CHAPTER 3

THE PERCEPTUAL LOOP THEORY

3.1 Introduction

This chapter explores Levelt's (1983; 1989) Perceptual Loop Theory in order to understand the process of self-monitoring and self-repair in spontaneous speech. Within the context of Levelt's theory, this chapter also examines previous studies that have examined the time intervals of these processes.

3.2 Self-Monitoring and Self-Repair

Thus far, it has been shown that self-repairs can be regarded as a manifestation of a "quality control" (Hieke, 1981, p. 148) mechanism present in the process of speech production to correct pre or postarticulatory errors. In order for self-repairs to take place, there must be an awareness that an error is about to be, or has been produced by the speaker. Hence, the concept of self-repair co-exists with the idea that self-monitoring occurs in the process of speech production. There have been attempts to explain the relationship between monitoring and self-repair in speech, particularly to account for:

(i) how and when errors are detected and corrected

(ii) how soon after error detection speech is interrupted.

One of the main theoretical models explaining how speech is monitored and repaired is the Perceptual Loop Theory (Levelt, 1983; 1989).

3.3 The Perceptual Loop Theory

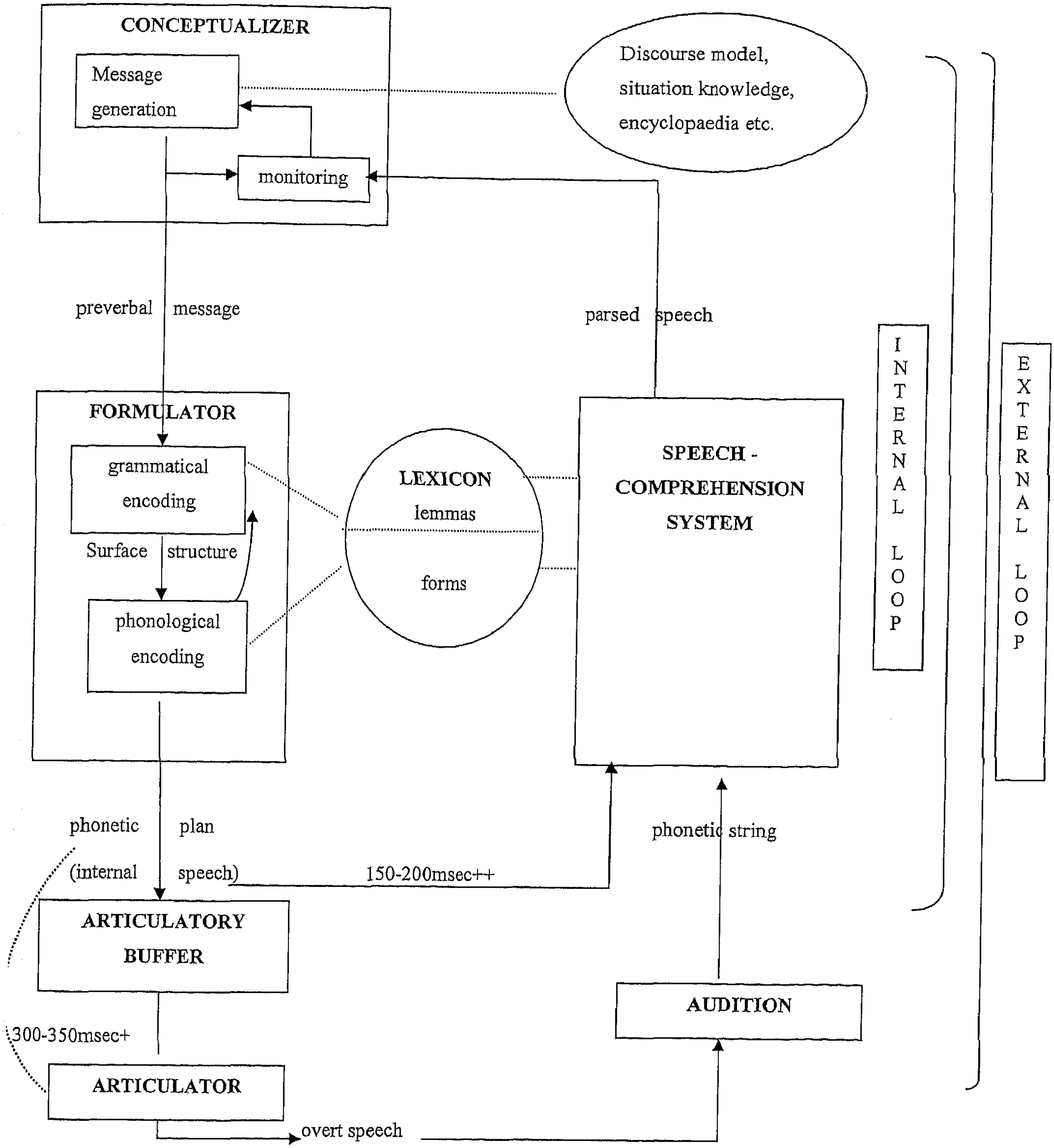
Based on a corpus of repairs made in the spontaneous speech of adult speakers of Dutch, Levelt (1983, 1989) formulated a theory to account for both monitoring and repairing in speech. The theory is based on the premise that speakers monitor their own speech just as they monitor the speech of others (Levelt, 1983; Levelt, Roelofs & Meyer, 1999). As was shown in the previous chapter, Levelt divided self-repairs into three major phases (See Figure 2.6):

(i) monitoring and interrupting speech whenever trouble is detected

(ii) hesitating and pausing (characterized by the use of silent or filled pauses)

(iii) repairing disfluent speech.

This process of self-monitoring and self-repair is accounted for in Levelt's (1983, 1989) model of speech production as shown in Figure 3.1, which gives a more detailed picture of the process of speech production shown previously in Figure 2.1.



Boxes = processing components; Circle and Ellipse = knowledge stores
+estimated time from delivery of phonetic plan to start of articulation
++ estimated time for recognition of internal speech

(from Levelt, 1989, pp. 9, 470 & 472)

Figure 3.1

A Blueprint for the Speaker

3.3.1 The Speech Production Process

Like Figure 2.1, Figure 3.1 also illustrates the process of speech production beginning from the conceptualization of the message to the articulation of the message. As can be seen from Figure 3.1, the generation of an idea or message of an intended utterance occurs at the Conceptualizer. The arrow from the message generation box and heading toward the monitoring box shows that at this stage of conceptualization, the message can be monitored for example, for appropriateness. The speaker might need to decide, for instance, on the right choice of word to express a particular idea based on his knowledge of the social rules governing language use.

If the process of monitoring at this stage finds the message to be inappropriate for some reason, a new message can be generated. A preverbal message that goes through this stage goes in as input into the Formulator, which turns this concept into a linguistic structure. This is done through the process of lemma selection, as explained in 2.2. A lemma is retrieved from the mental lexicon, where information about the lemma's meaning, syntactic, morphological and phonological features are available. Thus, the lemma can be grammatically encoded, producing a surface structure for the message. As explained in the previous chapter, this process involves, accessing information about grammatical form and the features associated with it such as person, number, tense and aspect (see Figure 2.2).

The surface structure is then phonologically encoded, where syllabification of the structure is thought to take place (Levelt, 2001; Levelt, Roelofs & Meyer, 1999). This process is shown in more detail in Figures 2.2 and 3.2. Both these figures show that at this stage, the individual sound segments that make up the intended word as well as its syllabic structure are thought to be put together. As shown in Figures 2.1, 3.1 and 3.2, the phonological word is then phonetically encoded, which results in a phonetic plan or "gestural" (Levelt, Roelofs & Meyer, 1999, p. 33) or "articulatory score" (Levelt, 2001). This plan specifies how the word is to be articulated by the speech organs. This phonetic plan is referred to as "internal speech". As indicated in Figure 3.1 internal speech then goes into the Articulator, and is realized as "overt speech" when the speech organs perform according to this score, which is audible to the speaker and to other listeners.

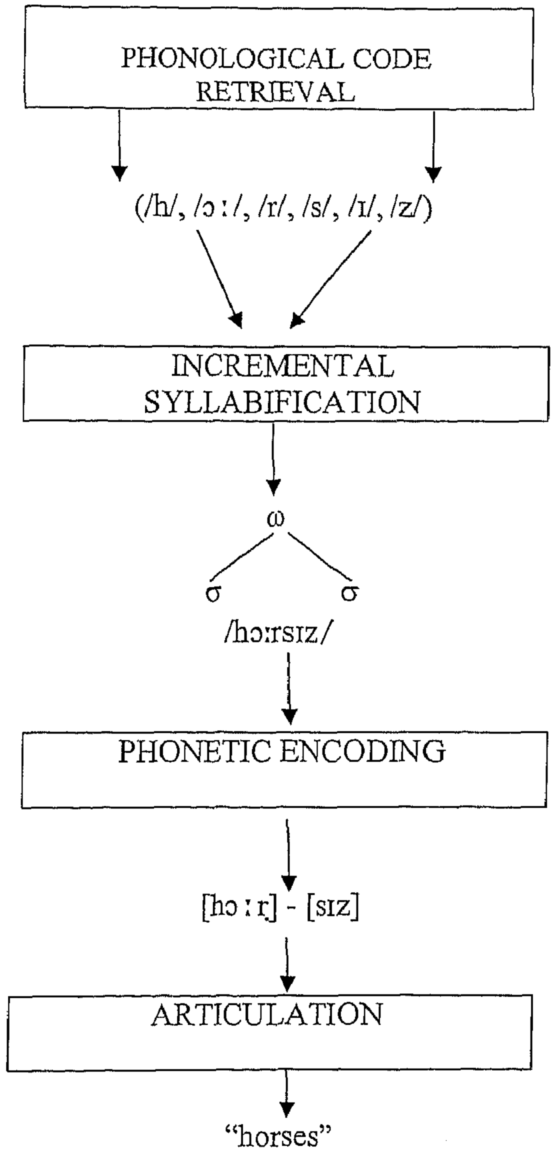


Figure 3.2

Form-Processing Stages from Levelt (1998)

Figure 3.1 shows that internal speech can also be monitored (this will be explored in the next section). Since the process of language production, from message conceptualization to the articulation of the message, occurs at an extremely fast pace, there is said to be an "Articulatory Buffer" (Levelt, 1989, p. 12) between the Formulator and the Articulator, where the phonetic plan is temporarily stored. While the Articulator is executing a phonetic plan, the next message to be articulated will be stored in this buffer, while waiting to be retrieved and executed (Levelt, 1989, p. 13).

3.3.2 The Process of Self-Monitoring

Once the articulators go into motion, audible speech is produced. According to Levelt's (1989) model, audible (see Audition in figure 3.1) self-produced overt speech goes through the Speech-Comprehension System, where it is processed in the same way we process other people's speech that we hear. This system also processes internal speech or prearticulatory speech as shown in Figure 3.1, and it is suggested that this is where error detection for both internal and overt speech occurs. Levelt, Roelofs and Meyer (1999, p. 33) illustrate this point with the following examples:

1. *Entrance to yellow ... er, to grey*
2. *We can go straight to the ye-... to the orange dot*

In the first example, it is suggested that overt speech is being monitored since the speaker can hear the error, that is, *yellow*. In contrast, in the second example, the speaker interrupts after the first syllable of the word *yellow*, suggesting that the error was already detected prior to its utterance. This detection is possible presumably because internal speech is being monitored.

Partially intercepted errors (Levelt, 1989, p. 474), such as in the second example above, which are auditorily perceived as fragments (Blackmer & Mitton, 1991; Nakatani & Hirschberg, 1994; Shriberg, 1994), and the interruption of speech by various forms of hesitation in speech (see example from Shriberg, 1999 in 2.9.1), are said to be indications of prearticulatory editing. Other evidence of prearticulatory editing comes from studies, where subjects reported that they detected errors in their inner speech (Dell & Repka, 1992). Postma and Noordanus (1996) also found that apart from normal speech production, subjects reported errors in silent, mouthed, and noise-masked speech production, implying that speakers need not hear their errors to correct them.

Such evidence supports Levelt's prearticulatory monitoring. According to Levelt (1983, p. 50), it is "assume[d] that repairing speech involves a perceptual loop: the self-produced inner or overt speech is perceived, parsed and checked with respect to intentional and contextual appropriateness, agreement of intended and delivered message, and linguistic correctness". In other words, as can be seen in Figure 3.1, both the internal and external loop feed input into the Speech-Comprehension System, which like the Formulator, has access to the mental lexicon, which presumably means that internal or external speech can be checked against information that is available in the lexicon. The output form the Speech-Comprehension System is parsed speech, which in a sense is the 'unpackaged' internal or external speech being processed. This speech is monitored, as shown in Figure 3.1.

If an error is detected, then decisions are made as to whether to interrupt speech or to go on. If a repair is to be made, then replanning begins at the Conceptualizer. Thus, the monitor is said to carry out two functions:

(i) a matching function, where inner and outer speech is compared to intentions and "criteria or standards of production" (Levelt, 1983, p. 50)

(ii) an awareness function, where a speaker is made aware of an error so that he may take corrective action.

During this process of monitoring, errors may be detected, resulting in speech being interrupted and hesitation produced, while repairs are planned. If an error is detected before the articulators have begun producing it, and speech is subsequently stopped, no overt error is audible. If the speaker hesitates through prolongations, silent and filled pauses, then these are all that are heard. This process is possible, if the error is detected while the message containing the error is being stored at the Articulatory Buffer (see Figure 3.1). However, if the signal to interrupt speech following error-detection reaches soon after or after the articulators have begun the production of the error, then we hear the error being cut-off mid production or hear speech interrupted soon after the error is produced, followed sometimes by hesitation and a repair to the error.

Menyhárt (2003) illustrates where these disfluencies may occur by adapting Levelt's model as shown in Figure 3.3. Disfluencies, such as grammatical, lexical, phonological, and pronunciation problems are said to originate at the corresponding level, while restarts, prolongation and within-word pauses seem to originate from word retrieval problems and the ordering of sound segments as shown in Figure 3.3. She also suggests that features that are generally considered as stalling devices, such as silent and filled pauses and repetitions, occur at the conceptual level. However, why this is so is not made clear.

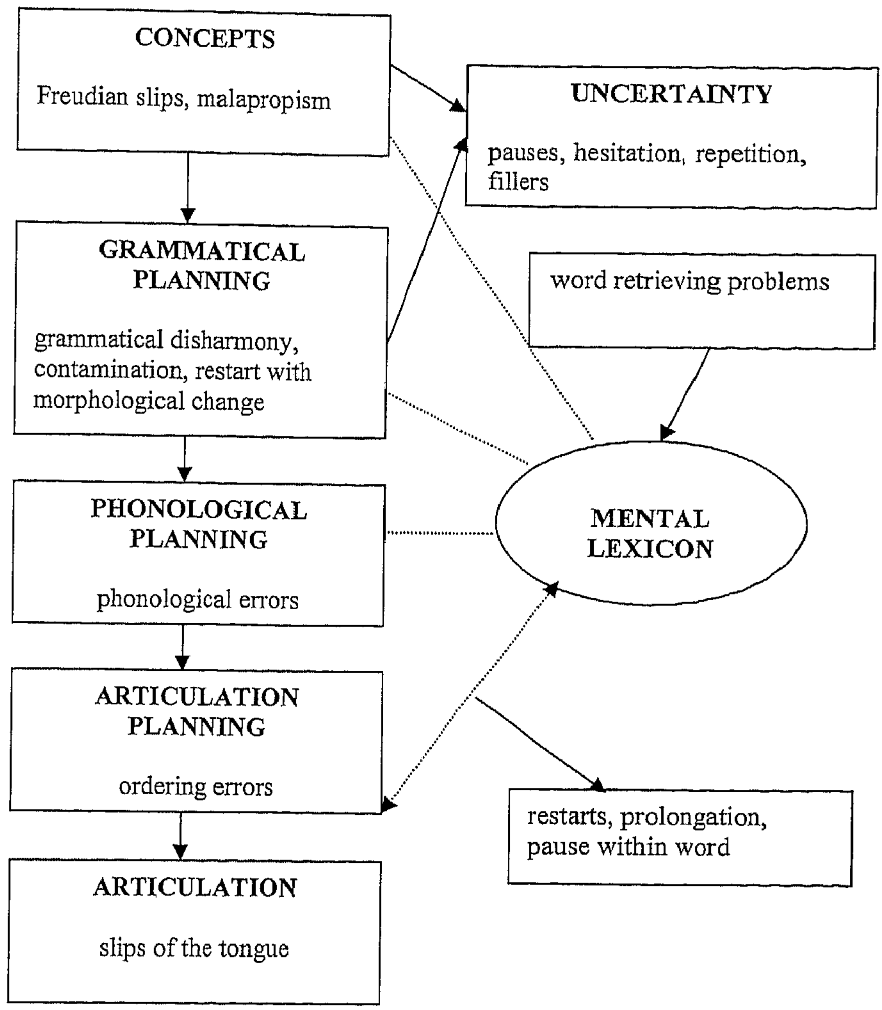


Figure 3.3

The Process of Speech Production and the Sources of Disfluency Phenomena from Menyhárt (2003, p. 45)

It should be noted that whilst errors may originate at the various levels, monitoring essentially only takes place when the utterance has been generated at internal speech or produced as external speech, as indicated in the monitoring model in Figure 3.1. It was first thought that the outputs at the other processing levels, for example, grammatical and phonological encoding could not be monitored in this model. In other words, only the final internal and overt speech (see Figure 3.1) can be monitored (Levelt, 1989).

However, a later study by Wheeldon and Levelt (1995) found that subjects seemed to be monitoring at a more abstract phonological level.

In one of the experiments carried out in their study, subjects first listened to a sound they had to monitor. They then listened to a word in English, and had to press a button if their non-verbal Dutch translation of that English word contained the sound that they were monitoring. The speakers repeated this experiment by counting aloud in Dutch. The reason for doing this was based on the theory that counting aloud amounted to articulatory suppression, which would impede the phonetic and articulatory processes, but not the phonological processes (Wheeldon & Levelt, 1995).

The speakers' responses were measured from the onset of the English word to the time they pressed the button when they detected the sound. A comparison of these two experiments showed that there was little difference in monitoring latencies when subjects counted aloud. This suggests that a more abstract phonological code was being monitored (Wheeldon & Levelt, 1995). However, Wheeldon and Levelt (1995) point out that there is no direct evidence to explain if the monitor has direct access to the process of phonological encoding prior to the phonetic plan.

3.4 A Comparison with Other Production Models

Nevertheless, in general, Levelt's monitoring model still sees monitoring as mainly taking place when elements have been put together, and not at individual levels of processing. This idea that only the end product of the processes is monitored differs from Laver's (1969, 1980) *Production Model* and MacKay's (1987) *Node Structure Theory* (NST).

3.4.1 Laver's Production Model

While the main processes in Laver's model, from the conceptualization to the articulation of the message, are similar to Levelt's, Laver's model, shown in Figure 3.4, differs in the way the monitor functions. As can be seen in Figure 3.4, in Laver's model, monitoring is a continuous process (Laver, 1969), occurring while speech is being planned, at each of the prearticulatory stages of speech production. The prearticulatory monitoring device shown to the left of Figure 3.4. is deemed to be external to the speech production process. It monitors and corrects errors as it is being formulated. The correction of errors necessitates going back to the component from where the error originates. As the linguistic elements of speech are being put together, any error that is detected would result in a decision to stop speech, and edit the erroneous portion, or otherwise (Laver, 1980). Thus, as with Levelt's model, errors can be detected before the error is articulated.

Figure 3.4 also shows that, similar to Levelt's model, speech is also monitored after it is articulated. Levelt's (1989, p. 468) criticism of such a model is that monitoring and editing at every processing level involves a "reduplication of knowledge" since both the element being monitored and the monitor is assumed to have the same information. In contrast, in Levelt's model (1989), apart from the conceptual message, it is the output of the Speech-Comprehension System, parsed speech that is checked (see Figure 3.1). Speech is not checked at every processing level, except perhaps at the abstract phonological level (Wheeldon & Levelt, 1995). Thus, less overt errors would be expected in Laver's model, since speech is monitored and repaired at all levels. This is because less overt errors would be expected to get through without being repaired to the later stages of production. In relation to this, it would be anticipated that repair-planning would take less time since it is done at the level of occurrence. At the same time, if speech is being subjected to that many checkpoints, this may delay the actual production of the message.

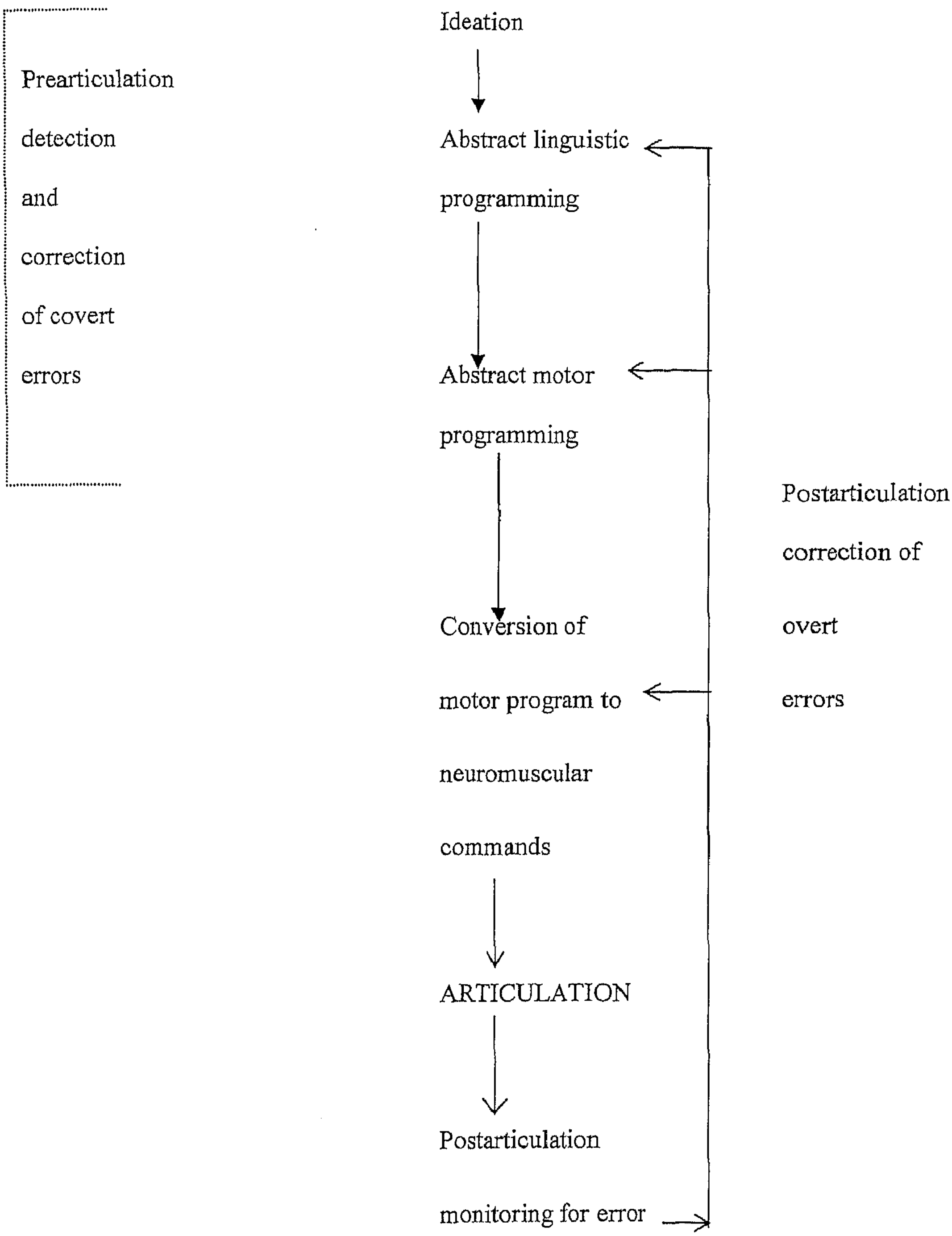


Figure 3.4

Schematic Outline of Neurolinguistic Functions from Laver (1980, p. 290)

3.4.2 MacKay's Node Structure Theory

MacKay (1987) developed the Node Structure Theory (NST) to account for perception and action in speech. Nodes refer to 'representational units' (MacKay, Wulf & Abrams, 1993, p. 625), which represent linguistic units like words, syllables, and phonological features. Nodes are said to be divided and hierarchically "organized into systems and subsystems that can be independently activated' (MacKay 1997, p. 62). MacKay (1992a, p. 42) explains that nodes, which form the basic components of the node structure theory, can be divided into three classes, depending on how they connect with each other. His description of these three classes of nodes is summarized in Table 3.1, where it can be seen that these three classes correspond to the three components of speech production in Levelt's (1989) model: the Conceptualizer, the Formulator and the Articulator. However, the processes of speech production differ in both models.

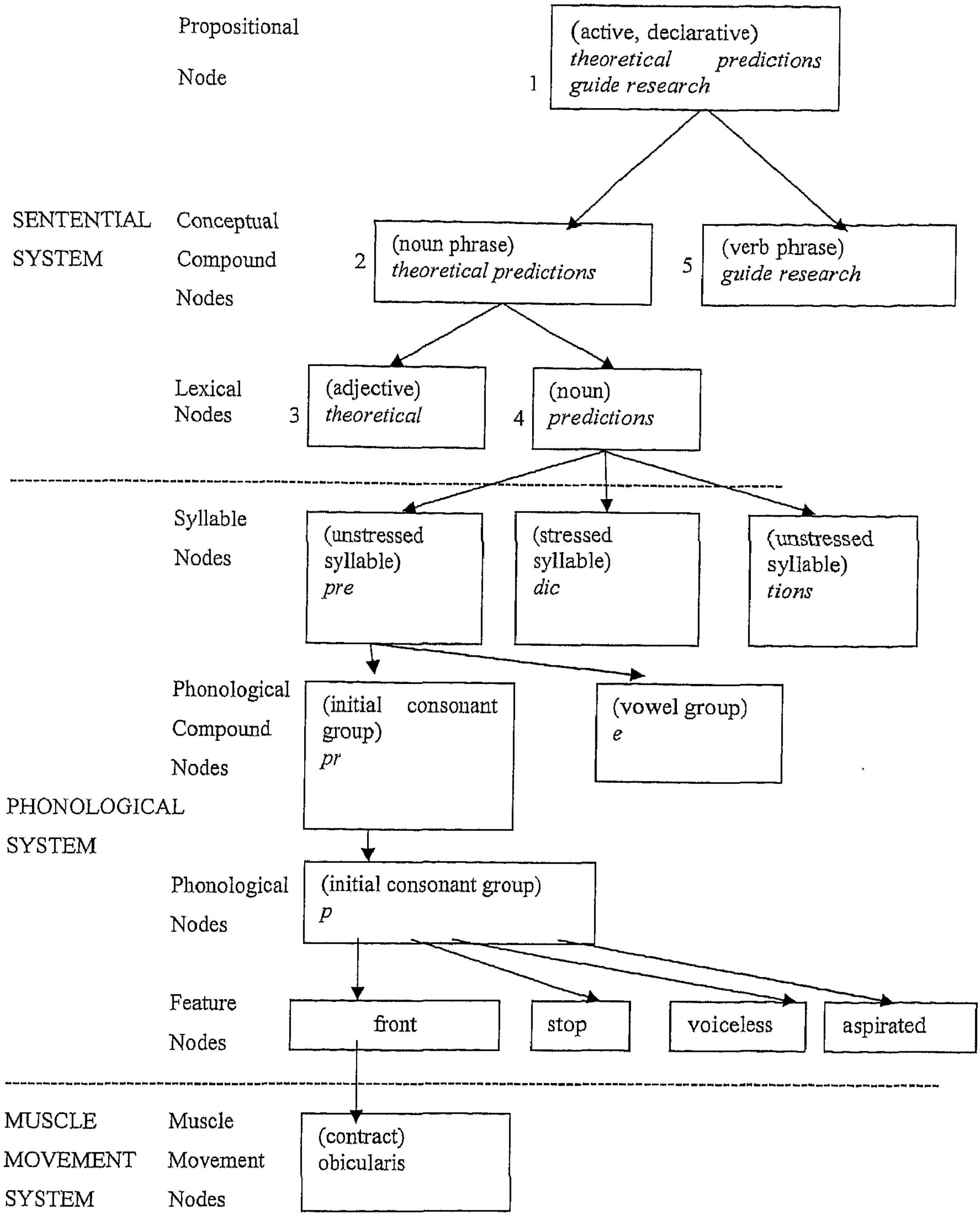
Table 3.1

Summary of the Three Classes of Nodes in the Node Structure Theory

Nodes	What they represent	Input-Output
Mental nodes	'higher level cognitive components common to perception and production' (e.g. phonological nodes, sentential nodes etc.)	'send "top-down" outputs to muscle movement nodes for use during production' 'send symmetrical top-down and bottom-up signals among each other' 'receive "bottom-up" inputs from sensory analysis nodes for use during perception'
Sensory analysis nodes	'Patterns of input' e.g. how speech inputs	send bottom-up outputs to mental nodes
Muscle movement nodes	'patterns of muscle movement' (e.g. respiratory, laryngeal, velar, articulatory organs)	receive top-down input from mental nodes for use during production

(from MacKay, 1992a, p. 42)

The mental nodes can be primed bi-directionally, that is top-down during production and bottom-up during speech perception. In contrast, the muscle movement nodes and sensory analysis nodes are unidirectional. The former is only activated during speech production, while the latter during speech perception. The activation hierarchy in the Node Structure Theory is illustrated in Figure 3.5. Basically there is a network of connections between nodes. The sentential system comprises the nodes for the concepts underlying words and phrases, while the phonological system comprises the underlying concepts, which represent syllables, consonant clusters, vowels and phonological features (MacKay, Wulf & Abrams, 1993, p. 625). The muscle movement system is responsible for the individual elements involved in actual production of speech.



N.B. the numbers show the order of activation; labels in nodes: italics = content, and parenthesis = sequential domain

(from MacKay, 1987 cited in MacKay 1992a, p. 43)

Figure 3.5

A Top-Down Hierarchy of Nodes

Figure 3.5 shows the top-down hierarchy of activation of nodes in the production of part of the sentence, *theoretical predictions guide research* (MacKay, 1992a, p. 43). It illustrates how the activation of the propositional node primes the two phrase nodes below it. The activation of the noun phrase node then primes the lexical nodes below it, which in turn prime the relevant nodes representing the phonological system below them, until finally the muscle movement nodes for the particular sounds are primed. This will result in overt speech. However, if the muscle movement nodes (see Table 3.1) are not activated, this will result in internal speech (MacKay, 1987, p. 101) since the articulators have not been set into action to produce speech.

According to MacKay (1987, p. 8), the five dynamic properties that nodes have are activation, priming, satiation, self-inhibition and linkage strength. The first, node activation, "is an all-or-non process that is self-sustained, like neural activation; it lasts for a specifiable period of time, regardless of whether the sources that led originally to activation continue to provide activation" (MacKay, 1992a, p. 42).

The second property, priming refers to "transmission across an active connection" (MacKay, 1992a, p. 45). Nodes become activated when they receive priming from a connected node. The third property, satiation, is a result of a node being repeatedly activated over a prolonged period of time (MacKay, 1992a, p. 46). As a result of satiation, nodes are less responsive to priming from connected nodes.

The fourth, self-inhibition is said to be the key player in the process of error-detection in the node structure theory (MacKay, 1992a, p. 45). Self-inhibition refers to "a brief period of reduced excitability which terminates node activation" (MacKay, 1992a, p.

43). Errors result in the production of novel units of speech, such as non-words, for example, words that are either lexically or phonologically permitted or appropriate in a particular language (MacKay, 1992a, p. 62). For example, if the consonant clusters *cp* and *sr* in the first two words of *cramped space* are erroneously activated, this would result in bottom-up priming, triggering awareness of these phonological errors (MacKay, 1992, p. 209). Thus, as MacKay (1987, p. 175) explains, "… error detection occurs almost as soon as the error itself …, and error correction can begin shortly thereafter, when the correct or primed-from-above node becomes activated". This is possible because monitoring and error detection take place "everywhere" (MacKay, 1992b, p. 215), unlike Levelt's model, where only formulated internal and audible external speech can be monitored. Further, Wheeldon and Levelt (1995), posit that the earliest stage at which monitoring takes place is at the abstract phonological level.

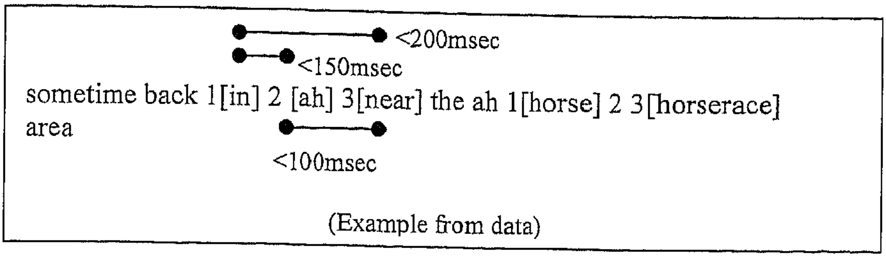
Thus, in the Node Structure Theory, error detection is a result of a prolonged activation of an uncommitted node, and as a result monitoring takes place within the mental node system. This means that error-detection and repair occurs at all levels in the Node Structure Theory (MacKay, 1992b). However, there is no external monitor, unlike in the previous two models. Because of the concept of priming and activation underlying the Node Structure Theory, it is predicted that "the probability of error detection will vary with the proximity of the units produced in error to the uncommitted node that they jointly prime" (MacKay, 1992a, p. 63). This theory also explains why fewer errors are expected at higher levels in the hierarchical system of nodes. This is said to be due to the higher activation levels between nodes at the lower levels within a system (MacKay, 1992a, pp. 51-52), which in turn, increases the probability of errors.

In short, in the Node Structure Theory, monitoring is inbuilt throughout the node system, allowing for error-detection and correction to occur at any level of this system. However, the Node Structure Theory does not provide for external monitoring of speech since "production units *are* perception units in the phonological and sentential systems... "(author's emphasis, MacKay, 1992b, p. 215). Although production and perception units are not the same in Laver's production model (see Figure 3.4), monitoring is also distributed throughout the speech formulation process. However, similar to Levelt's model, overt speech can also be monitored. The main difference in the Perceptual Loop Theory is that only formulated internal and auditorily audible speech can be monitored, and this is done through the Speech Comprehension System (see Figure 3.1). The different ways in which speech production is said to be monitored has implications on the observable time intervals related to error-detection, the interruption of speech and the production of repairs, and this will be discussed in the following sections.

3.5 Self-Monitoring and Self-Repair Intervals

The time intervals, measured in syllables, words or milliseconds (msec), involving an error, the point of interruption and the self-repair have been previously investigated to explain the nature of the process of self-monitoring and self-repair, particularly in relation to the extent to which current monitoring theories account for these intervals. The intervals, as shown in Figure 3.6, are:

- Error-to-Cut off
- Cut off-to-Repair
- Error-to-Repair



The interval between 1 (the error) and 2 (perceived interruption point) is the error-to-cut off time. The interval between 2 and 3 (the onset of repair) is the cut off-to repair time. The interval between 1 and 3 is the error-to-repair time. Note that there can be an editing phase between 2 and 3 containing a silent or filled pause. The time (msec) shows how short the intervals have been found to be.

(Adapted from Blackmer & Mitton, 1991)

Figure 3.6

Time Intervals in Self-Repairs

3.5.1 Error-to-Cut off Intervals

The existence of word fragments (see 2.8.1) and covert repairs (see 2.8.2) are generally considered as evidence of prearticulatory error-detection. This explanation is also used to support the speed at which speakers interrupted their utterances. For example, slightly more than 50% of the self-repairs in Levelt's corpus (1983) were interrupted within or immediately after the erroneous word. Nooteboom (1980), in his analysis of Meringer's (1908) corpus also found that speech was interrupted within five syllables, and within not more than five words.

Similarly, Blackmer and Mitton (1991) also found evidence of utterances being interrupted soon after an error. They measured the time interval between the onset of an error and the point at which speech was interrupted, that is, the cut off point (see Figure 3.6) and found that 14.5% of the overt repairs in their corpus were under 150msec, with

5.3% of them being under 100 msec for overt repairs. Instances of short error-to-cut off intervals were also found elsewhere in Hartsuiker and Kolk (2001), Kormos (2000) and Nakatani and Hirschberg (1994).

Studies focusing on second language (L2) production have also investigated intervals in self-repair. Kormos (2000) found that her Hungarian learners of English, who were divided into three proficiency groups, detected linguistic errors significantly earlier than inappropriate errors, presumably because the latter involved conceptual changes. van Hest (1996) also reported similar findings, although in her study, phonological errors had shorter error-to-cut off intervals than lexical and appropriateness errors.

Such findings are consistent with the Main Interruption Rule (see 2.7) (Nooteboom, 1980), that speech is interrupted immediately upon detection of an error. The term "immediately", however should be taken within the context of the estimated speed at which speech can be "instructed" to stop (Hartsuiker & Kolk, 2001, p. 118). It is said to take about 200msec to stop speech, that is, from the instructions to stop speech to the actual cessation (Ladefoged, Silverstein & Papcurn, 1973). Another caveat on immediate interruption of speech is placed by Levelt (1983), who says that speech tends not to be halted immediately in appropriateness-repairs (see 2.8.3) since the speaker has not really made an 'error' in such instances. The short error-to-cut off intervals (below 150msec) found in many studies suggest that error detection and the instructions to stop speech must have taken place prior to articulation, hence, supporting Levelt's Inner Loop Monitoring as well as internally distributed monitoring in Laver's Production Model and MacKay's Node Structure Theory.

3.5.2 Cut off-to-Repair Intervals

The time interval between the cut off point to the onset of the repair (see Figure 3.6) can provide information about the process of self-repair. It is not unusual for speakers to self-correct without the use of any hesitation device. As mentioned in Chapter 2, Nakatani and Hirschberg (1994), for example, found that only 9.4% of the repairs in their study had filled pauses. This in turn means that a short cut off-to-repair interval can be expected, and indeed, Blackmer and Mitton (1991) found short cut off-to-repair intervals in their study: 48.6% of the overt repairs were less than 100msec with 19.3% of them being 0msec. Further, they found that 24.8% of the repetitions in their corpus had 0msec cut off-to-repair intervals. Oomen and Postma (2001) also found 0msec cut off-to-repair intervals in 37% of their appropriateness repairs, and 7% of error repairs under normal speaking-rate conditions.

In relation to L2 production, however, Kormos (2000) found that only a small percentage (4.8%) of the repairs had short cut off-to-repair intervals (less than 150msec), and all of these repairs were made by speakers with advanced L2 proficiency. The implication is that L2 learners need more time to produce their repairs. This is supported by van Hest (1996), who in a comparison of L1 and L2 timings, found that cut off-to-repair intervals were longer in L2 compared to L1, meaning that her subjects took more time to produce a repair upon interruption in L2.

The fact that speakers produce repairs immediately upon interruption (in cases of 0msec cut off-to-repair intervals) or almost immediately after it has implications for the process of repair-planning. As maintained by Blackmer and Mitton (1991), such short cut off-to-repair intervals contradict the prediction that repair planning begins once an error as been detected, or upon or shortly after interruption (Laver, 1980; Levelt, 1983,1989). This is because such short or 0msec intervals do not provide time for repair-planning to take place if it is assumed that this process takes place from the point of interruption to the onset of the repair.

Blackmer and Mitton (1991) suggest, the repair must have been planned prior the point of interruption. Similarly, Nakatani and Hirschberg (1994) suggest that perhaps this interval, what they call the "Disfluency Interval" (see 2.9.1.1.2), need not correspond to re-planning time, implying that repair-planning may have begun earlier. Perhaps, as suggested in MacKay (1992b, p. 215), error detection and correction is seen as occurring "everywhere" or at all levels, rather than only at the levels suggested by Levelt (1989). This could account for such fast cut off- to-repair intervals.

In covert repairs cases where only hesitation was present, it is assumed that the error has been detected early enough for it to be prevented from being articulated. According to Levelt's model (1989, p. 473), given that it takes about 300-350msec from the delivery of the phonetic plan to the start of articulation and about 150 to 200msec for inner speech recognition (see Figure 3.1), the internal loop has between 0-100msec to instruct the Articulator to stop, if an error is detected in inner speech. This results in an interruption in the flow of speech through the use of hesitation devices such as silent pauses, filled pauses or lengthening of sounds. Thus covert repairs presuppose both prearticulatory error-detection and correction.

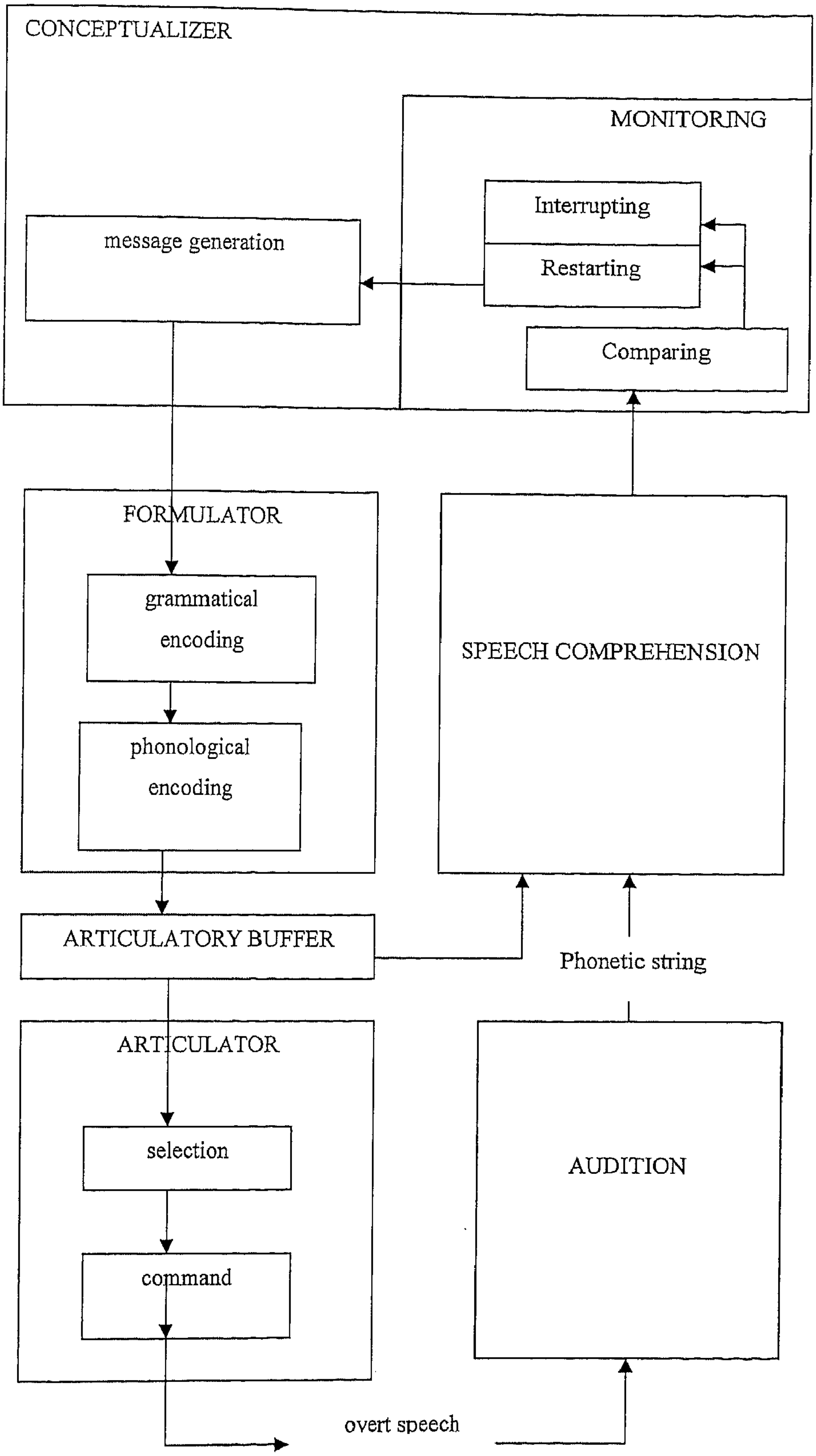
The assumption that repair-planning does not necessarily begin upon interruption is also supported by findings that speakers tend to stop their gestures before they interrupt their speech (Seyfeddinipur & Kita, 2001). Based on the assumption that "gesture and speech are semantically and temporally tightly co-ordinated" (Seyfeddinipur & Kita, 2001, p. 29), the cessation of gesture suggests that speakers were aware that they were going to stop speaking soon at that point in speech. However, they continued with their speech for an average of 240msec more (Seyfeddinipur & Kita, 2001). The implication of this is that perhaps repair-planning is taking place during this time. By the time or soon after speech is interrupted, the repair can be produced, thus producing short cut off-to-repair intervals.

3.5.2.1 Hartsuiker and Kolks' (2001) Monitoring Model

Hartsuiker and Kolk (2001) propose a modified model of the Perceptual Loop Theory to explain such intervals, since the latter does not appear to account for such fast intervals. Hartsuiker and Kolk (2001) do not consider the process of interrupting to be necessarily related to the process of repairing since interruption can occur for other reasons besides repair-planning. Thus, interrupting is seen as "an act of motor control that directly regulates the articulator" (Hartsuiker & Kolk, 2001, p. 134), and as a result, their Monitoring Model differs from Levelt's in two ways:

- The process of articulation is divided into two stages: a selection and command stage
- The monitor comprises three processing components: comparing, restarting and interrupting

In this model, the selection stage involves the selection or activation of motor programs, while the command stage "controls execution of the unit". As Figure 3.7 shows, the monitor performs three functions. Similar to the model in Figure 3.1, it first compares parsed speech with a target and triggers repair-planning, if an error is detected. However, in this model, two *parallel* processes (compared to Levelt, 1983; 1989) are said to take place immediately, that is, the process of interrupting and the process of restarting. Thus, upon error detection, the process of interrupting sends the stop signal to the Articulator, while the process of restarting or repair planning simultaneously begins.



(from Hartsuiker & Kolk, 2001, p. 117)

Figure 3.7
Hartsuiker & Kolks' Blueprint of Monitoring Model

Thus, the main difference between the model proposed by Hartsuiker and Kolk (2001) and Levelt's Perceptual Loop Theory, is that error-detection during the process of checking or comparing simultaneously triggers both the processes of repair-planning and speech-interruption as shown in Figure 3.8.

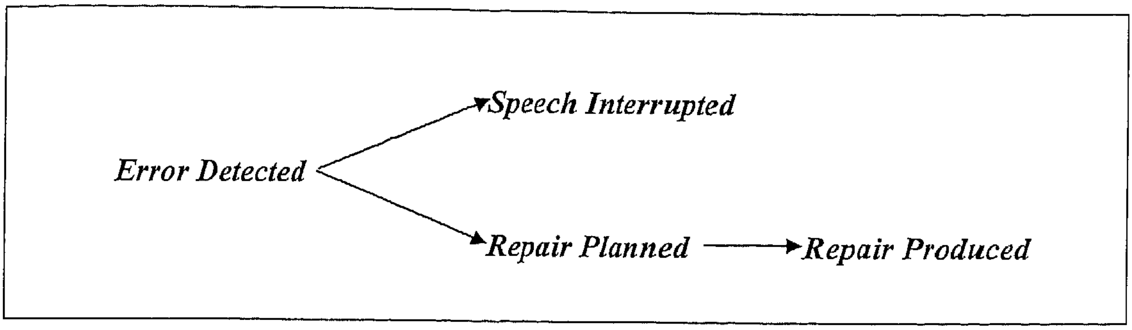


Figure 3.8

The Parallel Processes of Speech-Interruption and Repair-Planning

In other words, the process is not a serial process where error-detection results in the interruption of speech, which would then provide time for repair-planning as illustrated in Figure 3.9. This process would presumably mean that more time is needed to produce a repair upon error-detection since repair-planning only commences upon speech-interruption, and thus will not be able to account for faster cut off-to-repair times

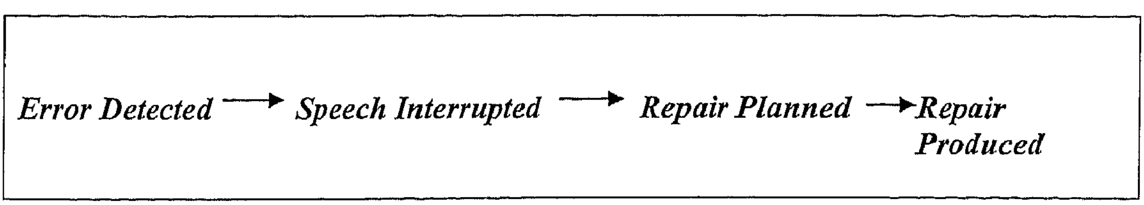


Figure 3.9

The Serial Processes of Speech-Interruption and Repair-Planning

3.5.2.2 Fast Cut off-to-Repair Intervals in Repeats

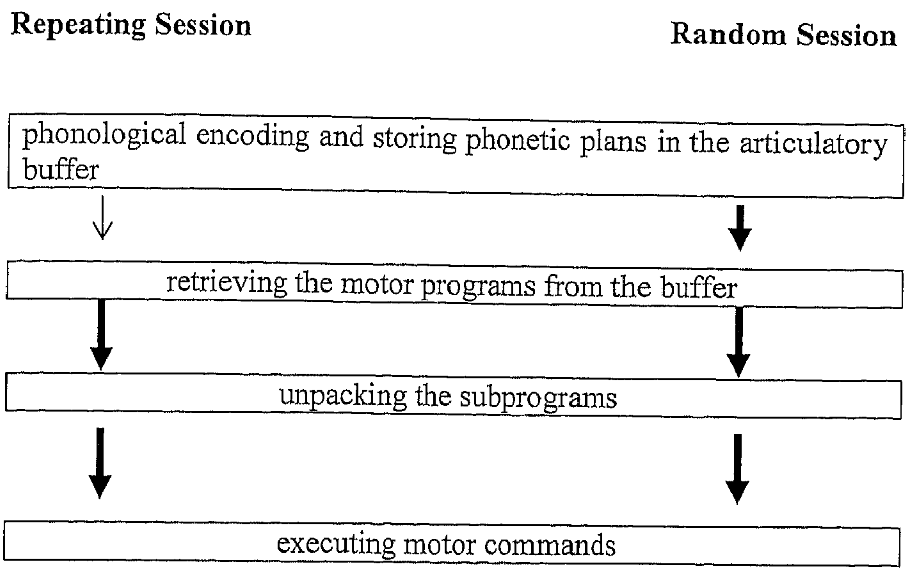
Another question related to fast cut off-to-repair intervals is whether repair-planning necessitates going back to the Conceptualizer, and reduplicating the process right up to articulation of the repair, as suggested in Figure 3.1. Blackmer and Mitton (1991) felt that, in relation to rapid repeats, two things could be happening. The first is that the process of producing the subsequent word or R2, and in some cases, further repeats of the word, is reduplicated, beginning from the Conceptualizer to the Articulator, but that this process is a rapid process.

The second possibility is that the Articulator (see Figure 3.1) "has an autonomous restart capacity" (Blackmer & Mitton, 1991, p.191). What is thought to happen is that while the articulators are producing a particular item, the following items may not be ready for articulation, presumably because the processes of conceptualization and formulation take time. Because of this, the word that is being produced is repeated, resulting in a "rapid restart" (Blackmer & Mitton, 1991, p. 191), which would then be followed by the following item when it is ready for articulation. This would explain why the cut off-to-repair intervals for repeats without editing terms in their study were significantly shorter than other covert repairs.

While Blackmer & Mitton's explanation was in relation to unbuffered speech (see Figure 3.1), Nota and Honda (2003) suggest that when a repeat is produced, the phonetic plan stored in the Articulatory Buffer is accessed, by-passing the Conceptualizer and other processes along the way. Their explanation is related to activity in the left anterior insula of the brain, which has apparently been found to be linked to the motor coordination of articulation (Dronkers, 1996; Dronkers, Redfern & Knight, 2000 cited in Nota & Honda, 2003). However, as Nota and Honda (2003) explain, while some studies attempting to link activation of this part of the brain with articulation found that it was activated during articulation, others did not. Nota and Honda (2003) hypothesized that such inconsistency in findings may be related to the experimental tasks involved, particularly in relation to whether the subjects of the experiments had to repeat their utterances or produce novel ones.

To test their hypothesis, Nota and Honda (2003) made their seventeen Japanese subjects repeat a nonsense CVCVCV word under two conditions. In one condition their subjects repeated the word (referred to as the "repeating session"). In another, the order of the syllables was randomly changed (referred to as the "random session"). Using functional Magnetic Resonance Imaging (fMRI), they measured the brain activity of their subjects during the two sessions. Their findings revealed that the left anterior insula was activated during the random but not the repeating session, where there was no brain activity recorded.

In relation to the speech production process, Nota and Honda (2003), felt that this showed that the repeated word did not go through another round of encoding process, but was merely retrieved form the Articulatory Buffer since it involved the production of the same word. This was why there was no brain activity in the left anterior insula during the repeating session. This of course might also explain why short error-to-cut off and cut off-to-repair intervals have been found in other studies. The brain activity during the random session indicated that for a new word the entire process is necessary. The different processes, leading to the articulation of repeats and novel utterances are illustrated by Nota and Honda (2003, p. 193) in Figure 3.10. This figure shows the processes put forward by Levelt (1989), but as the arrows indicate, for the production of repeats, the process goes only as far back as retrieving the phonetic plan from the Buffer.



(from Nota & Honda, 2003, p. 193)

Figure 3.10

Phases in the Execution of Articulation in a Repeating and Random Session

From experiments carried out by Nota and Honda(2003) and also the explanation of repeats being "rapid restarts" by Blackmer and Mitton (1991), it looks like the production of repeats involves a different process from the production of other self-repairs (see 2.9.2). This is understandable given that the repeated items themselves do not constitute an error and a subsequent repair like the other self-repairs. Further, unlike the other self-repairs, repeats act more like hesitation devices, much like filled pauses and prolongations. This perhaps raises the question of whether they should be considered as a type of repair or a type of hesitation.

3.5.3 Error-to-Repair Intervals

Blackmer and Mitton (1991) found that the error-to-repair intervals (see Figure 3.5) in 10% of the repairs in their study were less than 300msec. Consistent with studies to-involving L1 speech production, van Hest (1996) also found repairs with short error-repair intervals (below 200msec), and suggests, like Blackmer and Mitton (1991), that the planning of the repair is going on while speech is being produced.

Laver's production model (see 3.4.1) and the Node Structure Theory (see 3.4.2) both allow for distributed error-detection and correction, which could account for short error-to-repair times. Short error-to-repair intervals, are explained by the presence of the Articulatory Buffer. Blackmer and Mitton (1991) suggest that buffering provides time for inner speech to be monitored (see 3.3.1 and Figure 3.1), which is why the decision to interrupt speech is made, and the repair ready for articulation by the time or soon after the error is articulated. The Monitoring Model proposed by Hartsuiker and Kolk (2001) could also be used to explain fast error-to-repair intervals, where the process of interruption and repair-planning are triggered off simultaneously upon error-detection.

3.6 Chapter Summary

This chapter has shown that prearticulatory monitoring is accounted for by all the three models: the Perceptual Loop Theory, the Production based model, and the Node Structure Theory. However, while the Perceptual Loop Theory accounts for short error-to-cut off intervals (as low as less than 100msec) because of the presence of inner speech monitoring, it does not adequately account for short cut off-to-repair and error-to-repair intervals. This is essentially because it does not have distributed monitoring and editing devices throughout the speech formulation process such as in the Production based model or the bottom-up priming capability of the Node Structure Theory.

Further, the Perceptual Loop Theory assumes that repair-planning begins only upon interruption of speech which would not explain short error-to-repair intervals. This relationship between the processes of error-detection, interrupting speech and repair-planning is not fully explored in the other two models.

In view of these inadequacies, Chapter 7 will present and discuss results obtained from spontaneous speech to ascertain the extent to which the models reviewed in this chapter can be applied to explain the processes of error-detection, speech interruption and repair-planning.