Journal of Memory and Language 95 (2017) 19–35

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback

Sieb G. Nooteboom *, Hugo Quené

Utrecht University, Utrecht Institute of Linguistics OTS, Trans 10, 3512 JK Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 21 March 2016
revision received 22 December 2016

Keywords:
Speech errors
Self-monitoring
Repairs
Audition
Somatosensory

ABSTRACT

Two experiments are reported, eliciting segmental speech errors and self-repairs. Error frequencies, detection frequencies, error-to-cutoff times and cutoff-to-repair times were assessed with and without auditory feedback, for errors against four types of segmental oppositions. Main hypotheses are (a) prearticulatory and postarticulatory detection of errors is reflected in a bimodal distribution of error-to-cutoff times; (b) after postarticulatory error detection repairs need to be planned in a time-consuming way, but not after prearticulatory detection; (c) postarticulatory error detection depends on auditory feedback. Results confirm hypotheses (a) and (b) but not (c). Internal and external detection are temporally separated by some 500 ms on average, fast and slow repairs by some 700 ms. Error detection does not depend on audition. This seems self-evident for prearticulatory but not for postarticulatory error detection. Theoretical implications of these findings are discussed.

Introduction

The main questions

This paper is about self-monitoring for speech errors during speech production. We know that speakers often detect their own speech errors, because in spontaneous speech more than 50% of all speech errors against sound forms are repaired by the speaker (cf. Levelt, 1983; Levelt, 1989; Nooteboom, 1980; Nooteboom, 2005a). Also other types of speech errors are often repaired. This paper asks whether we can classify observed repairs into speech errors detected by self-monitoring before and after speech initiation, and if so, how we can distinguish between these two classes of repaired speech errors; whether there are two different processes for repairing a speech error, one leading to very fast and one leading to slow repairs; and to what extent the detection of speech errors by self-monitoring depends on auditory feedback.

Typical examples of repaired speech errors, taken from Blackmer and Mitton (1991), are the following:

"if Quebec can have a ba/ a Bill 101"

"behownd her/ behind her own closed doors"

The "/" in both cases indicates speech interruption, often followed by a silent interval. These two examples differ in an interesting way: In the first example the speech fragment containing the error "ba/ " is very short, and in many such cases fragments like these are shorter than a humanly possible reaction time. As pointed out by Levelt (1983) and Levelt (1989), who gave the example "v/ horizontal" in which the "v" is supposed to be the first speech sound of the word "vertical", this demonstrates that speech errors can be detected before speech initiation. However, the number of speech sounds spoken before interruption is not necessarily proof that the error was detected before speech initiation. We will call cases in which the error form is not fully spoken 'interruptions'.

In the second example we see that speech was only interrupted after both the word containing the error, "behownd" and the following word "her" were spoken. It is generally assumed that in such cases the speech error was detected by the speaker after speech initiation, via auditory perception of her or his own speech (Cf. Hartsuiker, Corley, & Martensen, 2005; Hartsuiker & Kolk, 2001; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; Hartsuiker, Kolk & Martensen, 2005; Nooteboom & Quené, 2008 and many others). Of course, a priori it is imaginable that also in this and similar cases the speech error was detected before speech initiation and the speaker just waited before interrupting the utterance, for example in order to gain time for planning a repair (cf. Seyfeddinipur, Kita, & Indefrey, 2008). We call such cases as "behownd her/ behind her own closed doors", in which the error form

* Corresponding author at: Cor Ruyslaan 20, 3584 GD Utrecht, The Netherlands.
E-mail address: S.G.Nooteboom@uu.nl (S.G. Nooteboom).

http://dx.doi.org/10.1016/j.jml.2017.01.007
0749-596X/

is fully spoken, 'completed', supposedly but not necessarily reflecting detection of speech errors in overt speech. Although we will initially distinguish between 'interrupted' and 'completed' spoken error forms, this initial classification will have to be replaced by another classification of repaired errors as probably detected in internal or in overt speech.

In this paper we focus on interactional segmental speech errors. Interactional errors are errors following from interaction between two different units in the speech program. Examples of interactional segmental errors include exchanges such as *Yew Nork* for *New York*, anticipations such as *Yew York* for *New York* and perseverations such as *New Nork* for *New York*. There are also segmental errors in which speech sounds are added or omitted under the influence of other speech sounds in the context. We will not consider additions and omissions, because we focus on experimentally elicited errors and we have not elicited additions and omissions. In this paper we will also not consider lexical, syntactic, semantic or appropriateness errors. There is no a priori reason to suppose that our results will also be valid for these other error categories. They probably are not, because temporal constraints on detecting and repairing segmental errors on the one hand and lexical, syntactic or semantic errors on the other hand appear to be rather different (cf. Nooteboom, 2005a). If indeed speakers can detect segmental speech errors both before and after speech initiation, this raises the question whether and how we can observationally distinguish between these two classes of repaired speech errors. This is the first question we will attempt to answer.

It has been shown, particularly with interrupted error forms, that frequently but not always, very short error-to-cutoff times are followed by very short cutoff-to-repair times, even of 0 ms (Blackmer & Mitton, 1991). Blackmer and Mitton concluded that in such cases a repair is available at the moment of speech interruption. This suggests that possibly there are two classes of repairs, distinguished by the moment the repair comes available to the speaker. If indeed this is the case, one may ask where this difference comes from. Thus the second question we will focus on is whether and how we can distinguish between fast and slow repairs, and if so, where this difference comes from. It seems reasonable to assume that there is an immediate connection with the detection of speech errors in internal versus overt speech. Later in this introduction we will explain why we think that after internal error detection it is often not necessary to plan a repair, whereas after external error detection often no repair is available, and a repair has to be planned in a time-consuming way.

Evidence in favour of the distinction between self-monitoring internal and self-monitoring overt speech is formed by demonstrations that the detection rate of speech errors by self-monitoring is affected negatively by loud masking noise (e.g. Lackner & Tuller, 1979; Oomen, Postma, & Kolk, 2001; Postma & Kolk, 1992; Postma & Noordanus, 1996). Because speech errors can be detected both before and after speech initiation, one would not expect that the error detection rate would drop to zero in the absence of auditory feedback. Effects of noise masking would be limited to error detection in overt speech, at least in as far as we assume that error detection in overt speech depends on hearing one's own voice, as is proposed by Levelt (1989) and Levelt et al. (1999). But some researchers believe that errors can be detected after speech initiation on the basis of somatosensory and / or proprioceptive feedback from the articulators (Hickok, 2012; Lackner, 1974; Pickering & Garrod, 2013). So far, the question to what extent error detection by self-monitoring overt speech depends on audition, remains unanswered. This is because we do not know which repaired speech errors are detected in internal and which in overt speech. If our attempt to distinguish between these two classes of repaired speech errors is successful, we can find out to what extent self-monitoring of overt speech depends on audition. This is the third main question we will try to answer in this paper.

The three main questions that we focus on in this paper are:

(1) Are speech errors detected by self-monitoring both before and after speech initiation, and if so, how can we distinguish between these two classes of detected speech errors?
(2) Can it be that there are two different processes for repairing a speech error, one leading to very fast and one leading to slow repairs?
(3) To what extent does the detection of speech errors by self-monitoring depend on auditory feedback?

Detection of speech errors before and after speech initiation

Theories of self-monitoring for speech errors are often classified as perception-based and production-based. For our purposes we consider so-called forward-modeling accounts of self-monitoring as a third category. The most influential theory of self-monitoring for speech errors is the perceptual loop theory proposed by Levelt (1989) and Levelt et al. (1999). In this theory both error detection in internal speech and error detection in overt speech employ the same speech comprehension system that is also employed in listening to other-produced speech. Internal speech is fed into the speech comprehension system directly, not following the route via articulation, acoustics and audition. It is assumed that errors made during the mental generation of speech, for example errors in phonological encoding, can be detected and repaired before speech initiation, leading to so-called covert repairs or 'prepairs' (cf. Postma and Kolk, 1992; Schlenk, Huber, & Wilmes, 1987). In this paper we will not consider covert repairs because we have no relevant observational evidence. Nevertheless errors detected in internal speech lend themselves to investigation because they are often articulated, leading to so-called early interruptions as in "if Quebec can have a ba/ a Bill 101". According to the perceptual loop theory errors can also be detected in overt speech, via audition and speech comprehension. For both the internal and external loop, the output of the speech comprehension system is fed into a centrally located monitor by which errors can be detected and repair planning initiated. Repair planning is supposed to start at speech interruption. It should be pointed out here that, although it may be convincingly argued that very short error-to-cutoff times necessarily correspond to speech errors detected in internal speech (simply because error-to-cutoff time is shorter than a humanly possible reaction time), this is not necessarily so for all interrupted error forms. The distinction 'interrupted' versus 'completed' does not necessarily correspond to the distinction 'internally' versus 'externally' detected. In this paper it will be attempted to find a way to tell at least statistically which repaired speech errors were detected in internal speech and which were detected in overt speech.

In production-based theories of self-monitoring it is assumed that there is some mechanism or mechanisms within the mental process of speech generation by which errors are detected. Examples are provided by Laver (1980; see also Schlenk et al., 1987), assuming special purpose editors within the speech generation system, and MacKay (1987), proposing that, because a speech error in some sense is a relatively new structure, it will cause prolonged activation of some node in the neural network generating speech; this prolonged activation will increase awareness and thereby lead to error detection. A different mechanism for error detection is proposed by Nozari, Dell, and Schwartz (2011). These authors made a model of error detection by conflict between simultaneously activated and competing units during speech coding. Interestingly, these production-based monitors are all directed at error detection in internal speech, before speech initiation. It is generally assumed

that speech errors can also be detected in overt speech, by auditory feedback.

This is different in forward modeling accounts of self-monitoring that explain self-monitoring for speech errors by assuming that corollary discharge signals representing the intended speech sounds can be compared with not only auditory but also somatosensory and/or proprioceptive feedback from the actual production of the intended sounds. This was first proposed by Lackner (1974). Interestingly, although Lackner assumes that auditory representation and perception can play a role in error detection, in his conception this is not necessary. Errors can also be detected during articulation by somatosensory and/or proprioceptive feedback from the articulators. Lackner does not distinguish between error detection in internal and overt speech. Hickok (2012) improves on this. His Hierarchical State Feedback Control model capitalizes on evidence that in planning motor tasks sensory target activity may be suppressed. This is equivalent to increasing the gain in non-targets, and may assist not only in preventing interference with preceding and following targets but also in detecting deviations from expectations, i.e. in detecting errors. The feedback loop enabling comparison between targets and execution is active from the early planning stage to the stage of actual motor activity. This would explain error detection both before and after speech initiation. These and similar ideas were further elaborated by Pickering and Garrod (2013).

All theories or models of self-monitoring for speech errors mentioned so far do not tell us how we could possibly distinguish observationally between repaired speech errors that are detected in internal speech and repaired speech errors that are detected in overt speech. However, we get a clue from a computational implementation by Hartsuiker and Kolk (2001) of the perceptual loop theory proposed by Levelt (1989) and Levelt et al. (1999). In this computational model assumptions of the perceptual loop theory concerning the timing of a number of stages in the production and perception of speech are combined with the proposal by Sternberg, Monsell, Knoll, and Wright (1978), Sternberg, Knoll, Monsell, and Wright (1988), Sternberg, Wright, Knoll, and Monsell (1980) that the processes of unit selection and command for articulation are serial. Also Hartsuiker and Kolk (2001) made an additional assumption, namely that planning a repair after error detection either in internal or in overt speech and executing a command to interrupt speech can be done in parallel. Both are supposed to start immediately after error detection. This replaces the assumption by Levelt (1989) and Levelt et al. (1999) that planning a repair starts at the moment of interruption and explains the observation by Blackmer and Mitton (1991) that a repair can be available at the moment of interruption. The model predictions with various parameter settings of the distributions of both error-to-cutoff times and cutoff-to-repair times were tested against actual such distributions reported by Oomen and Postma (2001). No good fit was obtained with assuming only internal or only external error detection. Both were necessary. The best fitting parameter setting of the model predicts that the delay between internal and external detection of speech errors is 350 ms.

From the Hartsuiker and Kolk model we infer that there may be a way to distinguish observationally between repaired errors detected in internal speech and repaired speech errors detected in overt speech. If it is correct that errors are detected at two stages of speech production, both before and after speech initiation, then we expect that underlying the actual distribution of error-to-cutoff times there are two distributions, one for internally and one for externally detected errors, the peaks of the two distributions being separated by some 350 ms. This is our first prediction:

(1) The distribution of error-to-cutoff times is bimodal, with the two peaks being separated by some 350 ms.

If this prediction is confirmed, we can possibly estimate the form of the underlying distributions and thus at least statistically distinguish between repaired errors detected in internal speech and repaired errors detected in overt speech.

Fast and slow repairs of detected speech errors

The computational model by Hartsuiker and Kolk (2001) does not only predict distributional aspects of error-to-cutoff times but also of cutoff-to-repair times. Cutoff-to-repair times can be very short, even 0 ms, not only for long error-to-cutoff times but also for very short error-to-cutoff times. This is explained in the Hartsuiker and Kolk model by the assumption that a repair can be planned during the 150 ms needed for executing an interruption command after error detection. Hundred and fifty ms is not much time for planning a repair. Therefore Hartsuiker and Kolk (2001) assume that the repair, i.e. the correct target form going to replace the error form, has been primed by the earlier phase of speech generation, when the error had not yet been made. Interestingly, within the model planning a repair has the same temporal properties for internal and external error detection. Apparently, it is assumed that priming of the correct target that is going to serve as a repair, is equally strong after error detection in internal and overt speech. From this one would predict that the delay of 350 ms of external error detection with respect to internal error detection, reflected in the error-to-cutoff times, carries over to the error-to-repair times (each error-to-repair time being the sum of the error-to-cutoff time and the cutoff-to-repair time). Therefore the model predicts that not only the error-to-cutoff times but also the error-to-repair times have a bimodal distribution with two peaks separated by 350 ms.

However, we see reasons to doubt the assumption that, apart from the 350 ms delay between internal and external detection, the temporal aspects of detection and repair are identical for the two classes of repaired speech errors. A first indication is the earlier mentioned assumption made by Hartsuiker and Kolk that the correct target that is going to serve as a repair, is primed by the earlier stages of speech preparation, and therefore not completely de-activated. The assumption is necessary because the assumed minimum time for speech interruption after internal error detection is only 150 ms. This is little time for planning a fully de-activated repair. But after external error detection the correct target form has, within the model, 350 ms more to be de-activated. So possibly, the degree of activation of a correct target form that is to serve as a repair is different between the two classes of repaired speech errors. This would cause a difference in the temporal aspects of repair planning between internal and external error detection. It may be even worse than this. There are a number of demonstrations that often interactional segmental speech errors are articulatory blends of two competing segments, an error segment and a correct target segment (Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; McMillan & Corley, 2010; Mowrey & MacKay, 1990). This suggests that when a segmental error is generated during phonological encoding, both the error form and the correct target form remain activated and in competition, generating an articulatory blend between the two forms. This is supported by our own observation that sometimes in segmental speech errors the onsets of an error form and a correct target form are rapidly alternating. Some Dutch examples are: *feit goud > gfgeitfout, tand veeg > tftantfeeg, bijl geit > gèbgbijlgeit, duit vast > dvduitvast, paf kies > puhkuhpfafkies*. These examples were taken from the speech errors elicited in two experiments described by Nooteboom and Quené (2008).

If occasionally, in the competition between error form and correct target form, a segmental error is generated, this error is planned to be spoken. However, the error can be rapidly detected in self-monitoring internal speech by comparing error form and the still available correct target form. Note that activation of the correct target form is sustained from the level of lexical selection, activation of the error form is not. Meanwhile a command to initiate speaking the error form has been issued, but following error detection a command to stop speech is also issued and speech is interrupted immediately after speech initiation. Thus the correct target form is immediately available as a repair, because it is sustained from the level of lexical selection. This would account for the observation by Blackmer and Mitton (1991) that there are many cases in which error-to-cutoff and cutoff-to-repair time are both very short. According to this view of the process of repairing segmental speech errors, after internal error detection even the 150 ms needed for execution of the interruption command would not be needed for planning a repair: No planning of a repair is necessary.

However, when a speech error is detected in overt speech, the correct target form that is going to serve as repair has, according to the Hartsuiker and Kolk model, 350 ms more time to be de-activated. Very likely much of the activation of the correct target form competing with the error form has fallen off at the moment the error is detected in overt speech. Therefore laborious replanning of the correct target form is necessary. In many cases this will take much longer than the 150 ms assumed in the Hartsuiker and Kolk model. Because in this view coming up with a repair after internal error detection is faster and coming up with a repair after external error detection is slower than predicted by the Hartsuiker and Kolk model, we expect that the difference between internally and externally detected errors in error-to-repair times is much greater than 350 ms. It has been pointed out to us that it has been shown in picture-naming experiments that even after 400 ms a word activated by an earlier picture is not fully de-activated because semantic and phonological properties still influence the naming latencies for the later picture (e.g. Hartsuiker, Pickering, & de Jong, 2005; Tydgat, Diependaele, Hartsuiker, & Pickering, 2012). Our main point, however, is that the degree of de-activation very likely is different between the two situations. Also, because of the considerable variation in error-to-cutoff times, often the time for de-activating the correct target form is much more than 400 ms.

These considerations lead us to expect that the difference between internally and externally detected errors in error-to-repair times is much greater than the difference between internally and externally detected errors in error-to-cutoff times. Thus our next prediction is:

(2) The difference between internally and externally detected speech errors is significantly greater in error-to-repair times than in error-to-cutoff times.

The role of auditory feedback in self-monitoring for speech errors

There are different opinions on how much self-monitoring for speech errors depends on auditory feedback. According to the perceptual loop model (Levelt, 1989; Levelt et al., 1999), although detection of speech errors in internal speech employs the same speech comprehension system as detection of speech errors in overt speech, audition is not involved. This stands to reason: Before speech initiation there is nothing to be heard. But as we have seen, most proponents of production-based self-monitoring for speech errors yet assume that detection of speech errors in overt speech depends on audition. The exception is formed by the proponents of forward modeling in speech production (Hickok, 2012; Lackner, 1974; Pickering & Garrod, 2013). These propose that speech errors can be detected after speech initiation not only from auditory but also from somatosensory and/or proprioceptive feedback from the articulators.

Evidence for the role of auditory feedback in self-monitoring mainly stems from experiments eliciting speech errors with and without loud masking noise: If the detection rate of speech errors suffers under loud masking noise this is taken as evidence for the importance of audition. Experiments of this nature have been reported by Postma and Kolk (1992), Postma and Noordanus (1996) and Oomen et al. (2001). Postma and Kolk (1992) found that loud masking noise among other things reduced the numbers of disfluencies and self-repairs. This points at the relevance of auditory feedback for self-monitoring. Postma and Noordanus (1996) asked their speakers to report their own errors during speeded production of tongue twisters under four different conditions, viz. silent, mouthed, noise-masked and normal auditory feedback. Errors were reported by pushing a button and describing the error each time an error was detected. They found that error detection rates were roughly equal in the first three conditions and higher with normal auditory feedback. Apparently, auditory feedback increases the detection rate. Oomen et al. (2001) compared detection rates by self-monitoring for speech errors in patients with Broca's aphasia and healthy controls, with and without loud masking noise. They found that patients with Broca's aphasia and healthy controls had comparable detection rates under loud masking noise, but that the patients with Broca's aphasia detected fewer errors than healthy controls under normal auditory feedback. They concluded that the patients with Broca's aphasia relied more than healthy controls on prearticulatory self-monitoring. The implication is that postarticulatory self-monitoring in healthy controls depends at least partly on audition. This agrees with Huettig and Hartsuiker (2010) who registered eye-movements while speakers named objects accompanied by phonologically related and unrelated written words. They found that these eye movements were driven by the perception of the speaker's own overt speech, not by inner speech. They concluded that self-monitoring of overt speech, but not of internal speech, is based on speech perception. Lind, Hall, Breidegard, Balkenius, and Johansson (2014), Lind, Hall, Breidegard, Balkenius, and Johansson (2015) demonstrated that speakers can react very rapidly to "speech errors" that, during a word production experiment, were inserted sneakily in their overt speech and that were accepted (or at least repaired) by the speakers as their own speech errors. They interpreted this as showing that error detection is mainly based on hearing one's own voice and that the assumption of error detection in internal speech is superfluous.

Lackner and Tuller (1979) reported an interesting experiment. They elicited segmental speech errors in sequences of four meaningless syllables, both with and without strong masking noise. Speakers had to report errors by pushing a button each time an error was detected. It was found that noise affected the detection of errors against the voiced-unvoiced distinction and against vowels but not of errors against place of articulation. This finding supports the proposal by Lackner (1974) that segmental errors of speech are often detected on the basis of somatosensory and/or proprioceptive feedback. That detection of errors against the voiced-unvoiced distinction and against vowels is affected by loud masking noise is explained by the observation that differences in articulator positions and in contact between articulators are much less conspicuous in the voiced-unvoiced and vowel oppositions than in the place of articulation opposition. Lackner and Tuller (1979) did not distinguish between internal and external detection of speech errors.

The available evidence leaves little doubt that audition can be involved in the detection of speech errors by self-monitoring overt

speech. However, it is difficult to know to what extent audition is important. This is so because researchers in the investigations mentioned so far had no way to know which repaired speech errors were detected in internal speech and which were detected in overt speech. If our attempt to separate between these two classes of repaired speech errors is successful, perhaps we can improve on this. The evidence provided by Lackner and Tuller (1979) that loud masking noise has no effect whatsoever on the detection of errors against place of articulation suggests that the role of audition in the external detection of errors at best is limited. But as they did not attempt to distinguish between internal and external detection, there remains a possibility that their speakers were mainly concentrating on prearticulatory error detection. In that case those who believe that self-monitoring overt speech depends on audition can still be right. If so, we have a rather strong prediction:

(3) If we elicit segmental speech errors and self-repairs both without and with loud masking noise, we will find that the rate of internal detection is the same for both conditions, but that the detection rate drops to virtually zero for detection in overt speech under loud masking noise but not in silence.

In order to test these various hypotheses, we ran two experiments eliciting interactional segmental speech errors and repairs on either the initial consonants or the vowels in CVC CVC utterances. In experiment 1 we elicited segmental interactions between consonants differing in place and/or manner of articulation. In experiment 2 we elicited interactions between consonants differing in place and/or manner articulation and also between consonants differing in the voiced-unvoiced distinction and between vowels.

Experiment 1

This experiment follows a classical SLIP technique (Spoonerisms of Laboratory-Induced Predisposition, cf. Baars & Motley, 1974; Baars, Motley, & MacKay, 1975). This works as follows: Speakers are successively presented visually, for example on a computer screen, with priming word pairs such as DOVE BALL, DEER BACK, DIM BOMB, followed by a target word pair BIN DOG, all word pairs to be read silently. On a prompt, for example a buzz sound or a series of question marks (“??????”), the last word pair seen, i.e. the target word pair, in this example BIN DOG, has to be spoken aloud. Interstimulus intervals are in the order of 1000 ms, as is the interval between the test word pair and the prompt to speak. Every now and then the speaker will mispronounce the target word pair BIN DOG as DIN BOG, as a result of phonological priming by the preceding word pairs. For the current purpose the technique has the advantage that speakers during or after each response to a stimulus have an occasion to make a repair. Such repairs resemble in their temporal course repairs of speech errors in spontaneous speech (cf. Nooteboom & Quené, 2008). A disadvantage of the technique is that, due to the necessary temporal pressure on the speakers in this task, there are relatively few external error detections. Therefore relatively many speakers are needed to achieve enough statistical power.

For this experiment we have the following predictions:

(1) Error-to-cutoff times are distributed bimodally with two peaks separated by some 350 ms.
(2) The difference between internally and externally repaired errors is considerably larger in error-to-repair times than in error-to-cutoff times.
(3) Detection rate in silence and under loud masking noise is equal for all errors detected in internal speech but drops to virtually zero for errors detected in overt speech under loud masking noise but not in silence.

Method

Speakers

There were 106 participating speakers, all students of Utrecht University varying in age from 17 to an exceptional 42 years. Average age was 23 years. Of these 106 speakers 85 were female and 21 were male. All speakers were native speakers of Dutch. No speaker had a speaking, hearing or not-corrected vision problem. Each speaker was paid € 5 for participation.

Materials

Two lists of stimulus items were prepared. Each stimulus item consisted of two Dutch CVC forms. In each list there were 32 test stimuli and 23 filler stimuli. For each test stimulus the targeted spoonerism was also used as a test stimulus (for example both *kaf piep* and *paf kiep* were test stimuli). Of these 32 test stimuli, 16 targeted interactions between consonants with only 1 feature difference (place or manner of articulation; similar) and 16 targeted interactions between consonants with 2 features difference (place and manner of articulation; dissimilar). Each test stimulus had 5 precursor word pairs to be read silently. Of these 5 precursor word pairs the last 3 were priming an exchange between the two initial consonants of the stimulus word pair. Each test stimulus was followed by a prompt to speak the last word pair seen. This prompt consisted of a sequence of 6 question marks. Table 2.1 gives a typical example of a test stimulus item together with its precursor word pairs, prompt and targeted spoonerism.

The test stimuli in list 2 were derived from those in list 1 by changing the two final consonants in the CVC forms. For example, *paf kiep,* eliciting the exchange *kaf piep* turned into *pap kier*, eliciting the exchange *kap pier*. This means that each related pair of test stimuli in List 1 (e.g. *paf kiep* and *kaf piep*) had a corresponding pair of related test stimuli in List 2 (e.g. *pap kier* and *kap pier*). The same precursors were used in both stimulus lists for these corresponding test items. Together these 4 stimuli constitute a set of 4 items (henceforth a stimulus item set) related across the two stimulus lists.

In addition to the 32 test stimuli in each list there also were 23 filler stimuli, 2 with 4 precursors, 2 with 3 precursors, 6 with 2 precursors, 4 with 1 precursor, and 9 with no precursor. The filler stimuli were intended to make the arrival of the question mark prompt unpredictable. The filler stimuli were identical in both stimulus lists. Each list was preceded by the same 7 practice items with varying numbers of non-priming precursors. The two lists of (test and filler) stimuli are given in Appendix A.

The masking noise used in the experiment (see procedure) was computer-generated so-called “pink noise” of 87 dB SPL(A) as mea-

Table 2.1
Example of a test stimulus item together with its precursor word pairs, the prompt for speaking the last word pair seen (see procedure) and the targeted spoonerism.

Precursor 1	bouw jool
Precursor 2	lijf deed
Precursor 3	koet pop
Precursor 4	kuur poet
Precursor 5	kas piet
Test stimulus	paf kiep
Prompt	??????
Targeted spoonerism	*kaf piep*

sured by a dB meter (Bruel & Kjaer 2230) within the shells of the earphones, i.e. noise with a power spectrum that decreases 3 dB per octave from low to high frequencies. In the actual noise the decrease with 3 dB per octave was applied between c. 200 and 18,000 Hz. Below 200 and above 18,000 Hz the intensity rolled off. This power spectrum is better suited than the white noise of 90 dB SPL used for example by Postma and Kolk (1992) for auditorily masking speech, particularly in the spectral region that is relevant for perceiving speech. This is so because more of the power is concentrated in the lower spectral regions, that are most relevant for speech intelligibility.

Procedure

Each speaker was tested individually in a sound-treated booth. The timing of visual presentation on a computer screen was computer controlled. Test and filler stimuli, each along with their priming or non-priming precursor word pairs, were presented in a random order that was different for each speaker, but that was the same, in terms of corresponding stimuli, for the two sessions of each speaker. All odd-numbered speakers were presented in the first session, with list 1 with auditory masking and then, in the second session, with list 2 without auditory masking. All even-numbered speakers were presented in the first session with list 1 without auditory masking, and then, in the second session, with list 2 with auditory masking. Precursor word pairs and target word pairs (filler or test) were presented consecutively, each word pair being presented for 900 ms with blank intervals of 100 ms in between. After the final word pair of each trial a "??????"-prompt, meant to elicit pronunciation of the last word pair seen (the test or filler stimulus containing the target word pair), was visible during 900 ms and was then immediately followed by a blank screen of 100 ms duration. The blank screen following the "??????" prompt was immediately followed by a cue consisting of the Dutch word for "correction", visible during 900 ms and again followed by a blank screen with 100 ms duration. Speakers were instructed to pronounce the last word pair seen before the "??????" prompt as rapidly as possible. They were strongly encouraged to speak as softly as possible without whispering, and, during the session with auditory masking with loud noise, to speak so softly that they could not hear their own voice. This was practiced during the practice items. The speakers were instructed to correct themselves immediately whenever they made an error. It was not necessary to wait for the "correction" prompt. After the correction period and a 100 ms resetting period, the first word pair of the following trial sequence was presented. All speech of each speaker was recorded with a Sennheiser ME 50 microphone in Audacity, and was digitally stored on disc in one of two tracks of a stereo file with a sampling frequency of 48,000 Hz. The resulting speech was virtually always clear, although mostly not very loud. For each speaker the experiment lasted roughly 20 min for the two sessions together.

Scoring the data

Responses to all test and filler items were transcribed either in orthography, or, where necessary, in phonetic transcription by the first author using the PRAAT computer program (Boersma & Weenink, 2016).

Responses were categorized as (i):

0. 'Fluent and correct responses' of the type BARN DOOR > BARN DOOR or BAD GAME > BAD GAME.
1. 'Completed spoonerisms' of the type BARN DOOR > DARN BORE or BAD GAME > GAD BAME.
2. 'Anticipations' of the type BARN DOOR > DARN DOOR.
3. 'Interrupted spoonerisms' of the type BARN DOOR > D. . .BARN DOOR. There were very few interruptions after the first vowel of the elicited spoonerisms (cf. Nooteboom, 2005b). All interruptions were included.
4. 'perseverations' of the type BARN DOOR > BARN BORE.
5. 'other errors' for example BARN DOOR > PARK DOOR; BARN DOOR > GIVE MAN; or BARN DOOR > BASK DOOM.
6. 'hesitations', such as BARN DOOR > . . ..BARN . . ..DOOR?
7. 'omissions', BARN DOOR > BARN; BARN DOOR>. . .DOOR; BARN DOOR>. . .. .

ii. When in the same response more than one error was made, for example an exchange between initial consonants being accompanied by a substitution of a vowel or a perseveration of a final consonant, or an anticipation in the initial consonant of the first word of the initial consonant of the second word accompanied by a lexical replacement of the second word by a completely different word, these errors accompanied by additional errors were coded in a separate column of an excel sheet as Add = 1.

In addition, responses were coded (iii) as valid or invalid responses. In this study valid responses were fluent and correct responses and all elicited interrupted or completed exchanges of the two initial consonants and anticipations of the initial consonant, without any further additional error. These errors were called valid errors. All other responses were considered invalid. The errors with additional errors were excluded because in these cases there was no way to know, if an error was detected, which error has triggered the detection. This implies that also single errors being perseverations of the initial consonant of the first word in the initial consonant of the second word were excluded. This was done because we wished to ensure that the timing of error detection and repair was always relative to the initial consonant of the first CVC word. Also the few cases of interrupted errors that were not repaired were considered invalid. Responses were also coded (iv) as being (a) repaired or not repaired after completion, or (b) repaired after being interrupted (i.e. interrupted any place before completion), (v) as to the duration in ms of the spoken response, i.e. error-to-cutoff time, (vi) as to the duration in ms of the cutoff-to-repair time (of course, only for repaired speech errors).

The voice lead of initial voiced plosives was not counted as belonging to the consonant duration. This was done because the duration of the voice lead is in Dutch extremely variable, between 0 ms and many hundreds of ms, and is perceptually not very functional (Van Alphen, 2004). However, because in plosives there is no sound during the closure whereas in fricatives there is sound during the closure, we have compensated for this difference by adding 52 ms to each duration of an error-to-cutoff time that started with a plosive and subtracting 52 ms for each cutoff-to-repair time for repairs starting with a plosive (see Appendix B for the argumentation resulting in the correction value of 52 ms for adjusting durations of durations of to-be-repaired spoken response starting with a plosive sound and cutoff-to-repair times in the case of repairs starting with a plosive sound).

Results of Experiment 1

Error rates and detection rates

In this experiment two times 55 stimuli (viz. 32 test stimuli and 23 filler stimuli) were presented to 106 participants. We thus obtained 11,660 responses, i.e. 6784 responses to test stimuli and 4876 responses to filler stimuli. Of the 6784 responses to test stimuli 5805 were fluent and correct and 979 contained one or more errors. A first breakdown of these responses to test stimuli is given in Table 2.2.

For our purpose we concentrate on single errors, because with multiple errors there is no way to know which of the errors triggered detection. Table 2.3 contains a classification of the types of

Table 2.2
First classification of response types, with examples

Response type	n	Example
Fluent and correct	5805	zoet veen > zoet veen
Multiple error	216	keus por > peul por
Single error	763	baan zoom > zaan boom
Total	6784	

Table 2.3
Numbers of single errors of different types.

Type of single error	n	Example
Exchange	261	boos del > doos bel
Interruption	94	tol veer > v..tol veer
Anticipation	36	duik bof > buik bof
Perseveration	18	duik bof > duik dof
Other single error	277	bak zoon > bok zoon
Hesitation	23	voet zeen > voet sssseenn.. zeen
Omission	54	bijt geen > ……; bijt geen > bijt……
Total	763	

single errors. Note than an exchange is considered as a single error in the first position.

We also focus on elicited exchanges, interruptions and anticipations in the first consonant of the first CVC word, because in this way we always know that detection by self-monitoring is triggered from the same position. So for our purpose there are 391 valid errors. It may be observed that the relative numbers of exchanges, interruptions and anticipations may differ from those found in other experiments, because these relative numbers appear to be rather sensitive to the particular oppositions involved in the elicited errors.

Error-to-cutoff times

We will now turn to testing our first prediction (prediction 1), viz. that error-to-cutoff times show a bimodal distribution, the peaks of the two underlying distributions being separated by at least 350 ms. There are 94 interrupted repaired errors and 24 completed repaired errors. We collapse these two categories of repaired errors to see how the error-to-cutoff times are distributed. The error-to-cutoff time is defined as the interval between the word onset of the error form, that in all cases began with the erroneous segment, and the moment speech stopped.

The error-to-cutoff interval times range from 49 ms to 1057 ms, with 20 cases shorter than 100 ms. The log-transformed error-to-cutoff interval times were analyzed in R (R Development Core Team, 2016) by means of uninformed mixture modeling (Fraley & Raftery, 2002; Fraley, Raftery, Murphy, & Scrucca, 2012), a family of clustering techniques that can be used to analyze an observed distribution into a mixture of multiple gaussian distributions, each having its own mean and standard deviation. As predicted, the optimal solution shows a mixture of two gaussians, illustrated in Fig. 2.1, with one peak corresponding to 139 ms (4.934, s = 0.228, 85 cases) and a second peak corresponding to 637 ms (6.456, s = 0.073, 33 cases; log-likelihood −123.9). This confirms our prediction that error-to-cutoff times are distributed bimodally with two peaks being separated by at least 350 ms. The separation is in fact close to 500 ms.

The bimodal mixture model was validated by means of two-stage bootstrapping over 200 replications (first over participants contributing error-to-cutoff observations, then over observations contributed by these bootstrapped participants; cf. Efron & Tibshirami, 1993; Nooteboom & Quené, 2008). Of the 200 bootstrap replications, only 1 resulted in a mixture model with one gaussian component (i.e. unimodal), the majority of 105 bootstrapped models had two gaussian components (i.e. bimodal),

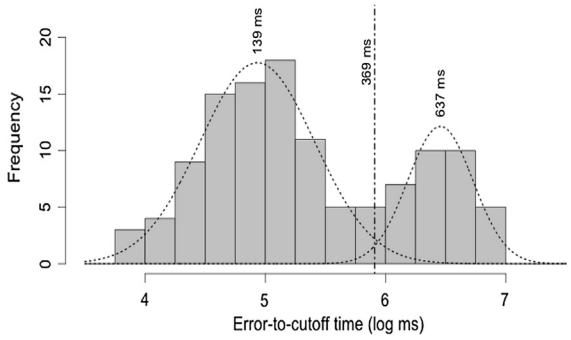

Fig. 2.1. Histogram of log-transformed durations of error-to-cutoff intervals, for N = 118 repaired errors. Distributions plotted with dotted lines indicate the estimated distributions from an uninformed gaussian mixture model (see text). The vertical dashed line indicates the interpolated boundary value (5.91 corresponding to 369 ms) between the two distributions.

and 94 had three or more components (i.e. multimodal). Over the 200 bootstrap replications, the median number of gaussian components is 2, with a 95% confidence interval of (2,8). The one single-component mixture model in the bootstrap validation involved 32 participants in this single component (which were bootstrapped from 62 contributing participants), the two-component models typically involved 13 participants for the smaller component, and the three-or-more mixture models in the bootstrap validation typically involved only 3 participants for their smallest component. Thus the three-or-more mixture models may have been overfitted to individual participants (over the 200 bootstrap replications, the number of components was indeed negatively correlated to the number of participants in the smallest component; $r_s = -0.86$, $p < 0.001$). In sum, the bootstrap validation clearly supports the two-gaussian mixture model.

This suggests that there are two distributions of error-to-cutoff intervals underlying the actual distribution, and that these two distributions are separated by c. 500 ms. Although the two presumed gaussian distributions overlap, the boundary value between the two distributions falls at 369 ms [exp(5.91)] according to the two-gaussian mixture model summarized above (this is approximately where the two gaussians intersect in Fig. 2.1).

In what follows we assume that in this experiment error-to-cutoff times shorter than 369 ms reflect cases where the error was detected internally, before speech initiation, and error-to-cutoff times longer than 368 ms reflect cases where the error was detected externally, after speech initiation. This boundary value probably is specific for this experiment. It cannot be generalized, because it depends heavily on the estimated form and the position on the time axis of the two underlying distributions. Clearly, the separation between 'internally' and 'externally' detected errors is only statistical, not absolute. To indicate this, we will from now on use single quotes for the terms 'internal' and 'external' when these refer to the two classes of repaired speech errors derived from the above or a similar analysis.

Error-to-repair times

The error-to-repair time is defined as the sum of the error-to-cutoff and the cutoff-to-repair times. The error-to-repair time is interesting because notably for 'externally' detected errors it provides an indication how much time is needed to plan a repair after error detection. For 'internally' detected errors this is not so easy because we have no observational evidence at what moment an error is made and detected. As we have seen in the introduction, in the computational model by Hartsuiker and Kolk (2001) all the timing leading to error detection and thereafter to a spoken repair is identical for 'internal' and 'external' error detection. The

only difference is the moment that the timing starts. This difference is, according to the model, 350 ms. Because there is no further difference involved in the timing of planning a repair, the difference of 350 ms of necessity carries over to the whole interval between error and repair. We have seen that in our data the difference between 'internally' and 'externally' detected errors in error-to-cutoff times is some 500 ms. If Hartsuiker and Kolk are correct in assuming that the timing for planning a repair is basically the same for errors detected 'internally' and 'externally', we expect that also the error-to-repair times show a difference of some 500 ms.

The optimal solution suggests that there is a clear separation between two gaussian distributions of repaired errors, one gaussian having a peak at 253 ms (5.531, s = 0.337, 94 cases) and another gaussian having a peak at about 970 ms (6.878, s = 0.037, 24 cases). In a sense the bimodality in Fig. 2.2 is trivial, because we know already that error-to-cutoff times, which constitute one of the components of error-to-repair times, are distributed bimodally. Of interest here is the wider separation between the peaks of the two gaussians, suggesting that there are two different mechanisms for repairing speech errors, one leading to fast repairs and one leading to slow repairs. The two peaks in the bimodal distribution differ by more than 700 ms. This suggests that the difference between 'internally' and 'externally' detected speech errors in error-to-repair times is greater than the difference in error-to-cutoff times, which we found to be some 500 ms.

In order to see whether indeed the difference between 'internally' and 'externally' detected speech errors is significantly greater in error-to-repair times than in error-to-cutoff times, we again interpreted all repaired speech errors with error-to-cutoff times shorter than 369 ms as detected 'internally' and all repaired speech errors with error-to-cutoff times longer than 368 ms as detected 'externally'. The difference of 136 ms between the averages of the error-to-repair times (430 ms) and of the error-to-cutoff times (294 ms) was removed first, by adding this difference to all error-to-cutoff values. This gave us two sets of values, one for normalized error-to-cutoff times and one for error-to-repair times with exactly the same mean. We then took the natural logarithm of all values in each set, in order to get less skewed distributions. For each of the two dependent variables the difference between 'internally' and 'externally' detected repaired errors was analyzed with a Welch *t* test for two samples. For normalized and log-transformed error-to-cutoff times the mean difference between 'internally' and 'externally' detected repaired errors is 1.028 with (0.928, 1.109) as 95% confidence interval. After backtransformation this difference corresponds to approximately 500 ms. For log-transformed error-to-repair times the mean difference between 'internally' and 'externally' detected repaired errors is 1.280 with (1.119, 1.440) as 95% confidence interval. After backtransformation this difference corresponds to approximately 600 ms. As the 95% confidence intervals do not overlap, the difference between error-to-cutoff times and error-to-repair times in the temporal gap between 'internally' and 'externally' detected repaired errors is significant with $p < 0.05$. This supports our prediction (2) that the difference between 'internally' and 'externally' detected errors is significantly greater in error-to-repair times than in error-to-cutoff times.

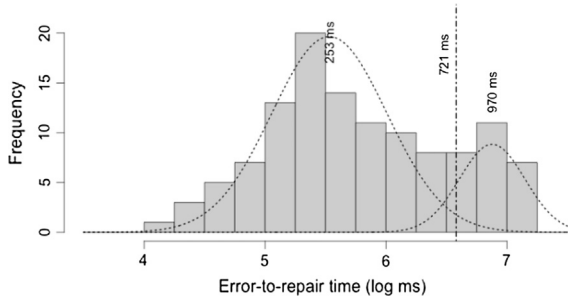

Fig. 2.2. Histogram of log-transformed error-to-repair times, for N = 118 repaired errors. Distributions plotted with dotted lines indicate the estimated distributions from an uninformed gaussian mixture model (see text). The vertical dashed line indicates the interpolated boundary value (6.58 corresponding to 721 ms) between the two distributions.

The role of auditory feedback in self-monitoring

Table 2.4 gives a breakdown of the 6196 valid responses (fluent and correct, and exchanges, interruptions and anticipations in the first consonants of the CVC CVC utterance).

Table 2.5 reports the numbers of valid errors broken down by no noise versus noise, similarity, undetected versus interrupted versus completed, and detected 'internally' versus detected 'externally'. Note that interrupted versus completed is replaced by detected 'internally' versus 'externally'; only the latter contribute to the totals. We refrain from analyzing the numbers of interrupted versus completed, because the numbers of repaired interruptions versus repaired completed are replaced with estimated numbers of 'internally' detected repaired errors and 'externally' detected repaired errors, to test the effects of noise and similarity on these categories of repaired errors.

For this experiment we have predicted (prediction 3) that the rate of 'internal' detection is equal in silence and under loud masking noise, but that the rate of 'external' detection but not of 'internal' detection drops to virtually zero under loud masking noise.

Detected errors with error-to-cutoff times shorter than 369 ms were classified as 'internal' detections, and detected errors with error-to-cutoff times of 369 ms or longer were classified as 'external' detections. The odds of 'internal' and 'external' detections were analyzed by means of a single multinomial logistic regression, validated by subsequent bootstrapping (over speakers, with 200 replications).The optimal model according to Likelihood Ratio Tests has only the similarity factor as a predictor; neither noise (p = 0.4652) nor the interaction of noise and opposition type (p = 0.2405) improved the multinomial model significantly. As one would expect, relative to errors involving similar consonants (place or manner of articulation), errors involving dissimilar consonants (place plus manner of articulation) have far higher odds of being detected 'internally' [β = +1.238, bootstrapped 95% C.I. (0.993, 1.614)] and also somewhat higher odds of being detected 'externally' [beta = +0.640, bootstrapped 95% C.I. (−0.081, 1.001)]. Contrary to expectation, the Likelihood Ratio tests indicate that for all errors, involving both similar and dissimilar consonants, the detection rate, either 'internally' or 'externally', is not affected by noise.

Discussion

The results of Experiment 1 show that error-to-cutoff times are distributed bimodally, with two peaks roughly 500 ms apart. This is somewhat more than the 350 ms predicted from the computational model by Hartsuiker and Kolk (2001). We interpret the bimodal distribution of error-to-cutoff times as meaning that there are two classes of repaired errors, those detected in 'internal' speech and those detected in 'external' or overt speech and that the delay of error detection in 'external' speech with respect to error detection in 'internal' speech is roughly 500 ms. Of course the number of observations on which our value of 500 ms is based is, particularly for the class of 'external' detections, not impressive. More data are needed.

From the Hartsuiker and Kolk model we would have predicted that the temporal separation between 'internally' and 'externally' detected errors in error-to-repair times is the same as the temporal

Table 2.4
Numbers of correct responses and valid errors against place and/or manner of articulation, separately for two noise conditions and for similar (place or manner of articulation) and dissimilar (place and manner of articulation) errors. For each row in the table the number of stimuli was 1696. Percentages are given between brackets.

Noise condition	Similarity	Correct	Valid errors	Total
No noise	Similar	1435 (93%)	108 (7%)	1543
No noise	Dissimilar	1497 (95%)	81 (5%)	1578
Noise	Similar	1409 (93%)	112 (8%)	1521
Noise	Dissimilar	1464 (94%)	90 (6%)	1554
Total		5805 (94%)	391 (6%)	6196

Table 2.5
Numbers of valid errors (i.e. exchanges, interruptions and anticipations in the initial consonants of the CVC CVC stimuli) separately for the two noise conditions, similar versus dissimilar and for undetected versus repaired interrupted versus repaired completed, and detected 'internally' versus detected 'externally'. Percentages are given between brackets. Italic values do not contribute to the totals.

Noise condition	Similarity	Undetected	*Repaired interruptions*	*Repaired completed*	Detected internally	Detected 'externally'	Total
No noise	Similar	82 (76%)	18 (17%)	8 (7%)	15 (14%)	11 (10%)	108
No noise	Dissimilar	0.50 (62%)	27 (33%)	4 (5%)	23 (28%)	8 (10%)	81
Noise	Similar	93 (83%)	14 (13%)	5 (4%)	14 (13%)	5 (4%)	112
Noise	Dissimilar	48 (53%)	35 (39%)	7 (8%)	33 (37%)	9 (10%)	90
Total		273 (70%)	94 (24%)	24 (6%)	85 (22%)	33 (8%)	391

separation in error-to-cutoff times. From our view that in the case of 'internal' error detection in general no planning of a repair is necessary, and that in the case of 'external' error detection planning a repair is relatively time consuming, we predicted that the temporal separation between 'internally' and 'externally' detected errors is greater for error-to-repair times than for error-to-cutoff times (our prediction 4). This is what we found. Obviously, planning a repair is a time-consuming affair: For 'externally' detected errors error-to-repair times run from 434 ms to 1373, with an average of 896 ms. If we would assume such a long time needed for planning a repair after 'internal' error detection, this could in no way account for very short cutoff-to-repair times accompanying very short error-to-cutoff times.

With respect to the role of auditory feedback in self-monitoring, we have seen that, where predictably errors involving dissimilar consonants have a higher detection rate than errors involving similar consonants, loud masking noise seems to have no effect on either the 'internal' or the 'external' detection rate of errors against place and/or manner of articulation. Also there is no effect of interaction between noise and similarity on detection rates. This runs contrary to our prediction that loud masking noise would virtually obliterate 'external' error detection. The absence of an effect of masking noise on error detection in overt speech supports the proposal by Lackner and Tuller (1979), that these errors are detected after speech initiation on the basis of somatosensory and/or proprioceptive feedback from the articulators. However, if Lackner and Tuller (1979) are right, we would predict that the detection of errors against the voiced-unvoiced distinction and against vowels would indeed be affected by loud masking noise. This will be tested in Experiment 2.

Experiment 2

Experiment 2 was set up first of all to see whether the results of Experiment 1 could be replicated. Secondly, we wanted to test the hypothesis that, where apparently in self-monitoring detection of segmental errors against place and/or manner of articulation does not depend on auditory feedback (see Experiment 1), detection of segmental errors against the voiced-unvoiced distinction and against vowels, does depend on auditory feedback. Therefore in Experiment 2, over and above the three hypotheses tested in Experiment 1, we have an additional hypothesis:

(4) The detection rate of 'externally' detected errors against the voiced-voiceless distinction and against vowels is higher in silence than under loud masking noise.

Method

The experimental technique employed in Experiment 2 was the same as in Experiment 1, with some minor differences explained below.

Speakers

There were 124 participating speakers, 103 females and 21 males, all students of Utrecht University, varying in age between 17 and an exceptional 51 years. Mean age was 23 years. Each speaker was paid € 5.00 for participation. We have tried to attract only speakers that had not participated in Experiment 1, but did not succeed. Fifty of the 124 speakers have participated in both experiments. We do not consider this a serious problem, because (a) Experiment 2 took place 8 months later than Experiment 1 and (b) most speakers were not aware of the purpose of the experiment.

Stimuli

Again we constructed two stimulus lists. In each list there were 32 CVC CVC test stimuli eliciting interactions between initial consonants differing in place and/or manner of articulation, 16 differing in one feature, 16 differing in two features. These 32 pairs of consonants targeted for interaction corresponded one-to-one to the interacting pairs of consonants used in Experiment 1. To the stimuli eliciting errors against place or manner of articulation we refer as opposition PM1. To the stimuli eliciting errors against place and manner of articulation we refer as opposition PM2. Each list also contained 16 CVC CVC test stimuli eliciting interactions between initial plosive consonants that only differed in the voiced-unvoiced distinction (VUV), and 16 CVC CVC test stimuli meant to elicit interactions between vowels (VOWEL). As in experiment 1, each test stimulus was preceded by 5 precursors the last three of which were chosen to elicit the desired interaction (see Table 2.1 in the description of Experiment 1). All test stimuli in List 2 were derived from the test stimuli in List 1, in most cases by changing the 2 final consonants and in a few cases by also changing the vowels, again with the 4 related stimuli across the two lists constituting a stimulus item set (See Method section Experiment 1). There were this time not 23 but 46 filler stimuli, 4 with 4, 4 with

3, 12 with 2, 8 with 1 and 18 with 0 precursors. The total number of stimuli per stimulus list was 110 stimuli, viz. 64 test stimuli and 46 filler stimuli. Filler stimuli were identical in the 2 lists. All CVC CVC test and filler stimuli in the two lists are given in Appendix A.

The masking noise in the experiment was computer generated. We used so-called "brown noise" of 87 dB SPL(A) as measured by a dB meter (Bruel & Kjaer 2230) within the shells of the earphones, i.e. noise with a power spectrum that decreases 6 dB per octave from low to high frequencies. The reason is that in Experiment 1 some speakers reported that they could hear some remnant of their own voice in the noise condition. In order to avoid distortion in the very low frequencies by the limitations of the headphones, the noise was also high-pass filtered with a cutoff frequency of 25 Hz (48 dB/octave roll-off). This power spectrum is better suited than the white noise of 90 dB SPL used for example by Postma and Kolk (1992) for auditorily masking speech, particularly in the spectral region that is relevant for perceiving speech. This is so because more of the power is concentrated in the lower spectral regions, that are most relevant for speech intelligibility.

Procedure

The procedure was the same as in Experiment 1. The only difference is that this time the masking noise was not recorded, and on a separate track a brief tone (1 kHz and 50 ms) was recorded with each stimulus, starting at the offset of the visual presentation of the "??????"-prompt. The experiment lasted c. 30 min for each speaker.

Scoring the data

Scoring the data was the same as in Experiment 1.

Results of Experiment 2

Error rates and detection rates

In this experiment two lists containing 46 filler stimuli and 64 test stimuli were presented to 124 participants, potentially giving $2 \times 124 \times 46 = 11,408$ responses to filler stimuli and $2 \times 124 \times 64 = 15,872$ responses to test stimuli. Due to technical problems of four speakers only one list was recorded. For two speakers this was in the No noise condition, for the other two it was in the Noise condition. Therefore the total number of responses to filler stimuli was $2 \times 122 \times 46 = 11,224$ and the total number of responses to test stimuli was $2 \times 122 \times 64 = 15,616$.

Of these 15,616 responses to test stimuli, 13,057 were fluent and correct and 2,559 contained one or more errors. Of these 2559 error responses 824 contained multiple errors and 1735 single errors. Table 3.1 gives a first classification of the kinds of responses made:

Table 3.2 contains a classification of the types of single errors.

For our purpose we concentrate on single errors, because with multiple errors there is no way to know which of the errors triggered detection. We also focus on elicited exchanges, interruptions and anticipations in the first CVC word, because, except for the vowel errors, in this way we always know that detection by self-monitoring is triggered from the same position. So for our purpose there are, including the vowel errors, 1033 valid errors.

Table 3.1
First classification of response types to test stimuli, with examples.

Response type	n	Example
Fluent and correct	13,057	teep tijk > teep tijk
Multiple error	824	bak zoon > zag boon
Single error	1735	dop tel > top del
Total	15,616	

Table 3.2
Numbers of single errors in responses to test stimuli of different types.

Type of single error	n	Example
Exchange	640	reeg rijp > rijg reep
Interruption	155	tel tom > to.. tel tom
Anticipation	238	bot pen > pot pen
Perseveration	133	duik bof duik dof
Other single error	342	dof bes > dof fes
Hesitation	43	buil pot > ..buil. . .pot . . .. .buil pot
Omission	184	reeg rijp > . . ..; zaal kin > zaal. . .. .
Total	1735	

Error-to-cutoff times

We will now turn to testing our first prediction, viz. that error-to-cutoff times show a bimodal distribution, the peaks of the two underlying distributions being separated by at least 350 ms, for the second time (cf. results Experiment 1). There are 153 interrupted repaired errors and 49 completed repaired errors. Because errors against vowels are in a different position than errors against consonants, we remove these errors against vowels. There remain 124 interrupted repaired errors and 39 completed repaired errors. We collapse these two categories of repaired errors to see how the error-to-cutoff times are distributed. As in Experiment 1, the error-to-cutoff time is defined as the interval between the word onset of the error form, that in all cases began with the erroneous segment, and the moment speech stopped.

The log-transformed error-to-cutoff times for the three oppositions PM1, PM2 and VUV were first analyzed with uninformed mixture modeling (see Fig. 3.1). As in Experiment 1, this analysis indicates two gaussians, one with a peak corresponding to 186 ms and one with a peak corresponding to 660 ms. The estimated temporal separation between two peaks is 474 ms. This is close enough to the value of roughly 500 ms found in Experiment 1 for the temporal separation between 'internally' and 'externally' detected errors to see this as a confirmation of what we found in Experiment 1. The bimodal mixture model was again validated by means of two-stage bootstrapping over 200 replications (first over participants contributing error-to-cutoff observations, then over observations contributed by these bootstrapped participants, as in Experiment 1). Of the 200 bootstrap replications, only 4 resulted in a mixture model with a single gaussian component (i.e. unimodal), 53 resulted in a model with two gaussian components (i.e. bimodal), the majority of 60 outcomes had three components, and the remaining 83 had four or more components. Over the 200 bootstrap replications, the median number of gaussian

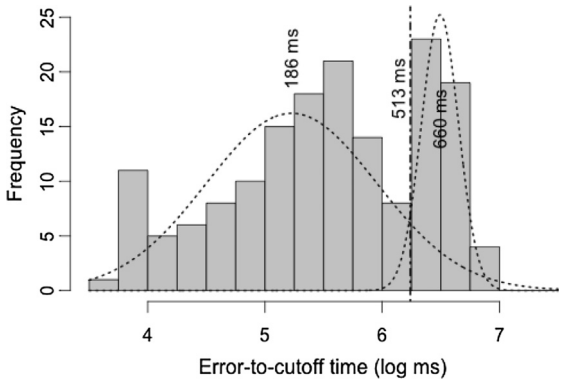

Fig. 3.1. Histogram of log-transformed durations of error-to-cutoff intervals, for N = 163 repaired errors against initial consonants. Distributions plotted with dotted lines indicate the estimated distributions from an uninformed gaussian mixture model (see text). The vertical dashed line indicates the interpolated boundary value (6.24 corresponding to 513 ms) between the two distributions.

components is 2, with a 95% confidence interval of (2,9). The 4 single-component mixture models in the bootstrap validation involved a median number of 45.5 participants in this single component (which were bootstrapped from 81 contributing participants), the two-component models typically involved 16 participants for the smaller component, and for the three-component and four-or-more-component models in the bootstrap validation these numbers were 10 and 4, respectively, for their smallest component. As in Experiment 1, over the 200 bootstrap replications, the number of components was indeed negatively correlated to the number of participants in the smallest component ($r_s = -0.87$, $p < 0.001$). In sum, the bootstrap validation supports a two-gaussian or three-gaussian mixture model, and it clearly does not support a single-gaussian (unimodal) distribution of error-to-cutoff times. We set a boundary between 'internally' and 'externally' detected errors at 513 ms [exp(6.24)] according to the two-gaussian mixture model summarized above (again approximately where the two gaussians intersect in Fig. 3.1). The 124 error-to-cutoff times for repaired interruptions range from 34 ms to 805 ms, with 29 values below 100 ms. The 39 error-to-cutoff times for repaired completed errors range from 291 to 983 ms, with 3 errors below 513 ms. Obviously, a number of repaired interruptions are classified as 'externally' detected and a number of repaired completed errors are classified as 'internally' detected.

Error-to-repair times

The error-to-repair time is again defined as the sum of the error-to-cutoff and the cutoff-to-repair time. We have seen that in our data the difference between 'internally' and 'externally' detected errors in error-to-cutoff times is about 500 ms. If Hartsuiker and Kolk are correct in assuming that the timing for planning a repair is basically the same for errors detected 'internally' and 'externally', we expect that also the error-to-repair times show a difference of some 500 ms. However, if we are correct in assuming that in the case of 'internally' detected errors there is no need for planning a repair and that in the case of 'externally' repaired errors the time needed for planning a repair is much more than the 150 ms allowed in the Hartsuiker and Kolk model, then we expect the two distributions of error-to-repair times, one for error-to-cutoff times shorter and one for error-to-cutoff times longer than 500 ms, to differ significantly more than the two estimated distributions of error-to-cutoff times.

The log-transformed error-to-repair times for the three oppositions PM1, PM2 and VUV, were again analyzed by means of unin-

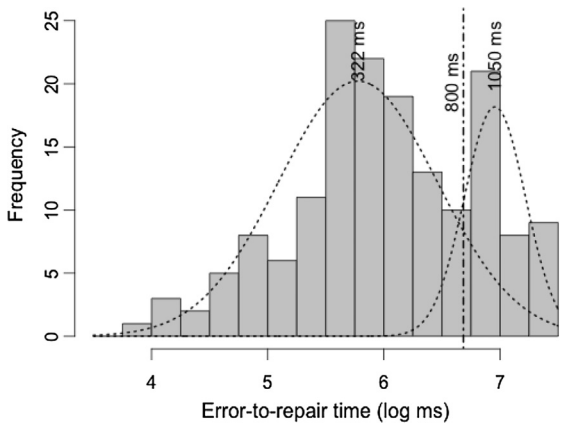

Fig. 3.2. Histogram of log-transformed error-to-repair times, for N = 164 repaired errors against consonants. Distributions plotted with dotted lines indicate the estimated distributions from the two-gaussian solution (see text). The vertical dashed line indicates the interpolated boundary value 6.68 corresponding to 800 ms between the two distributions.

formed mixture modeling (see Fig. 3.2). The optimal mixture model has a single peak only (unimodal) at 409 ms [exp(6.014), s = 0.610, BIC = −392], although the histogram in Fig. 3.2 seems to show two peaks, and although a two-gaussian (bimodal) distribution has an almost equally good fit [with peaks at exp(5.776) = 322 ms and exp(6.956) = 1050 ms; BIC = −396]. In a two-stage bootstrap validation over 200 replications of the mixture model (see above for details), 44 of the solutions were single-gaussian, the majority of 46 were indeed two-gaussian (bimodal), 30 were three-gaussian, and the remaining 80 involved four or more gaussians. In this case, then, the mixture modeling does not clearly suggest a bimodal distribution over a unimodal distribution, but does not contradict a two-gaussian solution either. The temporal difference between the two estimated gaussians of the bimodal mixture model is 728 ms, thus again, as in Experiment 1, roughly 700 ms. The boundary between the two gaussians suggested by the two-gaussian solution falls at 6.68 corresponding to 800 ms (see Fig. 3.2).

In order to see whether indeed the difference between 'internally' and 'externally' detected speech errors is significantly greater in error-to-repair times than in error-to-cutoff times, here we interpreted all repaired speech errors with error-to-cutoff times shorter than 513 ms as detected 'internally' and all repaired speech errors with error-to-cutoff times longer than 512 ms as detected 'externally'. The difference of 201 ms between the averages of the error-to-repair times (540 ms) and of the error-to-cutoff times (339 ms) was removed first, by adding this difference to all error-to-cutoff values. This gave us two sets of values, one for error-to-cutoff times and one for error-to-repair times with exactly the same mean. We then took the natural logarithm of all values in each set, in order to get less skewed distributions. For each of the two dependent variables the difference between 'internally' and 'externally' detected repaired errors was analyzed with a Welch *t* test for two samples. For normalized and log-transformed error-to-cutoff times the mean difference between 'internally' and 'externally' detected repaired errors is 0.799 with (0.736,0.862) as 95% confidence interval. After backtransformation this difference in error-to-cutoff times corresponds to approximately 500 ms. For log-transformed error-to-repair times the mean difference between 'internally' and 'externally' detected repaired errors is 1.268 with (1.130,1.406) as 95% confidence interval. After backtransformation this difference in error-to-repair times corresponds to approximately 700 ms. As these 95% confidence intervals do not overlap, the difference between effects is again significant with $p < 0.05$. We interpret our findings as confirming what we found in Experiment 1 and as showing that indeed, as predicted, the difference between 'internally' and 'externally' detected repaired errors is significantly larger in error-to-repair times than in error-to-cutoff times.

The role of auditory feedback in self-monitoring

Table 3.3 gives a breakdown of valid responses (fluent and correct, and exchanges, interruptions and anticipations in the position in which an interaction was elicited).

Table 3.4 concentrates on the 1033 valid errors in Table 3.3, classified as undetected, repaired interrupted versus repaired completed, and detected 'internally' versus detected 'externally'. Note that values in italics do not contribute to the totals because repaired interrupted and repaired completed were replaced by detected 'internally' and detected 'externally'.

In Experiment 1, we found contrary to expectation that 'internal' and 'external' detection rates were equal in the noise and no noise condition for all errors against place and/or manner of articulation. This supports the position of Lackner and Tuller (1979) that detection of segmental errors by self-monitoring can be based on somatosensory and / or proprioceptive feedback. We would not

Table 3.3
Numbers of correct responses and valid errors against VUV, PM1, PM2 and Vowels, separately for two noise conditions. For each row in the table the number of stimuli was 1936. Percentages are given between brackets.

Noise condition	Opposition	Correct	Valid errors	Total
No noise	VUV	1426 (85%)	253 (15%)	1679
No noise	PM1	1566 (93%)	120 (7%)	1686
No noise	PM2	1664 (96%)	78 (4%)	1742
No noise	Vowel	1647 (97%)	59 (3%)	1706
Noise	VUV	1532 (86%)	246 (14%)	1778
Noise	PM1	1678 (93%)	127 (7%)	1805
Noise	PM2	1772 (95%)	94.(0.5%)	1866
Noise	Vowel	1772 (97%)	56 (0.3%)	1828
Sum		13,057 (93%)	1033 (7%)	14,090

Table 3.4
Numbers of valid errors against the voiced-unvoiced distinction (VUV), place or manner of articulation (PM1), place plus manner of articulation (PM2) and Vowel, broken down by no noise versus noise, and undetected versus repaired interrupted versus repaired completed and 'internally' detected versus 'externally' detected. Percentages are given between brackets. The numbers in italics do not contribute to the totals.

Noise condition	Opposition	Undetected	*Repaired interrupted*	*Repaired completed*	'Internally' detected	'Externally' detected	Total
No noise	VUV	228 (90%)	18 (7%)	7 (3%)	16 (6%)	9 (4%)	253
No noise	PM1	91 (76%)	19 (16%)	0 (8%)	17 (14%)	12 (10%)	120
No noise	PM2	55 (70%)	21 (27%)	2 (3%)	18 (23%)	5 (7%)	78
No noise	Vowels	39 (66%)	12 (20%)	8 (14%)	11 (19%)	9 (15%)	59
Noise	VUV	224 (91%)	12 (5%)	10 (4%)	11 (4.5%)	11 (4.5%)	246
Noise	PM1	93 (73%)	29 (23%)	5 (4%)	21 (17%)	13 (10%)	127
Noise	PM2	64 (68%)	25 (27%)	5 (5%)	24 (26%)	6 (6%)	94
Noise	Vowel	37 (66%)	17 (30%)	2 (4%)	15 (27%)	4 (7%)	56
Sum		831 (80%)	153 (15%)	49 (5%)	133 (13%)	69 (7%)	1033

be surprised to see this finding confirmed. In Experiment 2 we have also made the prediction (prediction 4) that the detection rate of errors against the voiced-voiceless distinction and against vowels would be lower with loud masking noise than in silence. The reason is that Lackner and Tuller found that detection rate for these two classes of segmental errors was much lower under loud masking noise than in silence, presumably because somatosensory and proprioceptive feedback from the articulators is much less conspicuous.

As before we have tentatively separated 'internally' and 'externally' detected errors, classifying all error-to-cutoff times of 513 ms or shorter as 'internal' detections and all error-to-cutoff times longer than 512 ms as 'external' detections. The odds of 'internal' and 'external' detections were analyzed by means of a single multinomial logistic regression, with subsequent two-stage bootstrapping (over speakers and over valid errors, respectively), procedure as recommended by Shao and Tu (1995, p. 247 ff; cf. Nooteboom & Quené, 2008). The optimal model according to Likelihood Ratios Tests has only the opposition type as a predictor; neither noise ($p = 0.2396$) nor the interaction of noise and opposition type ($p = 0.3949$) improved the multinomial model significantly. Relative to items of PM1 (baseline), items of VUV (voiced-unvoiced) have far lower odds of being detected 'internally' [$\beta = -1.293$, s.e. 0.256, bootstrapped 95% C.I.(-1.450,0.,-0.781)], and also far lower odds of being detected 'externally' [$\beta = -1.004$, s.e.0.336, bootstrapped 95% C.I. (-1.110,-0.221)]. For errors against vowels, the odds of being detected 'internally' are somewhat higher than those for items of PM1 (place or manner of articulation) [baseline; $\beta = +0.455$, s.e. 0.279, 95% C.I. +(0.113,+1.047)], but the odds of being detected 'externally' are similar [$\beta = +0.286$, s.e.0.399, 95% C.I. (-0.228,+1.201)]. For errors against PM2 (place plus manner of articulation), the odds of 'internal' detection are again higher than those for items of PM1 [baseline; $\beta = +0.481$, s. e. 0.244, 95% C.I. (+0.308,+0.922)], but the odds of 'external' detection again are not [$\beta = +0.480$, s.e. 0.435, 95% C.I. (-1.056,+0.352)]. These results suggest that for all opposition types, the number of detected errors, whether detected 'internally' or 'externally', is not affected by noise. The results suggest also that errors against vowels are more often 'internally' detected and repaired than errors against consonants and that within the class of consonantal errors against place *plus* manner of articulation are most often detected and repaired followed by errors against place *or* manner of articulation, followed by errors against the voiced-unvoiced distinction. An obvious interpretation of this finding is that within the class of consonantal errors the most dissimilar ones have the highest detection rate and the most similar ones have the lowest. But most conspicuous is the absence of any effect of noise masking we predicted.

Discussion of Experiment 2

The results of Experiment 2 confirm what we found in Experiment 1, i.e. that error-to-cutoff times are distributed bimodally, with two peaks roughly 500 ms apart. This is somewhat more than the 350 ms predicted from the computational model by Hartsuiker and Kolk (2001). We interpret the bimodal distribution of error-to-cutoff times as meaning that there are two classes of repaired errors, those detected in 'internal' speech and those detected in 'external' or overt speech.

From the Hartsuiker and Kolk model we would have predicted that the temporal separation between 'internally' and 'externally' detected errors in error-to-repair times is the same as the temporal separation in error-to-cutoff times. From our view that in the case of 'internal' error detection in general no planning of a repair is necessary, and that in the case of 'external' error detection planning a repair is relatively time consuming, we predicted, contrastively, that the temporal separation between 'internally' and 'externally' detected errors is larger for error-to-repair times than for error-to-cutoff times. This is indeed what we found again in Experiment 2. Obviously, planning a repair is a time-consuming affair, as determined for 'externally' detected errors, taking 896 ms on average in Experiment 1, with a minimum of 434 ms and a maximum of 1373 ms, and taking 960 ms on average in Experiment 2, with a minimum of 370 ms and a maximum of

1813 ms. But such long times needed for planning a repair could hardly be assumed for errors detected in 'internal' speech. This could not account for very short cutoff-to-repair times accompanying very short error-to-cutoff times.

With respect to the role of auditory feedback in self-monitoring, we have seen that, where predictably errors against place plus manner of articulation (PM2) have a higher detection rate than errors against place or manner of articulation (PM1), loud masking noise has no effect on either the 'internal' or the 'external' detection rate of errors against place and/or manner of articulation. Also there is no effect of interaction between noise and similarity on detection rates. Following Lackner and Tuller (1979) in this respect, we propose that these errors against place and/or manner of articulation are detected 'externally' on the basis of somatosensory and/or proprioceptive feedback from the articulators. With respect to errors against the voiced-unvoiced distinction and against vowels we had predicted (prediction 4) that 'internal' detection of these errors would not be affected by loud masking noise but 'external' detection would, assuming with Lackner and Tuller (1979) that somatosensory and proprioceptive feedback from the articulators for these oppositions would not be very clear, and therefore auditory feedback would be needed. This prediction was not borne out by our data. Unexpectedly, there was no effect of noise condition on either 'internal' or 'external' detection of interactional segmental errors against the voiced-unvoiced distinction or against vowels. Apparently in our experiment detection of these errors in overt speech did not depend on auditory feedback.

General discussion

Detecting errors in 'internal' and overt speech

We have found in two experiments that error-to-cutoff times are distributed bimodally, with two peaks being roughly 500 ms apart. We have interpreted this as evidence that speech errors can be detected by self-monitoring both internal and overt speech and that the delay in detection in overt speech with respect to detection in internal speech is roughly 500 ms on average. Not all readers may be convinced. Seyfeddinipur et al. (2008) have suggested that speakers do not interrupt speech as soon as an error is detected, but may wait with speech interruption until a repair is available. They rejected the Main Interruption Rule proposed by Levelt (1989), stating that speakers interrupt their entire speech production upon detecting trouble. If they are right, then we could perhaps explain the bimodality of the distribution of error-to-cutoff times by assuming that all errors are detected in internal speech and that there are two classes of repairs, those that are available virtually immediately after error detection and those that take much time to plan (although we do not think Seyfeddinipur et al., 2008, proposed that all speech errors are detected in internal speech). But then the next question would be why there are these two classes of repairs. If all errors were detected internally and the time needed for planning a repair would be highly variable we have no reason to expect a bimodal distribution of error-to-cutoff times, nor of error-to-repair times. It is precisely the idea that errors can be detected on two different stages, both in internal speech and much later in overt speech, from which the bimodal distribution was predicted.

Lind et al. (2014) and Lind et al. (2015) have argued that the assumption that errors can be detected before speech initiation is superfluous. According to them all errors can be detected in overt speech. But if this is correct, there is no reason at all to expect a bimodal distribution. Of course, readers may not be convinced because the bimodal distribution itself is somewhat equivocal: the second peak in error-to-cutoff times, supposedly corresponding to externally detected errors, concerns relatively few cases in both experiments. Maybe those readers will have to wait until further evidence is obtained. But we observe that the bimodality of the error-to-cutoff times was foreshadowed by the clear bimodality of the distribution of numbers of segments of the error form spoken before interruption in a very similar experiment (Nooteboom, 2005b): Interruption occurred after one or two segments, or after completed utterances of six segments, with very few cases in between. This is difficult to explain without resorting to two different stages of error detection. Moreover, the bimodal distribution of error-to-cutoff times was clearly supported by bootstrap validations in both experiments. We conclude that there are two stages of error detection, one in internal speech and one in overt speech, and the two stages are temporally separated by some 500 ms. This delay of 500 ms for external as compared to internal error detection is somewhat more than the 350 ms predicted by the Hartsuiker and Kolk (2001). Liss (1998) suggested that external error detection in self-monitoring may lag 500 ms behind internal error detection, but this was meant specifically for her somewhat slow apraxic speakers. Our results basically confirm an important aspect of the Hartsuiker and Kolk computational model, viz. that there are two stages of error detection separated by hundreds of ms. In saying this, we do not necessarily imply that the perceptual loop theory is correct in assuming that all self-monitoring for speech errors, both internally and externally, employs the same speech comprehension system that is also employed in other-produced speech.

Repairing speech errors

Our results have confirmed that, as predicted, the difference between 'internally' and 'externally' detected errors is larger for error-to-repair times than for error-to-cutoff times. The difference between the two peaks is roughly 500 ms for error-to-cutoff times and roughly 700 ms for error-to-repair times. This also demonstrates that the temporal aspects of repair planning differ between 'internally' and 'externally' detected errors. In this respect our results offer a possible correction on the Hartsuiker and Kolk (2001) computational model. These findings support our proposal that in repairing speech errors during self-monitoring there are two classes of repairs, those that are rapidly available after error detection and those that have to be painstakingly planned. This difference is closely related to the difference between error detection in internal and overt speech, and particularly with the many hundreds of ms that error detection in overt speech comes later than error detection in internal speech. Our proposal that after internal error detection no repair planning is needed because the correct target is still active and available as a repair, in fact comes close to the proposal by Hartsuiker and Kolk (2001) that little planning time is needed because of priming by earlier mental stages of speech generation. Where we differ more conspicuously is in assuming that after error detection in overt speech, during the time needed to initiate speaking and to detect the overt error, the activation of the correct target form planned in parallel with the error form, has fallen off, and therefore no repair is available anymore. The reader may note that according to our results there is on average some 500 ms between error detection in internal and in overt speech. This is more than enough for the decay of activation of the correct target form. Because there is little remaining activation, planning a repair, i.e. re-planning the correct target, is time consuming. Our data tell us that planning a repair after external error detection varies from roughly 600 to roughly 1800 ms. This is considerably longer than the 150 ms allowed by the Hartsuiker and Kolk computational model. Such long error-to-repair times could hardly account for the fact that very short error-to-cutoff times are often combined with very short cutoff-to-repair times, the lat-

ter often of 0 ms. We conclude that, due to a difference in level of activation of the correct target form that is going to serve as a repair, on average repair planning is faster after internal than after external detection of speech errors

The role of auditory feedback in self-monitoring

The most surprising result in the current investigation is that we found no effect whatsoever of loud masking noise on the detection rate of speech errors. This is all the more unexpected because there are a number of convincing demonstrations of the relevance of auditory feedback for self-monitoring (Oomen et al., 2001; Postma & Noordanus, 1996). At first sight our result also seems in conflict with the demonstration by Huettig and Hartsuiker (2010) that eye movements are controlled by speech perception during self-monitoring overt speech but not during self-monitoring internal speech. But on reflection we see no reason why in their experiment eye-movements could not have been controlled by somatosensory and proprioceptive information from the articulators.

It has been suggested to us that the noise in our experiments did not sufficiently mask the overt speech to the speakers. This we reject. The shape of the noise was such that in the frequency band that is most relevant to speech intelligibility, 300–3500 Hz, the power of the noise in the part of the spectrum that is relevant to speech perception was comparable to what it would have been with white noise of 90 dB SPL overall, as used for example by Postma and Kolk (1992). Also most of our speakers complied with the request to speak softly, and claimed that they did not hear their own voice in the noise condition. Our own impression is that even if, by speaking very loud, one can hear remnants of one's own voice, these remnants are not intelligible speech. Apparently, self-monitoring for speech errors did not depend on audition in our experiments. From the experiment reported by Lackner and Tuller (1979) one might have expected that the detection of speech errors against place of articulation, and perhaps also manner of articulation, would not be affected by auditory feedback, but the detection of errors against the voiced-unvoiced distinction and against vowels would. Not so in our experiments. Does this mean that in our experiments all speech errors were detected in internal speech, before speech initiation? We do not think so. The bimodal distributions of error-to-cutoff and error-to-repair times clearly point at two different stages of self-monitoring for speech errors. We propose that all segmental speech errors, also those against the voiced-unvoiced distinction and against vowels, can be detected after speech initiation by somatosensory and/or proprioceptive feedback from the articulators.

Whether or not in a particular experiment there is also a contribution from auditory feedback, probably depends on task structure, very likely on the amount of time pressure on the speakers, but also on the type of errors. As detection of syntactic and/or semantic errors needs more verbal context than the detection of sound errors, it seems reasonable to expect that auditory feedback and verbal memory might be more important there. This might explain why for example Oomen and Postma did find an effect of masking with loud noise and we did not. These authors used tongue twister-like sequences eliciting speech errors and included both sound errors and semantic errors in their analysis. As to time pressure: In a SLIP experiment time pressure is considerable, and utterances are relatively short. This may have prevented our speakers from paying attention to their own voice in the no noise condition, instead concentrating on information from the articulators preceding acoustics and audition, which provides faster feedback. This potentially is different in experiments in which speakers produce tongue twister sentences and/or have to report self-detected speech errors by pushing a button. In those experiments speakers may have some more time to react to their own voice in self-monitoring. If this is indeed the case, we may control the contribution of auditory feedback by varying the task structure in future experiments.

The reader may have noticed that our proposal to distinguish between internal and external error detection would have been much stronger if we would have found a strong contribution of auditory feedback to external but not to internal error detection. This points at imaginable similar experiments in which the condition of loud masking noise is replaced with a condition in which somatosensory and/or proprioceptive feedback is blocked or diminished by local anesthesia. If our interpretation of the current experiments is correct, one should find an effect of the local anesthesia on the external but not on the internal detection rate. It is currently unclear to us whether such an experiment would be feasible.

Conclusions

The perceptual loop theory by Levelt (1989) and Levelt et al. (1999) states that there are two stages of self-monitoring for speech errors, one focusing on internal and one focusing on overt speech. The computational model of the perceptual loop theory by Hartsuiker and Kolk (2001) predicts that these two stages are separated by a delay of approximately 350 ms. Our experiments confirm the existence of two stages, and suggest that the delay may be somewhat longer, in the order of 500 ms. The Hartsuiker and Kolk model predicts that the difference between internally and externally detected errors in error-to-repair times is the same as the difference in error-to-cutoff times. Our results show that the difference between internally and externally detected errors in error-to-repair times is some 700 ms, i.e. significantly longer than the difference in error-to-cutoff times. We explain this by assuming that after internal error detection no repair planning is necessary, because the correct target form is still available, whereas after external error detection the activation of the correct target form has fallen off, so that a repair has to be planned with much time and effort. Our results also show that self-monitoring of overt speech for speech errors does not depend on auditory feedback. This supports the suggestion of proponents of forward modeling accounts of speech production that self-monitoring can employ somatosensory and proprioceptive feedback from the articulators.

Author Note

The authors are grateful to the Utrecht institute of Linguistics OTS for providing the technical facilities that enabled us to do the two experiments. In particular we are grateful to the technical staff of the institute. Alex Manus helped us with setting up Experiment 1, and Chris van Run did a fine job in re-writing the software and implementing the technical details for Experiment 2. Dr. Iris Mulders was extremely helpful in finding enough speakers. We are also grateful to Robert Hartsuiker and two anonymous reviewers for their many stimulating comments that helped improve the ms considerably. The raw data of the two experiments in .xlsx files can be found through the following web page: http://www.let.uu.nl/~Sieb.Nooteboom/personal/Experimentaldata.htm.

Appendix A

Table A
Stimulus word pairs, separately for Experiment 1 and 2, for Test (T) and Filler (F) stimuli, Similarity (Sim) in Experiment 1 (1 = Manner or Place of Articulation, 2 = 2 Manner and Place of Articulation, 4 = not specified for Fillers), and for Opposition (Opp) in Experiment 2 (1 = Manner or Place of Articulation, 2 = Manner and Place of Articulation, 3 = Voice, 4 = Vowel, 5 = not specified for Fillers).

	Experiment 1						Experiment 2					
	List 1			List 2			List 1			List 2		
Nr	Word pair	T/F	Sim	Word pair	T/F	Sim	Word pair	T/F	Opp	Word pair	T/F	Opp
1	paf kiep	T	1	pal kiem	T	1	paf kiep	T	1	pap kier	T	1
2	kaf piep	T	1	kal piem	T	1	kaf piep	T	1	kap pier	T	1
3	doos bel	T	1	doof bed	T	1	doos bel	T	1	dof bes	T	1
4	boos del	T	1	boof det	T	1	boos del	T	1	bof det	T	1
5	voet zeen	T	1	voer zeep	T	1	voet zeen	T	1	voet zeel	T	1
6	zoet veen	T	1	zoer veep	T	1	zoet veen	T	1	zoet veel	T	1
7	buik dof	T	1	buis dor	T	1	buik dof	T	1	buig dof	T	1
8	duik bof	T	1	duis bor	T	1	duik bof	T	1	duig bof	T	1
9	kam peen	T	1	kan peer	T	1	kat pees	T	1	kan peer	T	1
10	pam keen	T	1	pan keer	T	1	pad kees	T	1	pan keer	T	1
11	keus por	T	1	keur pol	T	1	kuil pop	T	1	keur pol	T	1
12	peus kor	T	1	peur kol	T	1	puil kop	T	1	peur kol	T	1
13	pit tos	T	1	pin tof	T	1	pip tol	T	1	pin tof	T	1
14	tit pos	T	1	tin pof	T	1	tip pol	T	1	tin pof	T	1
15	piek faam	T	1	pier faal	T	1	pier faal	T	1	pief faal	T	1
16	fiek paam	T	1	fier paal	T	1	fier paal	T	1	fief paal	T	1
17	zaan boom	T	2	zaag boot	T	2	zaan boom	T	2	zaal buit	T	2
18	baan zoom	T	2	baag zoot	T	2	baan zoom	T	2	baal zuid	T	2
19	geit been	T	2	gijn beet	T	2	geit been	T	2	gul bas	T	2
20	bijt geen	T	2	bijn geet	T	2	bijt geen	T	2	bul gas	T	2
21	tol veer	T	2	top veeg	T	2	tol veer	T	2	tol veen	T	2
22	vol teer	T	2	vop teeg	T	2	vol teer	T	2	vol teen	T	2
23	ken zooi	T	2	kef zoog	T	2	ken zooi	T	2	kin zog	T	2
24	zen kooi	T	2	zef koog	T	2	zen kooi	T	2	zin kog	T	2
25	ban zool	T	2	bak zoon	T	2	bal zuil	T	2	bak zoon	T	2
26	zan bool	T	2	zak boon	T	2	zal buil	T	2	zak boon	T	2
27	kaar zich	T	2	kaal zin	T	2	kaag zin	T	2	kaal zin	T	2
28	zaar kig	T	2	zaal kin	T	2	zaag kin	T	2	zaal kin	T	2
29	dol gaaf	T	2	dom gaar	T	2	dok gaar	T	2	dom gaar	T	2
30	gol daaf	T	2	gom daar	T	2	gok daar	T	2	gom daar	T	2
31	teem gaap	T	2	teef gaan	T	2	teef gaai	T	2	teef gaan	T	2
32	geem taap	T	2	geek taat	T	2	geef taai	T	2	geek taat	T	2
33	vaat tip	F	4	vaat tip	F	4	pak biet	T	3	pad bel	T	3
34	ros feil	F	4	ros feil	F	4	bak piet	T	3	bad pel	T	3
35	mom vit	F	4	mom vit	F	4	dop tel	T	3	dof tel	T	3
36	dijn koor	F	4	dijn koor	F	4	top del	T	3	tof del	T	3
37	git dek	F	4	git dek	F	4	bot pen	T	3	boen piet	T	3
38	rik loot	F	4	rik loot	F	4	pot ben	T	3	poen biet	T	3
39	wijn ruit	F	4	wijn ruit	F	4	buik pof	T	3	bus pof	T	3
40	kir waag	F	4	kir waag	F	4	puik bof	T	3	pus bof	T	3
41	haam lijp	F	4	haam lijp	F	4	pas bit	T	3	ban peer	T	3
42	ruik heem	F	4	ruik heem	F	4	bas pit	T	3	pan beer	T	3
43	rif weg	F	4	rif weg	F	4	doog teel	T	3	beur poos	T	3
44	was hef	F	4	was hef	F	4	toog deel	T	3	peur boos	T	3
45	loog haat	F	4	loog haat	F	4	bof pek	T	3	dik tof	T	3
46	ruin lies	F	4	ruin lies	F	4	pof bek	T	3	tik dof	T	3
47	vim kil	F	4	vim kil	F	4	buil pot	T	3	bier paal	T	3
48	woed looi	F	4	woed looi	F	4	puil bot	T	3	pier baal	T	3
49	ris meel	F	4	ris meel	F	4	zak ziel	T	4	boon beet	T	4
50	moet neut	F	4	moet neut	F	4	ziek zal	T	4	been boot	T	4
51	hoop laai	F	4	hoop laai	F	4	tijp teek	T	4	rijg reep	T	4
52	look haas	F	4	look haas	F	4	teep tijk	T	4	reeg rijp	T	4
53	jaag hof	F	4	jaag hof	F	4	tol tem	T	4	tal top	T	4
54	mik reeg	F	4	mik reeg	F	4	tel tom	T	4	tol tap	T	4
55	woef leen	F	4	woef leen	F	4	ken kif	T	4	zeeg zaan	T	4
56							kin kef	T	4	zaag zeen	T	4
57							zat zich	T	4	pit pel	T	4
58							zit zag	T	4	pet pil	T	4
59							teil teem	T	4	zeg zon	T	4
60							teel tijm	T	4	zog zen	T	4
61							tien toep	T	4	dos dep	T	4
62							toen tiep	T	4	des dop	T	4
63							wik wel	T	4	teel taan	T	4
64							wek wil	T	4	taal teen	T	4
65							vaat tip	F	5	vaat tip	F	5

(*continued on next page*)

Table A (*continued*)

	Experiment 1						Experiment 2					
	List 1			List 2			List 1			List 2		
Nr	Word pair	T/F	Sim	Word pair	T/F	Sim	Word pair	T/F	Opp	Word pair	T/F	Opp
66							ros feil	F	5	ros feil	F	5
67							vet pot	F	5	vet pot	F	5
68							puim boef	F	5	puim boef	F	5
69							wieg kuch	F	5	wieg kuch	F	5
70							maak juk	F	5	maak juk	F	5
71							mom vit	F	5	mom vit	F	5
72							dijn koor	F	5	dijn koor	F	5
73							git dek	F	5	git dek	F	5
74							rik loot	F	5	rik loot	F	5
75							wijn ruit	F	5	wijn ruit	F	5
76							kir waag	F	5	kir waag	F	5
77							haam lijp	F	5	haam lijp	F	5
78							ruik heem	F	5	ruik heem	F	5
79							gif dep	F	5	gif dep	F	5
80							ring loon	F	5	ring loon	F	5
81							wijf ruig	F	5	wijf ruig	F	5
82							kit waan	F	5	kit waan	F	5
83							haan lijs	F	5	haan lijs	F	5
84							ruis heet	F	5	ruis heet	F	5
85							rif weg	F	5	rif weg	F	5
86							was hef	F	5	was hef	F	5
87							loog haat	F	5	loog haat	F	5
88							ruim liep	F	5	ruim liep	F	5
89							rib wen	F	5	rib wen	F	5
90							wak hel	F	5	wak hel	F	5
91							loof haar	F	5	loof haar	F	5
92							ruin lies	F	5	ruin lies	F	5
93							vim kil	F	5	vim kil	F	5
94							woed looi	F	5	woed looi	F	5
95							ris meel	F	5	ris meel	F	5
96							moet neut	F	5	moet neut	F	5
97							hoop laai	F	5	hoop laai	F	5
98							look haas	F	5	look haas	F	5
99							jaag hof	F	5	jaag hof	F	5
100							mik reeg	F	5	mik reeg	F	5
101							woef leen	F	5	woef leen	F	5
102							ving kog	F	5	ving kog	F	5
103							deur bies	F	5	deur bies	F	5
104							deeg biet	F	5	deeg biet	F	5
105							baar vief	F	5	baar vief	F	5
106							vaam kien	F	5	vaam kien	F	5
107							hos gup	F	5	hos gup	F	5
108							hor weef	F	5	hor weef	F	5
109							heil noor	F	5	heil noor	F	5
110							riem lof	F	5	riem lof	F	5

Appendix B

In measuring acoustic durations of error-to-cutoff times and cutoff-to-repair times for utterances that may start with either a plosive or a fricative, there is a potential problem in that these durations differ systematically between utterances starting with a plosive and those starting with a fricative. The reason is that during the initial mouth closure there is a clear acoustic signal in the fricative but not in a plosive (except the highly variable voice lead in voiced plosives that was discarded in our measurements). In order to get a reasonable estimate of this systematic difference, we have made use of the circumstance that in Experiment 2 there were 6 pairs of CVC CVC stimuli where both members of the pair had been used as stimuli, viz. *baal zuid* vs *zaal buit, baan zoom* vs *zaan boom, bal zuil* vs *zal buil, bak zoon* vs *zak boon, ken zooi* vs *zen kooi,* and *kin zog* vs *zin kog*. For these 6 pairs of stimuli, we selected all those responses obtained in Experiment 2 where the same speaker responded fluently and correctly to both members of these pairs of stimuli (N = 1108 responses). The log-transformed response times were fed into a LMM with consonant class (plosive or fricative) as a fixed predictor, and speakers (n = 122) and stimulus pairs (n = 6) as random intercepts. The consonant class was also included as a random slope over speakers and over item pairs. The resulting LMM showed a significant effect of consonant class ($\beta = -0.0805$, s.e. 0.0088, $t = -9.2$, $p = .0004$). The back-transformed average of the response times for the utterance starting with plosive was 674 ms and the average for the response time starting with a fricative was 622 ms. Therefore in determining the error-to-cutoff times we added 52 ms if the erroneous fragment started with a plosive, and in determining the cutoff-to-repair times we subtracted 52 ms if the repair started with a plosive. Of course, the 52 ms correspond to an average difference. Therefore in a few cases subtracting 52 ms leads to a negative value of a cutoff-to-repair time. In those cases we have censored these negative values to 0 ms.

References

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

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.

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition, 39*, 173–194.

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer (Version 6.0.19)* [Computer program]. <http://www.praat.org/>.

Efron, B., & Tibshirami, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Fraley, C., Raftery, A. E., Murphy, T. B, & Scrucca, L. (2012). mclust: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Report No. 597, Department of Statistics, University of Washington. Available: <http://www.stat.washington.edu/mclust/>.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97*, 611–631.

Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics, 30*, 139–162.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes, 21*, 649–683.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition, 103*, 386–412.

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related Beply to Baars, Motley, and MacKay (1975). *Journal of Memory and Language, 52*, 58–70.

Hartsuiker, R. J., Kolk, H .H .J., & Martensen, H., (2005). Division of labor between internal and external speech monitoring. In R. Hartsuiker, Y. Bastiaanse, A. Postma, & F. Wijnen (Eds.), Phonological encoding and monitoring in normal and pathological speech (pp. 187–205). Hove: Psychology Press.

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., Pickering, M. J., & de Jong, N. H. (2005). Semantic and phonological context effects in speech error repair. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 921–932.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*, 135–145.

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*, 347–374.

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

Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing* (pp. 281–294). Hillsdale, N.J.: Erlbaum.

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* (pp. 287–306). New York: Academic Press.

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, Massachusetts: The MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' acceptance of real-time speech exchange Indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science, 25*(6), 1198–1205.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2015). Auditory feedback Is used for self-comprehension: When we hear ourselves saying something other than what we said, we believe we said what we hear. *Psychological Science, 26*(12), 1978–1980.

Liss, J. M. (1998). Error-revision in the spontaneous speech of apraxic speakers. *Brain and Language, 62*, 342–360.

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

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition, 117*, 243–260.

Mowrey, R., & MacKay, I. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America, 88*, 1299–1312.

Nooteboom, S. G. (2005b). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication, 47*, 43–58.

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 87–95). New York: Academic Press.

Nooteboom, S. G. (2005a). Listening to one-self: Monitoring speech production. In R. Hartsuiker, Y. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 167–186). Hove: Psychology Press.

Nooteboom, S. G., & 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., & Schwartz, M. (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., & Kolk, H. H. J. (2001). Phonological encoding and monitoring in normal and pathological speech. *Cortex, 31*, 209–237.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347.

Postma, A., & Kolk, H. H. J. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research, 35*, 337–544.

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.

R Development Core Team (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: <http://www.R-project.org>.

Schlenk, K., Huber, W., & Wilmes, K. (1987). "Prepairs" and repairs: Different monitoring functions in aphasic language production. *Brain and Language, 30*, 226–244.

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition, 108* (3), 837–842.

Shao, J., & Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica, 45*, 175–197.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typing. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in speech: Additional evidence. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

Tydgat, I., Diependaele, K., Hartsuiker, R. J., & Pickering, M. J. (2012). How lingering representations of abandoned context words affect speech production. *Acta Psychologica, 140*, 189–229.

Van Alphen, P. M. (2004). *Perceptual Relevance of Prevoicing in Dutch* Unpublished doctoral dissertation. Nijmegen, The Netherlands: Radboud University.

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.