Detection of errors during speech production: a review of speech monitoring models

Albert Postma*

Psychological Laboratory, Helmholtz Institute, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands

Received 24 March 2000; accepted 24 May 2000

Abstract

In this paper three theories of speech monitoring are evaluated. The perception-based approach proposes that the same mechanism employed in understanding other-produced language, the speech comprehension system, is also used to monitor one's own speech production. A conceptual, an inner, and an auditory loop convey information to a central, conscious monitor which scrutinizes the adequacy of the ongoing speech flow. In this model, only the end-products in the speech production sequences, the preverbal (propositional) message, the phonetic plan, and the auditory results, are verified. The production-based account assumes multiple local, autonomous monitoring devices, which can look *inside* formulation components. Moreover, these devices might be tuned to various signals from the actual speech motor execution, e.g. efferent, tactile, and proprioceptive feedback. Finally, node structure theory views error detection as a natural outflow of the activation patterns in the node system for speech production. Errors result in prolonged activation of uncommitted nodes, which in turn may incite error awareness. The approaches differ on the points of consciousness, volition and control, the number of monitoring channels, and their speed, flexibility, and capacity, and whether they can account for concurrent language comprehension disorders. From the empirical evidence presently available, it is argued for a central perception-based monitor, potentially augmented with a few automatic, production-based error detection devices. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Speech errors; Speech monitoring; Self-repair; Error detection

* Tel.: +31-30-2533657/2534281; fax: +31-30-2534511.

E-mail address: a.postma@fss.uu.nl (A. Postma).

1. Introduction

To err is human. To self-repair fortunately is also. Attempt the following task: type the word ‘repetition’ on your computer as fast as possible without looking at the screen. A mediocre typist like myself will likely (a) make a mistake – ‘repetition’ is a hard word to type, containing alternations of very similar letter clusters, (b) know that an error has occurred without having to look at the actual output, and (c) consider the ‘backspace’ key a blessing. Error correction takes place almost instantaneously, and, seemingly, automatically.

Self-repair refers to the correction of errors without external prompting, frequently within a short span of time from the moment of error occurrence. Self-repairing is typical for most human motor skills. Typing, driving, sports, musical exercises, and of course speaking, are full of corrective actions. Self-repairs imply the existence of specialized control devices or ‘monitors’ which verify the correctness of ongoing motor activity and response output. These monitors work on-line, that is during the course of motor activity within a limited time window, and employ multiple sources of information. As in the foregoing example, knowledge of results (e.g. inspecting the displayed words) may not be necessary for correction. Errors seem to be detected on some other basis, such as by tactile feedback (e.g. feeling that you have hit two keys simultaneously) or by some other more abstract means (such as somehow being aware that your key press program has included a double ‘p’ key instead of a repeated ‘t’).

This paper focuses on monitoring and error detection in speech production. Self-repair is a common event in both conversation and monologues. It has been estimated that more than 50% of speech errors are corrected by the speaker (Nootboom, 1980). Furthermore, one out of ten of all our utterances contains some sort of revision activity (cf. Nakatani & Hirschberg, 1994). Although a number of theoretical accounts of speech monitoring have been advanced, our understanding of the underlying processes is far from complete. The purpose of this paper is twofold. First, error detection and monitoring mechanisms as they have been proposed in the literature are reviewed. Three general solutions to the question of how monitoring is achieved are discussed: the perceptual loop theory (Levelt, 1983, 1989), production-based monitoring (Laver, 1973, 1980; Schlenk, Huber & Wilmes, 1987) and MacKay’s node structure theory (NST; MacKay, 1987, 1992a,b). The second goal of this paper is to explicate the differences between these three approaches and to formulate the research directions which will distinguish them, thus leading to a more inclusive theory of speech monitoring.

2. Organization of the speech production system

The translation of thought to articulated speech is complex, engaging a number of cognitive, linguistic, and motoric processes. Fig. 1 presents a coarse blueprint of the speaker (Dell, 1986; Garret, 1980; Levelt, 1989; Levelt, Roelofs & Meyer, 1999; Roelofs, 1996; Stemmer, 1985; Van Wijk & Kempen, 1987), extended with feed-

back mechanisms which may be important for speech monitoring. Speaking starts with conceptualization (planning an utterance's meaning and purpose). The conceptualizer delivers a propositional, preverbal message to the formulator. The formulator translates the preverbal message into a linguistic structure. This translation is hypothesized to require several steps. First, lemmas are selected that correspond to the conceptual elements included in the preverbal message. Second, syntax building

Fig. 1. Model of speech production, including the feedback loops and monitors which have been proposed in the literature to underlie speech error detection and self-repair.

procedures are activated which work with the syntactic information included in the lemmas (i.e. the meaning and syntactic parts of a lexical entry). Together, these steps (i.e. grammatical encoding) create a surface structure. This surface structure has to be further developed into a pronounceable phonetic plan. The morphological structures of the words in a sentence are made available. Next, the phonological structure is generated. Levelt and Wheeldon (Levelt & Wheeldon, 1994; Wheeldon & Levelt, 1995) suggest that the generation of phonological structure consists of two relatively independent mechanisms, segmental and metrical spellout. Segmental spellout derives a word's phonemic structure, its composition of consonants, consonant clusters, vowels, glides, etc. from the morphological structure. Metrical spellout is concerned with syllabic and metrical structure. Importantly, metrical spellout involves resyllabification which may cross lexical boundaries. Segmental and metrical information are combined in a step called segment-to-frame-association, in which phonemes are inserted into their syllabic slots (shown in Fig. 1 as the phonemic representation). Thus, during phonological encoding, first a syllabified phonological representation is constructed. Next, a context-dependent phonetic representation is derived from the phonological representation by looking up syllabic gestures in a mental syllabary (Levelt & Wheeldon, 1994). These gestures contain all the necessary detail to form the input for the articulatory apparatus.

One of the major problems in understanding speech production is how invariant, abstract, neatly ordered and segregated elements in the speech plan are translated to variable, context-dependent, distributed articulatory movements. One possible mechanism for this is the use of the aforementioned syllabic gestures, which define articulatory goals such as lip closure. It is left to the articulator how these goals are achieved. One prominent idea is that articulation is determined by co-ordinative structures, that is, groups of articulators that work in functional synchrony and conjunction, and seem to behave as a single unit (cf. Kelso, Tuller, Vatikiotis-Bateson & Fowler, 1984; Saltzman & Kelso, 1987). Afferent feedback is thought to play an important role, not, however, as in servo-mechanisms, but rather by tuning the parameters used by the co-ordinative structure (see also below).

The phonetic plan reaches the articulator (the system of co-ordinative structures) via a so-called 'articulatory buffer'. It frequently happens that formulation occurs while the articulator is still busy with a previous utterance. In order to cope with these asynchronies, parts of the phonetic plan not yet articulated are held active in the articulatory buffer.

3. What is an error? And what is a correction?

Obviously, we cannot begin to understand error detection without an idea of what an error is. Dictionaries customarily describe errors as 'things done wrong' or 'conditions of being wrong', as 'mistakes', or as 'blunders'. There are two senses in which things can be done wrong. First, actions can be incorrect with respect to some external criterion. This criterion reflects a generally agreed upon idea of well-formedness, successful performance, and/or error-free output. In speech production,

the external criteria employed are based upon the linguistic rules which apply to a given language. For example, the speech of aphasic patients is scored with respect to principles of syntactic and lexical correctness. Second, actions may be judged as errors with regard to some internal standard. A person's intentions form the starting point from which correctness and incorrectness have to be decided (cf. Dell, 1986; Reason, 1990; Senders & Moray, 1991).

A similar dichotomy of error has been proposed by Ohlsson (1996), who distinguished between objective and subjective views of error. In the former, an (idealized) external observer regards an error as an "action not on the shortest path to the desired end state". In the subjective view, actions are neither correct nor incorrect themselves. The conviction that a certain action is wrong arises from a conflict between what the actor *believes* to be true and what he or she perceives to be the case. Errors thus are defined with respect to an internal criterion. Hence, they are not only undesired, but also inhabitual, i.e. the actor could have done better (to his or her own standards, at least). Many actions are errors regardless of the perspective taken. Violations of a speaker's intentions will most often also trespass some external criterion of correctness (presuming that the speaker tries to follow the rules of a language). However, one can produce a message completely according to the intended format, while still violating certain general linguistic rules or norms (cf. Ohlsson, 1996), or one can produce linguistically correct nonsense. In speech production, it is the errors that are diagnosed with reference to an internal criterion that allow for successful self-repair. That is, it is subjective errors that constitute events for which control and correction by the speaker are possible (cf. Senders & Moray, 1991). It will be clear that this sense of error is crucial to the present discussion.

Another important distinction in errors is between response-selection and response-execution errors (Schmidt, 1982, 1988). In the former, a wrong action program is selected, which is, however, executed perfectly. In response-execution errors, in contrast, the correct program is selected but something goes wrong in its execution. For example, too much force may be exercised or the onset of a submovement may be mistimed. It is speculated that correction of selection errors is a central, attention demanding activity, working serially (only one correction at a time), and on a relatively slow time basis, and typically implying reprogramming. Correction of response-execution errors is faster, may occur in parallel, and is a less central, more autonomous mechanism. Correction of such errors occurs without notable consequences (i.e. correction consists of just a minor, peripheral adaptation).

Crucially important to the discussion of error monitoring mechanisms is the notion of feedback circulating in the speech production system. Feedback can serve three distinct forms of control: directive, tuning, and corrective.¹ *Directive control* refers to the notion that motor commands may depend directly on the sampling of the feedback. In other words, feedback drives the motor command pattern. Instructions to the motor output system derive from calculating the devia-

¹ Borden (1979) refers to the tuning functions as 'constructive', whereas the other two roles are said to be 'regulative'.

tions between the feedback on the current state and the desired goal state. Basically, this is how feedback was hypothesized to work in the traditional closed-loop control view (cf. Fairbanks, 1954; Kent, 1976). Although it is well established that many movements are not guided by closed-loop control, this type of control might be important in relatively slow movements (cf. Abbs & Eilenberg, 1976; Perkell, 1980; Perkell et al., 1997; Schmidt, 1982). Moreover, as will be elaborated later, it may play a role in spreading activation models of speech production (cf. Dell, 1986).

Feedback can also have a *tuning* function. This is a type of ‘off-line’ control, operating over longer periods of time after the movements in question have been performed. It is employed in learning a new motor skill and to adapt to temporary environmental changes. In other words, it calibrates or recalibrates the system (cf. Levelt, 1989; Neilson & Neilson, 1987; Perkell, 1980).

The *corrective* function of feedback refers to the on-line and post-hoc control of speech (motor) processing, and is aimed at detection and correction of errors which are not part of the closed-loop control cycle described above, but rather would lead to completely undesired output. Such errors violate the desired progress of an utterance, and their correction typically causes some form of disruption (e.g. the restart of planning and execution). Although the three feedback control functions are clearly difficult to differentiate, the concept of speech monitoring, as I use it here, is specifically concerned with corrective control, in particular with the detection and repair of response-selection errors.^{2,3}

4. Types of self-repair

The speech production literature makes clear that speakers can monitor their utterances for a multitude of distortions (see also Table 1). First of all, there are errors at the conceptual level (‘do I want to say this now and should I say it in this form?’), followed by what is coined an appropriateness repair. Second, speakers intercept and revise errors originating in the formulation stage, where things can go wrong in the lexical selection, syntactic construction, or sound form encoding. Third, the foci of monitoring may be directed towards suprasegmental characteristics, such as sound level or prosody (Cutler, 1983). In the first examples listed in Table 1, the incidents needing revision, as well as the actual revision, are always

² Of course in a motor execution error the deviation from a spatial-temporal goal may be relatively large. To the extent that correction of such anomalies causes disruption and reprogramming, these errors might also fall under the propagated concept of events to be monitored (cf. Schmidt, 1988).

³ In line with the distinction between directive, adaptive, and corrective role of feedback, Mattson and Baars (1992) discuss error minimizing mechanisms in speech production. They distinguish between boosting and editing. Boosting refers to the intricate wiring of speech production processes, which increases the chance on correct responses or ‘good errors’ (e.g. lexical errors). Editing presupposes a separate verifying mechanism, which lowers error rates or increases the interception of ‘bad errors’ (e.g. non-word errors). An example of boosting is the positive feedback described in more detail below. Boosting reflects directive control, while editing mainly entails corrective control. However, in as far as the editing mechanism works without real consequences for the planning and production of speech, it could serve a directive purpose as well.

more or less clear. Hence, the corrections are denoted as the overt self-repairs. Three things must be considered in overt self-repairs (Levelt, 1983): (a) the original utterance with the inconsistency which is needing repair, i.e. the reparandum; (b) the interruption, followed by a longer or shorter delay, termed the editing phase; and (c) the repair proper. In addition to overt repairs, speech might contain so called ‘covert repairs’. Here no overt error or observable correction is involved, but an interruption of the progress of speech occurs, typically in the form of hesitation, e.g. an excessive pause, a repetition, or an interjection (e.g. ‘eh’, ‘I mean...’). Although a clear reparandum is lacking in these cases, it is generally agreed that covert corrections

Table 1
Types of errors and self-repairs in speech production^a

Example	Type of error	Type of self-repair	Basic repair activities
<i>‘We start in the middle with – in the middle of the paper with a blue disc.’</i>	Conceptual error	Appropriateness repair – reformulation	Major retracing and revision based on syntactic structure utterance, leading to a grammatically well-formed continuation. Pronominalization is possible in the repair.
<i>‘John comes – uh – likes to come to the party.’</i>	Syntactic deadlock	Syntactic restructuring / reformulation	Retracting and reformulation based on syntactic structure.
<i>‘Left of purple is – uh – of white is purple.’</i>	Lexical error	Lemma substitution	Retrace span is morpho-phonologically determined. Substitute only erroneous lemma, keep the other elements.
<i>‘A unut – unit from the yellow dot.’</i>	Phonemic error	Phonemic error repair	Minor retrace and restart.
<i>‘...from my PROsodic – proSODic colleagues.’</i>	Prosodic error	Suprasegmental error repair	Retrace and restart, mainly of lexical stress errors for pragmatic or semantic reasons.
<i>‘willfiddily – fully’</i>	Morphemic error	Repair on the fly (no clear cut-off)	Replace, no retrace.
<i>‘...ideals by – uh – by va – voting for the Meech lake accord.’</i>	Phonemic error (preceded by hesitations)	Repair on the fly (no clear cut-off)	Minor retrace and restart.
<i>‘to a – uh – stapler.’</i>	Unknown	Covert repair	Postponing.
<i>‘to – to the right.’</i>	Unknown	Covert repair	Restart.

^a Examples are taken from Blackmer and Mitton (1991), Cutler (1983), De Smedt and Kempen (1987), Levelt (1983), Oomen and Postma (2000a), and Van Wijk and Kempen (1987).

Table 2

Overview of feedback loops in speech production, their presumed corrective functions, and temporal characteristics

Generic feedback category	Feedback loop	Alleged functional correlates	Detection time after error occurrence	Most likely moment of corrective intervention with respect to overt appearance of error
Intrinsic feedback (within CNS)	(1) Conceptual loop	Appropriateness monitoring	Immediate (cf. Blackmer & Mitton, 1991), but conceptual judgements may take time.	Rather slow. Installation of repair proper could be prolonged.
	(2) Lemma selection feedback	Verifying match between lemma and concept	Unclear. Presumably immediate.	Pre-articulatory and post-articulatory.
	(3) Syntactic construction feedback	Spotting syntactic deadlocks during incremental sentence planning	Unclear. Presumably immediate.	Pre-articulatory and post-articulatory.
	(4) Node activation feedback	Comparing top-down with bottom-up activation rates; signalling prolonged node activation; competition detector	At least two time steps in spreading activation cycle for monitors using positive feedback (cf. MacKay, 1992a,b; Postma & Kolk, 1993).	Pre-articulatory and post-articulatory.
	(5) Inner loop	Inner speech parsing	About 150 ms (cf. Levelt, 1989).	Effective lookahead range is at least 100 ms. This may become larger when articulatory buffering is present (cf. Levelt, 1989).
	(6) Information on articulatory timing	Spotting delays in arrival of new information in the articulatory system	Immediate.	Average segment durations and gaps lie between 100 and 250 ms (cf. Postma & Kolk, 1993). A whole repair cycle (including detection and repair) is performed in this period. This provides some estimation of the speed of the articulatory timing monitor.

Table 2 (*continued*)

Generic feedback category	Feedback loop	Alleged functional correlates	Detection time after error occurrence	Most likely moment of corrective intervention with respect to overt appearance of error
	(7) Efferent feedback	Comparing outgoing (efferent) signal strengths to pre-set criteria	Unclear. 50 ms prior to beginning of manual movements (Cooke & Diggles, 1984; Jaeger et al., 1979).	Unclear. In manual movements, repairs can be installed within 30 ms after overt error occurrence (Cooke & Diggles, 1984; Jaeger et al., 1979).
Response feedback (within PNS)	(8) Proprioception	Comparing alpha-gamma efferent signals to Ia afferent signals; sensing positions and motions of the articulators	Unknown.	25–70 ms in reflexive feedback loops after overt perturbation; 100–200 ms in longer feedback loops (Abbs & Gracco, 1983; Abbs et al., 1984).
External feedback (external)	(9) Tactile feedback	Feeling articulatory contacts	Unknown.	Post-articulatory. Tongue reaction times to tactile stimulation can be as fast as 136 ms (Siegenthaler and Hochberg, 1965).
	(10) Auditory feedback	Hearing and parsing own speech	Variable, but in general rather slow.	200 ms at fastest + delay for implementing interruption (cf. Levelt, 1989).
	(11) Knowledge of results	Noticing (hearing; seeing) that your message has not got through to others	Variable, extremely slow.	400 ms at fastest + delay for implementing interruption.

are made to anticipated errors, which may be comparable to the overt errors discussed above. In covert repairs, however, both the error and the repair remain hidden, and have to be inferred. There is ample evidence that speakers are capable of anticipating forthcoming mistakes, i.e. that they can inspect their speech programs prior to articulation (Blackmer & Mitton, 1991; Garnsey & Dell, 1984; Postma & Kolk, 1993). Speakers in a number of cases react without any delay to an overt error. In addition, the correction is executed without further waiting (suggesting it must have been ready before the interruption was made; Blackmer & Mitton, 1991). Furthermore, slips of the tongue can be reported in the absence of auditory feedback

or articulatory activities (Dell & Repka, 1992; Lackner & Tuller, 1979; Postma & Kolk, 1992a,b; Postma & Noordanus, 1996).

Covert repairs are often ambiguous and difficult to classify (e.g. does a long pause signal a real covert repair, whereas a short pause does not?). As such, certain classes of hesitations sometimes are considered covert repairs, and sometimes they are not regarded as real repair phenomena, but more as the direct result of difficulties in word finding or conceptual selection (i.e. they do not involve misselections but delays in selection). In turn, since the inciting error component remains hidden, covert repairs theoretically might be false alarms (Levelt, 1983). Van Hest (1996) addresses the difficulty in identifying a given disruption as a case of covert repair. She emphasizes two potential cues that a covert repair has occurred: the presence of an editing term, and the abruptness of the cut-off. In self-repair, the speech flow typically is interrupted rather suddenly, often in the form of a glottal stop. Other cues that a repair is taking place are signalled by prosody (e.g. the repair proper is often accented) and by pausal durations (Howell & Young, 1991; Levelt & Cutler, 1983; Nakatani & Hirschberg, 1994).

5. Levels of monitoring as proposed in the literature

As can be seen in Fig. 1 and Table 2 various feedback loops have been postulated that provide the means for error detection and correction. Some of them apply to overt repairs, whereas others apply specifically to covert repair. Table 2 lists 11 feedback loops which have been ascribed a corrective function in speech production. In short, they fall into three broad categories (cf. Borden, 1979; Schmidt, 1982): intrinsic, response and external feedback. Control based upon intrinsic or internal feedback exerts its influence before any movement is made. Response feedback is linked to the production of concrete motor output, such as proprioceptive feedback (Borden, 1979). External feedback is based upon the immediate results of the motor output. Below, I discuss the 11 feedback loops in the order in which they are presumed to play a role during the planning and articulation of an utterance. The numerical references are the same as used in Fig. 1 and Table 2.

[1] The first loop is between the conceptualizer and the preverbal message. Blackmer and Mitton (1991) and Levelt (1989) have labelled this the *conceptual loop*. Its function is appropriateness monitoring, that is, the interception of conceptual and semantic errors. In a way, the conceptual loop might be described as ‘thinking about one’s thoughts’. This implies metacognitive reflection, i.e. awareness of one’s goals and intentions as well as those of others (cf. Frith, 1992). Both Blackmer and Mitton (1991) and Van Hest (1996) report that conceptual errors are repaired significantly slower than lexical or phonological inadequacies, even though appropriateness monitoring commences as soon as the intention to speak is born. One reason for this could be that it is hard to reject a wrongly selected intention. Another reason could be that it takes rather long to install the repair (i.e. the selection of a new intention).

[2,3] Grammatical encoding follows the construction of the preverbal message. It chooses the lexical entries to express the concepts in the preverbal message, and

elaborates these entries within a syntactic frame. Here, we may find *lexicity and syntax monitors*. It has been proposed that both lemma selection and syntactic frame generation are controlled by either central or local monitors, which judge the suitability of selected elements and initiate correction if necessary (De Smedt & Kempen, 1987; Laver, 1980; Van Wijk & Kempen, 1987). One possible means by which the monitoring of syntactic construction might work is sensitivity to syntactic deadlocks during sentence planning (cf. De Smedt & Kempen, 1987). It is typically assumed (Kempen & Hoenkamp, 1987) that each conceptual element of the preverbal message is elaborated into a lexico-syntactic structure as soon as it becomes available. This is known as incremental planning. Sometimes incremental planning creates a situation in which syntactically correct continuation is impossible after a conceptual addition. The syntactic monitor would signal the syntactic deadlock and initiate the necessary correction, as in the following example: ‘John comes...uh...likes to come to the party.’ (from De Smedt & Kempen, 1987).

[4] In spreading activation accounts of speech planning processes in the formulator, nodes are organized hierarchically (e.g. lemma nodes, morpheme nodes, phoneme nodes) and are selected on the basis of their activation level. After selection, a node spreads activation to its subordinate connected nodes. Thus, the subordinate nodes get primed to some extent, which, in turn, guides their subsequent selection. Spreading activation accounts frequently include some type of monitoring mechanism, which I refer to as *node activation monitors*. Mattson and Baars (1992), for example, point out that one characteristic of an error situation is a high amount of competition amongst rival nodes from the same class. For some reason, an incorrect item is more activated than the intended one and gets selected. Hence, a competition detector could monitor the total amount of activation of items within the same class or pool, and respond with an error signal if a certain threshold is exceeded. Berg (1986a,b) and Schade and Laubenstein (1993) give a similar explanation. Due to the high amount of concurrent competition, an erroneously selected node will usually have a lower activation level than a correctly selected node. The monitor thus should react to suboptimal activation levels of selected units.

A different type of node activation monitoring mechanism has been proposed by Postma and Kolk (1993). One has to assume here that there is a recurrent connection from lower levels to preceding levels. If so, a primed subordinate node itself sends activation to connected superordinate nodes, thus creating positive feedback. By doing so, it eventually also boosts its own activation level, for the superordinate node will in turn spread an increased amount of activation to the lower nodes. Positive feedback has been proposed as an explanation of the so-called lexical bias effect (Baars, Motley & MacKay, 1975; Dell, 1985, 1986; Stemberger, 1985): that is, the tendency for sound errors to create actual words or morphemes.⁴

⁴ It should be mentioned here that the existence of a lexical bias effect in sublexical slips of the tongue is disputed. Also the idea that there is positive feedback from the phonological to the lexical stage and that it could serve to explain lexical bias has been criticized (see Baars et al., 1975; Del Viso, Igoa & Garcia-Albea, 1991; Levelt, Shriefers, Vorberg, Meyer, Pechman & Havinga, 1991; Nickels & Howard, 1995).

Lexical bias occurs because entries exist at the superordinate stage which propagate the error to the lower stage. Incorrect subordinates, which are not linked to an entry at the higher stage, and which thus do not constitute real words, create no positive feedback, and thus receive no recurrent top-down priming. The latter has two effects. First, it may prevent misselections of sublexical elements. That is, it may minimize the impact of non-lexical error noise. As such, positive feedback could exert a directive role. Second, when the error does become overt, the chance is higher that it will form a legal (but unsuitable) lexical combination. With some additions, positive feedback could also serve a corrective role. Postma and Kolk (1993) suggested that there might be an autonomous monitoring device checking the flow of information between nodes, and which works without conscious supervision. What will be clear from the foregoing is that even when processing is correct at a given stage, it can become flawed in a subsequent step. If there are positive feedback loops between stages as suggested by Dell (1985, 1986), these connections could be used ‘correctively’. For example, a localized monitor could compare the amount of activation a unit has sent downward to the amount it receives back. Too large a difference would prompt error detection.

MacKay (1992a,b) has also proposed a connectionist error detection account using positive feedback. Put simply, his idea is that erroneously selected units will always form a new combination or sequence in some respect (e.g. phonotactically, lexically, or semantically). As they return activation to superordinate stages after selection, they will boost a higher node which represents this rather novel combination (or is uncommitted). By inhibitory or self-inhibitory mechanisms in the model, a node’s activation level will be suppressed after activation. This does not apply to uncommitted nodes, however. Consequently after having been activated by the bottom-up priming of its erroneous subordinates, a novel, uncommitted node reveals prolonged activity. Error detection rests upon the awareness of this prolonged activity.

In Fig. 1, node activation monitoring is placed between the grammatical and the phonological encoding stages. In principle, node activation monitoring may occur throughout the system, not just at this particular location. In this sense, the lexicality and syntax monitor ([2,3]) could also be based upon similar node activation and positive feedback principles (see also MacKay’s NST below).

It is difficult to make any firm statements about the temporal characteristics of node activation monitoring. In the positive feedback accounts by MacKay (1992a,b) and Postma and Kolk (1993), it seems that at least two time steps have to pass within the spreading activation network before an error can be spotted. However, the relation between time steps in the network and real time is arbitrary. The only indication of the time course of positive feedback is the finding that lexical bias effects are reported when two-word utterances are produced within a deadline of 700 ms and not when the deadline is set at 500 ms (Dell, 1986, 1990).

[5] According to Levelt (1983, 1989) speakers monitor the phonetic plan prior to or during articulation as if they were listening to their own utterances. In other words, the speech comprehension system parses the final output of the speech planning process. This channel – the *inner loop* – is equated with the sensation of

inner speech. One of the elegant properties of Levelt's proposal is that no special additional monitor devices are needed for error detection. The speech comprehension system does what it always does only in this case with respect to its own current speech output.

The inner loop, like all other pre-articulatory monitors, gives speakers the opportunity to detect errors before they have been crystallized in some overt format. Levelt (1989) contends that this parsing takes about 150–200 ms after the phonetic plan is generated. Thus, an error can be intercepted 150 ms after its commission at the level of the phonetic plan. The articulator will not have executed the speech plan until after 200–250 ms. This leaves about 100 ms for detection and repair before overt realization. Even more time is available when the phonetic plan is buffered while awaiting its articulatory unfolding. Postma and Kolk (1993) have argued that the opportunity for pre-articulatory monitoring and timely detection and correction of errors will depend upon the time relations between the buffer and the articulation stage. Crucial, in this respect, are the size of the buffer and the speed of articulation. Liss (1998) indeed suggests that apraxic speakers might be relatively slow in self-repair because of damage to the articulatory buffer.

It has recently been suggested that the inner loop does not only assess the phonetic plan but is also or even exclusively applicable to the preceding phonemic representation (Levelt et al., 1999). Evidence for this is provided by Wheeldon and Levelt (1995), who had Dutch subjects silently translate English words into Dutch, and, while doing so, monitor for a certain target sound (in the Dutch targets). Monitoring latencies were linearly related to the target's positions in the word, suggesting that speakers may become aware of the contents of the phonemic codes while constructing these codes. Since latencies were not affected by the duration of segments in the spoken translations, it was argued that the inspected codes were phonemic and not the subsequent phonetic representations. Thus, speakers should be able to detect potential errors on-line during the construction of phonemic representations.

[6] Blackmer and Mitton (1991) propose that a special restart routine is located between the articulatory buffer and the articulation stage. As discussed elsewhere (Postma & Kolk, 1993), this routine can be thought of as a *buffer-articulation timing monitor*, which is sensitive to the timing of new material to be articulated. If no new input is available at the moment that the articulator has finished a stretch of a speech program (and the articulator still is set for further motor performance), there clearly is an error. The articulator is supposed to possess an autonomous restart capability which deals with these events. It reacts by executing the old program once again. Such is most likely to occur with high speaking rates, which presumably prohibit buffering and thus increase the chance of mistimings. This results in rapid repetitions of short speech fragments.

[7] The translation of more or less abstract program codes in actual motor output opens several information channels for monitoring. Amongst them is the 'efferent feedback'. Wundt and James already debated vigorously the question of whether we may in some way "sense the energy or commands which leave our brain" (cf. Festinger & Canon, 1965). Terms like 'efferent copy', 'corollary discharge' or

‘central efference monitoring’ were used to denote these sensations.⁵ The basic idea is that (copies of the) commands leaving the central nervous system and travelling to the periphery are sent to some comparison centre, and may then be used for some sort of control in addition to the control they already exert by energizing the effector organs.⁶ A number of possible roles of *efference monitoring* have been proposed. First, it is thought to prepare the system for a pattern of incoming (sensory) information. In the present terminology, this would be a tuning function. Second, it can be used to check afferent – e.g. proprioceptive or muscle reflex – feedback. In case of discrepancies, errors can be detected and corrected. The speed at which such a correction takes place is rather high (<90 ms) (cf. Schmidt, 1988). Thus, potentially, motor execution can be adjusted without interruption. In essence, this function then seems to be directive. A further basis for directive control has been suggested. The efferent copy may be compared to some internal standard. As such, an even faster way of verifying that ongoing motor commands are correct is possible. Corrections may be issued before the errant movements are initiated. Kelso (1982) calls this the central efference monitoring concept.

In addition to tuning and directive roles, efferent information may serve a corrective role. Lackner and Tuller (1979) had subjects report their own speech errors in a noise-masked condition (thus without auditory feedback). The fact that indeed many errors were reported under these circumstances was taken as an indication that subjects were using efferent information for monitoring. Lackner and Tuller’s line of argument is not compelling, however. Error detection responses under noise might derive from other types of feedback than efferent feedback, such as inner loop information.

Installation of corrective activities may begin as soon as 50 ms prior to the beginning of movement in manual responses. Effectively, this leaves overt corrections as fast as 30 ms after overt error occurrence (cf. Cooke & Diggles, 1984; Jaeger, Agarwal & Gottlieb, 1979).

[8,9] Once effectors in the articulatory apparatus start moving, they produce several sources of feedback. Proprioception concerns the sensing of where your limbs (and in case of speech articulators) are and where they are moving to. Receptors in the muscles, joints, and skin transmit proprioceptive signals. Specifically, the reflexive muscle spindle feedback loops in the alpha-gamma motor systems are important to motor control (cf. Rosenbaum, 1991; Schmidt, 1982, 1988). These loops constitute a very short-wired circuitry between efferent and afferent neurons, providing information about the amount of stretching and loading of muscles. They allow very fast, reflex-like corrections of muscle activities, for example as in the ‘knee jerk reflex’, requiring no attentional supervision (cf. Schmidt, 1988). In general, muscle-spindle feedback is associated with motor activities such as posture control and reaching, but it also supports speech motor gestures (cf. Levelt, 1989).

⁵ Kelso (1982) differentiates between these three possibilities. I will not go further into the exact differences here.

⁶ As a potential distinction with the concept of ‘feedback’, efference is often called feedforward. That is, it already exerts its influence before output is generated.

Proprioception thus could function directly – perhaps in combination with efferent copies (see above) – in correcting execution errors. Reflexive adjustments are engaged by this feedback system in reaction to sudden, external disruptions, like the blocking of a movement, the adding of a load, or the encounter of unexpected strong resistance, as in skiing over bumps. Schmidt (1987) refers to these adjustments as automatic corrections. In speech, it has been claimed that muscle-spindle feedback is used to bring the articulators to specific target positions (cf. MacNeilage, 1970; see also Levelt, 1989). In addition, slower proprioceptive feedback loops exist which might underlie directive control in slow, continuous movements (cf. Schmidt, 1988). For example, when you make a slow pointing movement towards a target some distance away, you may use proprioception (as well as other types of feedback, such as vision) to keep yourself on track. The question remains whether slow proprioceptive feedback control applies to speech production.

Proprioception and the somewhat later tactile feedback (i.e. touch sense from mechanoreceptors situated mainly in the ora mucosa, reporting on articulatory contact, such as the tongue against the teeth) could also be used to *tune* the system. They are important in learning a new motor skill (cf. Borden, 1979). Moreover, they can help determine the starting positions of the articulators. Hence, co-ordinative structures have been hypothesized to be tuned by proprioceptive and tactile information. The initial model which drives the co-ordinative structures is updated by proprioceptive and tactile feedback (cf. Levelt, 1989). If the vocal tract is substantially changed by means of pipes, bite blocks, or dental surgery, proprioception and taction form the basis for a new internal model employed by the co-ordinative structures (or by some alternative effector system). Finally, one can speculate about a corrective function here, presuming the existence of *proprioception and taction monitors*. Lackner and Tuller (1979) claim that speakers can detect response-selection errors by means of proprioceptive and tactile information.

A recent study by Postma and Noordanus (1996) is relevant to the foregoing. We asked subjects to rehearse tongue twisters and stop and report each error they made in four conditions: silent, mouthed, noise-masked, and normal auditory feedback. Interestingly, the number of reported errors did not differ in the first three conditions, whereas it increased in the latter. If motor execution feedback (proprioceptive, tactile or efferent) is used for error detection, it should have caused a difference between the silent condition, in which motor movements were made, and the mouthed and noise-masked conditions, which did involve motor movements. The foregoing example is not completely compelling, however. Because we asked our subjects to react with a deliberate, conscious error-detection response, we may have excluded more autonomous forms of speech monitoring. The proprioception and taction monitor could work autonomously and automatically. We will return to this possibility later.

How fast do the proprioceptive and tactile loops work? Reactions based upon proprioceptive signals after external perturbation of limbs vary between 30 and 200 ms, as indicated by EMG data (Schmidt, 1988). Abbs and Gracco (Abbs & Gracco, 1983; Abbs, Gracco & Cole, 1984) demonstrated adjustments in upper lip movements between 25 and 70 ms after perturbation of the lower lip. Reaction times to

tactile stimulation were examined by Siegenthaler and Hochberg (1965). They reported an average latency of 136 ms of tongue responses upon stimulation of the lip.

[10] Originally, audition was thought to have a directive function in closed-loop control of speech production (cf. Fairbanks, 1954; Kent, 1976). Because audition arguably is too slow to have a direct impact on speech commands, and because with noise masking speech still remains reasonably intelligible, this view is unlikely. More plausible is the option that auditory feedback is employed for tuning. It is crucially important during speech acquisition. Children born deaf have severely impoverished speech, whereas people who become deaf at a later age remain able to speak fairly well (cf. Maassen, 1985). Perkell et al. (1997) argue that auditory feedback enables the learning of a robust internal model which, once it is established, provides on-line control of articulatory movements in achieving output goals. In addition, auditory monitoring of the acoustic environment can assure intelligibility by facilitating more rapid adjustments of the ‘postural’ parameters underlying average sound level, speaking rate, and prosody. Neilson and Neilson (1987) suggest that audition plays a similar role in intermittent updating of system parameters.

Another important function of audition is *correction*, carried out by what is called the *auditory loop*. Hearing your own speech makes you aware of errors. Levelt (1983, 1989) claims that auditory loop repairs proceed in the same way as the repairs originating from the inner loop [6]. That is, auditory feedback is sent through one’s own speech comprehension system, parsed, and consciously monitored. Recognition of words when listening to running speech appears possible about 200 ms after word onset (Marslen-Wilson & Tyler, 1981, 1983, December). This sets a lower bound for error detection by the auditory loop at 200 ms after the articulation of the slip. A constant delay may be added to this needed for making the necessary interruption.

[11] A speaker’s goal is not simply the production of a series of sounds, but rather the communication of information. A speaker generally wants to convey some meaning or intention. In this respect, he or she is eager to observe the outcomes of utterances upon the environment. In turn, these outcomes may be used for control purposes. *Knowledge of results* is the term coined for the post-response information actors receive about the success of their performance. Knowledge of results is important in mastering new skills (Schmidt, 1988). Thus, it likely serves a tuning purpose: learning to speak may depend upon observing that a message is understood. In addition, knowledge of results may support the development of error detection mechanisms (Schmidt & White, 1972). Knowledge of results may also serve a corrective function: seeing or being alerted that someone does not understand you can make you aware of an error. Customarily, these events are denoted as ‘other-installed repairs’. Schegloff, Jefferson and Sacks (1977) list a number of differences between self- and other-repairs. Other-repairs tend to be delayed until the next turn in conversation. The error interruption is also different. In self-repairs it may vary from abrupt breaks to prolongations, sound interjections, and ‘ehs’. In other-repairs, it typically concerns a ‘what, where, who, or when’ question. Finally, a self-repair usually is handled efficiently and finished within a single sentence. Other-repairs generally need several turns before the result is achieved.

Knowledge of results may connect to the speech comprehension system in the same way as the auditory loop: you hear and parse other utterances in order to realize that the other is signalling some misunderstanding or error in your conversation. In addition, knowledge of results can be based on visual information (i.e. on seeing that your message does not get through). Compared to the other monitoring loops, feedback on knowledge of results is rather slow. First, there is the time needed by the other (e.g. listener) to decode what was said and detect the error. This is limited by the processing speed of the auditory signal (i.e. at least 200 ms). Next, the listener has to formulate and articulate the other-repair. In turn, this message has to be decoded by the original speaker, adding another 200 ms at least to the processing time.

6. Types of monitors: the distinctions between the perceptual loop theory, production-based speech monitoring, and the node structure theory

In the previous section, 11 feedback mechanisms were discussed. It remains unclear whether all of them fulfil a corrective role, instead of being restricted to tuning or directive control. Specifically, this might apply to speech motor execution feedback (i.e. efferent, proprioceptive and tactile feedback). In order to address this question, we need to consider in more detail precisely how monitor mechanisms work. At present, there appear to be three viable approaches to speech monitoring. The most influential of these views of speech monitoring thus far has been Levelt's perceptual loop theory (Levelt, 1983, 1989), which assumes that only certain end-products in the speech production flow are monitored. Moreover, these end-products are analyzed in essentially the same way as are the utterances of others, that is with the speech comprehension system. According to Levelt (1983, 1989) the inner and auditory loop [6,10] feed information to the comprehension system, which parses this information and conveys it to a central, conscious monitor. Levelt includes the conceptual loop [1] also in his perceptual loop theory of speech monitoring. As part of the conceptualizer, the conceptual loop has direct access to this monitor. It makes sense to also include knowledge of results [11] in the perception-based monitoring approach.

The production-based approach holds that in self-repairing, speakers have direct access to various processing components in the speech production flow (cf. Levelt, 1983). In other words, intermediate aspects of speech planning, viz. components *inside* the formulator, are accessible for monitoring (cf. Laver, 1973, 1980; Schlenk et al., 1987). Production-based monitoring may also apply to speech motor execution feedback. In essence, production monitors are special purpose editors that form integral parts of the production system. They comprise channels [2] through [4] and [6] through [9] in the left of Fig. 1.

MacKay's node structure theory (NST) (MacKay, 1987, 1992a,b) accounts for error detection in terms of the natural outflow of activation patterns in the node system (see discussion of the node activation monitoring [4]) used for speech production (as well as for speech perception). Errors invariably concern the activa-

tion of units that are novel at some level in the speech production hierarchy. This novel combination results in the prolonged engagement of uncommitted nodes, which in turn incites awareness of the (erroneous) code and thereby error detection. MacKay's view to some extent overlaps with production-based monitoring. It supposes distributed means for error detection, and applies to the same levels in the formulator as the production monitors. In contrast to the production-based approach, no special device for the actual monitoring seems necessary in the NST. Prolonged activation of uncommitted units automatically leads to error detection.

In order to compare these three approaches and to formulate the experiments that provide further insight into the mechanisms of speech monitoring, I will elaborate upon what I think are the most crucial differences (see Table 3).

Table 3
Differences between perception- and production-based speech monitors, and the node structure theory of monitoring

	Perception-based monitoring	Production-based monitoring	Node structure theory
Location	Central	Distributed	Inherent
Awareness	Conscious	Automatic, reflex-like	Subsequent awareness is implied and perhaps necessary for error detection
Number of levels	3–4	8 or more	Matches the number of levels in the (internal) node system
Capacity	Limited by attentional resources	Autonomous resources	No resource limitations
Aspects of speech flow scrutinized	Flexible	Fixed	Activation of uncommitted nodes, signalling 'novelty' in the selection process
Temporal characteristics	Relatively slow	Fast	Fast
Repair proper	Elaborated revision, co-ordinated with original utterance in order to aid the listener	Simple retrace and restart /postpone; repair 'on the fly'	Not discussed; potentially the responsibility of the subsequent awareness mechanism
Speech comprehension	Related to comprehension skills	No relation	Damage to the node system inflicting comprehension deficits will cause comparable monitoring disorders

6.1. *Location: central versus distributed*

The perceptual loop theory postulates a single, central monitor within the conceptualizer. It receives information from three (or four) channels. The situation is less clear for the production-based approach. One variant of such an approach assumes several distributed (independent) monitors throughout the speech production system (Laver, 1980). In other production-based approaches, however (De Smedt & Kempen, 1987; Van Wijk & Kempen, 1987), a single monitor is presumed to examine both end-products and intermediate results. That is, only the monitor loops are distributed. The monitor itself is central. As we will see below, this option is not very likely. As mentioned above, in NST monitoring can be regarded as a distributed process. Prolonged activity of uncommitted nodes can occur anywhere throughout the node system (i.e. wherever an error occurs).

6.2. *Error awareness*

Does error detection require conscious awareness? Rabbitt (1990) argues that, in motor skills, error signalling and correction may vary from being fast and automatic to being slower and more deliberate. The ‘automaticity’ of error correction responses is evidenced by the fact that they apparently cannot reliably be consciously suppressed. The critical question is, does such also apply to speech repair?

As a natural consequence of being located within the conceptualizer, the perception-based monitor can be described as a more or less conscious, deliberate correction device. Levelt (1989) claims that “self-corrections are hardly ever made without a touch of awareness”. Production-based monitors, however, do not need to work consciously (cf. Berg, 1992; Laver, 1973; Nooteboom, 1980; Postma & Kolk, 1993). They might function autonomously and largely subconsciously. Awareness has a central position in NST. MacKay (1992a,b) argues that awareness following prolonged node activation enables error detection. Thus, a speaker should always be aware of each slip she or he intercepts.

Bearing upon the awareness issue is the study of the development of metalinguistic abilities. In growing up, children develop a variety of metalinguistic skills, amongst which is spontaneous self-correction (Clark, 1978). There is discussion of whether self-correction implies linguistic awareness, i.e. the ability for explicit reflection on some part of your linguistic activities. Some authors appear to hold the view that failure causes awareness (Clark, 1978; Levelt, Sinclair & Jarvella, 1978). Others, however, point out that the speech monitoring underlying self-repair is only dimly conscious or even pre-conscious. Karmiloff-Smith (1986) argues that, in children, speech repair precedes explicit metalinguistic awareness and conscious access to what precisely is wrong and why.

Levelt (1983, 1989) has criticized production-based theories for postulating monitors which look *inside* processing components or modules. According to modular views of human cognition (Fodor, 1983) and general psycholinguistic beliefs, the principle of information encapsulation applies to speech production. That is, linguis-

tic mechanisms such as formulation and articulation work without central supervision, and in a more automatic fashion. They function in parallel, and do not share resources (see also below). Hence, only the end-products of these processes are accessible for attention, not their intermediate results. In other words, the processes themselves are cognitively impenetrable, and can not be inspected by an attentional monitor. The seriousness of Levelt's criticism is weakened, however, if the production-based monitor is not attentional or conscious, but an autonomous, special purpose device. Furthermore, one of Levelt's own findings suggests that the phonemic representations *can* be monitored by speakers (Wheeldon & Levelt, 1995). Because this phonemic code is not an end-product, allowing it to be monitored counters the idea that speakers can not inspect intermediate processing results (see also Kolk & Postma, 1996). Recent modifications of the perceptual loop theory hold that the relation between speech production and perception is more interactive than originally assumed. That is, self-perception (i.e. speech monitoring) may take place at different points during the construction of a speech plan (Kempen, 2000; Levelt et al., 1999).

It is often conveyed as an anecdote that self-repairs occur on their own without any awareness (cf. Laver, 1973). In line with these casual observations, Postma and Noordanus (1996) had speakers report their own errors by pressing a response button. Occasionally, the errors were accompanied by a self-repair. Interestingly, in a number of cases this error–self-repair combination was not followed by the obligatory button press. This could have happened if the self-repair was executed automatically, without conscious awareness. However, an alternative explanation is that subjects temporarily forgot to obey the task instructions (cf. Rabbitt, 1990).

6.3. *Three levels or more?*

The perceptual loop approach conceives of a limited number of monitoring channels. These channels are related to perceptual mechanisms and the conscious attention system. Production-based monitors would strongly increase the number of channels for monitoring. Hence, the two theories place different restrictions on the types of errors subject to self-correction. Perception-based monitors allow for conceptual errors, and all errors violating linguistic orthodoxy in some sense, given that one can also detect them in other-produced speech. Production monitors, in contrast, are sensitive to several additional error types. For example, they would catch asynchronies between speech planning and execution ([7]), or they could intercept various deviations in motor execution (provided that these deviations do not simply constitute response-execution errors under directive control). Bearing upon the number of levels issue is an intriguing case discussed by Mowrey and MacKay (1990). One of their subjects produced an articulatory abnormality (as revealed by articulatory measures) without any perceptual or acoustic consequences. The speaker followed this utterance by overdue hesitation and greater deliberateness of the next phrase, thus signalling his or her awareness of the anomaly. Mowrey and MacKay (1990) conclude that this "...suggests strongly that non-auditory self-monitoring feedback mechanisms exist in the motor control of speech" (e.g. scanning

efferent, proprioceptive or tactile feedback). In contrast with this conclusion is the finding discussed above that speakers report as many errors in a condition without motor movements (i.e. silent speech) as in conditions with motor movements (i.e. mouthed and noise-masked speech; Postma & Noordanus, 1996). This suggests that there is no effective monitoring of speech motor movement feedback, or at least that there is no error awareness at this level.

In the NST, error monitoring is an inherent property of the mental node system. MacKay (1992a) distinguishes the following levels: propositional nodes, conceptual compound nodes, lexical nodes, syllable nodes, phonological compound nodes, phonological nodes and feature nodes. Prolonged activation of these so-called content nodes enables error detection. Hence, there seem to be at least seven levels for monitoring. An interesting feature of the NST is that all monitoring resides within the mental node system. There is no separate, external loop for detecting speech errors. Consequently, a speech condition which has no auditory feedback (such as silent speech or noise-masked speech) should yield as many error interceptions as a condition with normal auditory feedback. There is strong evidence, however, for reduced error detection in conditions which lack auditory feedback (Dell & Repka, 1992; Postma & Kolk, 1992b; Postma & Noordanus, 1996; see also Levelt, 1992). This fact poses a problem for NST.

6.4. *Capacity*

Being a conscious, attentional function the perception-based monitoring can be thought of as resource-limited (Postma, 1997). This would agree with the notion of a restricted number of monitoring channels. You cannot divide your attention over too many different sources of information at the same time. On the other hand, autonomous production-based monitors, such as the lexicality monitor and the buffer-articulation timing monitor, would be less hampered by capacity restrictions. They might possess their own specialized resources (cf. Rabbitt, 1990). In contrast to the perception- and production-based approaches, the NST posits no resource limitations at all. Because the different approaches make different assumptions about the capacity limits, it seems worthwhile to examine what happens to self-repair patterns when central attentional resources are reduced (e.g. by means of a dual task). It has been suggested that shutting down one channel (e.g. the auditory loop) may increase error detection by means of the remaining (pre-articulatory) channels (cf. Postma & Noordanus, 1996). Schizophrenic patients with positive symptoms are assumed to have defective internal monitoring channels (viz. the inner loop). Leudar, Thomas and Johnston (1994) observed that these patients indeed made fewer within-error-word self-repairs than controls, but more frequently produced corrections which started after the erroneous word had been completed. Assuming that the former repairs are pre-articulatorily based, and the latter post-articulatorily, it seems that one is able to direct more resources to one channel, the auditory loop in this case, when the internal channels are not functioning properly. An experimental study directly addressing capacity limitations in monitoring was conducted by Jou and Harris (1992). They

assessed the effect of a secondary task (mental arithmetic) on speech production. The absolute number of self-repairs was the same in single task (speech production only) and dual task conditions. Unfortunately, the percentage of errors repaired (i.e. number of self-repairs divided by number of errors) was not computed. This measure gives a more direct estimate of monitoring accuracy. In a recent paper, Oomen and Postma (2000b) indeed found a dual task condition to reduce the percentage of errors repaired, although not dramatically. Together, the foregoing notions favour a resource-limited conception of monitoring and, as such, appear most in line with the perception theory.

6.5. *Error detection criteria*

One of the most difficult questions regarding speech monitoring is what the criteria used by the proposed monitors to detect errors are. In the perception theory, it is stated that we use the same capabilities to detect our own errors as the ones we use for judging other produced speech (though without specifying what these criteria would be).⁷ One of the main objections to production theories is that their criteria involve a reduplication of knowledge (and of perception at large; Levelt, 1983, 1989). That is, whenever an intermediate processing component is monitored, the monitor would access the same kind of knowledge as engaged by that processor in the first place. For example, let us suppose that the lemma selection process retrieves its items from a mental lexicon or dictionary. The lexicality monitor would, in turn, have to access this same mental dictionary to do its work.

Reduplication of knowledge in monitoring criteria indeed seems a serious objection against the production theory. However, it remains to be seen whether production monitors necessarily work this way. It could be that at least some of the proposed production monitors do not (Postma, 1997). For example, in the version of the positive feedback monitor proposed by Postma and Kolk (1993), the monitors are rather simple devices which simply compare top-down and bottom-up activation rates to some adjustable criterion. Similarly, Blackmer and Mitton's timing monitor (Blackmer & Mitton, 1991) reacts to very elementary features of the information flow.

The prime criterion applied in the NST is that of 'pertinent novelty'. An error will always constitute a novel combination at some level in the activation flow in the node system. If someone is pronouncing 'crawl srace' instead of 'crawl space', this includes activation of a novel consonant group ('sr' is a phonotactically unfamiliar sequence in English). 'Cool tarts' instead of 'tool carts' is novel only at the superordinate propositional level. According to MacKay (1992a,b), the number of levels between the level at which the error occurs and the superordinate level at which the convergently primed uncommitted nodes are located predicts the probability of interception. In the above example of 'crawl srace', activation of the uncommitted

⁷ As Levelt (1992) acknowledges, this still leaves us with a big problem: how do listeners detect linguistic ill-formedness?

‘sr’ phoneme compound node will prompt quick error detection, while ‘cool tarts’ instead of ‘tool carts’ can only be detected several levels higher at the propositional level.

A further issue related to the criteria engaged by monitors concerns their versatility. Elsewhere (Postma, 1997), I have speculated that whereas the criteria in production monitors are relatively fixed, they are flexible in perception monitors. Several studies by Baars and Motley have shown that the type of errors intercepted might vary with the conditions tested. For example, subjects reading lists of all real words were more likely to show the lexical bias effect than when reading lists containing non-words (Baars et al., 1975). The existence of a pre-articulatory editor was hypothesized which would be more sensitive to lexical status in the first case than in the latter. Similar context-dependent semantic and syntactic biases have been induced experimentally (Motley, 1980; Motley, Camden & Baars, 1982). At first sight, this flexibility is most readily explained in terms of variations in the perception monitor: speakers may concentrate more upon lexical, or syntactic correctness, or upon semantic suitability depending upon the particular speaking situation. Levelt (1989, 1992) emphasizes that monitoring fluctuates with the distribution of attentional resources. Speakers thus scrutinize different things in different situations. Motley, Camden and Baars (1982), however, offer a more mechanical explanation. Situational context typically primes a whole domain of relevant items or nodes. Editing occurs on the basis of sensing mismatch between an encoded element and the primed pool of relevant items. Hence, a selected item which is incorrect but situationally related will have a higher probability of being accepted as correct than a contextually inappropriate item.

Taken a bit further, one might suppose that not just the criteria can be adapted, but also completely new channels for monitoring are sometimes employed by the perception monitor. For example, as part of certain stuttering therapies, one may force speakers to focus upon proprioception and tactile information (channels normally not engaged in perception monitoring; e.g. Webster, 1980). Doing so, they can become aware of and correct errors in motor execution.⁸ It is difficult to imagine how production monitors could suddenly tune in to completely new information sources (this would imply the creation of a completely new monitor), or shift between criteria which have totally different contents. Of course some variation in production monitoring is possible. For example, error sensitivity might vary. In the positive feedback monitor proposed by Postma and Kolk (1993), such would be a simple function of changing the critical ratios of bottom-up to top-down activation rates. The NST might incorporate situational flexibility in the manner proposed by Motley et al. (1982), as discussed above. It is difficult, however, to see how flexibility might involve monitoring of new information sources other than those wired within the mental node system, such as monitoring of tactile feedback or knowledge of results.

⁸ Of course, a very flexible perception monitor might be difficult to discern both theoretically and empirically from the possibility of multiple production monitors.

6.6. *Temporal characteristics*

A serious criticism of the production theory states that inspection of the production processes themselves would seriously hamper the progress of the speech flow. For example, in the model by Laver (1980), a processing component can only become active after the results of the preceding stage have been approved by the local monitor. Blackmer and Mitton (1991) distinguish between hold-up (like Laver's) and flow-through monitors (further processing of material is allowed while checking the material). Hold-up monitors would indeed strongly obstruct the flow of speech production. However, there are no compelling or logical reasons why production monitors can not be of the flow-through type.

Blackmer and Mitton (1991) point out that one important means to learn more about speech monitoring and to dissociate the production-based approach from the perception theory includes a closer examination of the temporal features of self-repairs. Two temporal aspects appear critically relevant: the time at which errors are detected, and the speed with which the subsequent correction is executed. From their analysis of a corpus of self-repairs made in a natural communicative setting, Blackmer and Mitton reach several conclusions.

1. Interruption after error detection is not necessarily immediate. This follows from the fact that error-to-cut-off times can be practically zero as well as cut-off-to-repair times (proper). Both observations suggest that speakers must have detected the error and formulated its revision well before its overt appearance.
2. The former point thus implies that replanning may take place simultaneously with speaking, and not just after articulation has been halted.
3. Fast overt error-to-cut-off and cut-off-to-repair times confirm the existence of pre-articulatory monitoring. Moreover, this can *not* be explained by hold-up production monitors, which only allow errors to be detected and corrected covertly, or to pass through to the overt speech (in which case they might be detected post-articulatorily, but then the repair should be rather slow).
4. In the perception theory, fast self-repairs can only occur when buffering (e.g. of the phonetic plan) is present. If no buffering takes place (e.g. articulation rate is high), error-to-cut-off and cut-off-to-repair intervals should become prolonged.
5. If rapid multiple adjacent sound repetitions (e.g. 'I-I-I- went away') are covert repair phenomena – as hypothesized – this poses a problem to the perception theory. Only by assuming that the first repetition is a reaction to an error, and all the consecutive repetitions are collateral exaggerations, might save a perception monitor account of fast repetitions.

In short, it appears that error detection compared to the moment of error incubation is relatively slow in perception monitors. Production monitors, on the other hand, can spot an error as soon as it arises (see also Table 2). In the perception theory, error detection in relation to the moment of its overt occurrence depends upon the availability and capacity of the articulatory buffer. Increasing output speed (i.e. articulation rate) diminishes buffering (Blackmer & Mitton, 1991; Levelt, 1989;

Van Hest, 1996). Each new output from the phonological encoder is not held ready in the buffer for a limited period of time prior to articulation, but is articulated as soon as it becomes available. Consequently, error-to-cut-off and cut-off-to-repair times would increase. In contrast, production monitors, however, are not critically dependent upon buffering. Effectively, their functioning might co-vary with the increase in speech rate. Hence, when one starts talking faster, either the error-to-cut-off time is unchanged, or, if it does increase, the cut-off-to-repair intervals will be reduced. In NST, speaking faster causes an elevation of the speed at which the timing nodes responsible for the selection of content nodes (i.e. the speed at which the next level of the speech hierarchy is filled in) work. One might presume that all of the activation flow in the node system is speeded up in a similar fashion. Hence, prolonged activation of uncommitted nodes, inciting error detection, might also occur sooner when the speaking rate is faster. In other words, NST makes the same predictions regarding temporal characteristics of error detection as does the production-based approach.

In line with the production-based monitoring and NST account, we recently observed that when speakers speed up, both error-to-cut-off and cut-off-to-repair times decrease (Oomen & Postma, 2000a). Subjects had to describe a path through a network of ordinary objects on a computer screen. The path was indicated by a red dot moving through the network at two different speeds. The highest speed made subjects increase their output rate from 3.6 to 4.5 syllables per second. Both error-to-cut-off and cut-off-to-repair times speeded up accordingly (the former by an average of 0.2 syllables or 140 ms, the latter by about 100 ms; both measures computed for formulator error repairs). Hence, monitoring seems to adjust its speed of error detection and repair to the faster speech output rate. This rather surprising finding is problematic for the perception-based monitoring approach.

6.7. Repair proper

The step following error detection is planning and executing the repair proper. In the perception monitor, all replanning has to begin at the first component in speech production: the conceptualizer (cf. Blackmer & Mitton, 1991). In the production-based approach, the local production monitor may restart the processing module in which the error originated without having to return to the start of the sequence. Blackmer and Mitton (1991) have exploited this idea by equipping the articulator with an autonomous restart capacity in case of timing errors in the arrival of new phonetic materials. Their explanation of rapid sound repetitions (see the foregoing point) illustrates this possibility.

Kempen and colleagues (De Smedt & Kempen, 1987; Van Wijk & Kempen, 1987) distinguish two basic repair strategies: reformulation and lemma substitution. They suggest that these strategies are triggered by different kinds of errors. Reformulation is called for when in order to create a new content, the syntactic structure of the utterance has to be revised. In lemma substitutions the syntactic tree does not have to be revised to express the new, updated meaning. Lemma replacement suffices. The linguistic unit of central importance in reformulations is the major

syntactic constituent. In lemma substitutions it is the phonological phrase. In addition, Van Wijk and Kempen (1987) propose a third repair strategy involved in phonological errors. It is characterized by “the near absence of delayed interrupts and of backtracking beyond the reparandum” (Van Wijk & Kempen, 1987).

If multiple distinct repair strategies exist, different types of monitors (e.g. production versus perception) could stand at their basis. In this respect, it could be worthwhile to consider the elaborateness of the repair part. In Postma (1997), I have suggested that production monitors may underlie a simple, rather straightforward repair mode. Blackmer and Mitton (1991) propose that in fast repairs – repairs ‘on the fly’ – such an automatic repair form could be present. Repairs ‘on the fly’ are executed in a way that is similar to incremental processing and addition of new coordinations in normal (error-free) utterances. Perception-based repairs, in contrast, would engage more elaborate replanning and revisions. Furthermore, we may speculate that they involve consideration of the listener and general communicative setting. That is, the repair should be such that one’s audience easily understands what sort of error was made and how the utterance was intended. In a similar vein, Berg (1992) has pointed out that both productive and perceptual constraints characterize speech repair. Productive constraints are such that a speaker strives for maximal fluency and lowest effort in communicating a message. Perceptual constraints dictate that an error and its correction do not interfere with a listener’s decoding process.

The organization of the repair proper is not extensively treated in NST. It seems to be the prime responsibility of the awareness system which is involved in error detection. As such, planning the revision might proceed in the same way as described in the perceptual loop theory.

6.8. Speech monitoring and speech comprehension skills

An issue which is of critical importance to theories of speech monitoring is how they view the relation between speech monitoring and speech comprehension skills in neurological patients. The perception theory predicts that self-repair behaviour and comprehension skills should be correlated. The reason is obvious. Both the inner and the auditory loop pass through the speech comprehension system (see Fig. 1). In fact, the monitor analyses the parsed output from the comprehension system for error detection. Accordingly, Goodglass and Kaplan (1972) associated breakdowns in internal language monitoring with damaged language comprehension mechanisms. Similar claims were advanced by Crosson (1985) and Marshall and Tompkins (1982). In contrast, Schlenk et al. (1987) argue that language comprehension and self-monitoring need not be related. They divided self-repairs in aphasic patients into two broad classes: pre-articulatory repairs, signalled by hesitation phenomena, such as filled pauses and repetitions, and post-articulatory corrections, i.e. repairs of overt errors. They showed that aphasic patients with comprehension difficulties had significantly lower post-articulatory repair rates, while pre-articulatory monitoring appeared normal. Schlenk et al. (1987) concluded that dissociable production and perception monitors might underlie these differential effects on repair types. Accord-

ing to these authors, pre-articulatory repairs can be installed by both production-based monitoring and perception-based monitoring – the inner loop – and thus can correlate with either comprehension skills or production skills (e.g. reflected by the number of errors). Furthermore, MacKay (1992b) points out that the perception theory predicts that patients with word sound deafness (i.e. mishearing auditory information) will reveal deviant monitoring behaviours in speech production – such as making many pseudo-corrections – which is, however, not the case. Likewise, McNamara, Obler, Au, Durso & Albert (1992) suggest that their observations of disordered self-repair in patients with Parkinson's disease, a clinical group with no history of language comprehension deficits, also imply that speech monitoring and comprehension need not be related. Maher, Rothi and Heilman (1994) studied a patient with a clear lack of awareness of self-produced errors despite relatively well-preserved auditory comprehension. Nickels and Howard (1995) also reported no correlation between auditory comprehension and self-corrections and interrupted responses in a group of aphasic patients.

The foregoing dissociations between language comprehension ability and self-repair in speech production seem problematic for the perception-based monitoring theory. However, it could be that patients still are able to perform a single (i.e. comprehension) task fairly well but not speech production and monitoring simultaneously (see Lebrun, 1987; Shuren, Smith Hammond, Maher, Rothi & Heilman, 1995).⁹ Hence, more critical are cases of defective language comprehension with normal speech repair. Marshall, Rappaport and Garcia-Bunuel (1985) describe a patient with severe auditory agnosia and defective auditory comprehension, whose correction of phonemic errors in self-produced speech was surprisingly good.

In short, the perception theory predicts that reduced language comprehension abilities will always be accompanied by defective self-repair. Under the production theory, comprehension and repair during production vary independently. NST holds that the mental node system (e.g. semantic, lexical, and phonological nodes) is shared by perception and production. Peripheral disturbances to the auditory system (e.g. the sensory nodes) might hamper speech comprehension, while they need not affect self-repair patterns. However, intrinsic damage to the mental node system would inflict comprehension deficits which are mirrored by impairments in speech production and monitoring. Nickels and Howard (1995) argue quite strongly against 'shared input and output lexicons'. Central deficits in auditory comprehension need not correlate with phonological errors in naming. More importantly, Nickels and Howard (1995) describe certain aphasic patients who, despite making multiple production errors, still attempted to correct almost all of their naming errors.

7. General conclusions on speech error detection

As attested by self-repairs in speech and other motor skills, we keep close watch

⁹ Levelt (1989, 1992) also points out that attentional requirements in speech production and comprehension are quite different. Since in his view speech monitoring is an attentional function, detection of self-produced errors does not have to resemble detection of other-produced errors.

on the quality of our behaviour. The inclination to monitor is so strong that we even do it when instructed not to or when it is counterproductive (Rabbitt, 1990; Rabbitt & Rodgers, 1977). How are we able to detect our own errors? How do we decide that something we have just said or are in the act of saying is faulty in some way, and how do we know what the correct form should be? One of the central problems in accounting for error detection and self-repair is to avoid the tendency to treat monitors as homunculi judging our performances. As a solution, monitors have been advanced which serve quite limited purposes. Speaking generally, monitors apply restricted criteria to diagnose correctness of selective aspects of planning and motor output. The criteria used by monitors should be different from the instructions driving the output system (cf. Norman, 1981; Schmidt, 1982) as well as from the idealized (output) code which the system strives for. As MacKay (1987) points out, if the monitoring code and the ideal output code were identical, somewhere within the system the correct code would exist. It then of course is curious why the system did not use this code in the first place. Hence, monitors should possess independent references of correctness. In the foregoing overview various speech production monitors, each employing its own set of criteria, have been examined. These criteria range from (prolonged) activation levels and time asynchronies to high-level judgments about linguistic orthodoxy and semantic appropriateness. Three monitor theories have been compared: the perceptual loop theory, the production-based monitoring approach, and NST. The perception monitor has been described as a central attentional controller situated at the top of the speech production hierarchy, employing language comprehension skills and verifying three (final) stages in the speech flow (the conceptual, inner speech, and auditory loop). Production monitors are distributed throughout the speech production system. In principle, each speech processing component and numerous aspects of the articulation phase can be monitored. As such, production monitoring could underlie the occurrence of the fast, seemingly automatic repairs we all sometimes experience, whereas the perception monitor might be responsible for more deliberate forms of correction. An attractive property of the NST is that no special monitor device is needed, but that error detection is an automatic outcome of the way spreading activation is thought to occur between the classes of nodes within the system. The major task now is to implement the theory in a running computer simulation to see whether its central organizational principles hold up and whether it can successfully be fitted to speech error and self-repair data.

Recent work has begun to unravel the neurophysiological correlates of central (i.e. perception-based) error detection mechanisms. Various studies have demonstrated that when subjects in a forced-choice situation press the wrong button, this typically is accompanied by an error-related negativity in their ERPs: the ERN (Bernstein, Scheffers & Coles, 1995; Dehaene, Posner & Tucker, 1994; Scheffers, Coles, Bernstein, Gehring & Donchin, 1996). The ERN onset is too short (100 ms after EMG activity onset) for sensory feedback and must reflect some internal monitor. Scheffers et al. (1996) link it to executive control of human behaviour, while Dehaene et al. (1994) situate the error detection mechanism in the frontal areas, particularly the anterior cingulate SMA. The frontal lobes have often been

cited to form the neurological basis for self-reflection and monitoring (Wheeler, Stuss & Tulving, 1997). In addition, McGuire, Silbersweig and Frith (1996) found that the auditory monitoring of speech invokes higher activity in the bilateral, temporal cortex. Frith (Frith, 1994; Frith & Done, 1988) has argued that the reciprocal interactions between frontal and posterior (hippocampal) areas may form the basis for the cognitive mechanism of metarepresentation involved in action monitoring. It may be of interest to see whether error interception activities which are not assumed to be centrally governed engage different neuroanatomical areas. As such, the study of clinical, neurological groups with deviant monitor behaviours could be of interest. As already noted, the study of self-repair patterns in aphasics appears worthwhile. Moreover, following up on the proposals by Frith (1987, 1992), Leudar, Thomas and Johnston (1992, 1994) obtained evidence for defective internal self-monitoring of speech in schizophrenics. In contrast, their usage of external (i.e. auditory) feedback was quite good. Further exploration is needed to establish whether it is the conceptual loop (as hypothesized by Frith), the inner loop, or some other pre-articulatory (production) monitoring component which is disordered.

The evidence bearing upon the three monitoring accounts is summarized in Table 4. Both production-based monitoring and the NST have met with mixed support. Small but significant reductions in self-repair rates under dual task conditions clearly counter the idea that monitoring is a completely autonomous, self-contained process, as proposed by both approaches. More importantly, Postma and Noordanus (1996) pointed out that monitoring most likely does not use speech motor execution feedback, but instead a substantial role is played by the auditory loop. The latter is particularly damaging to NST. Patient data further seem to discredit some of the claims of NST. As such, the present evidence favours the perception theory. There is one difficulty, however, for the perceptual loop theory: the finding that an increased speech rate does not lead to slower repairs (Oomen & Postma, 2000a). In addition, there are two possibilities which need further exploration. First, there is the possibility that in some cases speech repair works automatically without awareness. Second, it is critically important to see whether there are more patient cases of disordered speech comprehension with normal speech repair, presumably for certain forms of errors. If these possibilities are further corroborated, production-based monitoring becomes more plausible alongside perception-based monitoring. Hence, perception- and production-based monitoring may complement each other. A combination of monitoring mechanisms might account for certain intriguing patterns of results, such as, for example, observing certain classes of repairs to occur automatically, while self-repair in general still seems resource-limited, or finding patients with disordered comprehension correcting normally certain classes of their own errors. Nickels and Howard (1995) plead for a combination of comprehension and production-based monitoring. In specific circumstances, aphasic patients might abandon comprehension monitoring for strategic reasons and rely solely on more automatic production monitoring. Several new questions arise from this point of view. Where exactly are production monitors found? Is each processing level attached to a local monitor? If not, what determines the existence of such a

Table 4
Evaluation of the existing empirical support for the various accounts of speech monitoring^a

	Perception-based monitoring	Production-based monitoring	Node structure theory	Comments
Location	+		+	It might be argued that the complexity of error detection is best managed by a central, high-level error detection device.
Awareness		+		There are some indications for automatic, subconscious error detections both in speech production and other motor skills.
Number of levels	+	–	–	No difference in reported error rates between silent speech on the one hand and mouthed and noise-masked speech on the other hand suggests that there is no effective monitoring at the motor level. More importantly, the best error detection occurs when auditory feedback is present, counter to NST.
Capacity	+	–	–	Capacity restrictions lower the number of self-repairs, though not dramatically.
Aspects of speech flow scrutinized	+		+	Context effects in error detection seem best accounted for by perception-based monitoring, followed by NST.
Temporal characteristics	–	+	+	Observation by Oomen and Postma (2000a) of faster error-to-cut-off times with increased speaking rate counters the perception-based monitoring account.
Repair proper				Both simple and elaborate revisions exist. The former seem in line with production-based monitoring, the latter with more centrally governed speech repair.
Speech comprehension	±	±	–	Link between disordered speech comprehension and self-repair deficits has been found, corroborating perception-based monitoring account. The critical test of disordered comprehension and normal speech repair has been reported once, however, in line with production-based monitoring. NST seems not very likely because damage to the mental node system still can leave error signalling unaffected.

^a + means support; – and – – indicate minor and major grounds for rejection; ± applies to mixed support. A blank is given when no convincing evidence is available yet.

control mechanism? Do comparable control mechanisms exist in other motor skills? As production monitors are dedicated specialists, they may be exclusively related to a fixed aspect of speech (motor) production. Most plausibly, certain stages in the formulator may become equipped with their own control mechanisms through the abundant generation of linguistic structures. On the other hand, the central (perceptual) monitor guarding the progress of one's utterances may be the same which checks the throwing of darts or the pressing of a button in a decision situation.

Acknowledgements

I am very grateful to Addie Johnson and Claudy Oomen for their valuable comments on this paper. Also, I thank Loekie Elbers, Geert Panhuysen, and Frank Wijnen for their careful reading of a previous version of the text.

References

- Abbs, J. H., & Eilenberg, G. R. (1976). Peripheral mechanisms of speech control. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press.
- Abbs, J. H., & Gracco, V. L. (1983). Sensorimotor actions in the control multi-joint speech gestures. *Trends in Neuroscience*, 6, 391–395.
- Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior*, 16 (2), 195–231.
- Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382–391.
- Berg, T. (1986a). The aftermath of error occurrence: psycholinguistic evidence from cut-offs. *Language and Communication*, 6 (3), 195–213.
- Berg, T. (1986b). The problems of language control: editing, monitoring, and feedback. *Psychological Research*, 48, 133–144.
- Berg, T. (1992). Productive and perceptual constraints on speech error correction. *Psychological Research*, 54, 114–126.
- Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). "Where did I go wrong?" A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, 21 (6), 1312–1322.
- Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39, 173–194.
- Borden, G. J. (1979). An interpretation of research on feedback interruption in speech. *Brain and Language*, 7, 307–319.
- Clark, E. V. (1978). Awareness of language: some evidence from what children say and do. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language*. Berlin: Springer-Verlag.
- Cooke, J. D., & Diggles, V. A. (1984). Rapid error correction during human arm movements: evidence for central monitoring. *Journal of Motor Behavior*, 16 (4), 348–363.
- Crosson, B. (1985). Subcortical functions in language: a working model. *Brain and Language*, 25, 257–292.
- Cutler, A. (1983). Speakers' conceptions of the function of prosody. In A. Cutler, & D. R. Ladd, (Eds.), *Prosody: models and measurements*. Berlin: Springer-Verlag.
- Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, 5, 303–305.
- Dell, G. S. (1985). Positive feedback in hierarchical connectionist models: application to language production. *Cognitive Science*, 9, 3–24.

- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5, 313–349.
- Dell, G. S., & Repka, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental slips and human error: exploring the architecture of volition*. New York: Plenum Press.
- Del Viso, S., Igoa, J. M., & Garcia-Albea, J. E. (1991). On the autonomy of phonological encoding: evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research*, 20, 161–185.
- De Smedt, K., & Kempen, G. (1987). Incremental sentence production, self-correction, and coordination. In G. Kempen (Ed.), *Natural language generation: recent advances in artificial intelligence, psychology, and linguistics*. Dordrecht: Kluwer.
- Fairbanks, G. A. (1954). A theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders*, 19, 133–139.
- Festinger, L., & Canon, L. K. (1965). Information about spatial location based on knowledge about efference. *Psychological Review*, 72 (5), 373–384.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*, 17, 631–648.
- Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum Associates.
- Frith, C. D. (1994). The theory of mind in schizophrenia. In A. S. David, & J. C. Cutting (Eds.), *The neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum Associates.
- Frith, C. D., & Done, D. J. (1988). Towards a neuropsychology of schizophrenia. *British Journal of Psychiatry*, 153, 437–443.
- Garnsey, S. M., & Dell, G. S. (1984). Some neurolinguistic implications of prearticulatory editing in production. *Brain and Language*, 23, 64–73.
- Garret, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk*. London: Academic Press.
- Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia, PA: Lea and Febiger.
- Howell, P., & Young, K. (1991). The use of prosody in highlighting alterations in repairs from unrestricted speech. *The Quarterly Journal of Experimental Psychology*, 43A (3), 733–758.
- Jaeger, R. J., Agarwal, G. C., & Gottlieb, G. L. (1979). Directional errors of movement and their correction in a discrete tracking task. *Journal of Motor Behavior*, 11 (2), 123–133.
- Jou, J., & Harris, R. J. (1992). The effect of divided attention on speech production. *Bulletin of the Psychonomic Society*, 30, 301–304.
- Karmiloff-Smith, A. (1986). From meta-processes to conscious access: evidence from children's meta-linguistic and repair data. *Cognition*, 23, 95–147.
- Kelso, J. A. S. (1982). Concepts and issues in human behavior. In J. A. S. Kelso (Ed.), *Human motor behavior: an introduction*. Hillsdale, NJ: Erlbaum.
- Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10 (6), 812–832.
- Kempen, G. (2000). Human grammatical coding. Manuscript submitted for publication.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201–258.
- Kent, R. D. (1976). Models of speech production. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press.
- Kolk, H., & Postma, A. (1996). Stuttering as a covert repair phenomenon. In R. F. Curlee, & G. M. Siegel (Eds.), *Nature and treatment of stuttering*. Boston, MA: Allyn and Bacon.
- Lackner, J. R., & Tuller, B. H. (1979). 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 Garret*. Hillsdale, NJ: Lawrence Erlbaum.

- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley.
- Laver, J. D. M. (1973). The detection and correction of slips of tongue. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence*. The Hague: Mouton.
- Laver, J. D. M. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen, and hand*. New York: Academic Press.
- Lebrun, Y. (1987). Anosognosia in aphasics. *Cortex*, 23, 251–263.
- Leudar, I., Thomas, P., & Johnston, M. (1992). Self-repair in dialogues of schizophrenics: effects of hallucinations and negative symptoms. *Brain and Language*, 43, 487–511.
- Leudar, I., Thomas, P., & Johnston, M. (1994). Self-monitoring in speech production: effects of verbal hallucinations and negative symptoms. *Psychological Medicine*, 24, 749–761.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). The perceptual loop theory not disconfirmed: a reply to MacKay. *Consciousness and Cognition*, 1, 226–230.
- Levelt, W. J. M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of semantics*, 2, 205–217.
- Levelt, W. J. M., & Wheeldon, L. R. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269.
- Levelt, W. J. M., Sinclair, A., & Jarvella, R. J. (1978). Causes and functions of linguistic awareness in language acquisition: some introductory remarks. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language*. Berlin: Springer-Verlag.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production. *Psychological Review*, 98 (1), 122–142.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, 22, 1–75.
- Liss, J. M., (1998). Error-revision in the spontaneous speech of apraxic speakers. *Brain and Language*, 62, 342–360.
- Maassen, B. (1985). *Artificial corrections to deaf speech*. Unpublished doctoral dissertation, University of Nijmegen, The Netherlands.
- MacKay, D. G. (1987). *The organization of perception and action: a theory for language and other cognitive skills*. New York: Springer.
- MacKay, D. G. (1992a). Awareness and error detection: new theories and research paradigms. *Consciousness and Cognition*, 1, 199–225.
- MacKay, D. G. (1992b). Errors, ambiguity, and awareness in language perception and production. In B. J. Baars (Ed.), *Experimental slips and human error: exploring the architecture of volition*. New York: Plenum Press.
- MacNeilage, P. F. (1970). Motor control and serial ordering in speech. *Psychological Review*, 77, 182–196.
- Maher, L. M., Rothi, L. J. G., & Heilman, K. M. (1994). Lack of error awareness in an aphasic patient with relatively preserved auditory comprehension. *Brain and Language*, 46, 402–418.
- Marshall, R. C., & Tompkins, C. A. (1982). Verbal self-correction behaviors of fluent and nonfluent aphasic subjects. *Brain and Language*, 15, 292–306.
- Marshall, R. C., Rappaport, B. Z., & Garcia-Bunel, L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language*, 24, 297–313.
- Marslen-Wilson, W. D., & Tyler, L. K. (1981). Central process in speech understanding. *Philosophical Transactions of the Royal Society London, Series B*, 259, 297–313.
- Marslen-Wilson, W. D., & Tyler, L. K. (1983). Reply to Cowart. *Cognition*, 15 (1–3), 227–235.
- Mattson, M. E., & Baars, B. J. (1992). Error-minimizing mechanisms: boosting or editing? In B. J. Baars (Ed.), *Experimental slips and human error: exploring the architecture of volition*. New York: Plenum Press.
- McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, 119, 907–917.

- McNamara, P., Obler, L. K., Au, R., Durso, R., & Albert, M. L. (1992). Speech monitoring skills in Alzheimer's disease, Parkinson's disease, and normal aging. *Brain and Language*, 42, 38–51.
- Motley, M. T. (1980). Verification of "Freudian slips" and semantic prearticulatory editing via laboratory-induced sponnerisms. In V. A. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen, and hand*. New York: Academic Press.
- Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 21, 578–594.
- Mowrey, R. A., & MacKay, I. R. A. (1990). Phonological primitives: electromyographic speech error evidence. *Journal of the Acoustical Society of America*, 88 (3), 1299–1312.
- Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, 95 (3), 1603–1616.
- Neilson, M. D., & Neilson, P. D. (1987). Speech motor control and stuttering: a computational model of adaptive sensory motor processing. *Speech Communication*, 6, 325–333.
- Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: comprehension, monitoring and lexicality. *Cortex*, 31, 209–237.
- Nooteboom, S. G. (1980). Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen, and hand*. New York: Academic Press.
- Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, 88 (1), 1–15.
- Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, 103 (2), 241–262.
- Oomen, C. C. E., & Postma, A. (2000a). Effects of increased speech rate on monitoring and self-repair. *Journal of Psycholinguistic Research* (in press).
- Oomen, C. C. E., & Postma, A. (2000b). Limitations in processing resources and speech monitoring. Manuscript submitted for publication.
- Perkell, J. S. (1980). Phonetic features and the physiology of speech production. In B. Butterworth (Ed.), *Language production*. London: Academic Press.
- Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, 22, 227–250.
- Postma, A. (1997). On the mechanisms of speech monitoring. In W. Hulstijn, H. Peters, & P. van Lieshout (Eds.), *Speech production: motor control, brain research and fluency disorders*. Amsterdam: Elsevier.
- Postma, A., & Kolk, H. H. J. (1992a). Error monitoring in people who stutter: evidence against auditory feedback defect theories. *Journal of Speech and Hearing Research*, 35, 1024–1032.
- Postma, A., & Kolk, H. H. J. (1992b). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research*, 35, 537–544.
- Postma, A., & Kolk, H. H. J. (1993). The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, 36, 472–487.
- Postma, A., & Noordanus, C. (1996). The production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, 39 (4), 375–392.
- Rabbitt, P. (1990). Age, IQ and awareness, and recall of errors. *Ergonomics*, 33, 1291–1305.
- Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology*, 29, 727–743.
- Reason, J. T. (1990). *Human error*. Cambridge: Cambridge University Press.
- Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, 35, 854–876.
- Rosenbaum, D. A. (1991). *Human motor control*. San Diego, CA: Academic Press.
- Saltzman, E., & Kelso, J. A. S. (1987). Skilled action: a task-dynamic approach. *Psychological Review*, 94 (1), 84–106.
- Schade, U., & Laubenstein, U. (1993). Repairs in a connectionist language-production model. In R. Kohler, & B. B. Rieger (Eds.), *Contributions to quantitative linguistics*. Dordrecht: Kluwer.
- Scheffers, M. K., Coles, M. G. H., Bernstein, P., Gehring, W. J., & Donchin, E. (1996). Event-related brain potentials and error-related processing: an analysis of incorrect responses to go and no-go stimuli. *Psychophysiology*, 33, 42–53.

- Schlenk, K., Huber, W., & Wilmes, K. (1987). "Prepairs" and repairs: different monitoring functions in aphasic language production. *Brain and Language*, 30, 226–244.
- Schmidt, R. A. (1982). More on motor programs. In J. A. S. Kelso (Ed.), *Human motor behavior: an introduction*. Hillsdale, NJ: Erlbaum.
- Schmidt, R. A. (1987). The acquisition of skill: some modifications to the perception-action relationship through practice. In H. Heuer, & A. F. Sanders (Eds.), *Perspectives on perception and action*. Hillsdale, NJ: Erlbaum.
- Schmidt, R. A. (1988). *Motor control and learning: a behavioral emphasis*. Champaign, IL: Human Kinetics.
- Schmidt, R. A., & White, J. L. (1972). Evidence for an error detection mechanism in motor skills: a test of Adams' closed loop theory. *Journal of Motor Behavior*, 4 (3), 143–153.
- Senders, J. W., & Moray, N. P. (1991). *Human error: cause, prediction and reduction*. Hillsdale, NJ: Erlbaum.
- Shuren, J. E., Smith Hammond, C., Maher, L. M., Rothi, L. J. G., & Heilman, K. M. (1995). Attention and anosagnosia: the case of a jargon aphasic with unawareness of language deficit. *Neurology*, 45, 376–378.
- Siegenthaler, B. M., & Hochberg, I. (1965). Reaction time of the tongue to auditory and tactile stimulation. *Perceptual and Motor Skills*, 21, 387–393.
- Stemberger, J. P. (1985). An interactive activation model of language production. In A. Ellis (Ed.) *Progress in the psychology of language* (Vol. 1). London: Erlbaum.
- Van Hest, G. W. C. M. (1996). *Self-repair in L1 and L2 production*. Tilburg: Tilburg University Press.
- Van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: evidence from experimentally elicited corrections. *Cognitive Psychology*, 19, 403–440.
- Webster, R. L. (1980). Evolution of a target-based behavioral therapy for stuttering. *Journal of Fluency Disorders*, 5, 303–320.
- Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 311–334.
- Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychological Bulletin*, 121 (3), 331–354.