Bochumer
Linguistische
Arbeitsberichte
5

Monitoring via the Perceptual Loop: Is the Inner Loop Based on Perception or Production?

Melanie Zarges

Bochumer Linguistische Arbeitsberichte

BLA

Herausgeber: Stefanie Dipper & Björn Rothstein

Die online publizierte Reihe "Bochumer Linguistische Arbeitsberichte" (BLA) gibt in unregelmäßigen Abständen Forschungsberichte, Abschluss- oder sonstige Arbeiten der Bochumer Linguistik heraus, die einfach und schnell der Öffentlichkeit zugänglich gemacht werden sollen. Sie können zu einem späteren Zeitpunkt an einem anderen Publikationsort erscheinen. Der thematische Schwerpunkt der Reihe liegt auf Arbeiten aus den Bereichen der Computerlinguistik, der allgemeinen und theoretischen Sprachwissenschaft und der Psycholinguistik.

The online publication series "Bochumer Linguistische Arbeitsberichte" (BLA) releases at irregular intervals research reports, theses, and various other academic works from the Bochum Linguistics Department, which are to be made easily and promptly available for the public. At a later stage, they can also be published by other publishing companies. The thematic focus of the series lies on works from the fields of computational linguistics, general and theoretical linguistics, and psycholinguistics.

Band 5 (August 2011)

Herausgeber: Stefanie Dipper
Sprachwissenschaftliches Institut
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum

Björn Rothstein
Germanistisches Institut
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum

Erscheinungsjahr 2011
ISSN **2190-0949**

Melanie Zarges

“Monitoring via the Perceptual Loop: Is the Inner Loop Based on Perception or Production?”

2011

Bochumer Linguistische Arbeitsberichte

(Bla 5)

Inhaltsverzeichnis

1 Abstract 4
2 Introduction 4
3 The Perceptual Loop Theory 5
3.1 Monitoring by Means of Perceptual Loops According to Levelt (1983) 6
3.2 Proposed Amendment and Adjustments 8
4 The Dissociation of the Inner and the Outer Loop 9
4.1 Is There an Inner Loop? 10
4.2 Do the Inner and the Outer Loop Work in the Same Way? 12
5 The Inner Loop: Perception or Production? 14
6 Conclusion 19
7 References 23

1 Abstract

To monitor one's speech means to check the speech plan for errors, both before and after talking. There are several theories as to how this process works. We give a short overview on the most influential theories only to focus on the most widely received one, the Perceptual Loop Theory of monitoring by Levelt (1983). One of the underlying assumptions of this theory is the existence of an Inner Loop, a monitoring device that checks for errors before speech is articulated. This paper collects evidence for the existence of such an internal monitoring device and questions how it might work. Levelt's theory argues that internal monitoring works by means of perception, but there are other empirical findings that allow for the assumption that an Inner Loop could also use our speech production devices. Based on data from both experimental and aphasiological papers we develop a model based on Levelt (1983) which shows that internal monitoring might in fact make use of both perception and production means.

2 Introduction

The term 'monitoring' refers to the ability to assess speech, produced both by oneself and others, for errors. Speakers are able to monitor for several aspects of their speech production, such as linguistic correctness, including phonetics or syntax (am I producing this correctly?), conceptual sufficiency (is this what I meant to say?), and appropriateness in context (does this get my message across to the listener?). Self-monitoring also usually involves some kind of repair after error detection (Levelt, 1983).

Several theories of monitoring have been proposed, the three most influential of which all assume different kinds of monitoring devices (see Postma, 2000, for review). Monitoring is presumed to either work via production, or via feedback of activation, or via perception.

The production monitoring account proposes monitoring devices on every level of language production, which are able to look inside every single production process (but see Nozari et al., 2011, for a very recent account of a central conflict-based production monitor), permitting the speaker to stop upon detecting an error This would then lead to a restarting of the whole procedure. There are several difficulties with this theory. First, a detected error would cause immediate cessation, leading only to covert repairs, meaning corrections which do not become apparent in articulated speech, apart from possible dysfluencies in speech production. Only mistakes not detected by the monitor would become overt, and after hearing an error one could then retrace and correct the message. This is not in line with empirical data on monitoring latencies, as overt errors are often interrupted too fast to be explained by a reaction to hearing oneself. (Hartsuiker & Kolk, 2001; Levelt, 1983) Also, since the production monitoring account proposes a specified monitor for every production level, in order to find a mistake the monitor would need to have access to the same information as the production process. This would lead to either assuming that monitoring and production processes compete for the same input, which would be unproductive, or lead to the assumption of an unnecessary double realization of information for every single layer of production, which would be inefficient (Levelt, 1983). But again, see Nozari et al. (2011) for a different account of a potential production monitor.

Another proposed possibility for monitoring uses feedback of activation. On the basis of the Node Structure Theory (MacKay, 19992a, b, reviewed in Postma, 2000), MacKay assumes that, normally, those nodes belonging to the word meant to be produced should hold the most activation. It is usually presumed in such models that some activation will also flow towards related nodes, leading to the possibility that an error ensues when the wrong node accidentally achieves a higher level of activation than the intended one. This might happen on any level of speech production, explaining phonological errors as well as morphological, syntactic, or semantic ones. The monitor in such a model would consist of a warning system going off, if activation for the wrong node becomes too strong. The uncommitted node gaining too much activation would make the speaker aware of something going wrong, resulting in a conscious cutoff and repair. But if monitoring was based on activation, then auditory feedback, meaning listening to one's own speech, should not have any effect. However, empirical data show that auditory feedback improves monitoring accuracy, which cannot be explained by a simple feedback account (Oomen & Postma, 2001).

The most influential monitoring theory to date (Postma, 2000) is the Perceptual Loop Theory, originating with Levelt, (1983), which assumes a central comprehension-based monitor. This paper will mostly concentrate on this theory, and discuss the empirical data which do or do not support it.

3 The Perceptual Loop Theory

The Perceptual Loop Theory, according to Levelt (1983), assumes a central comprehension monitor within the limits of a serial incremental model of speech production. Basically, this amounts to the hypothesis that monitoring one's own speech is like monitoring other-produced speech. Levelt gives two reasons for this theory: first, it is more economical for a speaker to use the perception mechanisms already present to also listen to his or her own speech, both overtly and internally. This avoids an unnecessary double realization of information. For the second argument, Levelt draws upon the theory that speakers can only gain access to their intent and the end products of their speech planning processes, not to the planning process itself. This assumption does not allow for a production monitor.

3.1 Monitoring by Means of Perceptual Loops According to Levelt (1983)

Levelt's model of speech production, as reviewed in Levelt (1983), assumes several levels on which the different components of spoken language are selected, activated, and transmitted. The procedure starts with constructing the message, which means creating and ordering the speaker's intended meaning. A great deal of attention is focused on this process, which makes it subject to the working memory and to a first monitoring loop, the conceptual loop.

The intended message is then sent to the formulator, where the corresponding lemmata are selected and encoded into a phonetic string. This string is equivalent to the "inner speech", and is accessible to monitoring by means of the inner loop. Levelt assumes that this monitoring is happening while the phonetic string is being buffered, waiting to be dispatched to the articulator and converted into overt speech. This is, again, subject to a monitoring loop, the articulatory or outer loop.

Perceiving language involves sending it by way of the speech comprehension system to the parser. The parser assimilates both other- and self-produced overt speech, as well as the speaker's inner speech. The concept of the parser is somewhat imprecise, but it seems to convert the phonetic string into a representation of phonetic, morphological, syntactic, and semantic information to be fed back into the conceptualizer (Huettig & Hartsuiker, 2010). Here, both kinds of input are compared to the intended message.

In sum, the monitor, according to Levelt, has three loops. It first draws on the intended message, the concept of the speaker, in what is called the conceptual loop. The monitor also has access to, firstly, the pre-articulatory phonetic string, our inner speech, which is sent through the inner loop, and finally the overt speech, which is checked by means of the articulatory or outer loop. Both the inner and the outer loop utilize parsed speech from the speech comprehension system as input.

The monitor itself has two functions. It matches the speaker's intended message to the actually perceived speech and checks for linguistic correctness and continuity. It also makes the speaker aware of mistakes. Should the monitor detect an error, it sends a warning signal to the working memory, where at least part of the intended message is still stored. Now the speaker has to decide whether they want to adjust their message, meaning keeping to the existing plan but changing a wrong word, or whether they have to stop language production and start over with a new speech plan.

As for the time frame, the perceptual loop theory draws on the main interruption rule (as reviewed in Levelt, 1983), according to which speech is interrupted as soon as the monitor becomes aware of a mistake, regardless of word boundaries. However, Levelt notes that this only seems to be the case for words which actually contain an error. He therefore differentiates between two kinds of repairs, actual speech errors and appropriateness repairs, the latter meaning an adjustment to the message to improve its comprehensibility or to better project the speaker's intent to the listener. His revised main interruption rule states that, while erroneous words will be halted immediately, interruption in appropriateness repairs will usually not take place until the end of a word. The monitor works on a flow-through basis, meaning that while speech is being monitored, temporarily remaining in the articulatory buffer, the production process moves on. Thus, even an internally detected error might become overt, but will be interrupted sooner than monitoring of overt speech can explain (Hartsuiker & Kolk, 2001).

To summarize, he most important assumptions of Levelt's (1983) monitoring theory are a central perception-based monitor with three loops, which draw on conceptualised, inner, and outer speech, the last two of which are already parsed. Correction works through planned intent still active in working memory and is a conscious process demanding attention. Therefore, monitoring should be subject to limitations on resources. Since self-monitoring is achieved by the speech comprehension system, it should work just like monitoring the speech of others. This is shown in Figure 1.

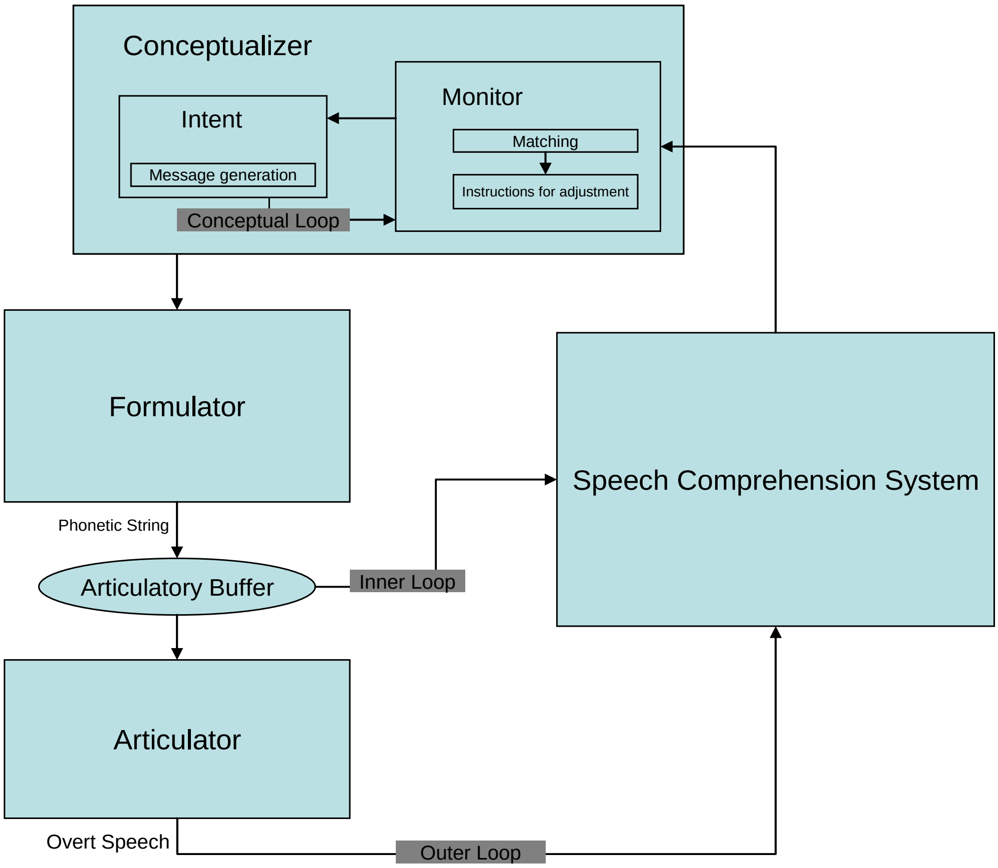

Figure 1. The Perceptual Loop Theory according to Levelt (1983).

3.2 Proposed Amendment and Adjustments

Many aspects of the perceptual loop theory have been questioned and tested, and some have been revised. One of these aspects is the input to the inner loop. While Levelt (1983) assumes that there is a phonetic string being fed to the internal monitoring system, the current opinion tends towards a more abstract phonological representation (Hartsuiker & Kolk, 2001; Roelofs et al., 2007; Slevc & Ferreira, 2006; Wheeldon & Levelt, 1995).

One of the first studies designed to specify the nature of the input to the inner loop was carried out by Wheeldon and Levelt (1995). They asked native speakers of Dutch to silently translate English words into Dutch and monitor their translation for a previously specified phoneme. Participants were asked to press a button to indicate whether the translation of an English word contained said phoneme. In one of the experiments, participants were also instructed to count out loud while translating. This articulatory suppression task is supposed to interfere with phonetic, but not with phonological encoding (Wheeldon & Levelt, 1995). The additional counting task showed almost no effect on monitoring accuracy, indicating that the inner loop does not rely on a phonetic string, while the fact that participants were able to monitor their silent translations for single phonemes shows that they must be able to draw on some kind of abstract speech code prior to articulation. Wheeldon and Levelt concluded that internal speech monitoring is based on a phonological, rather than a phonetic, representation.

The perceptual loop theory has also been implemented and tested as a computational model on the basis of existing empirical data (Hartsuiker & Kolk, 2001), which has led to some revisions. While Levelt (1983) assumes that the monitor has two functions, comparing and making adjustments, Hartsuiker and Kolk argue for a finer distinction between interruption and repair.

Part of the original Perceptual Loop Theory is the main interruption rule, which states that the utterance of erroneous words is stopped as soon as possible and repaired directly afterwards. In a serial model of speech production, this would mean that articulation would first need to be halted, and only after stopping all production processes would the repair be planned and formulated. This would lead to a much longer cutoff-to-repair time, meaning the time interval between interrupting and restarting the revised speech plan, than has been found in empirical studies (Hartsuker & Kolk, 2001). A modified main interruption rule, however, can explain these data points. A necessary assumption is the distinction between interruption and repair. Not all interruptions of speech are followed by repairs, or even due to speech errors. Therefore, the idea of adjusting an utterance in the Levelt model needs to be specified further and separated into interrupting and restarting speech planning.

With interruption and repair as two different processes, the main interruption rule can now be reformulated as stating that detecting an error results in an immediate interruption, stemming from a monitor signal directly to the motor system of the articulator, while at the same time a repair may already be planned in the formulator. According to Hartsuiker and Kolk (2001), a repair can therefore be constructed parallel to the interruption signal, leading to the very short repair-to-cutoff times which have been observed.

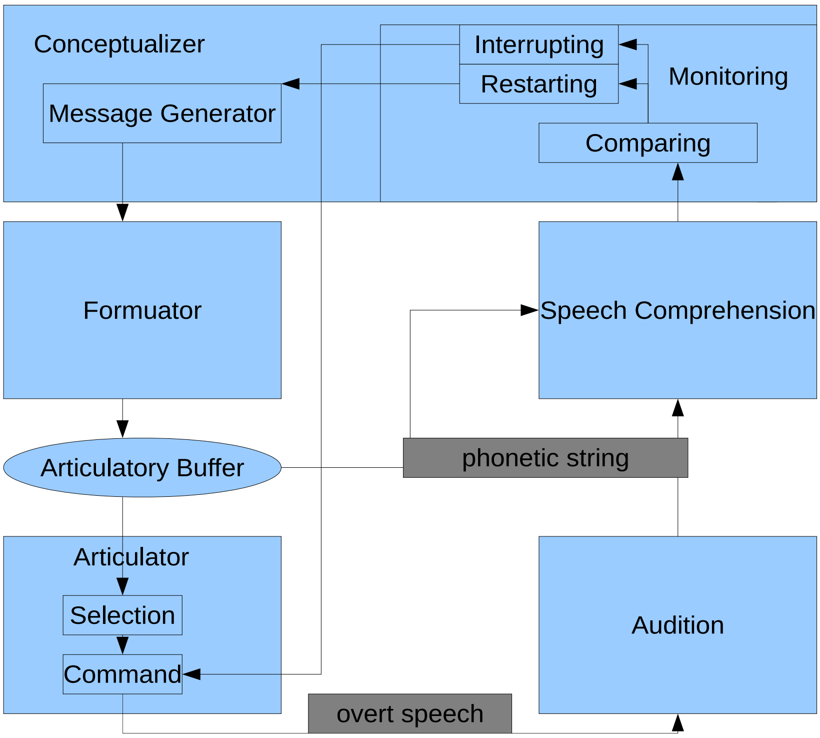

Figure 2. The Perceptual Loop Model, following Hartsuiker & Kolk (2001).

In sum, the Perceptual Loop Theory as implemented by Hartsuiker and Kolk presumes a monitor which has three functions, namely comparing, interrupting, and restarting speech planning, the last two of which can run parallel to each other. This is shown in Figure 2.

4 The Dissociation of the Inner and the Outer Loop

There has never been any doubt that speakers monitor their own overt speech for errors (Özdemir et al., 2007). Usually, some kind of pre-articulatory monitoring is also assumed. The question remains of how this internal monitoring actually works. This section will give a short overview of the empirical data that support internal monitoring, and then consider the assumption of the Levelt model of 1983 that this inner monitor works just like the monitoring of overt speech.

4.1 Is There an Inner Loop?

An important assumption of the perceptual loop theory is that of an internal monitoring device working by way of the speech comprehension system, akin to hearing an inner voice talk inside your own head (Levelt, 1983). It has been shown that the input into the inner monitor seems to be more abstract than a phonetic code, however, and is most probably derived at a phonological level (Roelofs et al., 2007; Wheeldon & Levelt, 1995). Over the years, the existence of an internal monitor has been shown in a number of studies, using different methods and paradigms.

Postma and Noordanus (1996) had participants report their errors while reciting tongue-twister sentences under different speech conditions, for example in silent or noise-masked speech. For the latter condition, participants were asked to wear headphones playing white noise while producing and monitoring sentences meant to elicit errors, ruling out any auditory feedback. The results of the study show that participants had no problems detecting errors in silent, mouthed (articulator movements without sound), and noise-masked speech, proving that auditory feedback is not necessary for error detection. It does, however, seem to improve monitoring accuracy, as more mistakes were detected when overt speech was available. This agrees with the assumptions of the perceptual loop theory, as an additional outer loop would help detect more errors.

A computational implementation of the perceptual loop theory (Hartsuiker & Kolk, 2001) shows that an inner loop is needed to account for existing empirical data, especially with regard to error-to-cutoff times. This term designates the time needed to stop articulation after an error has been detected; in many cases this is a very short time span, shorter than any humanly possible reaction time. Stopping the error therefore must have been planned before articulation, which would imply that the error would need to have been detected prior to articulation as well.

A different kind of simulation has been carried out by Slevc and Ferreira (2006), who tried to emulate the monitoring system by means of a stop-paradigm task. Here, participants of the study were asked to name pictures, unless they were presented with a signal, at least 300 ms after picture onset, telling them to stop. Planning a single word is estimated to take about 200 - 250 ms (Levelt, 1983). Slevc's and Ferreira's task is meant to mimic the function of the monitor, which also needs to stop speech already in the course of production, if an error is signalled to have been found. Participants were able to halt a significant number of their productions, as proposed by the perceptual loop theory. Also, stopping was more difficult if the stop signal consisted of a word that was phonologically similar to the picture name, speaking for a phonological representation as input to the inner monitor.

More evidence for an internal monitor comes from studies employing inner phoneme monitoring tasks. Participants were asked to silently produce words, and to give a signal when their silent rendition contained a previously specified phoneme (Özdemir et al., 2007; Roelofs et al., 2007; Wheeldon & Levelt, 1995). The ability to detect and react to single phonemes in an internal representation of speech clearly speaks for the existence of a pre-articulatory monitoring feature. Also, participants showed sensitivity to the distance of the target phoneme from the point of the word where it diverges from all other words in the language, the uniqueness point (Özdemir et al., 2007). Effects of the uniqueness point are usually observed in perception studies, indicating the existence of a perceptual inner loop.

Internal monitoring not only works with phonological criteria, but also on a word-level basis, which has been shown in several studies on the lexical bias effect (Hartsuiker et al,. 2004; Nooteboom & Quené, 2008; Oppenheim & Dell, 2008). This effect refers to the tendency of phonological speech errors to more often result in the production of real words than of non-words, as long as the context requires real word production. As shown by Hartsuiker et al. (2004), there is no lexical bias effect in a context where participants are only asked to produce non-words, supporting the idea of a monitor which can apply a lexicality criterion to filter out non-words, but which is also able to adapt to circumstances such as participants wanting to produce nonsense words. This is in line with the assumption of the perceptual loop theory that monitoring is a process which requires attention and is subject to central resources, and therefore able to adapt to diverse situations by applying different kinds of criteria to the monitoring loops.

More evidence for an inner loop comes from studies based on the SLIP-technique. This method uses phonological priming to elicit spoonerisms, e.g. *darn boor* being produced instead of *barn door* (Baars & Motley, 1974). It has been shown that participants make mistakes, or slips, more easily if the prime's target spoonerism is not a non-word, displaying the lexical bias effect. Participants produce significantly fewer slips if the target word is a taboo word or a word which is especially emotionally charged. Hamm et al. (2004), for example, tested anorexic girls with stimuli meant to elicit target words with reference to their disorder. Under time pressure, significantly more target words were produced, while in stress-free conditions fewer of these words were produced than unrelated spoonerisms, which the SLIP paradigm usually generates. Apparently, test participants were able to, and did, filter out words related to their eating disorder as long as they had the time and resources to monitor for them. This, again, is in line with a pre-articulatory monitoring loop able to focus on different criteria, here maybe emotional valence. Also, this supports the idea of Levelt (1983) that inner speech is held in an auditory buffer while the inner loop runs its course. Faster stimuli renditions would lead to faster articulation, therefore also decreasing buffer time, leaving less time for the inner monitor to filter out words it would normally prevent from being uttered.

In sum, there is a lot of data supporting the idea of an internal monitoring loop which has access to a phonological representation of pre-articulatory speech, but which can also monitor for other criteria, depending on context and circumstances such as lexicality or emotionality. This seems to be in line with the perceptual loop theory, which proposes an inner loop working by means of the speech comprehension system. However, as will be shown in the next section, a different explanation for these points of data is also possible.

4.2 Do the Inner and the Outer Loop Work in the Same Way?

One of the major assumptions of the perceptual loop theory is that both the inner and the outer loop work by means of the speech comprehension system, so they should work exactly alike, except that the inner loop would, of course, be faster, since it has earlier access to information. There is, however, some data which point towards the different monitoring loops also serving different purposes.

Oomen and Postma (2002) tested monitoring performance under resource limitation by using a dual task paradigm. Participants were asked to describe a network of pictures and lines by relating the path a red ball took along the connecting lines. In one experimental condition, they had to additionally tap a random sequence on a keyboard. This random generation task was meant to decrease the availability of central resources, as deliberately keeping a sequence of movements random needs a lot of attention (Oomen & Postma, 2002). The results of the study show that, when faced with a dual task, monitoring accuracy decreases, as predicted by the perceptual loop theory (Levelt, 1983). To test the theory of a perception monitor, the same experiment was conducted again, only with a perception task instead of the production task. Now participants were instructed to listen to another person describing the movements of the red ball on the network and press a button upon detecting a mistake. While monitoring accuracy was better without the dual task for both conditions, there were some unexpected differences in the monitoring latencies.

Error-to-cutoff times in the perception task, for which only the outer loop was available, were slower for the dual task condition, while error-to-cutoff times in the production experiment sped up if there was an additional task. A limitation of resources seems to lead to less attention being available for monitoring, which might result in a shorter monitoring span. If participants tried to speed up their monitoring to accommodate the accessibility of fewer resources, they would possibly become aware of a smaller number of errors, but those errors that were detected would be discovered sooner, leading to a shorter error-to-cutoff time. This idea is supported by a similar study, also using network descriptions where pictures to be named were indicated by a moving red dot, to analyze the effects of time pressure on monitoring (Oomen & Postma, 2001). The method used was similar to the study described above, only that the red dot connecting the pictures moved at either a normal or a relatively fast speed, and participants were instructed to keep pace with their descriptions. The results showed that error-to-cutoff times were faster when participants had to speed up their language production, showing that less available time for monitoring leads to faster error detection times. However, this only explains the decrease in error-to-cutoff times for the production experiment in Oomen and Postma (2002), it does not account for the slower times in the perception experiment.

Assuming that the inner and outer loop both work by means of the perception system, it seems strange that the monitoring system would work differently in a production than in a perception task. Apparently, monitoring other produced speech is not like monitoring one's own speech, indicating a discrepancy between the inner and the outer loop. An alternative explanation, given by Oomen and Postma (2002), would be that participants were able to focus their monitoring behaviour depending on the given task. Under stress, for example when faced with an additional task, they might have relied more on the inner loop, which is faster than the outer loop. This still does not explain why people would not try to speed up their outer loop monitoring by examining other produced speech also for a shorter time, maybe relying more on early syllables or word onsets, especially since this tactic apparently works for internal monitoring.

More evidence for a discrepancy between the inner and the outer loop comes from a study by Slevc and Fereira (2006), in which a stopping task was used to simulate monitoring functions. Participants were asked to name pictures as quickly and accurately as possible. In some trials they would be given either no signal or a signal containing the name of the given picture, either as a spoken or as a written word, in which case they were to keep on talking. In other trials they would be given a signal differing from the word they were currently producing, in which case they were asked to stop production immediately. The main finding was that phonological similarity of stop-signal and target word made it harder to stop speech production, while semantic relation had no effect.

Monitoring accuracy in this experiment was higher, meaning that more trials were successfully stopped, when the stop or go signal was visually presented as a written word, than with auditorily presented stop or go signals. This would mean that focussing on an inner loop not only decreases error-to-cutoff times, but that excluding the outer loop apparently actually increases monitoring accuracy. It appears that participants did not read the stimulus silently and then monitored their own produced reading, as there is no reason why this would yield a better result than overtly produced speech, unless people are more efficient at monitoring their own speech than monitoring other-produced speech, which, again, would imply that monitoring oneself is not like monitoring others, and that the inner and the outer loop do not work in the same way.

There is also some evidence for the assumption that the outer loop is more sensitive to some mistakes than the inner loop. Oppenheim and Dell (2008) found an effect of lexical bias in overt and inner speech, while in the same task a phonemic similarity effect was found in overt speech only. The same error production task leading to different results for inner and overt speech leaves the possibility of either different monitoring systems for, or different representations of, internal and overt speech. The authors argue in favour of a surface-impoverished representation of internal speech. This issue will be taken up again later.

5 The Inner Loop: Perception or Production?

So far it has been shown that empirical data support an internal monitor as proposed by the perceptual loop theory, but this monitor apparently does not, as so far predicted, work in the same way as the outer loop. It seems silly to contest that the outer loop works by means of the speech perception system, unless there are two different kinds of systems for perceiving our own overt speech and other-produced overt speech, which would presuppose two different kinds of perception systems, a double representation of a lot of information; this would be a waste of resources and is therefore an unlikely assumption. If the outer loop works by means of the speech comprehension system and the inner loop does not function like the outer loop, it seems logical to presume that maybe the inner loop works by means of the speech production system.

Data supporting a production-oriented inner monitoring account have been obtained by studying patients suffering from aphasia or other diseases involving language deficits, such as Alzheimer's or Parkinson's disease. Marshall et al. (1998) found partially severely impaired monitoring in patients with jargon aphasia, who showed relatively good comprehension skills, but whose production system was apparently damaged. One patient was able to detect errors in his own speech when repeating spoken stimuli, but not in a picture-naming task. As long as he was relying mostly on his speech comprehension system and only had to imitate sounds, he was able to spot mistakes, but when he had to produce names, calling on both a conceptual and a phonological layer of speech production, he was unable to monitor his own speech. Incidentally, the connection between those layers was supposed to be the most severely damaged area of his speech production (Marshall et al., 1998).

If the patient had trouble connecting lemmata and their phonological representations in order to form a phonetic plan, and assuming that the inner monitor works with an input of phonological representation, it would make sense to conclude that the patient was not able to perform inner monitoring to detect mistakes in the picture-naming task. Repeating speech would bypass the conceptual level and the lemma retrieval, as this task only requires the emulation of sounds. A phonological code was generated and articulated, and monitored relatively successfully by the outer loop. This allows for the hypothesis that the patient's intact comprehension skills allowed him to draw upon a perception monitor by way of the outer loop, but his impaired production did not permit a production-based inner loop to function properly. Note, however, that if no phonological representation of picture names was obtained, there also would have been no input for a perception-based inner loop.

More striking results have been obtained by a study on patients with either Alzheimer's or Parkinson's disease (McNamara et al., 1992). Patients with Parkinson's disease usually show dysfluencies in speech such as, for example, stuttering, and they tend to have problems with managing attention. Their language comprehension, though, is mostly intact. In an experiment in which they were asked to describe the Cookie Theft picture (of the BDAE (Goodglass & Kaplan, 1972), cited in: McNamara et al., 1992) Parkinson patients showed very low monitoring skills. They made three times as many errors as the participants of a control group, indicating a production impairment, and detected and attempted to repair significantly fewer errors, showing damaged monitoring. This study shows again that patients with good comprehension skills can still display monitoring impairments, indicating that the monitor cannot rely on perception only. It should be noted that Hartsuiker and Kolk (2001) argue that these patient data are not convincing enough to refute the notion of a purely perceptually based monitor, as the participants in question could have a number of additional impairments in any of the sub-processes assumed by the Perceptual Loop Theory, such as impaired production processes leaving dissatisfying input for the inner loop, leading to poor monitoring while comprehension remains unaffected.

More compelling evidence against a perception-based inner monitor comes from several studies showing the reverse effect, namely intact monitoring skills in patients with impaired comprehension. Schlenck et al. (1987) observed poor outer loop monitoring in patients with Broca’s and Wernicke’s aphasia, as both patient groups made few attempts to repair their own overt mistakes. This was expected, as especially Wernicke’s aphasics suffer from severe comprehension impairments. Conversely, Schlenck et al. found a number of what they called “prepairs”, defined as “various forms of linguistics searching behaviour” (Schlenck et al., 1987, p. 226) which indicates trouble in language production. Hartsuiker and Kolk (2001) equate this with the more widely used term of covert repairs. It seems that even aphasics with bad comprehension skills still exhibit a quite active inner monitor, as Wernicke’s aphasics showed as many pre-articulatory repairs as the other aphasic groups, to whom were ascribed better comprehension skills. This points towards a production-based inner monitoring device.

Marshall et al. (1985) report the case of a patient exhibiting similar abilities and dysfunctions. Their patient was suffering from a severe auditory agnosia, which would exclude the possibility of outer loop monitoring. The patient did not understand spoken language, but spoke fluently. Also, when producing speech, she was still aware of her mistakes and frequently tried to correct even phonological errors. This effect could not have been achieved by her damaged speech comprehension system, again speaking for an internal loop which is governed by production processes.

It might be interesting to note that the patient was able to detect errors, but repairing them was very difficult for her. She obviously managed to interrupt her speech, but the replanning of her utterances seemed impaired, as an attempted repair often resulted in an even more erroneous word than her original output. This could actually be a point in favour of the perception monitoring account, as it might be presumed that, upon detecting an error, her inner loop tried to gain access to her impaired comprehension system, leaving the monitor to deal with probably even more confusing input than the original error, which might have had an influence when she tried to replan her message. This is, however, highly speculative. Hartsuiker and Kolk (2002) assume that the relatively good reading comprehension skills of the patient might point towards a deficit early in the hearing process, leaving some part of the comprehension system and maybe the parser intact. While this is surely a possibility, it does also not explain the problems with replanning and repairing utterances which the patient exhibited.

In sum, neuropsychological studies reporting on data obtained by working with patients with Parkinson's disease, or with aphasics, show that preserved comprehension skills do not avert monitoring deficits, and, on the other hand, that the monitor can still work even when the comprehension system is severely impaired. Both findings support the idea of a production-based inner loop.

The idea of a production monitor might also shed new light on the data obtained by the stop-signal-paradigm study by Slevc and Ferreira (2006). Excluding the auditory loop by showing participants pictures as stop signals instead of using auditory stimuli was found to increase monitoring accuracy for the phonologically similar condition. The idea of a perception monitor can yield no reason why there should be a difference in stopping accuracy between auditory and visual signals, as both inputs should be made available to the monitor through the same comprehension system. In fact, other data (Postma and Noordanus, 1996) have shown that the auditory loop might be more sensitive to detection of speech errors than the inner loop. The difference in the two studies lies in the task: Postma and Noordanus asked participants to produce tongue twister sentences in silent, mouthed, noise-masked, and normal auditory feedback conditions. Most errors were detected and repaired in the normal auditory feedback condition, the only one including an auditory loop.

Slevc and Ferreira, however, did not try to elicit speech errors to look at correction rates; they tried to simulate an inner monitoring task by asking participants to try and stop a planned message during production using perceived signals. It seems logical that, in a condition where people were asked to produce sentences and report on their mistakes, an additional monitoring loop would increase the chance of detecting an error, no matter if these loops were comprehension or production based. In fact, an inner loop working by means of production, supported by a perception-based outer loop, might even enhance the monitoring chances by being able to set different foci more easily than if both had to use the same route. But in the stop-signal experiment, there was never a need to report on errors. Excluding the outer loop might have made stopping to phonologically similar signals easier because, assuming a production-based inner loop, using pictorial stimuli erased the need for the comprehension system to come into play at all. This is because, most likely, the names of the pictures were not encoded phonologically.

Comprehending the pictures for the naming task was not supposed to draw on the speech perception system. The written stop signal probably did not pass the speech comprehension system, either. Assuming an inner loop that goes directly from the speech production system to the conceptualizer would eliminate the need to pass through the comprehension system.

Inside the monitor, the produced speech plan, probably its phonological representation (see Roelofs et al., 2007; Wheeldon & Levelt, 1995), would need to be compared to the written stop signal. An auditory stop signal would necessarily have to pass the speech comprehension system, but there is no reason why the written stimuli would need to leave the conceptualizer at all. It would not have to be phonologically planned, explaining also why the phonological similarity would not have much of an influence on stopping accuracy. So an auditory stop signal would have to be processed by the comprehension system, while, at the same time, the production system works on the picture naming task, leading to a splitting of attention and a need for more central resources, which is known to affect the monitor (Hamm et al., 2004; Oomen & Postma, 2002). Assuming a production-based inner monitor, a visual stop signal would only need to activate the speech production system, obviating the need for split attention.

In Oomen and Postma (2002), it was shown that limitations in resources affect monitoring. The authors speculated that added stress makes participants rely more on inner loop monitoring, which is faster than the outer loop. Input for the inner loop is available sooner than overt speech in a self-monitoring task, giving the inner loop a time advantage, but Oomen and Postma also demonstrated that limiting resources through an additional task leads to faster error-to-cutoff times, in other words, the dual task condition sped the monitor up. It was also demonstrated in an earlier study (Oomen & Postma, 2001) that increasing speed of language production also leads to faster error-to-cutoff times. Maybe the limitation of resources and consequent stress on the test participants in Oomen and Postma (2002) also led to a faster overall language production. In experiment two of the same study, participants were given a perception task. Monitoring latencies were slower for the double task condition, leading the authors to believe that a limitation of resources decreases the speed of the comprehension system. If a limitation of resources slows the comprehension system down, but speeds up the production system – possibly drawing on resources normally used for comprehension processes and utilizing them to enhance production – and if such a limitation of resources also speeds up the monitor, than this would be another point of evidence in favour of a production-based inner loop.

Further evidence of this comes from a recent eye-tracking study (Huettig & Hartsuiker, 2010). Based on the fact that perceiving other-produced speech apparently drives eye movements towards phonologically-related written distractors, the authors conducted one production and one perception experiment, with the expectation that a perception monitor would lead to similar results in both conditions, but with different latencies. In the production experiment, participants were asked to name simple line drawings which were presented to them at the same time as three written distractor words. These were either all unrelated, or there was one phonologically similar word (e.g. heart (picture), harp (phonological distractor), couch and window (unrelated distractors)), or there was one semantically related word (e.g. bed (picture), stool (semantically related distractor), violin and hat (unrelated distractors)). While naming the pictures, eye movements were tracked. For the perception task, participants were looking at the same stimuli while passively listening to an auditory recording of the picture names.

In view of the fact that in the production experiment the inner loop has access to phonological information faster than in the perception variant, when the auditory stimulus first would have to be comprehended, Huettig and Hartsuiker expected eye movements in the production experiment to be driven towards the phonological distractor words about 50 or 55 ms before auditory onset.

Results showed that, in both experiments, participants looked significantly more often at the phonologically related distractor object than at the unrelated words. Semantic relation had no effect on eye movements. Latencies for both experiments were nearly identical, showing that in the production experiment, the inner loop did not direct looks at the phonologically related words before the onset of overt speech. Instead, participants seemed to wait until they were listening to their own articulated speech before they moved their eyes towards the distractor. As it has been shown that perception drives eye movements towards phonologically similar stimuli, these results do not support a perception-based inner loop. If the inner loop were based on perception, it should have activated the speech comprehension system, which should have led sooner to more looks towards the phonologically related item. The authors conclude from this that planning speech meant for articulation without an explicit self-monitoring task does not involve a perception-based inner loop.

Overall, there is strong evidence from several patient studies and one eye-tracking study with healthy participants in favour of a production-governed monitoring of internal speech. This is supported further by evidence from a simulation task of (inner) monitoring (Slevc & Ferreira, 2006) and data on the limitation of cognitive resources and their effects on monitoring behaviour (Hamm et al, 2004; Oomen & Postma, 2001, Oomen & Postma, 2002).

There is, however, also one study presenting evidence which points towards a purely perception-based monitor. Özedmir et al., (2007) showed that the uniqueness point of individual words impacts on internal monitoring. The uniqueness point is defined as the moment where the phoneme that distinguishes a word from all other words in a language becomes available. It has been shown to have the reliable effect of speeding up word recognition in perception experiments, while there are no known effects on speech production (Özdemir et al., 2007). In an inner phoneme monitoring task, a facilitating effect of distance from the uniqueness point was found, while there was no effect in a picture naming task. Huettig and Hartsuiker (2010) take issue with the assumption that a task such as inner phoneme monitoring uses the same monitoring processes as monitoring in naturally produced speech, or a more accurate simulation of this, which these authors believe to have achieved by means of picture naming tasks, without an additional and explicit monitoring task. This issue will be revisited later.

6 Conclusion

The Perceptual Loop Theory going back to Levelt (1983) is the monitoring theory which is most in line with a majority of the existing empirical data up to date (but see Nozari et al., 2011, for discussion). It can explain internal speech monitoring, error-to-cutoff times being faster than human reaction times, some effects of resource limitation, an adaptive monitoring system able to focus on different kind of errors (explaining context effects on lexical bias), and, assuming the modified main interruption rule proposed by Hartsuiker and Kolk (2001), cutoff-to-repair times too fast for the replanning of an utterance to have started only after interruption. But the perceptual loop theory cannot explain all of the existing data. As has been shown, there are some aspects of internal monitoring which do not support the idea of a monitoring system relying only on the speech comprehension system. There seem to be some points of data which can be better explained by a production monitor.

The idea of a production monitor has mostly been rejected because, so far, production monitoring has entailed not a central, but several, probably unlinked, monitoring devices on all levels of speech production. This would only explain the stopping of production once an error was found, and then restarting the process again. How the monitor detects an error was not clear, and a repair parallel to stopping production, which seems to be the only way to explain observed cutoff-to-repair times, would not have been reasonable. The activation-feedback account of monitoring, on the other hand, can only explain how errors are detected, but not why or how a repair works, other than through awareness.

Several authors have remarked on the possibility of an interaction of a perceptual monitor and the feedback account (Hartsuiker et al., 2004). Oppenheim and Dell (2008), for example, propose that inside the production system, as modulated by Dell's spreading-activation model (Dell, 1986), some features of inner speech could be blocked. This would prevent activation to flow, for example, from a lexical to the phonemic level, leaving inner speech impoverished. This impoverished representation would then be fed by way of the inner loop into the speech comprehension system, explaining why Oppenheim and Dell found lexical bias in both inner and overt speech, as lexical information would have been activated and available to the inner loop, while there was no phonemic similarity effect for inner speech, since the phonemic level held no activation. This would, however, lead to the assumption of at least two different kinds of inner speech.

In order to generate the phonetic string of overt speech, the phonemic level needs to gain some activation. Therefore, Oppenheim and Dell presume a difference between inner speech intended to be articulated and silent inner speech. Huettig and Hartsuiker (2010) comment on the possibility that the inner loop is production-based when speech is meant to be articulated, while there might be an additional inner loop to be used for consciously listening to inner speech, for example when asked to explicitly monitor inner speech.

It seems very likely that an explicit internal monitoring task leads to the focussing of attention on inner speech. Maybe this directs the resources normally used for production to activate the perception system, because now it seems important to actually listen to internally produced speech. This would mean that “the little voice inside their head” (Oppenheim & Dell, 2008:529) most people hear while thinking is only there because they listen for it. Paying extra attention to inner speech might send it through the comprehension system, while in a normal speech production situation this would take too much time and resources. This would lead to the assumption of two inner loops, one perception- and one production-based, which can be activated according to specific situations. Figure 3 attempts to visualize this.

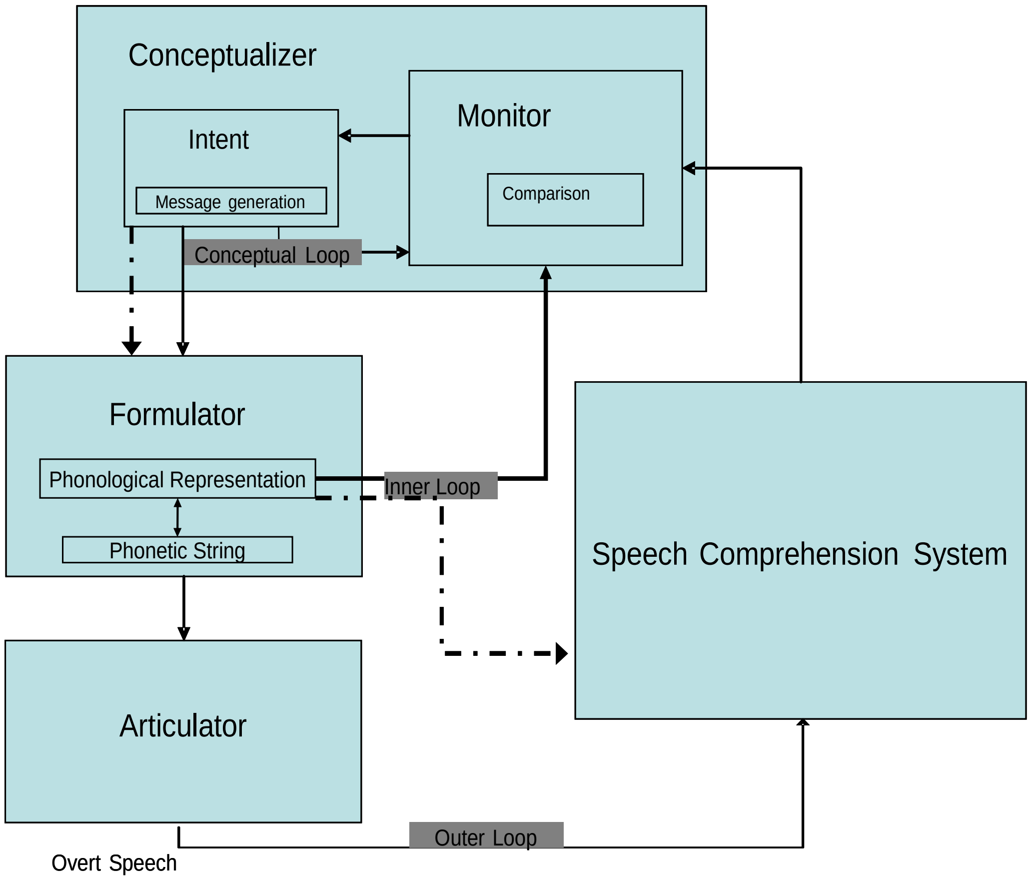

Figure 3. Monitoring model following Levelt's (1983) perceptual loop theory; assuming two inner loops, one production-based for inner speech meant for articulation (indicated by the continuous arrow), one perception-based for silent, consciously comprehended speech (indicated by the dashed arrow).

Assuming such a model still leaves a number of questions open. Presuming an interactive feedback account of monitoring or some other form of production monitoring, a warning signal would be sent by the monitor if, for example, a wrong phoneme were to acquire too much activation while in the production system. This still does not explain how the system giving out the alarm signal receives the information which phoneme is the wrong or right one. This could be solved by assuming that every speech plan is habitually monitored, stopping the need for a feedback-produced warning signal. Interactive feedback would then result in misled activation being detected by the monitor. The monitor needs, obviously, to have a function of comparing an intended plan to the currently produced one. But, assuming that the inner production loop works on the basis of phonological representations, how can it compare the phonological plan to an intended plan that would first have to be generated? Even assuming that the monitor picks out mistakes on the basis of wrong activation and feedback, how does it know which nodes should have been activated? In other words, what exactly does the monitor compare the speech plan to?

A perception-based inner loop has the advantage of parsed speech. Parsed speech is supposed to hold the information of all linguistic production levels, phonological, morphological, syntactic, and semantic (Huettig & Hartsuiker, 2010). But how does the parser break down information so that even the most abstract parts of it, like single phonemes, are available for a comparison? In a phoneme monitoring task, there is a prespecified phoneme as input in the monitor, and produced speech can be compared to this information. But how does the monitor decide that a speaker made a phonological mistake without previous input to the monitor?

The monitor is supposed to be located inside the conceptualizer, so it has access to our terms and definitions, but do these include phonological representations? The ability to monitor other-produced speech for phonological errors indicates that the speech production system does not need to be engaged to get access to phonological information. Roelofs et al. (2007) attempted to find out whether phonological representations are shared by the perception and the production system or whether they are just closely linked. They came to the conclusion that the inner loop is a link between phonological representations of the perception and the production system, although they could not exclude the possibility of shared representations. An account of shared representations would actually be able to explain monitoring behaviour quite well. Assuming that the monitor has access to a system of shared phonological information as well as to the conceptual system shared by production and perception would allow the monitor to draw on some information to which to compare the phonological plan, just as it seems to compare the generated message to the still memorized intent inside the conceptualizer (Levelt, 1983).

A production monitoring account presuming a monitoring device for every production process has been rejected because of its uneconomic multiple realization of information and because it would not be able to explain observed error-to-cutoff times. Assuming one production-based inner loop would still allow for a central flow-through monitor, explaining obtained latencies, even those a perceptual inner loop cannot account for. Adding the idea of being able to consciously switch to a perceptual inner loop would additionally explain the differences in experiments using explicit inner monitoring tasks (i.e. Özdemir et al., 2007; Roelofs et al., 2007; Wheeldon & Levelt, 1995) and those trying to simulate more natural speech production (Huettig & Hartsuiker, 2010; Slevc & Ferreira, 2006).

Intact prearticulatory monitoring in aphasic patients with impaired comprehension could also be due to the production-based inner loop still working, while comprehension-based monitoring fails. Impaired monitoring in patients with

intact comprehension would be more difficult to explain, but it has been argued above that these problems might not represent monitoring deficits, but rather a failure to produce the right input for the monitoring system.
Overall, an inner loop working through both production and perception according to the situation seems to explain existing data better than the respective one-sided accounts. The problem of a possible redundancy in information storage remains, but so far even authors arguing in favour of a perceptual monitor insist on a separate phonological representation for production and perception (Roelofs at al., 2007). How a double inner loop would be more redundant than this seems unclear.

There remain a number of open questions about the working of the monitor *per se*. How does the monitor compare for correctness, what does the monitor compare its input to, and what is the actual difference between parsed speech and an unparsed inner speech plan? Another aspect of interest for further research is the idea of two different kinds of inner speech, or the two different inner monitoring routes. Although the possibility has been mentioned by several studies (Huettig & Hartsuiker, 2010; Özdemir et al., 2007; Oppenheim & Dell, 2008), it remains yet to be seen whether empirical evidence for or against this suggestion can be obtained.

7 References

Baars, B. J., Motley, M. T. (1974). Spoonerisms: Experimental elicitation of human speech errors. *Journal Supplement Abstract Servise, Fall 1974. Catalog of Selected Documents in Psychology, 3,* 28-47.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psycological Review, 93,* 283-321

Goodglass, H, Kaplan, E. (1972). *The assessment of aphasia and related disorders.* Philadelphia: Lea & Febiger. As cited in: McNamara et al. (1992).

Hamm, S., Junglas, K., Bredenkamp, J. (2004). Die Zentrale Exekutive als präartikulatorische Kontrollinstanz. *Zeitschrift für Psychologie, 212 (2),* 66-75.

Hartsuiker, R. J., Kolk, H. H. J. (2001). Error Monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42,* 113-157.

Hartsuiker, R. J., Corley, M., Martensen, H. (2004). The lexical bias effect is modulated by context, but the standard monitoring account doesn’t fly: Related beply to Baars et al. (1975). *Journal of Memory and Language, 52,* 58-70.

Huettig, F., Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes, 25 (3),* 347-374.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41-104.

MacKay, D. G. (1992a). Awareness and error detection: new theories and research paradigms. *Consciousness and Cognition, 1*, 199-225. As cited in: Postma, 2000.

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. As cited in: Postma 2000.

Marshall, R. C., Rappaport, B. Z., Garcia-Bunuel, L. (1985). Self-Monitoring behaviour in a case of severe auditory agnosia with aphasia. *Brain and Langauge, 24,* 297-313.

Marshall, J, Robson, J., Pring, T., Chiat, S. (1998). Why does monitoring fail in jargon aphasia? Comprehension, judgement, and therapy evidence. *Brain and Language, 63,* 79-107.

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.

Nooteboom, S., Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language, 58,* 837-861.

Nozari, N., Dell, G. S., Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63,* 1-33.

Oomen, C. C. E., Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research, 30 (2),* 163-184.

Oomen, C. C. E., Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes, 17 (2),* 163-184.

Oppenheim, G. M., Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phoneme similarity effect. *Cognition, 106,* 528-537.

Özdemir, R., Roelofs, A., Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition, 105,* 457-465.

Postma, A., Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech, 39 (4),* 375-392.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition, 77,* 97-131.

Roelofs, A., Özdemir, R., Levelt, W. J. M. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology, Learning, Memory and Cognition, 33 (5),* 900-913.

Schlenck, K., Huber, W., Willmes, K. (1987). „Prepairs“ and repairs: Different monitoring functions in aphasic language production. *Brain and Language, 30,* 226-244.

Slevc, L. R., Ferreira, V. S. (2006). Halting in single word production: A test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language, 54,* 515-540.

Wheeldon, L. R., Levelt, W. J. M (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language, 34 (3),* 311-334.