-one (see Fig. 12.2), which enables simultaneously occuring priming to converge and summate. Lower level nodes have an additional advantage in this summation process because convergent priming arrives either simultaneously or very closely in time at lower level nodes. For example, the priming from feature nodes converges simultaneously or nearly simultaneously on a segment node, whereas the priming from lexical nodes converges on a phrase node sequentially and over a period of a many hundreds of milliseconds. Without themselves becoming activated, lower level nodes can pass on enough temporally summated priming to enable higher level nodes to reach the minimum criterion required for multiplication to threshold (activation). Higher level nodes on the other hand must become activated during perception in order to transmit sufficient priming to enable the highest level nodes to become activated via priming multiplication.

The second structural basis for the principle of higher level activation is the fact that in general, lower level connections have greater linkage strength than higher level connections (see MacKay, 1982). Greater linkage strength means that low-level bottom-up connections pass on priming extremely efficiently. Thus, even when unactivated, lower level nodes transmit enough (second-order) priming to enable their connected (higher level) nodes to become activated via multiplication of priming.

By way of illustration, consider the hierarchy of bottom-up connections for the word *practice* in Fig. 12.2. To facilitate exposition, assume that the sensory analysis nodes representing the acoustic input provide the equivalent of first-order priming to the phonological feature nodes. Without becoming activated, each feature node therefore passes on somewhat weaker (second order) priming to their connected segment nodes. However, because each segment node receives simultaneous bottom-up priming from at least four feature nodes, this 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 pass on this summated priming to their connected phonological compound and syllable nodes, and again because of convergent summation, favorable timing, 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.

Why Lower Level Activation is Undesirable

A comparison of the costs of activation versus priming illustrates why unnecessary (i.e., lower level) activation is undesirable in everyday perception. Unlike priming, which is automatic and parallel or simultaneous, activation is non-automatic, sequential, and time consuming: The activation mechanism (sequence node) must first receive a build-up of priming and then become activated via a pulse from the timing node. The sequence node must then activate its most primed content node via multiplication of priming. Activating more than one node at a time in a domain is virtually impossible and the rate of activation must not be so fast as to induce errors (see MacKay, 1982). This adds further to the temporal inefficiency of activation relative to priming and suggests that if activation is unnecessary, it should not occur. And because activating lower level (e.g., phonological) nodes is unnecessary (see foregoing), the principle of higher level activation postulates that only higher level nodes normally become activated during everyday perception.

The Optimal Level for Activation

Although activation incurs costs (not just time costs as discussed earlier but probably effort costs as well), activation is necessary for perceptual awareness, which becomes especially desirable at the highest possible levels to ensure appropriate or adaptive action (benefits). Consequently, there must be some optimal level where activation becomes cost effective: Below the optimal level, costs of activation (time and effort) outweight benefits, and above the optimal level, benefits of activation (perceptual awareness and adaptive action) outweigh costs. Two factors determine this optimum level; prior practice (linkage strength) and the time characteristics of convergent priming, and together these factors suggest that lexical content nodes represent the typical level where activation must first occur in the case of adults perceiving common words. On the one hand, unaided bottom-up priming *to* lexical content nodes probably surpasses the minimum level required for activation via multiplication (discussed earlier). On the other hand, the priming passed bottom-up *from* lexical content to phrase nodes may typically fall below this minimum criterion because of the poor temporal summation and weak linkage strength of the connections to phrase nodes. As a result, unless lexical nodes become activated, no higher level nodes whatsoever can be activated, making perceptual awareness impossible. In short, the principle of higher level activation must begin with lexical content nodes during everyday speech perception.

This is not to say that lower level nodes *cannot* become activated. As MacKay (in preparation) points out, the cost-effective level for activation varies with attention and the context or perceptual situation. For example, if an input is especially degraded or unfamiliar, activating phonological or even sensory analysis nodes may become necessary in order to provide sufficient bottom-up priming to enable higher level nodes to become activated via multiplication of priming.

Evidence for the Principle of Higher Level Activation

During *production*, phonological nodes must invariably become activated if the phonemes of a word are to be produced in proper serial order. However, during *perception*, phonological nodes become primed but not activated under the principle of higher level activation.[3] In addition to the logical arguments discussed previously, at least four lines of empirical evidence support this principle of higher level activation, as discussed below.

Perception of the Distal Stimulus. As expected under the principle of higher level activation, we normally perceive not the proximal stimulus or pattern of sensory stimulation but the distal stimulus or higher level conceptual aspects of an input. This phenomenon is of course not limited to speech but applies more generally to all perceptual systems, including vision and audition. For example, we perceive an object such as a lamp at some distance from ourselves but fail to perceive the disparate retinal images that provide the sensory basis for that perception. Similarly, we hear the sound of a car's horn as coherent and localized in space but fail to perceive the complex sensory events underlying this perception, e.g., differences in time of arrival of the sound to the two ears (see Warren, 1982). The reason is that priming from the sensory analysis nodes which represent these sensory events is passed on so automatically and so effectively that full-fledged activation and perception normally never occur at that level.

Phonemic Restorations. The phonemic restoration phenomenon provides further support for the principle of higher level activation. When an extraneous 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). For example, when subjects

[3]One should not infer here that we are more aware of phonological segments in production than perception (because phonological nodes normally become activated during production but not during perception). Activation is only one of several conditions necessary for awareness.

listen to a sentence containing the word *legi*lature,* where the *s* has been masked by a cough*, the word sounds intact, and the missing *s* sounds as real and as clear as the remaining acoustically present phonemes (Warren, 1970). The subjects somehow synthesized the missing *s* and when informed that the cough replaced a single speech sound were unable to identify what sound is missing.[4]

The question of how this perceptual synthesis occurs is readily answered 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 word *legi*lature, legislature*(noun) will acquire greatest priming: even though the *s* has been obliterated in the acoustic waveform, no other node in the (noun) domain is likely to acquire as much top-down (contextual) and bottom-up 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, the *s* node can therefore become activated, causing clear perception of the obliterated *s*.

Consider now the cough and why it isn't perceived in its true (isomorphic) position in the sequence of speech sounds. The cough* is represented by nodes that are unconnected to the speech perception nodes—there are no nodes and serial order rules for representing the vowel group *i**, syllable *gi**, word *legi*lature* or lexical concept "legi*lature." This explains why the cough is poorly localized with respect to the speech sounds and why (in part) the cough seems to coexist in a separate perceptual space from the sentence (see Warren & Warren, 1970). Nonspeech noises are perceived via separate content and sequence nodes in a different perceptual system, analogous in some ways to a separate sensory system, even though speech and nonspeech noises share the same basilar membrane.

The Recognition of Segments Versus Syllables. The time it takes subjects to identify a segment versus a syllable within a sequence of nonsense syllables provides further support for the principle of higher level activation. The original experiment by Savin and Bever (1970) can be used for purposes of illustration because subsequent studies have replicated their basic findings and come to the same basic conclusion (see Massaro, 1979). Savin and Bever (1970) had subjects listen to a sequence of nonsense syllables with the aim of detecting a target unit as quickly as possible. There were three types of targets; an entire syllable e.g., *splay,* the vowel within the syllable, i.e., *ay,* or the initial consonant of the syllable, i.e., *s.* The subjects pressed a key as soon as they detected their target and the surprising result was that reaction times were faster for syllable targets than for segment targets, either the initial consonant or the vowel in the syllable. Syllables apparently become perceived first, with perception of phonemes coming later.

[4]Warren introduced an ingeneous control procedure to ensure that the missing sound was truly missing. In the control procedure, a word such as *legislature* was "mispronounced" as e.g., *legiklature,* and the mispronounced segment was replaced by the cough*. The results showed that again the subjects synthesized the missing *s.* Clearly then, this synthesis does not derive from coarticulatory cues present within the acoustic waveform because these cues differ for *s* versus *k.*

These findings cannot be explained if all nodes in an input hierarchy must become activated, or if lower level activation is always necessary for higher level activation. However, the principle of higher level activation readily explains these findings: The subjects activated only the higher level (syllable) nodes on first encounter with the nonsense syllables, enabling immediate perceptual recognition of the syllable targets. Perception of the segment targets required an extra step: activation of segment nodes via multiplication of priming from the appropriate sequence node.

Aphasic Deficits. The principle of higher level activation predicts an asymmetry in the effects of brain damage on perception versus production (see MacKay, 1985) that receives support from recent studies of expressive and receptive deficits in aphasic, that is, the principle of higher level activation predicts that some lesions will impair only production whereas other lesions will impair both perception and production. Specifically, if a lesion selectively damages content nodes, then both production and perception will suffer under the theory. Selective damage to *sentential* sequence and/or timing nodes will likewise impair both production and comprehension, although the patient may still be able to produce and recognize *phonological* components. However, selective damage to *phonological* sequence and/or timing nodes will severely impair production, but will leave perception intact. The reason is that *phonological* sequence and timing nodes are unnecessary for the perception of common words, because phonological content nodes don't normally become activated (the principle of higher level activation). Moreover, for patients with intact comprehension, the theory predicts a specific type of production deficit involving the sequencing and timing of phonological components (see MacKay, 1985).

PERCEPTUAL ERRORS

As Freud (1901) and Meringer and Mayer (1895) pointed out, perceptual errors provide a means of (1) infering the otherwise hidden mechanisms of everyday speech perception, and (2) "testing" existing theories of perception, because theories that are incapable of explaining the errors that occur are incomplete or inadequate as accounts of the mechanisms underlying "veridical" perception. In this section I show how the node structure theory explains the regularities in perceptual errors that have been observed so far. I then outline some predicted regularities for future test.

Regularities in Perceptual Errors

In misperceptions collected from everyday speech, the misperceived units range in scope from distortions of entire words and phrases (e.g., *popping really slow* misperceived as *prodigal son*) to single features (e.g., *pit* misperceived as *bit*), and middle components of a word are more likely to be misheard than those at the ends (Browman, 1980). However, about 85% of all misperceptions are simple word substitutions: The listener mishears one word as another. This preponderance of word substitutions suggests that the word constitutes a particularly important unit in speech perception. Indeed, misperceptions almost never give rise to nonwords, as might be expected if words and phonemes had identical status as units in everyday speech perception. The principle of higher level activation readily explains why words predominate over nonwords in perceptual substitutions.

Slips of the ear sometimes resemble slips of the tongue (see Browman, 1980). By way of illustration, consider the misperception of *carcinoma* for *Barcelona* in the case of an individual who is for the moment concerned or preoccupied with this particular disease. Such substitutions represent a perceptual analogue of the Freudian slip and receive a parallel explanation under the node structure theory (See MacKay's (1982) explanation of Freudian slips). In this particular example, the perceptual substitution occurs because priming for *carcinoma*(noun), arising from the preoccupation (top-down) and from aspects of the acoustic stimulus (bottom-up) exceeds the priming for *Barcelona*(noun) arising from the input itself. As a consequence, the extraneous node *carcinoma*(noun) becomes activated under the most-primed-wins principle.

However, the similarities between errors in production versus perception are less striking than the asymmetries. One of these asymmetries bears on the distinction between mental, muscle movement, and sensory analysis nodes (see Fig. 12.1). For example, by masking incoming speech sounds, environmental noises can cause misperceptions but not misproductions, an asymmetry that follows directly from the independence of sensory analysis and muscle movement nodes. Less obviously but for the same reason, whole classes of production errors are absent from perception. An example is stuttering, a class of production errors that simply never occurs in perception. No one misperceives someone to say *pppplease* when they in fact said *please*. This asymmetry suggests that stuttering may usually originate within the system of muscle movement nodes (see MacKay & MacDonald, 1984) that are independent of both mental and sensory analysis nodes for perceiving speech.

The structure of bottom-up versus top-down connections contributes another set of asymmetries to errors in production versus perception. Thus, ambiguity causes problems for perception (see MacKay, 1970a) but not for production, because the top-down connections to lexical nodes are unique: The top-down connections from a node such as *the tall crane*(noun phrase) go to the node representing either one or the other of the meanings of *crane* (i.e., *crane 1*, the mechanical hoist, or *crane 2*, the bird) but not to both. By way of contrast, bottom-up connections are nonunique, so that an ambiguous word such as *crane* primes both *crane* 1(noun) and *crane* 2(noun).

On the other hand, psychological synonymy (see MacKay, 1973b) causes errors in production but not perception. Blends are the production errors. An example is *sotally*, a combination of the words *solely* and *totally* that occured in the context "He was sotally (solely/totally) responsible for that." Under the theory, blends occur whenever two nodes in the same domain, e.g., *solely*(adverb) and *totally*(adverb) receive exactly equal priming and via multiplication of priming become simultaneously activated. Lower level components of either one word or the other then become activated automatically depending on which one receives the most priming (see MacKay, 1973b), giving rise to errors such as *sotally*. Errors resembling blends never occur in perception because bottom-up priming from an acoustic input uniquely primes one or the other of the nodes representing synonyms.

Predicted Asymmetries between Perceptual Versus Production Errors

The node structure theory predicts three systematic differences that remain to be tested between slips of the tongue versus slips of the ear.

The Phonological Similarity Prediction. The node structure theory predicts that phonological similarity will play much more of a role in misperceptions than misproductions. As discussed earlier, production errors *sometimes* involve similar sounding words, but under the node structure theory, misperceptions should *virtually always* involve similar sounding words such as *carcinoma* and *Barcelona*. The reason is that bottom-up priming converges and summates to such an extent on the input side, that misperceptions must incorporate most of the phonological components of the actual input. However, during *production*, bottom-up priming only converges on nodes which are undergoing self-inhibition as a result of recent activation. As a consequence, only divergent bottom-up priming can cause phonologically similar word substitutions during production, and because divergent priming is second-order and relatively weak, the theory predicts that production errors will less frequently involve similar sounding words.

The Sequential Domain Prediction. The theory predicts that sequential class will play more of a role in misproductions than misperceptions. In production errors, words almost invariably substitute words from within the same sequential domain or syntactic class e.g., nouns interchange with other nouns and never with verbs or adjectives (see Fromkin, 1973; MacKay, 1979), but the theory predicts that misperceptions will frequently violate this syntactic class constraint. An example violation is the misperception of "descriptive" (an adjective) as "the script of" (a determiner, a noun, and a preposition, respectively) (from Browman, 1980).

In production, output alternatives are limited to a single sequential domain: Errors can only occur when an extraneous node in the same sequential domain as the intended word achieves greatest priming when the sequence node becomes activated. However, in perception, listeners cannot know with certainty what chunk of the acoustic waveform constitutes a word let alone what sequential class the word belongs to. Perceptual alternatives are not confined to a single sequential domain: Many sequence nodes become primed to some extent and an extraneous sequence node can sometimes receive greatest priming and become activated in violation of the syntactic class constraint. Thus, in the earlier example, *descriptive*(adjective) may receive and pass on less priming to its sequence node than *the*(determiner), *script*(noun) and *of*(preposition), in part because of the lexical frequency of *the* and *of* but perhaps also because of top-down (expectation) priming of *script*(noun). As a consequence, the extraneous sequence nodes DETERMINER, NOUN, and PREPOSITION become activated rather than ADJECTIVE and cause the observed violations of the syntactic class constraint.

The Sequential Error Prediction. Sequential errors involve the misordering of words or speech sounds just uttered or about to be uttered and are quite common in production. An example at the phonological level is the misproduction *coat thrutting* for *throat cutting* (from Fromkin, 1973). An analogous example at the sentential level is *We have a laboratory in our computer* for *We have a computer in our laboratory* (from Fromkin, 1973).

The node structure theory predicts many more sequential errors for production than for perception. One reason is that temporal sequences must be "constructed" during production but come built in during perception. Another reason is that the convergent bottom-up summation of priming that occurs in perception but not in production strongly constrains perceptual errors to resemble the actual input, and prevent sequential errors in perception which involve phonologically dissimilar words such as *laboratory* and *computer* in the earlier example.

ASYMMETRIES BETWEEN PRODUCTION AND PERCEPTION

I now examine how the node structure theory explains some additional, already observed asymmetries between perception versus production.

The Maximal Rate Asymmetry. The most striking difference between speech perception and production is that we can perceive speech at a faster rate than we can produce it. Foulke and Sticht (1969) summarized the evidence on the perception of speech sounds accelerated by means of computers that sample and compress acoustic signals without introducing pitch changes. The data indicate that compressed speech becomes difficult to comprehend and remember but remains highly intelligible up to 400 words per minute or about 30 ms per phoneme. In contrast, *producing* speech of equivalent intelligibility at remotely comparable rates is well beyond human capacity.

What accounts for this maximal rate asymmetry? An explanation in terms of the time and effort required to physically move articulators such as the jaw has several problems. One is the data on the rate of internal speech, which of course involves no movement of the articulators whatsoever. For example, MacKay (1981) had subjects produce sentences internally as rapidly as possible, pressing a key as they began and finished each sentence. Then after a 20-sec. pause the subjects repeated the same sentence internally, and so on, for a total of 12 practice trials at maximal rate. As expected, the maximal rate of internal speech increased systematically with practice but approached asymptote at about 100 ms per phoneme after seven practice trials. Although this asymptote is considerably faster than the maximal rate for producing these same sentences aloud and with equivalent amounts of practice, it nevertheless remains much slower than the 30 ms per phoneme rate for the perception of compressed speech. Because no movement of the articulators takes place during internal speech, this remaining difference indicates that muscle movement factors cannot completely explain the maximal rate asymmetry. However, the differences between perceptual vs. production processes discussed here provide a straightforward explanation. For example, activation takes time, and because not all nodes become activated during perception (the principle of higher level activation) perception can proceed at a faster rate than production. Similarly, the convergent summation of priming that occurs in perception but not production means that when it occurs, activation can proceed more quickly during perception than during production (see MacKay, 1982).

The Listening Practice Asymmetry. MacKay and Bowman (1969) reported a "conceptual practice effect" with an interesting but asymmetric counterpart on the perceptual side. The 1969 subjects were German-English bilinguals, who were presented with sentences one at a time and simply produced each sentence as rapidly as possible. An example is "In one corner of the room stood three young men." Following a 20-sec. pause, the sentence was presented again, for a total of 12 repetitions of the same sentence. Half the sentences were in English and the other half in German. The independent variable was practice and the dependent variable was the time to produce the sentence.

Practice had a significant effect: The rate of speech was 15% faster for the last 4 than the first 4 practice trials even though the subjects were attempting always to speak at their maximum rate. After practicing the sentence 12 times, the subjects received a "transfer" sentence in their other language, which they also produced at maximal rate. This transfer sentence was either a nontranslation (unrelated to the original sentence in meaning, syntax, and phonology) or a translation (with identical meaning but different phonology and word order from the original). The surprising finding was that subjects produced the translations significantly faster than the nontranslations (2.44 sec. per sentence vs. 2.24 sec. per sentence), indicating an effect of practice at the lexical concept level and above (see MacKay, 1982, for details and explanation).

Consider now the *perceptual* analogue of this conceptual practice effect. A *listening practice* condition was presented to determine whether repeated *listening* to a sentence leads to a conceptual facilitation effect similar to the physical practice condition discussed earlier. Twelve German-English bilinguals listened to a tape recording of the sentences repeated 12 times at maximal rate by the previous subjects. To ensure that subjects in this listening practice condition were paying attention to the input, they were instructed to indicate whether changes or errors occurred from one repetition to the next.

A transfer phase, identical to that in the physical practice condition, followed the 12 listening practice trials. During this transfer phase, subjects in the listening practice condition produced out loud and at maximal rate a sentence that was either a translation or a nontranslation of the sentence they had heard repeated 12 times, and as before, production times were faster for translations (2.31 sec.) than nontranslations (2.57 sec.), a 10% facilitation comparable to the 8% for physical practice. This facilitation effect is readily explained under the mental node hypothesis and suggests that repeated listening may suffice to develop the mental skill underlying the conceptual practice effect.

However, there was an asymmetry: Production times in the listening practice condition were longer than those in the physical practice condition, both for nontranslations (6% longer) and for translations (3% longer). One explanation of this statistically reliable difference is that subjects were less highly motivated during listening practice than during physical practice. However, we cannot yet rule out the more interesting possibility that physical practice is genuinely superior to listening practice. Listening practice may strengthen bottom-up connections while leaving top-down connections relatively weak, thereby facilitating performance less than physical practice, even for the highest conceptual levels of the skill.

The Word Production Asymmetry. Differences between perception versus production in the node structure theory help explain the word production asymmetry: the fact that we can usually recognize and understand a word long before we can use it in everyday speech production (Clark & Hecht, 1983). The main requirement for *recognizing* a word in the theory is the existence of previously formed bottom-up connections from phonological and sensory analysis nodes. However, *producing* the word requires the formation of several additional types of connections, which together may delay development of the production vocabulary. One is top-down connections from the lexical content node to the appropriate phonological and muscle movement nodes. Another is inhibitory connections between sequence nodes, which enable sequential production.

The Detection-Correction Asymmetry. Differences between detecting self-produced versus other-produced errors illustrate an additional asymmetry between perception versus production: Whereas speakers correct their own errors with about equal frequency across units of different size (Nooteboom, 1980), listeners detect errors involving larger units such as words with much higher frequency than errors involving smaller units such as phonemes and phoneme clusters (Tent & Clark, 1980). Thus listeners most easily detect errors which cause an obviously deviant meaning, whereas speakers detect all types of errors with equal sensitivity (see Cutler, 1982). This equi-sensitivity for the speaker is readily explained in the node structure theory: Output errors occur when an inappropriate node at any level in a hierarchy becomes activated, no matter what the size of the surface units involved (segment, segment clusters, syllables, or phrases), which predicts that speakers will perceive and correct errors about equally often for small versus large units (all other factors being equal).

The Missing Feedback Effect. Verbal transformation experiments illustrate another interesting asymmetry between perception versus production: the missing feedback effect. In the typical verbal transformation study, subjects listen to an acoustically presented word repeated via tape loop for a prolonged period and report hearing changes in the stimulus (Warren, 1968). For example, after many repetitions, subjects might misperceive the word *police* as *fleas, please* or *fleece*. These illusory changes are explained as follows under the node structure theory. When the word *police* is presented for the first time, it strongly primes the lexical concept node *police*(noun), and primes other nodes such as *fleas*(noun) and *fleece*(noun) to a lesser extent (depending on their phonological similarity to *police*), so that only *police*(noun) becomes activated under the most-primed-wins principle. However, after *repeated* presentation, the nodes for *police* become fatigued and respond less strongly, so that eventually *police*(noun) acquires less priming than say, *fleece*(noun), which therefore becomes activated under the most-primed-wins principle. The result is illusory perception of *fleece*, along with increasingly rapid perceptual shifts between the various other alternatives (*fleas, please, police, fleece*).

Consider now the missing feedback effect. Lackner (1974) had subjects repeat a word every 500 msec for several minutes and then listen to a recording of their own output. The subjects experienced the usual verbal transformations when they subsequently *listened* to the tape recording of their own output but reported no perceptual transformations whatsoever when *producing* the word. This asymmetry is curious because the acoustic events at the ear are identical in the two conditions. Why don't self-produced auditory inputs cause verbal transformations? Lackner (1974) and Warren (1968) attribute the missing feedback effect to a corollary discharge that accompanies the motor command to produce a word. This corollary discharge cancels or inhibits the external (proprioceptive and auditory) feedback resulting from producing the word, so that self-produced auditory inputs fail to induce the perceptual changes that occur in the typical verbal transformation experiment. However, the corollary discharge hypothesis has difficulty explaining both the many interactions between speech perception and production (discussed earlier) and aspects of Lackner's own data, namely that no *production errors* resembling the perception errors occurred when the subjects were repeating the words.

The node structure theory explains all of these phenomena by means of a common mechanism: the self-inhibition that follows activation of mental nodes. Under the theory, the mental nodes for producing and perceiving a word such as *police* are identical. As a consequence, when someone *produces* the word *police*, auditory feedback returns as priming to some of the just-activated nodes that produced it but arrives during their self-inhibitory phase and therefore has no effect. This explains why self-produced repeated inputs fail to cause verbal transformations, the missing feedback effect.

Consider now the absence of output errors during repeated *production* of a word. Repeating the word *police* causes fatigue of the corresponding mental nodes, but because top-down priming is unique (as aforementioned), only *police*(noun) and no other lexical node receives systematically increasing priming and becomes activated under the most-primed-wins principle. This reduces the likelihood of *production* errors resembling the ones that occur in perception and explains this additional asymmetry between the perception versus production of speech.

CONCLUSIONS

The asymmetries discussed earlier present problems for theories that assume a symmetric relation between processes for perception and production (see Gordon & Meyer, 1984). Identical components can represent both perception and production (the mental node hypothesis) in these theories, but perceptual processes are simply the reverse of the corresponding production processes, like the bidirectional reactions in chemical formulas (see Fodor, Bever, & Garrett, 1974).

This symmetry assumption is both simple and attractive: It enables researchers to devote all their efforts to studying perception, because solving the

problem of perception also solves the problem of production under the symmetry assumption. Indeed, the appeal of the symmetry assumption may explain why psycholinguistics has until recently focused almost exclusively on perception rather than production or the relation between the two (see Fodor, Bever, & Garrett, 1974). However, the asymmetries discussed earlier indicate that studies of perception are by themselves insufficient and that comparisons of perception and production are both necessary and theoretically important. To further stimulate such comparisons, I conclude with a summary of the processes in the node structure theory that are asymmetric between production versus perception. Although some of these asymmetries have already received mention under other headings, summarizing them here provides a sharp contrast with "symmetrical" theories of the relation between production and perception.

The Sequential Activation Asymmetry. The logical order for activating mental nodes in production and perception is asymmetric under the node structure theory. By way of example, the numbers in Fig. 12.2 indicate the order in which the sentential nodes illustrated there must be activated during production (1, 2, 3, 4, 5), and the symmetry assumption requires the reverse order of activation during perception (5, 4, 3, 2, 1). However, this reverse order is logically impossible. If perception is to be error free and activation takes place during perception, the sequence must be something like 3, 4, 2, 5, 1 (where the numbers represent the corresponding nodes in Fig. 12.2 and left to right represents order of activation).

Convergent Versus Divergent Connections. Another important asymmetry between top-down versus bottom-up processes in the node structure theory stems from the fact that top-down connections are divergent or one-to-many in nature whereas bottom-up connections are convergent or many-to-one. The many implications of this asymmetry have already been discussed.

The Principle of Higher Level Activation. Under the principle of higher level activation, only higher level nodes become activated during everyday perception. Because of their linkage strength, timing characteristics, and convergent summation, lower level nodes pass on priming so efficiently that activation is unnecessary. This stands in contrast with production, where every node in the hierarchy must become activated if the output is to occur in proper serial order.

The Uniqueness Asymmetry. Top-down priming is generally unique: Only a single node in a given domain normally receives first-order top-down priming at any given time. Bottom-up priming, on the other hand is generally nonunique: more than one node in a domain normally receives first-order priming at any one time.

In conclusion, a great deal remains to be done to test and further develop the node structure theory. I am also aware of the sketchy and incomplete treatment here of various complex and sometimes controversial issues. Moreover, the sketchiness sometimes reflects multiple causes; both lack of space and lack of available data. For example, consider the perception and production of different types of phonological features. Like Cooper, Billings, and Cole (1976), Meyer and Gordon (1983) observed interactions between perceiving and producing the voicing feature but Gordon and Meyer (1984) found no such interactions for place of articulation. Cooper, Billings, and Cole (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, with no intervening mental nodes for representing 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: Omitting the issue was multiply determined.

Consider now the relation between the node structure theory and other theories of speech and cognitive skill. Like the motor theory of speech perception, the node structure theory gives speech a special place among systems for perception and action. Speech stimuli can be self-produced whereas one rarely produces stimuli such as the visual world (except marginally in drawing or writing). The node structure theory also makes speech special by incorporating a speech mode of perception, which is distinct from other perceptual modes: One and the same stimulus can be processed in the speech mode by activating the sequence nodes for the phonological system or in an auditory concept mode by activating the sequence nodes for the auditory concept system. The staggering degree of practice that speech normally receives (see MacKay, 1981, 1982) also makes speech special in the theory, as does the self-inhibitory mechanism that content nodes for speech require to deal with self-produced feedback. However, speech is not *fundamentally special* in the theory because similar node structures and degrees of practice can in principle be achieved for other perceptual and motor systems (see MacKay, 1985). Moreover, although different perceptual modes (systems) differ in nodes and perhaps also node structures (pattern of connections), they do not differ in fundamental principles of activation.

The node structure theory also bears a general resemblance to recent theories of word recognition (McClelland & Rumelhart, 1981) and of typing (Norman & Rumelhart, 1983) and participates in the current trend toward a focus on dynamic or process issues, in addition to static or structural ones. Like other recent theories, the node structure theory is concerned with underlying mechanisms and has the potential for mapping psychological constructs onto neuroanatomical ones (see MacKay, 1985), an exciting prospect because, as Ojemann (1983) points out, some such mapping seems essential for cracking the code of the brain.

However, the node structure theory provides some genuinely new mechanisms, units, and predictions. For example, the most-primed-wins principle bears a general resemblance to a principle built into McClelland and Rumelhart's (1981) theory of word recognition, but in order to produce speech as well as to recognize words, the node structure theory incorporates a fundamentally different (and much simpler) mechanism for achieving this principle.

The node structure theory also summarizes a wealth of results that in large measure are not accounted for in other theories and that have previously been considered to fall within four separate areas of inquiry: neuropsychology, psycholinguistics, cybernetics, and motor control. To complete the unification of these areas requires a great deal more theoretical and empirical work, and to achieve our ultimate goal, a unified theory of skill (speech being only one, albeit highly proficient example), will engage the field for many decades to come.

ACKNOWLEDGMENTS

The author thanks Robert Bowman for assistance in running the experiment and Drs. Dom Massaro, Herbert Heuer, Douglas Mewhort, Richard Warren, and Bill Cooper for comments on an earlier draft. This chapter was completed while the author was a member of the Perception and Action Research Group at the Center for Interdisciplinary Research (Z.i.F.) at the University of Bielefeld. The author acknowledges the support of Z.i.F. with appreciation.

REFERENCES

Browman, C. P. (1980). Perceptual processing: Evidence from slips of the ear. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (p. 213–230). New York: Academic Press.

Clark, E. V., & Hecht, B. F. (1983). Comprehension, production and language acquisition. *Annual Review of Psychology, 34*, 325–49.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. New York: Harcourt, Brace, Jovanovich.

Cooper, W. E., Billings, D., & Cole, R. A. (1976). Articulatory effects on speech perception: A second report. *Journal of Phonetics, 4*, 219–32.

Cooper, W. E., Blumstein, S. E., & Nigro, G. (1975). Articulatory effect on speech perception: A preliminary report. *Journal of Phonetics, 3*, 87–98.

Cooper, W. E., & Nager, R. M. (1975). Perceptuo-motor adaptation to speech: An analysis of bisyllabic utterances and a neural model. *Journal of the Acoustic Society of America, 58*, (1), 256–265.

Cutler, A. (1982). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 7–28). Amsterdam: Mouton.

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

Foulke, E., & Sticht, T. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin, 72*, 50–62.

Freud, S. (1901). *Psychopathology of everyday life* (A. A. Brill, trans. 1914). New York: Penguin.
Fromkin, V. A. (1973). Appendix. In F. A. Fromkin (Ed.), *Speech errors as linguistic evidence.* Paris: Mouton.
Gordon, P. C., & Meyer, D. E. (1984). Perceptual-motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception and Performance, 10,* (1), 153–178.
Klapp, S. T., Anderson, W. G., & Berrian, R. W. (1973). Implicit speech in reading reconsidered. *Journal of Experimental Psychology, 100,* 368–74.
Kozhevnikov, V. A., & Chistovich, L. A. (1965). *Speech articulation and perception.* Washington: *Joint Publications Research Service.*
Lackner, J. (1974). Speech production: Evidence for corollary discharge stabilization of perceptual mechanisms. *Perceptual and Motor skills, 39,* 899–902.
Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (p. 112–136). New York: Wiley.
Leeper, R. A. (1936). A study of a neglected portion of the field of learning: The development of sensory organization. *Journal of Genetic Psychology, 46,* 42–75.
Liberman, A. M., Cooper, F. S., Harris, K. S., & MacNeilage, P. F. (1962). A motor theory of speech perception. In *Proceedings of the speech communication seminar.* Stockholm: Royal Institute of Technology.
Lisker, L. (1978). *Rapid* vs. *rabid:* A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research, 54,* 127–132.
McClelland, J. C., & Rumelhart, D. E. (1981). An interactive-activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.
MacKay, D. G. (1970a). Mental diplopia: Towards a model of speech perception at the semantic level. In I. D'Arcais & W. J. M. Levelt (Eds.), *Recent advances in psycholinguistics* (p. 76–100). Amsterdam: North–Holland.
MacKay, D. G. (1970b). How does language familiarity influence stuttering under delayed auditory feedback? *Perceptual and Motor Skills, 30,* 655–669.
MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology, 3,* 210–227.
MacKay, D. G. (1973a). Aspects of the theory of comprehension, memory and attention. *Quarterly Journal of Experimental Psychology, 25,* 22–40.
MacKay, D. G. (1973b). Complexity in output systems: Evidence from behavioral hybrids. *American Journal of Psychology, 86*(4), 785–806.
MacKay, D. G. (1978). Speech errors inside the syllable. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments.* Amsterdam: North–Holland.
MacKay, D. G. (1979). Lexical insertion, inflection and derivation: Creative processes in word production. *Journal of Psycholinguistic Research, 8,*(5), 477–498.
MacKay, D. G. (1981). The problem of rehearsal or mental practice. *Journal of Motor Behavior, 13*(4), 274–285.
MacKay, D. G. (1982). The problems of flexibility, fluency and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89*(5), 483–506.
MacKay, D. G. (1985). A theory of the representation, organization and timing of action with implications for sequencing disorders. In E. A. Roy (Ed.), *Neurological studies of apraxia and related disorders.* Amsterdam: North–Holland.
MacKay, D. G. (1986). Self-inhibition and the disruptive effects of internal and external feedback in skilled behavior. In H. Heuer, C. A. M. Brunia, J. A. S. Kelso, & R. A. Schmidt (Eds.), *The generation and modulation of action patterns.* Journal of Experimental Brain Research Supplement. Berlin: Springer.
MacKay, D. G. (in preparation). *The organization of perception and action: Fundamentals of theoretical psychology.*

MacKay, D. G., & Bowman, R. W. (1969). On producing the meaning in sentences. *American Journal of Psychology, 82*(1), 23–39.
MacKay, D. G., & MacDonald, M. (1984). Stuttering as a sequencing and timing disorder. In W. H. Perkins & R. Curlee (Eds.), *Nature and treatment of stuttering: New directions.* San Diego: College–Hill.
Massaro, D. W. (1979). Reading and listening. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (pp. 331–354). New York: Plenum.
Meringer, R., & Mayer, K. (1895). *Versprechen and Verlesen.* Stuttgart: Goschensche Verlagsbuchhandlung.
Meyer, D. E., & Gordon, P. C. (1983). Dependencies between rapid speech perception and production: Evidence for a shared sensory-motor voicing mechanism. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance X.* Hillsdale, NJ: Lawrence Erlbaum Associates.
Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance. Slips of the tongue, ear, pen and hand.* New York: Academic Press.
Norman, D. (1981). Categorization of action slips. *Psychological Review, 88*(1), 1–15.
Norman, D. A., & Rumelhart, D. E. (1983). Studies of typing from the LNR research group. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typewriting* (pp. 45–66). New York: Springer–Verlag.
Ojemann, G. A. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *The Behavioral and Brain Sciences, 2,* 189–230.
Porter, R. J., & Lubker, J. F. (1980). Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motor linkage in speech. *Journal of Speech and Hearing Research, 23,* 593–602.
Savin, H. B., & Bever, T. E. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior, 9,* 295–302.
Spoer, K. T., & Smith, E. E. (1973). The role of syllables in perceptual processing. *Cognitive Psychology, 5,* 71–89.
Straight, S. (1980). Auditory vs. articulatory phonological processes and their development in children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1. Production.* New York: Academic Press.
Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review, 77,* 234–249.
Tent, J., & Clark, J. E. (1980). An experimental investigation into the perception of slips of the tongue. *Journal of phonetics, 8,* 317–25.
Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15,* 49–74.
Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin, 70,* 261–270.
Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392–3.
Warren, R. M. (1982). *Auditory perception: A new synthesis.* New York: Pergamon Press.
Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American, 223,* 30–36.
Woodworth, R. S. (1938). *Experimental Psychology.* New York: Holt.