D. G. MacKay (1992)

Constraints on

Theories of Inner Speech

In D. Reisberg (ed.)

Auditory imagery (pp 121-149)

Erlbaum; Hillsdale, N. J.

6 Constraints on Theories of Inner Speech

Donald G. MacKay
University of California, Los Angeles

Many people report that they can produce speech without moving their lips, and are consciously aware of this internal speech when they solve problems, read, write, or plan their everyday activities (Weisberg, 1980). The experience of inner speech is virtually universal among adults and has played a major role in psychological theory (Dell & Repka, in press). Some psychologists have viewed inner speech as identical to thought (e.g., Watson, 1950), while others have viewed inner speech as a side effect or necessary concommitant of thought (e.g., Sokolov, 1972). Psychologists have also viewed the long-term storage of information as dependent on inner speech or covert rehearsal, as when one silently repeats a telephone number in order to facilitate later recall (Atkinson & Shiffrin, 1968). Few, if any, psychologists probably continue to view the relation between inner speech, memory, and thought in just these ways, but all agree on the importance of inner speech, which, under various names, remains a central construct in psychological theories. For example, units resembling those required for producing inner speech are said to underlie writing (Ellis, 1988), typing (MacKay, in press), and the rehearsal and short-term storage of verbal materials (Baddeley, 1990).

This chapter reviews some fundamental phenomena that must be addressed in theories of inner speech. I begin with the problem of representation: What is the nature of the units underlying the production and experience of inner speech? Three basic representational constraints are addressed:

1. The units involved in inner speech: Internal speech errors, the time to produce sentences internally, and transfer of practice effects indicate that like

overt speech, inner speech involves phonological, morphological, lexical and phrase level units.

2. The nonarticulatory nature of inner speech: Neurolinguistic disorders and experiments on internally produced tongue twisters indicate that movements, however small, of the lips, tongue, velum, and other muscle systems are irrelevant to inner speech. Even the lowest level units for inner speech are highly abstract, representing a range of cognitive activities involving very different muscles (e.g., speech, writing, and typing) or no muscles whatsoever (e.g., language comprehension).

3. The nonauditory nature of inner speech: Theories postulating an auditory code for the "inner ear" must address methodological and theoretical criticisms discussed here, and must explain various differences between internal speech perception and the acoustics of overt speech.

Relations between the perception and production of overt and inner speech pose additional constraints addressed here. An example is: *The irrelevance of motor activity to internal speech:* Viable theories of inner speech must explain why the electromyographic activity that accompanies inner speech is unnecessary for either acquiring or experiencing inner speech. In addition, the theory must explain a set of dissociations between motor abilities and the experience of inner speech. Traditional theoretical approaches to relations between the perception and production of internal speech are discussed with respect to such criteria, and are found wanting.

The role of inner speech in the perception and immediate recall of visual stimuli suggests further constraints for theories of internal speech, as do relations between memory and rehearsed or repeated internal speech. Examples are:

1. *Unrehearsability:* Theories of inner speech must explain why language is sometimes rehearsable, and sometimes not. How, for example, can someone unfamiliar with German accurately rehearse a German word such as /gelb/, but not a German trilled /r/.
2. *Rehearsal and volition:* Theories of inner speech must explain why inner rehearsal is sometimes involuntary and difficult to control.
3. *Effects of overt versus internal rehearsal on speech production:* An interesting set of similarities and differences in the effects of overt versus internal rehearsal on overt speech production provide additional constraints on theories of inner speech.

This chapter begins by reviewing available data bearing on four fundamental issues that theories of inner speech must address: (a) What is the nature or representational character of internal speech?; (b) what is the relation between the perceptual and generative components of internal speech?; (c) how does internal speech relate to overt speech?; (d) and what role does internal speech

play in cognitive processes such as the perception of visual stimuli and memory for verbal materials? Although some of these issues are far from resolved, the data reviewed here suggest important constraints that theories of internal speech, present and future, must address.

THE NATURE OF INTERNAL SPEECH

Phenomenal reports of inner speech generally include two components: a "generative" component (i.e., people report selecting the words of inner speech in a way that seems to resemble overt speech production) and an "auditory" component (i.e., people report hearing internally produced words in their "mind's ear"). These "perceptual" and "generative" components may in reality be largely inseparable (see MacKay, 1987, pp. 1–38), but are treated separately here for historic and didactic reasons.

The Generative Component of Internal Speech

As discussed below, direct evidence for the generative component of inner speech comes from the one observable aspect of mental imagery in general, and inner speech in particular: the electromyographic (EMG) activity in the muscles that occurs during mental imagery and covert rehearsal (Jackson, 1930; Sokolov, 1972). A major, as yet unsolved theoretical issue connected with this generative component is how speakers are able to intentionally evoke internal speech.

What Kinds of Units are Involved in Internal Speech? Evidence on the nature of the units involved in inner speech comes from two sources. One is the errors that speakers detect during internal speech. In preliminary observations, Meringer and Meyer (1895) reported several instances of mental errors detected in their own internal speech and noted that these errors closely resembled those that occur during overt speech. Extending these observations, Dell (1978, 1980) had subjects produce tongue twisters such as "Unique New York" from memory at fixed rates, either aloud or mentally, and report the errors that they detected. The same types of errors were reported during internal speech as during overt speech, usually anticipations, perseverations and reversals of phonological components, but sometimes also anticipations, perseverations or reversals of lexical and morphological components. The recent and more extensive study of Dell & Repka (in press) used the same procedures as Dell's earlier studies and also observed identical types of errors in inner and overt speech. This correspondence of the units underlying inner and overt errors indicates that like overt speech, inner speech involves phonological, morphological and lexical units.

The second source of evidence (MacKay, 1981) indicates that inner and overt speech share additional units at still higher levels. MacKay had subjects practice producing identical sentences as rapidly as possible, either overtly or silently to themselves without moving their lips. The dependent variable was speech rate, and subjects in both conditions timed themselves by pressing one key as they began to say the sentence and another key as they finished. Results of this procedure indicated that both internal and overt speech improved with practice and reached asymptote after about the same number of practice trials.

Then, by adding a transfer paradigm to this procedure and by using German-English bilinguals as subjects, MacKay demonstrated that internal and overt practice caused equivalent improvement in the ability to produce sentential components such as phrases. Specifically, after practicing producing a sentence at maximal rate twelve times in one language, either internally or overtly, the subjects overtly produced a transfer sentence that was either a word-for-word translation or a nontranslation of the practiced sentence in their other language. The results (see Table 6.1) indicated that maximal speech rate was faster when the transfer sentences were translations rather than nontranslations. Moreover, degree of transfer was equivalent for the internal versus overt practice conditions. Because this transfer effect could only be occuring at lexical and phrase levels, and not at the phonological or muscle movement levels (which are completely different for the two languages and provide no basis for transfer), this finding indicates that inner and overt speech involve identical lexical and phrase units. In short, the units for producing inner and overt speech seem to be identical at all levels.

Is the Generative Component Articulatory or Phonological? A great deal of evidence indicates that the generative component of internal speech and rehearsal involves an underlying code that is phonological rather than articulatory in nature. By standard definition, a generative component with articulatory characteristics represents the activities of particular muscles for the lips, tongue, velum, and other articulatory, laryngeal and respiratory organs, whereas a generative component with phonological characteristics represents not particular muscles,

TABLE 6.1
The Time to Produce a "Transfer" Sentence, Following Production of a Semantically Equivalent Sentence, Either Overtly (Physical Practice) or Internally (Internal Practice), in the Other Language of Bilinguals

	Nature of the Transfer Sentence		*Facilitation*	
Practice Condition	*Nontranslation*	*Translation*	*Time Difference*	*%*
Physical practice	2.44 (.26)	2.24 (.23)	.20	8
Internal practice	2.19 (.31)	1.96 (.25)	.23	11

Note. Time in seconds with standard deviations in brackets.

but more abstract units. For example, the same abstract phonological units could and probably do play a role in overt articulation (MacKay, 1987, pp. 7–38), writing (Ellis, 1988) and typing (MacKay, in press), activities that involve completely different sorts of muscles. Indeed, a strong case can be made that the same abstract phonological units also underlie the comprehension of spoken language, an activity that does not involve muscles of any sort (see MacKay, 1987, pp. 62–125).

Studies of internal speech errors (Dell, 1978, 1980; Dell & Repka, in press; Meringer & Meyer, 1895) provide one source of evidence for the phonological rather than articulatory character of the lowest level units for internal speech. As already noted, errors reported during internal speech involve phonological components, but not phonetic, articulatory, or muscle movement components. Strictly articulatory errors (e.g., the slurring of speech sounds commonly seen in the production of overt speech) have never been reported for everyday internal speech.

Moreover, experimental studies (Dell, 1978, 1980) indicate that anticipations, perseverations and reversals occur with the same absolute frequency in overt and inner speech. This additional resemblance suggests that contrary to popular belief, overtly produced tongue twisters result in errors at the phonological level but not at the articulatory or muscle movement level. And because the tongue did not move during the inner speech of Dell's subjects, tongue twister errors must have nothing to do with the tongue. Like the term *auditory imagery* in its current applications to inner speech, "tongue twisters" may be misnamed; they are more accurately described as "phonological twisters."

Patients who are speech-impaired (dysarthric) or congenitally speechless (anarthric) due to brain damage affecting peripheral control of the articulatory musculature provide further evidence for the phonological rather than articulatory nature of internal speech and rehearsal. Wilson & Baddeley (cited in Baddeley, 1990, pp. 86–87) tested the memory of a dysarthric patient who could comprehend language and communicate using a simple keyboard device, and showed that this patient was virtually normal on a wide variety of memory tasks involving inner speech or rehearsal. However, this patient completely lacked the capacity to articulate, indicating that articulatory activity is unnecessary for the normal functioning of internal speech and rehearsal. Nor is articulatory activity necessary in order for children to learn to rehearse subvocally. Bishop & Robson (1989) showed that anarthric children who are incapable of articulation *from birth* are nevertheless virtually normal on a wide variety of memory tasks involving inner speech or rehearsal. On the basis of such evidence, Baddeley (1990) argued that the representation underlying rehearsal and short-term storage of verbal information that he had formerly called the "articulatory loop" was really a "phonological loop," a fundamental theoretical change.

In summary, phenomena that seem to be articulatory in origin, for example, the errors arising during rapid production of phonological twisters, are in fact phonological rather than articulatory. Moreover, available evidence indicates that the generative component of internal speech, which seems on the surface to be articulatory in nature, in fact involves a phonological representation. Finally, activity of the speech musculature is unnecessary for the normal functioning and acquisition of the processes underlying inner speech and rehearsal.

The Perceptual Component of Internal Speech

Phenomenology of the "Inner Ear." Evidence for the so called auditory aspects of inner speech consists largely of the phenomenal experience of an inner voice without the occurrence of vocal output or of environmental input. The convincing nature of this seemingly auditory experience has led some psychologists to include inner speech within the category of "auditory imagery" and to suppose that an auditory or acoustic representation underlies the perception of inner speech. However, phenomenal experience or introspective report provides a shaky basis for specifying the representational character or theoretical nature of internal speech. The problem is that what seems phenomenally to be auditory often is not. For example, visual events can play a role in determining what people perceive phenomenally as a strictly auditory speech experience, and vice versa.

By way of illustration, consider the McGurk effect. McGurk and MacDonald (1976) seated subjects in front of a video monitor and had them listen to and observe a video recording of a person saying simple CV syllables such as *pa, ba, ta,* or *da*. Their task was simply to verbally identify the syllable that they *heard*. Unbeknownst to the subjects, the auditory syllables had been dubbed-in in synchrony with the speaker's lip movements, and on some trials differed from the speech sounds that these lip movements normally give rise to. For example, the acoustics of a person saying /ta/ might be dubbed in to synchronize with the visual lip movements of a person saying /pa/. This audio-visual conflict condition showed that visual features such as lip closure exerted a strong effect on what syllable the subjects reported hearing. With a conflict between the visual lip movements for /pa/ and the acoustics for /ta/, subjects usually reported hearing the visually based alternative, /pa/, rather than the auditorily based alternative, /ta/. Moreover, these subjects were quite surprised to discover that the /pa/ that they "heard" in this condition changed perceptually to /ta/ whenever they altered the visual input by complying with the experimenter's instructions to close their eyes.

How can a nonauditory event (visual lip movements) unconsciously influence a perception that subjects are convinced on the basis of phenomenal experience has auditory origins? The McGurk effect indicates that the seemingly auditory quality of overt speech perception is not necessarily auditory in origin and cannot be attributed solely to events within auditory or acoustic systems. And what holds for perception of overt speech holds also for perception of internal speech:

The seemingly auditory quality of our internal speech cannot be automatically attributed to events within an auditory or acoustic system, or even, as we will see, to *any* strictly sensory system. Thus, because the term auditory imagery used throughout this volume suggests that general auditory or acoustic representations are involved, a theoretically neutral term such as speech imagery is perhaps more appropriate for describing the seemingly auditory experience that arises during internal speech.

Does an Auditory System Underlie Internal Speech Perception? It has often been suggested that internal speech is auditory in nature and takes place within the same system that images pure tones, music, and environmental sounds such as a barking dog or a running faucet. For example, Baddeley and Logie (this volume) maintain that environmental sounds automatically access the store for digits, music, and internal speech. One plausible implication of this hypothesis is that normal levels of background music and noise should cause massive interference with speech perception. However, massive interference has never been observed. Even with complex tasks involving the use of verbal memory and reasoning, interference from music and noise has been difficult to demonstrate. Indeed, improvement sometimes occurs (Wilding, Mohindra & Breen-Lewis, 1982); and when interference has been found, the effect has been slight and difficult to replicate rather than massive. For example, Salame and Baddeley (1989) had difficulty replicating their own demonstration that music interferes with the encoding and recall of visually presented digits. Moreover, the music that sometimes did introduce interference was of an especially raucous and distracting sort (e.g., Offenbach's Cancan and Ravel's Bolero). And the weak and difficult to replicate interference effect depended on presenting the music at 75dB *on the average*, much louder than the normal level of speech (about 45–55 dB). There is reason to believe that such music at such levels of amplification could interfere with language comprehension because of effects on timing mechanisms that are shared by the otherwise independent systems for speech, and many other cognitive systems, including those for auditory and musical cognition (MacKay, 1987, pp. 90–111). This being the case, amplified Offenbach may also interfere with tasks that are otherwise unrelated to audition, auditory imagery, or even language (e.g., rotating a mental image of Texas or solving visual analogies). Demonstrating that such interference does not in fact occur is necessary for accepting the hypothesis that interfering effects of music are auditory in nature or specific to a common system for analyzing "pure speech auditory images" and "pure tone auditory images."

In summary, the seemingly auditory quality of inner speech is not necessarily auditory in origin and, without more solid evidence, cannot be attributed to events within systems that are strictly auditory or acoustic in nature. Like producing inner speech, perceiving inner speech may involve a phonological code.

Differences Between Internal Speech and Overt Acoustics. Many aspects of the acoustics of overt speech are normally absent from our awareness of self-produced internal speech. To illustrate, consider loudness and fundamental frequency, integral characteristics of the acoustics of overt speech. Unlike overt speech perception, awareness of the loudness and fundamental pitch of words produced internally is normally absent. Moreover, speakers normally fail to note the absence of these omnipresent characteristics of overt speech. If forced to characterize their own internal voice on these dimensions, they might say that their internal speech has neutral loudness and fundamental pitch. Consistent with such introspections, Intons-Peterson (this volume) presents experimental evidence suggesting that loudness is an attribute of sounds that is not specified in auditory imagery.

Such observations suggest that there exists a separate system for analyzing concepts related to acoustic aspects of speech such as loudness, intonation, sex of speaker, and speaker emotion or mood, and that this "auditory concept system" can operate in parallel with systems representing the phonological and sentential components of inner speech. Figure 6.1 illustrates the relations between these sytems for everyday language perception, which can be conceptualized as follows: An acoustic analysis system (see Fig. 6.1) feeds an initial analysis of speech in parallel to a phonological system and to an auditory concept system that categorically codes, for example, emotional content, intonational

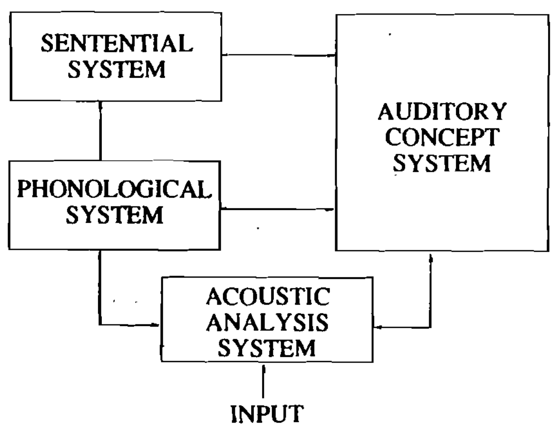

FIG. 6.1. Possible relations between the systems analyzing the acoustics, phonology, syntax and auditory concepts related to speech (e.g., loudness, sex of speaker, speaker identity, emotional content, and intonation). These same acoustic analyses and auditory concept systems may or may not analyze the acoustics for environmental concepts such as a barking dog and a car's horn.

category, loudness, sex of speaker, and speaker identity, including the distinction between self-produced versus other-produced speech.

Representations in the auditory concept system are hierarchically related to, but fundamentally different from, those in the acoustic analysis system. For example, whereas the auditory concept system would represent speaker sex directly and categorically (as male vs. female), the acoustic analysis system would represent speaker sex in terms of a variety of acoustic properties, one of which is the speaker's average fundamental frequency. And whereas the auditory concept system would represent intonation directly and categorically (e.g., as interrogative vs. declarative), the acoustic analysis system would represent intonation in terms of a variety of acoustic properties, one of which is the relative rise or fall in fundamental frequency during the final word or syllable of an utterance (see Levelt, 1989, pp. 312–317). Similarly, whereas the auditory concept system would represent the emotional attitude of a speaker directly and categorically (e.g., as excited vs. calm), the acoustic analysis system would represent emotional attitude via many acoustic properties, one of which is the relatively high pitched or shrieky voice quality of an excited speaker (see Levelt, pp. 102).

Representations in the auditory concept system are also fundamentally different from those in the phonological and sentential/syntactic systems. The phonological system represents syllables and speech sounds and their order of occurence in words, while those in the sentential/syntactic system can generate language-specific propositions to represent any information whatsoever. Unlike auditory concepts, however, propositions are noncategorical. Thus, whereas the auditory concept system would represent speaker identity categorically (e.g., familiar vs. unfamiliar), the sentential/syntactic system would represent speaker identity in terms of propositions such as "My son is talking to me," or, "Kenny MacKay is talking to me."

Unlike the phonological and sentential/syntactic systems, the internal structure of the auditory concept system in Fig. 6.1 is relatively unexplored, and its exact limits remain to be determined. For example, the auditory concept system may or may not also analyze the concepts underlying identification of environmental sounds such as a barking dog or a car's horn, and may or may not also analyze musical concepts. If not, then the auditory concept system in Fig. 6.1 requires a more specific label, say, "speech concept" system, to distinguish it from, say, "music concept" systems. For musicians, the complexity of these music concept systems might rival those for language itself.

Perhaps there are even separate, parallel systems for more specific *speech* concepts such as speaker identity, intonation, and emotional attitude. And perhaps the acoustic analysis system also has a complex internal structure with separate and parallel subsystems for representing phonetics as opposed to, say, the acoustics of music or of environmental sounds. Figure 6.1 captures none of these yet to be explored possibilities.

However, the analysis in Fig. 6.1 is consistent with observations indicating that internal speech produced in one's own voice behaves differently from internal speech produced in the voice of someone else, as when imagining the voice of a friend or of a famous figure such as Margaret Thatcher or John Kennedy. Geiselman and Glenny (1977) had subjects rehearse words either in their own voice or in a familiar but imagined male or female voice. A surprise recognition test for these words was then presented, and, as might be expected, recognition was superior when the voice used during rehearsal and recognition matched in sex. Interestingly, however, no such interaction was observed for words that subjects had rehearsed in their own voice: Recognition performance was no better when the words to be recognized were spoken by someone of same sex as the subject. This finding suggests that one's own inner voice is neutral with respect to sex (see Geiselman & Glenny), and the as yet unanswered theoretical question is why. The reason within the present analysis is that one's own inner speech is generated within the phonological and sentential systems of Fig. 6.1, and both of these systems are neutral with respect to loudness and fundamental frequency. However, imagining the sound of another person's voice requires both of these systems, *plus* the auditory concept system, operating in parallel (see Fig. 6.1). For as we have seen, the auditory concept system is required to represent speaker sex and loudness, and this added information can help speakers discriminate the traces for internally generated versus overtly perceived speech (but not always; see, e.g., Johnson & Raye, 1981; R. E. Anderson, 1984).

Such observations contradict the common assumption (see e.g., Baddeley & Logie, this volume) that there exists a single seat or locus for the inner ear, and that this locus is the same for all imagery with phenomenally auditory origins. Different aspects of the same speech signal such as phonology, loudness, fundamental frequency and intonation may come together in lower level systems such as the acoustic analysis and muscle movement systems for producing speech (MacKay, 1987, p. 16) but may be analyzed within parallel but separate higher level systems that represent self-produced and imagined inner speech differently. Although boundaries of these higher level systems remain to be explored, it currently seems unwise to assume that auditory events for speech, music, and environmental sounds are analyzed at all levels within a common system.

RELATIONS BETWEEN INTERNAL AND OVERT SPEECH PERCEPTION AND PRODUCTION

Relations between the perception and production of internal and overt speech have been the focus of a great deal of research. Five examples from this research tradition are discussed here, from both contemporary literature and classical literature of the 1920s and 1930s.

Processing Differences Between Internal and Overt Speech

Just as internal speech lacks characteristics of the overt speech signal, internal and overt speech differ in their processing characteristics, in both perception and production. For example, the generation of internal speech takes much longer than the perception of otherwise identical overt speech (MacKay, 1987, p. 114). Similarly, it takes longer to generate an image of, say, a letter (about 2 sec) than to recognize the corresponding visually presented letter (about 500 msec; Cocude & Dennis, 1986). Finally, the maximal rate of internal speech is much faster than the maximal rate of overt speech, all other factors being equal. Although an early comparision of the rate of internal and overt speech using data from a single subject (Landauer, 1962) failed to obtain a statistically reliable difference, subsequent and more systematic tests by R. A. Anderson (1982), MacKay (1981), Marshall and Cartwright (1978, 1980), and Weber and Castleman (1970) have uniformly found that internal speech procedes more rapidly than overt speech. Faster rates for internal versus overt production have also been observed for other highly practiced skills, for example, imagining oneself tying a shoelace versus actually tying a shoelace (Annett, 1988). The more rapid rate of internal speech suggests a possible benefit of internal rehearsal relative to overt rehearsal (see MacKay, 1981 for others) and requires explanation in theories of internal and overt speech.

Effects of Internal and Overt Speech Production on Perception

The issue of whether and how internal and overt speech production influences ongoing speech perception is currently rather controversial. To illustrate, consider the current, apparently unrecognized conflict between the findings of Lackner (1974) and Reisberg, Smith, Baxter & Sonenshine (1989) on the verbal transformation effect (VTE). The VTE refers to the fact that perception changes when an acoustically presented word is repeated via tape loop for prolonged periods (5 sec to several minutes). After hearing the word *pace* repeated for 3 minutes, for example, subjects report hearing words such as *face, paste, base, taste,* or *case,* and the number of perceptual forms and the rate of perceptual change from one form to another increases systematically as a function of time or repetitions (Warren, 1968).

The conflict between Lackner and Reisberg et al. arises from a variant of the standard VTE experiment discussed above. The variant involves repeated production of a word that results in a phenomenon known as the missing feedback effect: In a very well controlled experiment, Lackner found that the perceptual changes that occur when listening to a repeating word *fail to occur* when the subjects themselves are producing the repeated word; the auditory feedback that accompanies repeated *production* of a word fails to trigger verbal transformations. Lackner's subjects repeated a word every 500 msec for several minutes and later listened to a tape recording of their own output over earphones in a soundproof booth. The subjects experienced the usual transformations when *listening* to the tape recording of their own output, but, for some reason, experienced almost no perceptual transformations when *producing* the word.

This missing feedback effect is empirically interesting because the acoustic events at the ear are identical when hearing the input during versus after production. The missing feedback effect is also theoretically interesting because it bears on theories of the relation between speech perception and speech production. Lackner (1974) attributed the missing feedback effect to a corollary discharge or efference-copy that accompanies the motor command to produce a word. This corrollary discharge cancels or inhibits the external (proprioceptive and auditory) feedback resulting from producing the word, so that the on-line auditory input during production fails to bring about the fatigue induced perceptual changes that are the hallmark of the VTE. An unsolved problem in Lackner's account is why no *production errors* resembling the perceptual errors occurred when subjects actively repeated a word: Why doesn't fatigue also induce production errors?

The recent work of Reisberg et al. (1989) on the VTE also bears on these issues but has received empirical and theoretical interpretations that are quite different from those of Lackner. For example, Reisberg et al. claimed that VTEs do not differ for the standard perception procedure versus production procedures resembling those discussed above, although aspects of their data seem to contradict this empirical claim and support the findings of Lackner: In one experiment using the standard VTE perceptual procedure, 98% of Reisberg et al.'s subjects perceived a particular transform (e.g., *dress* for the repeating word *stress*) during 2 minutes or less of repetition, whereas 20% fewer subjects (i.e., 78%) perceived this transform when they repeatedly produced the word *stress*. In another similar experiment, 100% of the subjects hearing *stress* in the perception condition reported the transform *dress*, and 100% of the subjects in the production condition reported perceiving *dress* while repeatedly producing *stress*. Thus, a major difference remains in the results of the production condition in Lackner (few transforms of any type) versus Reisberg et al. (100% and 78% transforms of a particular type). Whereas Reisberg et al.'s perception subjects reported only 1.0 to 1.28 times as many transforms as their production subjects, Lackner's perception subjects reported about 15 times as many transforms as his production subjects.

The low probability of transforms in Lackner's (1974) production condition is almost certainly not attributable to reduced subject expectations of transforms. Because Lackner's subjects were instructed to monitor for and report deviations in perceived vowel quality, however small, as they repeatedly produced a word, the instructions surely led them to expect changes in their repeated productions.

Nor is the difference attributable to the procedure in Reisberg et al.'s (1989) perception condition of having an experimenter produce the words repeatedly and perhaps variably from one production to the next. In a pilot study, Reisberg et al. found no difference between conditions where an experimenter said the word repeatedly, where the subject said the word repeatedly, and where the word was generated repeatedly via computer (not unlike Lackner's tape loop procedure). Nor is the difference between the two studies readily attributable to differences in scoring procedures. Reisberg et al.'s scoring procedure was, if anything, more stringent than Lackner's so that, arguably, their production condition should have resulted in fewer rather than more transformations. In short, reasons for the differing results in these two studies are currently unknown and difficult to imagine, and the issue of the effects of language production on language perception in the verbal transformation task remains unresolved.

However, in an additional experiment, Reisberg et al. found a systematic relation between the VTE and "degree of enactment" on the part of subjects who themselves produced a word or phrase repeatedly either aloud (overtly), whispering, or silently (mouthing). Overtly articulated repetitions were said to be highly enacted, whispered repetitions less enacted, and silently mouthed repetitions still less enacted. Thus, progressively fewer subjects reported perceiving the preselected transform when they repeatedly spoke (85%), whispered (68%), or silently mouthed (53%) the word or phrase. Reisberg et al. concluded that the verbal transformation effect reflects an interpretive process that varies in direct proportion with the degree of enactment that accompanies a repeating stimulus originating via auditory input or via feedback from production. Other conditions in Reisberg et al. suggested that this repeating stimulus need not originate externally, via sensory channels, but could also originate internally, as when subjects imagine hearing themselves or a friend saying a word repeatedly.

Is Motor Activity Necessary for Perception of Internal Speech?

The classical hypothesis that motor activity is necessary for perceiving internal speech and other forms of imagery has been extensively examined and conclusively rejected. Early experiments showed that microscopic muscular movements, invisible to the naked eye, occurred during internal speech and other forms of imagery, and interestingly, this same EMG activity invariably precedes by a few milliseconds the full blown muscle activity that occurs during normal movements (e.g., Schmidt, 1982). From these observations it was hypothesized that EMG activity triggers sensory feedback that once was deemed essential for thinking and imagery (see e.g., Weisberg, 1980).

Consistent with the hypothesis that EMG activity triggers feedback necessary for centrally generated imagery, EMG responses were initially found to be localized or specific to the type of images experienced, rather than general or nonspecific in nature (Jacobson, 1930, 1931; Max, 1937): The pattern of EMG activity during imagery and during overt performance of the same action seemed to be identical. However, other studies reviewed by Feltz and Landers (1983) have raised questions about whether the EMG innervations associated with imaging occur more generally throughout the body than do corresponding overtly produced movements.

Even more damaging to the EMG-feedback hypothesis, a large number of early studies indicated that EMG activity was not necessary for internal imagery, and specifically not for the seemingly auditory imagery that often accompanies internal speech (Sokolov, 1972), silent reading (Pintner, 1913), and problem solving (Weisberg, 1980). For example, paralysis (Smith, Brown, Tolman & Goodman, 1947) and anesthetization of the lips and tongue (Dodge, 1896, as cited in Weisberg), and other forms of interference with EMG activity in speech muscles leave the ability to generate internal speech unimpaired (Sokolov, 1972). Such findings are inconsistent with the original hypothesis that low level motor activity and sensory feedback are necessary for perception of internal speech.

Also contrary to the feedback hypothesis, and to the more general hypothesis that EMG activity is essential for imagery, are the data on congenital and acquired dysarthria discussed above. If control of inner speech is fundamentally phonological rather than articulatory, and does not depend on peripheral musculature, as the data of Wilson and Baddeley (in Baddeley, 1990) suggest, then peripheral feedback from the musculature is also irrelevant to the control of inner speech and rehearsal. Similarly, if inner speech can be learned without use of the peripheral musculature, as the data of Bishop and Robson (1989) suggest, then feedback from peripheral musculature is also irrelevant to the acquisition of inner speech and rehearsal.

Dissociations Between Motor Activity and Perceptual Experience

Motor activity and perceptual experience can be dissociated because the production and perceptual experience of internal speech is unimpaired when the ability to move the corresponding muscles is prevented, for example, by motor paralysis or brain damage. The "fis phenomenon" (Smith, 1973) represents another such dissociation. Young children often have difficulty producing the muscle movements for a speech sound in certain phonetic environments even though they can perceive and presumably imagine these speech sounds perfectly well. For example, a child might say *fis* instead of *fish* but nevertheless be able to perceive the distinction between *fis* and *fish:* Thus, if an adult imitates the child by saying "O.K., here's your fis," the child will strenuously object, "No, no: FIS, FIS," indirectly indicating perception of the fis-fish distinction. Moreover, if a tape recording of the child saying the word "fis" (instead of "fish") in isolation is later played to the child, the child will perceive "fis" rather than "fish" with high probability.

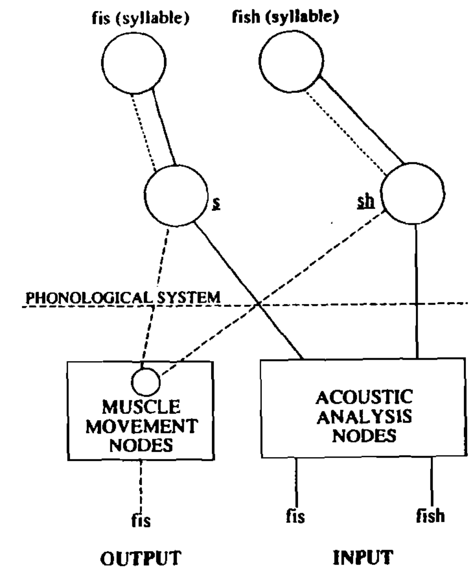

FIG. 6.2. An analysis of the "fis phenomenon" in children. Solid lines represent bottom-up connections. Broken lines represent top-down connections.

Figure 6.2 illustrates a general framework for explaining the "fis phenomenon" and other dissociations between motor abilities and perceptual experience. The input, either "fish" or "fis," is accurately analyzed in an acoustic analysis system and in a phonological system that plays a role in both perceiving and producing the distinction between these sounds. The problem arises in a subordinate system that represents the muscle movements for producing /s/ and /sh/ in this context: The phonological units for both /sh/ and /s/ have been mapped onto the units for producing /s/ within this muscle movement system, so that the child produces "fis" instead of "fish," the correctly intended and executed output *at the phonological level.*

A phenomenon that resembles the *fis* phenomenon in certain respects can be observed in the internal speech of adults who speak with a foreign accent or in a dialect that differs from the speech that they hear on a daily basis. For example, I speak a standard Canadian dialect with a distinctive "clipped" pronounciation of words containing the dipthong "ou" (e.g., *out, about, south,* etc). Despite having lived in California for two and a half decades, I fail to produce the standard American versions of these "ou" words in normal, everyday language production. Even with slow and deliberate attempts, I am unable to adequately produce an American "ou." Moreover, I normally fail to distinguish between the Canadian and American versions of "ou" in my everyday perception of spoken American, or in my internal and external speech, I automatically produce the

Canadian "ou" and perceive the American "ou" without becoming aware of the acoustic difference between the two. Here, then, is another aspect of speech acoustics (Canadian "ou") that is absent from internal speech *and overt perception,* as if acoustic characteristics of the American "ou" have been mapped onto a single abstract phonological unit that represents both (see Fig. 6.3). As a result, both perception and internal speech are neutral with respect to these acoustic properties. However, this single abstract "Canamerican ou" unit has been mapped onto a single pattern of muscle movements that correspond to the Canadian rather than the American "ou" (see Fig. 6.3), so that only the Canadian "ou" is overtly produced.

This accent phenomenon further illustrates how internal speech perception (the neutral "ou") can differ from acoustics of overt speech. Such differences between internal and overt speech are not limited to special dialects or accents, but are quite general in nature: As we have seen, many aspects of overt speech perception are absent during self-produced internal speech, as if inner speech involves the phonological system but not the acoustic analysis and/or muscle movement systems (see Fig. 6.1–6.3).

Are Perceptual Systems Completely Independent from Output Systems?

Within the framework (Fig. 6.2–6.3) for explaining dissociations between motor abilities and perceptual experience, some of the systems for perception and

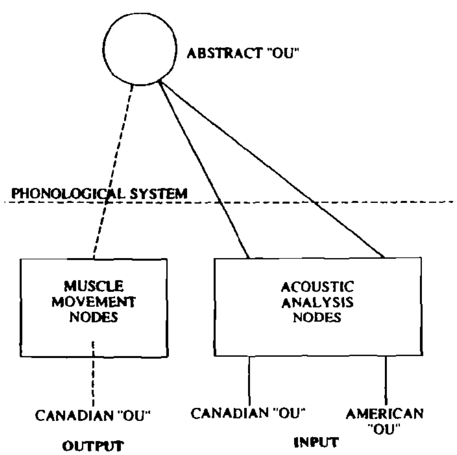

FIG. 6.3. An analysis of an "accent phenomenon" in the internal speech of an adult Canadian preceiving and producing American "ou" in words such as *out, south, house,* etc. Solid lines represent bottom-up connections. Broken lines represent top-down connections.

production of speech are shared rather than separate. MacKay (1987) reviewed a wide range of data for such shared perception/production systems. This evidence suggests that language perception and production, including the inner ear and inner voice, share the same phonological and sentential systems. However, the idea of a shared perception/production system for phonology has not gone unchallenged. Baddeley (1990) and Baddeley and Salame (1986) argued from dual task performance that input systems are separate from output systems at the phonological level. Their simultaneous tasks were comprehension of visually presented sentences and repeated internal or overt articulation of a word or syllable such as *the, the, the, the, the. . . .* The results indicated very little interference between these two tasks: Sentences were understood almost as well when subjects were repeating the syllable as when they were not. On the basis of this relative lack of interference, Baddeley concluded the following:

1. Input systems for analyzing the sentences must be separate from output systems for producing words or syllables;
2. An articulatory loop for rehearsing a word such as *the* must be separate from the systems for analyzing visually presented sentences resembling "moast peepul seemd tue bee aybul tue heer thuh werds eevan wen thay wurr seying thuh";
3. The inner voice must be separate from the inner ear because saying *the* repeatedly had little effect on the task of deciding whether visually presented nonwords are or are not phonologically compatible with real words, as is the case for the words "moast peepul seemd tue bee aybul tue heer thuh wirds internully."

However, all of these conclusions are open to question on several grounds. One is that repeating a syllable such as *the* may occupy only a small subcomponent of the phonological and other systems for comprehending sentences (see MacKay, 1987, pp. 52–55). This being the case, little interference would be expected in these tasks, even if (a) input systems for analyzing sentences and output systems for producing the word *the* are shared; (b) there is no articulatory loop that is separate from the phonological units for producing or repeating a word such as *the;* and (c) the inner voice is not completely separable from the inner ear.

TRADITIONAL ACCOUNTS OF RELATIONS BETWEEN INTERNAL AND OVERT SPEECH

The traditional theoretical approach to the perception of internal speech incorporates the idea of a "double agent," an internal speaker who speaks and an internal listener who listens. This "double agent assumption" is so common and accepted as to have become built into the everyday meaning of terms such as "speak" and "talk" in English and many other languages: The term "speech" implies talking to someone, so that internal speech must refer to talking to oneself, a hypothetical duplicate of the self who is listening rather than talking.

Many examples of the double agent assumption could be cited from the recent literature. Baddeley's (1986, 1990) articulatory/phonological loop theory represents an example that has addressed the issue of imagery during covert rehearsal, but it has so far failed to address the issues related to language production and perception that are of interest here. I therefore examine Levelt's (1989) recently proposed Perceptual Loop theory of how we generate and comprehend internal speech as an illustration of the "double agent" view.

Levelt argued that language production proceeds top down through a hierarchy of semantic nodes and phonological nodes. The lowest level nodes in this phonological production system are linked to the phonological perceptual system via two pathways: an internal "loop" that is used for perceiving internal speech, and an external "loop" that includes the muscle movement system, the auditory system, and a separate phonological system for perceiving both overt and internal speech. Under the double agent assumption, then, systems for producing speech are separate from comprehension systems, which also monitor internally and externally generated versions of the output for errors. How this second agent (the language comprehension system) "knows" that the production system has made an error, substituting *table* for *chair*, for example, is unspecified in Levelt's theory.

Reisberg et al. (1989) raised the double agent problem to another level by assuming that the internal listener (Levelt's language perception system) is capable of becoming an internal producer of auditory images: That is, Reisberg et al. proposed that auditory images for verbal materials can be generated internally using mechanisms that are independent of the usual mechanisms for producing language. Why we need such a listener-production system as well as the traditional production-production system, and how this duplicate listener-production system differs from the normal production system for language and speech pose additional problems that this view must solve. The answer of Reisberg et al. that a duplicate production system is needed for imagining and anticipating strictly auditory sounds that we cannot produce (say, the honking of a goose) may apply to "pure auditory imagery" but runs into difficulties when applied to inner speech because, as we have seen, the production of inner speech is fundamentally phonological rather than auditory in nature.

Levelt's support for his so-called Perceptual Loop theory is weak and open to alternate interpretations that apply also to other versions of the double agent assumption. The main source of support for Levelt's theory is an effect of auditory masking on the detection of a particular type of experimentally induced speech error in a study by Lackner and Tuller (1979). Lackner and Tuller had subjects repeat experimentally constructed phonological twisters such as pi-di-ti-gi at a controled rate for 30 sec and push a button every time they noticed making an error. In one (nonmasking) condition the subjects produced the phonological twisters without masking, and in another (auditory masking) condition, they heard white noise that masked their auditory feedback. Subjects detected errors involving substitutions of the place of articulation feature (e.g., ti-di-ti-gi instead of pi-di-ti-gi) no more often in the masking than in the nonmasking conditions (116 vs. 98). However, they detected errors involving substitutions of the voicing feature (e.g., di-di-ti-gi instead of pi-di-ti-gi) more often in the masking than in the non-masking conditions (252 vs. 175). To explain why voicing errors are easier to detect in the masking condition, Levelt argued that masking suppresses use of the external (auditory) loop, leaving only the internal loop (from the phonological production system to the phonological perceptual system). To explain why the masking effect was specific to voicing errors, Levelt argued that voicing (unlike place of articulation) depends on a small production difference that translates into large acoustic effects. Levelt then argued that the large acoustic effects in voicing errors are easiest to pick up using the external loop (acoustic analysis system) that happens to be suppressed in the masking condition, thereby making voicing errors more difficult to detect.

These arguments seem tenuous on several counts: One is that contrary to Levelt's assumption, there are as many production differences between voiced and unvoiced speech sounds as there are perceptual differences (Lisker, 1978). Moreover, comparing the "size" of articulatory and acoustic differences for different phonological features or for different values of the same phonological feature is like comparing eggs to chickens: To make sense, the comparison requires a theory of the relation between the two, and if such a theory were available, the notion of "size" would almost certainly be irrelevant (as is the case for the theory relating eggs and chickens).

Another problem is that important aspects of Lackner and Tuller's (1979) data find no explanation in Levelt's theory. For example, voicing errors were not only detected more often in Lackner and Tuler's masking condition, they also *occurred* much more often than place of articulation errors (427 vs. 214) in both the masking and nonmasking conditions. This additional finding is difficult to explain in Levelt's theory.

Another implication of Levelt's Perceptual Loop theory is that self-produced phonological errors should be detected more quickly and easily than word errors (all other factors being equal). The reason is that word errors involve units that are higher in the output hierarchy and thus further from the perceptual monitor for detecting them; thus, more time would be needed to reach the monitor for detecting word errors than for detecting phonological errors. Although further data are required for resolving this issue, available data do not support this prediction: Nooteboom (1980) reported that lexical errors are as easy to detect as phonological errors. This finding is difficult to explain without further assumptions in Levelt's theory.

Another problem for Levelt's Perceptual Loop theory concerns the nature of mental errors, that is, errors that occur during internal speech. Under the Perceptual Loop theory, overt speech should enable superior error detection (all other factors being equal) because the external loop in overt speech allows a second opportunity for detecting errors that is absent during internal speech. However, Dell's (1978, 1980) data do not support this prediction and are difficult to explain without further assumptions in Levelt's theory.

In conclusion, the traditional concept of an internal listener that monitors self-produced outputs derives from a figure of speech that may represent a poor foundation for building a theory of the structures and mechanisms underlying language perception and error monitoring. Moreover, the internal listener concept is functionally questionable: The "double agent" approach to comprehension of internal speech must address the fundamental issue of why speakers must independently "listen to" the meaning and sound of what they are saying internally when they know all along the meaning and sound of what they are saying.

THE ROLE OF INTERNAL SPEECH IN VISUAL WORD PERCEPTION

Just as nonauditory events can influence a seemingly auditory experience, as in the McGurk effect, nonvisual events can influence a seemingly visual experience. In particular, phonological processes that are involved in internal speech can influence the perception of visual stimuli. Unlike the McGurk effect, effects of internal phonology on visual perception are not new. One of the earliest themes of research in cognitive psychology was to show that a phonological code resembling inner speech contributes to tasks that involve visual stimuli, for example, silent reading (Pintner, 1913), visual word and letter detection (MacKay, 1972), and immediate recall of visual letter strings (Sperling, 1960). For example, MacKay demonstrated that phonological factors play a role in the ability to detect letter strings presented briefly via tachistoscope. Neither the stimulus nor the response was auditory, phonetic, or phonological in nature: Subjects were instructed simply to write down exactly what they saw.

Subjects were informed that the letter strings would consist of either correctly or incorrectly spelled words, and on half the trials they were told what word would be presented (correctly or incorrectly spelled). Two types of misspellings were presented: phonologically compatible misspellings, which can be pronounced in the same way as the original word (e.g., *werk* for *work*), and phonologically incompatible misspellings, which require a different pronunciation from the original word (e.g., *wark* for *work*). The data showed that phonologically incompatible misspellings were easier to detect than phonologically compatible ones, indicating that a phonological code must play a role in detecting visually presented letter strings under these conditions. This same effect had been observed earlier (MacKay, 1969) for subjects attempting to detect misspellings embedded in briefly presented sentences such as "Nobody knew thet the werk was compleated on the new buildung." Interestingly, however, it was shown (MacKay, 1972) that with tachistoscopic presentation of individual words, the difference between phonologically compatible and incompatible misspellings only emerges when subjects are verbally warned of what word will be presented. With no advance warning, the same phonologically incompatible strings are no easier to detect than phonologically compatible ones. This finding indicates that like inner speech, the phonological processes that underlie detection of visual letter strings are neither simple nor completely automatic in nature.

THE ROLE OF INTERNAL SPEECH IN IMMEDIATE RECALL OF VISUAL INPUTS

Sperling (1960) and Conrad (1964) have provided widely cited evidence that the code underlying immediate recall of visual letter strings is not visual but acoustic or phonological in nature and resembles inner speech. For example, Conrad showed that the pattern of errors that subjects make in immediate recall of visually presented consonant sequences resembled the pattern of perceptual errors that they make in identifying the same syllables presented auditorily against a background of white noise. Conrad and Hull (1964) extended this finding by showing that sequences of "acoustically similar" consonants (e.g., D, V, T, P, C) are harder to recall than sequences of "acoustically different" consonants (e.g., F, Y, D, R, K). Interestingly, however, Wickelgren (1965) and others have demonstrated that "acoustic similarity" only disrupts ability to recall the *order* of the letters and, if anything, tends to *facilitate* recall of the letters themselves. This finding presents a challenge for the hypothesis of Baddeley (1990) and Baddeley and Logie (chapter 8, this volume) that "acoustic similarity" reduces recall by impairing our ability to discriminate between similar traces. Impaired trace discrimination might reduce recall of the letters themselves and perhaps also their order (exactly how remains to be specified by Baddeley & Logie). But without additional assumptions, the trace discrimination hypothesis cannot explain why short-term recall is impaired for order but *not* items in strings of phonologically similar leters. Again, the exact nature of phonological processes underlying cognitive acts such as the recall of visual letter strings is not simple, not strictly visual, and not yet explained.

In summary, aspects of the code underlying perception and immediate recall of visual letter strings is not visual, and like inner speech, may be phonological in nature.

MEMORY, REHEARSAL AND INTERNAL SPEECH

The topics of memory, rehearsal and internal speech are closely related because the covert rehearsal that occurs, for example, when one silently repeats a telephone number for later recall, seems indistinguishable from internal speech (see also Baddeley and Logie, this volume). Indeed, rehearsal and its effects on memory may represent one of the main functions of internal speech. Accounts of the relation between memory, rehearsal, and internal speech are of course available (see e.g., Baddeley and Logie, this volume), and to review here the evidence that is consistent with these theories would be redundant. However, it is worth noting that we are a long way from theoretical consensus on relations between memory, rehearsal, and internal speech. For example, as Baddeley (1990, p. 72) notes, the well known data advanced in support of his own theory are "capable of being explained in several other ways."

What follows are some additional and less widely recognized constraints on extant and future theories of memory, rehearsal and internal speech.

Unrehearsability

Certain types of information, such as a particular smell, seem difficult to call up and rehearse or imagine in detail. Why are some types of information rehearsable, for example, familiar words or sentences, whereas other types of information are unamenable to rehearsal without extensive training, for example, isolated pure tones (Wickelgren, 1966)? It is not that phonology per se is easily rehearsed: For example, sentences heard in a foreign language are unrehearsable even if the phonology of these sentences is compatible with English phonology. Nor does rehearsability depend critically on storage capacity or stimulus complexity per se: Relatively simple stimuli, for example, a single phoneme such as a German trilled /r/, cannot be accurately rehearsed by someone unfamiliar with German.

Perhaps the main determinant of rehearsability is prior practice. If a behavior such as a trilled /r/ is so unfamiliar that the appropriate muscle movements have not been learned, internal rehearsal will be of little help in the overt expression of the behavior (see MacKay, 1981). The language memory literature has been able to overlook this limitation of internal rehearsal because the internally rehearsed behavior at issue (skilled language production) is a special case: Over the course of a lifetime the muscle movements for producing familiar words and syllables in one's native language have become highly practiced (see MacKay, 1981, 1982). In general, however, internal rehearsal tends mainly to benefit either simple skills that have been practiced since early childhood, or the complex skills of virtuoso performers, for example, professional musicians and sports players who have extensively practiced all levels of the skill (MacKay, 1981). Conversely, as

Feltz and Landers (1983) pointed out, unless subjects have some prior experience in a task, little or no effect of internal rehearsal is found.

Rehearsal, Volition and Awareness

As Baddeley and Logie (this volume) point out, conscious experience is one of the defining characeristics of inner speech and auditory imagery in general. However, like consciousness itself, the inner speech that occurs during rehearsal does not require conscious or intentional initiation. Although rehearsal can be and often is voluntary, it is not *necessarily* voluntary. For example, speakers are often unable to voluntarily remove an internally recurring phrase from awareness (Bargh, 1990; Reisberg, 1989): The phrase continues to repeat as if it were being rehearsed involuntarily. The partially involuntary nature of inner rehearsal may have contributed to Parkin's (1987, p. 11–12) observation that, "it is a perverse fact about human memory that we often remember things we would rather forget and forget things we want to remember." The involuntary repetitions seen in compulsive behaviors likewise suggest that repetition is not always under voluntary control.

Effects of Overt and Internal Rehearsal on Behavior

Studies of overt and internal rehearsal have focussed mainly on long-term memory tasks. However, interesting effects of overt versus internal rehearsal have also been observed for aspects of behavior such as errors and the maximal rate of speech. As reviewed here, these effects provide important new constraints on theories of inner speech.

Improvement Following Internal Versus Overt Rehearsal. Effects of internal and overt rehearsal differ in interesting and counterintuitive ways. When subjects (MacKay, 1981) practiced producing identical sentences either overtly or internally at maximal rate, internal speech initially improved faster with practice, but reached asymptote after about the same number of practice trials as overt speech. These results are shown in Fig. 6.4 (right panel), where the average production time (in secs per sentence) is plotted on log-log coordinates. As can be seen, both internal and overt speech improved as a function of practice, which in itself is interesting because the subjects were trying to speak at their maximum rate throughout. The regularity of the functions in Fig. 6.4 is also noteworthy. The practice function for the overt rehearsal condition is completely linear, as would be expected under the power law that describes most learning curves.

However, an important irregularity distinguished the results for internal versus overt rehearsal: Learning resulting from internal rehearsal was more rapid than

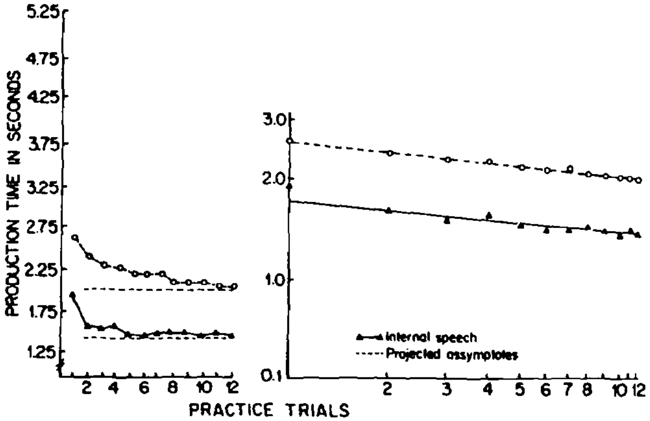

FIG. 6.4. Left panel: Time to produce identical sentences (in seconds) overtly (circles) versus internally (triangles) as a function of practice. Adapted from MacKay (1982). Right Panel: Log-log replot of the left panel data.

could be expected from log-log linearity on the initial practice trial (see Fig. 6.4). This "upward deviation" from log-log linearity indicates that internal speech improved faster than did overt speech following that initial trial.

Such upward deviations from log-log linearity are not limited to inner speech, but have been observed in other cognitive skills as well. For example, Snoddy (1926) reported an initial upward deviation from log-log linearity in the practice function for a mirror tracing task. Subjects watched their actions reversed in a mirror as they used a pencil to trace the outline of a visual pattern such as a star. The time per pattern (corrected for errors) is shown in Fig. 6.5 (as replotted in Newell & Rosenbloom, 1981). As can be seen there, production times improved rapidly with practice over the first four trials and then settled into log-log linearity for the remaining 55 trials. Explaining such deviations are a major challenge for theories of internal and overt rehearsal involving a wide range of cognitive skills (see MacKay, 1982).

Correlated Rates for Internal and Overt Rehearsal. As can be seen in Fig. 6.4, internal speech proceeded more rapidly than overt speech across all 12 practice trials in MacKay (1981) (see also Anderson, 1982; Marshall & Cartwright, 1978, 1980; and Weber & Castleman, 1970). However, for a given subject, maximal rates of internal and overt rehearsal were highly correlated. Subjects who produced sentences quickly during overt rehearsal (MacKay, 1981) also produced them quickly during internal rehearsal, as might be expected if internal and overt rehearsal involved many of the same components (i.e., the Phonological and Sentential systems in Fig. 6.1–6.3).

Transfer Effects from Internal to Overt Speech. By adding a transfer paradigm to the procedure discussed above and by using German-English bilinguals as subjects, MacKay (1981) demonstrated that internal and overt rehearsal cause equivalent improvement in the ability to produce the phonological and sentential components of a sentence. Specifically, the subjects practiced producing a sentence in one language at maximal rate either internally or overtly and then produced a transfer sentence that was either a word for word translation or a nontranslation of the practiced sentence in their other language. The results (see Table 6.1) indicated that the maximal rate of speech was faster when the transfer sentences were translations rather than nontranslations. Moreover, the degree of transfer following internal and overt practice was equivalent.

Differing Effects of Overt Versus Internal Rehearsal on Errors. Dell (1978, 1980) reported that various types of errors (anticipations, perseverations, and reversals) occurred with identical absolute frequency in inner and overt speech. However, Dell and Repka's (in press) more recent study of effects of internally and overtly rehearsing phonological twisters gave a slightly different pattern of results. Dell and Repka's subjects reported inner slips less frequently, and more often in syllable-, word-, and phrase-initial positions, relative to overt slips. Moreover, overt rehearsal or repetitions of a phonological twister reduced the probability of errors during subsequent overt production of the twister, but inner rehearsal failed to reduce errors when a twister subsequently was articulated overtly.

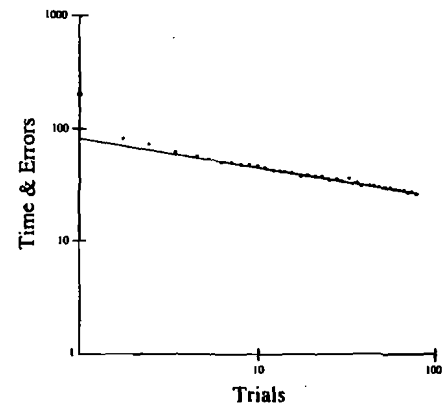

FIG. 6.5. Production times (corrected for errors) as a function of practice (log-log coordinates) in a mirror tracing task (from Snoddy, 1926, modified from the replot of Newell and Rosenbloom, 1980). Filled circles represent the "upward deviation" from log-log linearity (see text for explanation).

To explain these new results, Dell and Repka invoked an earlier suggestion of Vygotsky and others that speakers are capable of abbreviating their inner speech by omitting noninitial segments in syllables and words, especially words occupying noninitial positions in a phrase. Dell and Repka then explained the differences between their results and those of Dell (1978, 1980) in terms of individual differences in this hypothesized process of phonological abbreviation. Whereas subjects in Dell and Repka were all undergraduates, most of the subjects in Dell were psychology graduate students engaged in the study of short-term memory. Unlike undergraduates, these graduate students would have been quite knowledgeable about internal speech and its theoretical importance, and so may have been less likely to abbreviate their internal speech than the undergraduates in Dell and Repka. However, further research on these issues is clearly needed.

SUMMARY AND CONCLUSIONS

This chapter has reviewed a number of fundamental constraints or requirements that theories of inner speech must address. The present list of constraints is undoubtably incomplete, and will surely grow as the field progresses. Contained within the list, however, are *sine qua non* requirements for a viable theory: Theories of inner speech which fail to capture these constraints can be considered incomplete or inadaquate, and the traditional "double agent" approach to explaining relations between perception and production of internal speech seems to fall within this category.

What sort of theory will be needed for explaining the fundamental phenomena of inner speech? The theory must postulate a hierarchy of units, including units representing phrases, words, morphemes, and above all, phonological components. In producing inner speech, these units must be activated in sequence, but without activating muscle movement units for overt articulation.

The theory of inner speech must explain why some aspects of language are impossible to generate internally, why inner speech is sometimes involuntary and difficult to control, and why effects of overt and internal rehearsal on speech production are similar in some respects and different in others. However, the theory of inner speech must not give a central role to articulatory units or abilities, or to electromyographic activity within articulatory, laryngeal and other speech muscles. Nor is it necessary for the theory of inner speech to postulate a strictly *auditory* code for the "inner ear."

In addition to providing a standard against which to evaluate current theories, the criteria outlined here can be used to develop new and more adequate theories. Indeed, I myself hope to use these criteria in extending my own theory of language perception/production (MacKay, 1987) to cover inner speech. A final question concerns the generality of the present criteria. If they apply to other

inner skills besides language production/perception, the present chapter may provide a rough outline for what a general theory of imaging must eventually explain.

ACKNOWLEDGMENTS

The author thanks D. Reisberg, D. Burke, and two anonymous reviewers for valuable comments on an earlier version of this chapter.

REFERENCES

Anderson, R. A. (1982). Speech imagery is not always faster than visual imagery. *Memory and Cognition, 10,* 371–80.

Anderson, R. E. (1984). Did I do it or only imagine it? *Journal of Experimental Psychology: General, 113,* 594–613.

Annett, J. (1988). Motor learning and retention. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Volume II* (434–440). Chichester, England: Wiley.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Baddeley, A. (1986). *Working memory.* Oxford University Press.

Baddeley, A. (1990). *Human memory: Theory and practice.* Boston: Allyn & Bacon.

Baddeley, A., & Salame, P. (1986). The unattended speech effect: Perception or memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12*(4), 525–529.

Bargh, J. A. (1990). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Ulleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford.

Bishop, D. V. M., & Robson, J. (1989). Unimpaired short-term memory and rhyme judgement in congenitally speechless individuals: Implications for the notion of "articulatory coding." *Quarterly Journal of Experimental Psychology, 41a,* 123–141.

Cocude, J., & Dennis, R. W. (1986). The time course of mental imagery: Latency and duration of mental images. In D. G. Russel & D. F. Marks (Eds.), *Imagery 2.* Dunedin, New Zealand: Human Performance.

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology, 55,* 75–84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55,* 429–432.

Dell, G. S. (1978). Slips of the mind. In M. Paradis, (Ed.), *The Fourth LACUS Forum.* (pp. 69–75). Columbia, SC: Hornbeam Press.

Dell, G. S. (1980). *Phonological and lexical encoding in speech production: An analysis of naturally occurring and experimentally elicited speech errors.* Unpublished doctoral dissertation, University of Toronto, Toronto, Canada.

Dell, G. S., & Repka, R. J. (in press). Errors in inner speech. In B. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition.* New York: Plenum.

Dodge, R. (1896). Motoric speech production. Cited in Weisberg, R. W. (1980). *Memory, thought and behavior.* New York: Oxford University Press.

Ellis, A. W. (1988). Normal writing processes and peripheral acquired dysgraphias. *Language and Cognitive Processes, 3*(2), pp. 99–127.

Feltz, D., & Landers, D. (1983). The effects of mental practice on motor skill learning and performance: A meta-analysis. *Journal of Sport Psychology, 5*, 211–220.

Geiselman, R. E., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition, 5*, 499–504.

Jacobson, E. (1930). Imagination of movement involving the skelatal muscles. *American Journal of Physiology, 91*, 567–608.

Jacobson, E. (1931). Imagination, recollection, and abstract thinking involving the speech musculature. *American Journal of Physiology, 97*, 200–209.

Johnson, M. K., & Raye, C. L. (1981). Reality monitioring. *Psychological Review, 88*, 67–85.

Lackner, J. (1974). Speech production: Evidence for corollary discharge stabilization of perceptual mechanisms. *Perceptual and Motor Skills, 39*, 899–902.

Lackner, J. R., & Tuller, B. H. (1979). The role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 281–294). Hillsdale, NJ: Lawrence Erlbaum Associates.

Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills, 15*, 646.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: M.I.T. Press.

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.

MacKay, D. G. (1969). Phonetic factors in the perception and recall of spelling errors. *Neuropsychologia, 6*, 321–325.

MacKay, D. G. (1972). Input testing in the detection of misspellings. *American Journal of Psychology, 85*, 121–128.

MacKay, D. G. (1981). The problem of rehearsal or mental practice. *Journal of Motor Behavior, 13*, 274–285.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89*, 483–506.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer-Verlag.

MacKay, D. G. (in press). Spontaneous and evoked slips of the pen, tongue, and typewriter. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C. W. Wallesch (Eds.), *Linguistic Disorders and Pathologies*. Berlin: Walter de Gruyter & Co.

Marshall, P. H., & Cartwright, S. A. (1978). Failure to replicate a reported implicit/explicit speech equivalence. *Perceptual and Motor Skills, 46*, 1197–1198.

Marshall, P. H., & Cartwright, S. A. (1980). A final (?) note on implicit/explicit speech equivalence. *Bulletin of the Psychonomic Society, 15*, 409.

Max, L. (1937). An experimental study of the motor theory of consciousness IV: Action current responses in the deaf during awakening, kinesthetic imagery and abstract thinking. *Journal of Comparative Psychology, 24*, 301–344.

Meringer, R., & Meyer, K. (1895). *Versprechen und verlesen*. Stuttgart: Behrs Verlag.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–56). 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* (pp. 87–95). New York: Academic Press.

Parkin, A. J. (1987). *Memory and amnesia: An introduction*. New York: Basil Blackwell.

Pintner, R. (1913). Inner speech during silent reading. *Psychological Review, 30*, 129–153.

Reisberg, D. (Nov., 1989). *Anticipations and after-thoughts: How far does the "present" extend?* Paper presented at the 30th Annual Meeting of the Psychonomics Society, Atlanta, GA.

Reisberg, D., Smith, D., Baxter, D., & Sonenshine, M. (1989). Enacted auditory images are ambiguous; Pure auditory images are not. *Quarterly Journal of Experimental Psychology, 41A*, 619–181.

Salame, P., & Baddeley, A. (1989). Effects of background music on short-term memory. *Quarterly Journal of Experimental Psychology, 41A*, 107–122.

Schmidt, R. A. (1982). *Motor control and learning: A behavioral emphasis*. Champaign, Ill.: Human Kinetics.

Smith, N. V. (1973). *The acquisition of phonology: A case study*. New York: Cambridge University Press.

Smith, S. M., Brown, H. O., Tolman, J. E. P., & Goodman, L. S. (1947). Lack of cerebral effects of D-tubocurarine. *Anesthesiology, 8*, 1–14.

Snoddy, G. S. (1926). Learning and stability. *Journal of Applied Psychology, 10*, 1–36.

Sokolov, Y. N. (1972). *Inner speech and thought*. New York: Plenum.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*, 1–29.

Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin, 70*, 261–270.

Watson, J. B. (1950). *Behaviorism*. New York: Norton.

Weber, R. J., & Castleman, J. (1970). The time it takes to imagine. *Perception & Psychophysics, 8*, 165–168.

Weisberg, R. W. (1980). *Memory, thought and behavior*. New York: Oxford University Press.

Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *American Journal of Psychology, 78*, 567–574.

Wickelgren, W. A. (1966). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology, 6*, 13–61.

Wilding, J., Mohindra, N., & Breen-Lewis, K. (1982). Noise effects in free recall with various orienting tasks. *British Journal of Psychology, 73*, 479–486.