Monitoring for speech errors has different functions in inner and overt speech

Sieb Nooteboom
Utrecht University

In this paper it is argued that monitoring for speech errors is not the same in inner speech and in overt speech. In inner speech it is meant to prevent the errors from becoming public, in overt speech to repair the damage caused by the errors. It is expected that in inner speech, but not in overt speech, more nonword errors are detected than real-word ones, and that overt repairs of errors detected in inner speech differ from overt repairs of errors detected in overt speech in that they have shorter offset-to-repair times, are spoken with raised instead of lowered intensity and pitch, and are less often accompanied by editing expressions. These hypotheses are tested against a collection of experimentally elicited spoonerisms and a collection of speech errors in spontaneous Dutch. The hypotheses are basically confirmed.

1. Introduction

Baars, Motley and MacKay (1975) suggested that in speech production there is pre-articulatory editing of inner speech, during which speech errors are detected and repaired before they surface in overt speech. This assumption was needed to explain the phenomenon of so-called lexical bias in segmental speech errors. Lexical bias is the phenomenon that segmental speech errors more often create real words than non-words, other things being equal. Lexical bias has been attested both in experimentally elicited speech errors (e.g. Baars et al. 1975; Dell 1986; Nooteboom 2005b; Nooteboom & Quené 2008) and in spontaneous speech errors (e.g. Dell & Reich 1981; Nooteboom 2005a; Hartsuiker, Anton-Mendez, Roelstraete & Costa 2006; but see Garrett 1976 and Del Viso, Igoa & Garcia-Albea 1991). The idea is that nonword errors are more often detected, rejected and repaired in inner speech than real-word errors, causing a greater frequency of real-word errors in overt speech. The need for such covert, rapid and fluent pre-articulatory editing was rejected by Stemberger (1985) and Dell (1986). They explained lexical bias from immediate feedback of activation between phonemes and lexical units within the mental speech production system proper. However, Levelt (1989) and Levelt, Roelofs and Meyer (1999) rejected feedback, and revived the original explanation by Baars et al. More recently, Hartsuiker, Corley and Martensen (2005) provided evidence that lexical bias in segmental speech errors is caused both by immediate feedback of activation between speech sounds and word forms, as proposed by Dell (1986), and by covert monitoring of inner speech, applying a criterion of lexicality, as proposed by Levelt et al. (1999). This was further confirmed by Nooteboom and Quené (2008) who in some conditions demonstrated a positive lexical bias in early interrupted speech errors such as *b.. dark boat*, but in other conditions found a reverse lexical bias in such early interruptions. Importantly, interruptions as these must have been reactions to inner speech, because speech fragments like *b..* are shorter than a humanly possible reaction time. Therefore the speaker cannot have reacted to detecting the error in his overt speech. The positive lexical bias in interrupted speech errors cannot have been caused by covert monitoring, because repairs of such interrupted speech errors are made overtly, not covertly, and thus such positive lexical bias provides evidence for feedback. The reverse lexical bias that was also found cannot be explained from feedback, because feedback of the kind proposed by Stemberger (1985) and Dell (1986) never leads to a reverse lexical bias. Such reverse lexical bias thus provides evidence for an active monitor, detecting and rejecting nonwords in inner speech more frequently than real words. These findings bring into focus the psychological reality of an active monitor, covertly detecting, rejecting and sometimes repairing speech errors in inner speech. Apparently, lexical bias in segmental speech errors results both from feedback of activation between speech sounds and word forms in the mental production of speech, and from a criterion of lexicality used in monitoring inner speech for speech errors. It should be said, though, that this lexicality criterion only modulates a relatively high level of detection of both real-word errors and nonword errors. Obviously, error detection does not depend exclusively on a lexicality criterion (cf. Nooteboom & Quené 2008).

Kolk (1995) has suggested that if at the moment of repair the word's lemma is still active, it is sufficient to start the compilation of the form once more. This would explain the rapidity of the repair. The implied alternative would be that the correct target is accessed for a second time all the way from the monitor which is supposed to reside at the conceptual level (cf. Levelt et al. 1999). But in Kolk's formulation it is obviously assumed that the correct target was not available at the moment of error detection. If so, it remains unclear how segmental real-word errors can be rapidly detected. It seems more reasonable to assume that error form and correct target are simultaneously available and competing for the same slot in inner speech. Normally, the most activated form would win, but then the monitor can detect the difference and decide that an error has been made. The assumption that error form and correct target are simultaneously competing for the same slot in inner speech is supported by the relatively frequent occurrence of speech errors such as *gfgfg..feit goud* in experiments eliciting consonant exchanges (Nooteboom 2007). Obviously in this case *feit* and *goud* are competing for the same slot. Also I found a few interesting cases where the participant very softly whispered the correct target and then immediately, with no silent interval, spoke the elicited spoonerism aloud. This also suggests that correct target and speech error are competing for the same slot in inner speech.

At first sight the supposed monitoring of inner speech for speech errors looks very similar to what happens in monitoring overt speech for speech errors. Such overt errors are not always but in c. 50% of the cases detected and then overtly repaired by the speaker (Nooteboom 1980). This leads to such sequences as *bark boat.. dark boat*. Indeed Levelt (1989) and Levelt et al. (1999) have suggested that there is only one monitor checking speech for errors. This monitor employs the same speech comprehension system that is also involved in perceiving other-produced speech. The comprehension system receives two forms of input, one being inner speech before it is articulated. This closes the internal perceptual loop. The other input is overt speech processed by the auditory system. This closes the external perceptual loop. The suggestion is that, apart from the input, the monitoring process is identical in both cases.

The idea that monitoring of inner and monitoring of overt speech are identical also seems implied by three prerequisites for a satisfactory account of any monitoring bias, formulated by Hartsuiker (2006), who stated as prerequisites that (1) the proposed account poses functional monitoring criteria; (2) the bias can be altered by manipulations affecting monitoring performance; (3) the monitoring bias occurs also in perception. The third prerequisite suggests that the same bias that is assumed to operate in monitoring inner speech should also be found in speech perception, and of course, one can perceive both speech produced by others and speech produced by oneself. If one thinks of the latter, Hartsuiker's third prerequisite suggests that if one assumes that monitoring inner speech employs a criterion of lexicality causing lexical bias in overt speech errors, then one would predict that in monitoring overt speech for speech errors also more nonwords than real words are detected, rejected and repaired. Hartsuiker also predicts, of course, that a similar criterion operates in the perception of other-produced speech. The suggestion clearly is that finding a lexical bias in the detection of overt speech errors, either produced by the speaker himself or by another speaker, would support the assumption of a lexicality criterion being employed by the monitor operating on inner speech.

However, it might be argued that self-monitoring of inner speech and self-monitoring of overt speech have different functions and different time-constraints. If so, it might be expected that the strategies involved are not necessarily identical, and this may lead to different behavioural patterns. Monitoring inner speech for speech errors probably attempts to prevent errors in inner speech from becoming public, preferably without any perceivable effects on the fluency of the speech produced, or at least with so little perceivable effect as possible. For this reason, it is reasonable to assume that detecting and repairing errors in inner speech is under some time pressure. Monitoring overt speech for speech errors obviously comes too late to prevent those errors from becoming public. There is no reason to suppose that detecting and repairing overt speech errors is under time pressure. Rather the speaker should take his or her time to make clear to the listener that an error has been made and to prevent damage to communication by this error. These considerations lead to some testable predictions.

If indeed monitoring for speech errors in inner speech strives towards making the error as little noticeable as possible for the listeners, one would expect that speakers hasten to repair the error. A speaker may repair the detected error before it is spoken at all. If so, there is no overt repair. But, probably due to the time constraints on monitoring, rather often a speaker inadvertently initiates speaking the error detected inner speech, and then rapidly interrupts his or her speech. This leads to errors of the type *s..fat soap*. The basic assumption here is that both the interruption and the following repair are intended to make the error as little noticeable as possible. This leads to time pressure. This is potentially different for errors that are only detected after they have been spoken. The full error has been made and very likely also heard by the listener. In this case it is more relevant for the speaker to let the listener know that an error has been made and that in the interpretation the error has to be replaced by the repair. Here there is not necessarily any time pressure on the speaker. Now it has recently been shown that under time pressure, other things being equal, (unless error and target are phonetically very similar) nonword errors are more often detected than real-word ones, but if the time pressure is removed nonword errors are equally often detected as real-word ones (Nooteboom & Quené 2008). If this result is valid, one expects more nonword errors to be detected than real-word ones in normal monitoring of inner speech, but not in normal monitoring of overt speech. So here is a first hypothesis to be tested.

A second hypothesis is concerned with the offset-to-repair times. If indeed speakers hasten to cover up any errors that are detected in inner speech but yet inadvertently initiated, as in *s..fat soap*, one expects offset-to-repair times to be generally short in such errors, whereas offset-to-repair times following errors detected in overt speech are probably much longer, because in these cases the speaker has no reason to hurry, and may need time to plan the repair in such a way that the listener knows that an error has been made and should be replaced with the repair in order to arrive at the intended interpretation. Of course, as has been pointed out by Blackmer and Mitton (1991), if the error is detected in inner speech and yet, inadvertently, spoken as an initial fragment, not only the interruption but also the repair may have been planned before speech was initiated. In such cases one may find offset-to-repair times of 0 ms.

A third hypothesis is concerned with speech prosody. If indeed speakers have a tendency to cover up the overt consequences of errors detected in inner speech, they may be expected not only to have very short offset-to-repair times, but also to make the inadvertent error fragment less noticeable by speaking the rapidly following repair with more intensity and on a higher pitch than the error fragment itself was spoken. For errors detected in overt speech the situation is different. It is too late for any cover up anyway, and the error may potentially lead to a wrong interpretation by the listener.

Thus here the error must remain relatively noticeable, and the repair should stand out as a repair by a prosody that is markedly different from both the error and the regular correct responses. This could be achieved by speaking the repair with somewhat less intensity and on a lower pitch than the error itself.

Finally, a fourth hypothesis relates to the editing expressions, such as *sorry*, *oh!*, *er*, *correction*, or *hahaha*, that may accompany a repair. If speakers tend to cover up the overt consequences of errors detected in inner speech, they very likely will not use such editing expressions in their overt repairs of such errors. But in making repairs of errors detected in overt speech, such editing expressions would be functional (Levelt 1983), and one may thus expect the frequent use of such editing expressions.

From the assumption that monitoring inner speech and monitoring overt speech have different functions, one can thus derive several hypotheses relating to different behavioural patterns for the two types of repair. However, to test these hypotheses, it should be clear which overt repairs are reactions to error detection in inner speech and which are reactions to error detection in overt speech. It is not self-evident that this is always easy to do. Although in cases like *s..fat soap* it is very unlikely that the interruption results from error detection in overt speech in view of the very short duration of the erroneous speech fragment, in cases with interruptions that come somewhat or much later it is not clear whether these are relatively late reactions to errors in inner speech or whether these are reactions to errors in overt speech, or perhaps to both simultaneously. Hartsuiker and Kolk (2001), using a computational model of self-monitoring including both an internal an external perceptual loop, have demonstrated that empirical distributions of overt offset-to-repair times cannot be simulated satisfactorily with only an external loop. A considerable contribution of the internal loop is needed in order to get a reasonable fit of the empirical distributions of overt offset-to-repair times. Hartsuiker, Kolk and Martensen (2005), using an equation with two unknowns for estimating the relative numbers of detected and undetected speech errors in inner from those in overt speech, found that the accuracy of the internal perceptual loop is considerably better than that of the external perceptual loop. The accuracy of the latter may in some conditions even be zero. The corollary of this is that many overt interruptions and repairs are reactions to the detection of errors in inner speech. However, it is reasonable to assume that the later the interruption comes after the error has been made the greater the probability that the error was detected in overt speech. Below we will assume that early interrupted speech errors of the type *s..fat soap* or *sa..fat soap* are errors detected in inner speech and that repaired completed speech errors of the type *sat soap... fat soap* or *sat foap...fat soap* are errors detected in overt speech (for evidence that this assumption at least is highly plausible see Nooteboom 2005b).

Obviously, that an interruption was planned in inner speech does not necessarily mean that the repair was also planned in inner speech. However, if one finds an early interruption such as *s..fat soap* with an offset-to-repair time of zero ms, or not much longer, it seems almost certain that the repair was planned before speech was initiated (Blackmer & Mitton 1991). Experimental evidence that often repairs are planned before the speech is interrupted is obtained by Hartsuiker, Catchpole, De Jong and Pickering (2008). But if the offset-to-repair time is many hundreds of ms, it is not unlikely that, although the interruption was already planned before speech initiation, the repair was only planned after the overt speech fragment was spoken (and perhaps re-detected in overt speech). Likewise, if there is a repaired completed exchange with an offset-to-repair time of 0 ms, this may be a case where the error had been detected and the repair had already been planned before speech initiation, whereas with much longer offset-to-repair times this is much less likely. The point here is that if repairs of early interruptions are interpreted as reactions to error detection in inner speech, and repairs to completed speech errors as reactions to error detection in overt speech, this should be seen as probabilistic. It seems unlikely that a 100% separation between repairs as reactions to error detection in inner speech and repairs as reactions to error detection in overt speech can be achieved. Some repairs of interrupted speech errors may be reactions to overt speech (although the interruptions themselves were reactions to error detection in inner speech) and some repairs of completed speech errors may be reactions to error detection in inner speech. This implies that individual cases have little to say about the validity of the earlier hypotheses. These can only be tested statistically.

Before the earlier mentioned hypotheses are put to the test, it may be good to go back for a moment to the original hypothesis by Baars et al. (1975), viz. that there is pre-articulatory editing of inner speech, covertly and fluently detecting, rejecting and repairing speech errors before they are spoken. Obviously such cases are not observable. This makes it difficult to ascertain the reality of covert and fluent editing of inner speech. Yet there is at least one convincing demonstration of such covert and fluent editing. This is found in data published by Motley, Camden and Baars (1982) who demonstrated activity of the monitor by measuring Galvanic Skin Responses to experimentally elicited but suppressed taboo words in otherwise perfectly fluent and correct speech. The earlier mentioned model by Hartsuiker and Kolk (2001) allows such covert and fluent editing. However, Hartsuiker, Kolk and Martensen (2005) assume that, although sometimes covert repairs may leave no observable traces, often covert repairs lead to observable disfluencies. Unfortunately it is not easy to investigate unobservable covert repairs, and it is also not easy to know whether specific disfluencies stem from covert repairs of speech errors or from other causes. Therefore the remainder of this paper will focus on overt interruptions and repairs, being either reactions to speech errors in inner speech or to speech errors in overt speech. This will be done both for experimentally elicited speech errors and for spontaneous speech errors.

The following four hypotheses will be tested, under the assumption that most early interruptions and their repairs are reactions to error detection in inner speech and most repaired completed speech errors are reactions to error detection in overt speech:

1. Other things being equal, there more potential nonword than potential real-word errors are early interrupted (Nooteboom 2005b; Nooteboom & Quené 2008). However, there are equally many nonword and real-word repaired completed speech errors.
2. Overt repairs following interrupted speech errors have shorter offset-to-repair times than overt repairs following completed speech errors.
3. Overt repairs following interrupted speech errors have more intensity and higher pitch than the errors themselves. Overt repairs following completed speech errors have less intensity and lower pitch than the errors themselves.
4. Overt repairs following interrupted speech errors are much less often accompanied by editing expressions than overt repairs following completed speech errors.

2. The data: Elicited speech errors and errors in spontaneous speech

The speech errors and their interruptions and repairs used in this paper have been described in earlier publications, but the relevant patterns in the data described below are mostly new. The corpus of speech errors in spontaneous Dutch was described in Nooteboom (2005a). The corpus of experimentally elicited Dutch speech errors was described partly in Nooteboom (2005b) and partly in Nooteboom and Quené (2008).

2.1 Elicited speech errors

These errors stem from classical SLIP (Spoonerisms of Laboratory-Induced Predisposition) experiments, meant to elicit consonant exchanges like *fine book* becoming *bine fook* (nonword error) or *cool tap* becoming *tool cap* (real-word error). The SLIP technique basically works as follows: Subjects are successively presented visually, for example on a computer screen, with word pairs such as *dove ball, deer back, dark bone, barn door*, to be read silently. On a prompt, for example a buzz sound or a series of question marks ("?????"), the last word pair seen (the test word pair as opposed to the biasing word pairs), in this example *barn door*, 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 a word pair like *barn door* will be mispronounced as *darn bore*, as a result of segmental biasing by the preceding word pairs.

Below the interrupted elicited exchanges such as *da..barn door*, are interpreted as reactions to error detection in inner speech, the repaired completed elicited exchanges such as *barn door* >> *darn bore* are interpreted as reactions to error detection in overt speech. The particular selection of cases used to test the hypotheses is explained and argued separately for each hypothesis.

2.2 Speech errors in spontaneous speech

The collection of speech errors in spontaneous speech described in Nooteboom (2005a) in fact consists of two separate corpora. One of these, the AC/SN corpus, does not contain any repairs, and therefore is useless for the current purposes. The other, the Utrecht corpus, contains speech errors together with any repairs, noted down between 1977 and 1982 by staff members of the Utrecht Phonetics department on the initiative of the late Anthony Cohen, at that time professor of Phonetics in Utrecht. Of the 2500 errors in the collection 1100 are segmental syntagmatic errors. These will be used here. As these errors were noted down in standard orthography, or sometimes in phonetic script, and the sound has not been recorded, obviously there are no known offset-to-repair times. The moment of stopping for making a repair can of course be estimated from the amount of speech material spoken after the error has been made. It was found that after a segmental error only very rarely a speaker speaks more than the error form itself before stopping. This is different for lexical errors where in many cases the speaker stops only after a few more words; cf. Nooteboom (2005a). So we can distinguish between two classes of repairs: The class of repairs following an early interruption of the error form, as in *b.. thicker bush* or *bi.. thicker bush*, and a class of repairs following the completed error form as in *bicker...thicker bush.* All other cases are so rare that they are statistically irrelevant.

3. Testing the hypotheses

3.1 *Hypothesis 1: Other things being equal, there are more nonword than real-word errors early interrupted. There is no such difference between nonword and real-word repaired completed speech errors.*

This hypothesis will first be tested on elicited speech errors and then on spontaneous speech errors.

3.1.1 *Elicited speech errors*

For the current purpose the speech errors are taken from two published experiments as reported in Nooteboom (2005b) and in Nooteboom and Quené (2008). The first publication describes only a single classical SLIP experiment, the latter described two such experiments. However, of these two experiments for testing this hypothesis only the data of Experiment 1 will be used. The reason is that in Experiment 2 time pressure was on purpose artificially removed, whereby the conditions for normal monitoring of inner speech were changed. In order to obtain a reasonably large set of experimental data the data of Nooteboom (2005b) and Experiment 1 of Nooteboom and Quené (2008) are collapsed. These experiments are in all relevant aspects comparable.

As earlier indicated, each early interruption, whether or not followed by an overt repair, is interpreted as a reaction to error detection in inner speech and each overt repair of a completed consonant exchange is interpreted as a reaction to error detection in overt speech. It should be noted that interruption itself marks an error as detected, whether or not it is repaired later. For the completed exchanges in a SLIP experiment, the only evidence of detection by the monitor is the presence of a repair. For the main pattern in the data see Table 1.

Table 1. Numbers of unrepaired completed exchanges, early interruptions, and repaired completed exchanges, from two SLIP experiments combined (see text)

	unrepaired completed exchanges	**Interruptions**	**repaired completed exchanges**
Real-word errors	74	65	19
nonword errors	41	104	10

The assumption made here is that unrepaired completed exchanges have not been detected by the speaker at all, interruptions have been detected in inner speech, and repaired completed exchanges have been detected in overt speech. The point of interest is whether in inner speech, in overt speech, or in both, nonword errors are detected more frequently than real-word ones. To find out, the distributions of real-word and nonword errors for interruptions and for repaired completed errors are compared with the same distribution for the undetected errors. Obviously, there may be uncertainty about the lexicality of an error when it has been early interrupted. However, these errors were made in two conditions, one condition eliciting real-word errors, the other eliciting nonword errors. It is assumed here that the interrupted errors conform to the eliciting condition. The data clearly show that the distribution over real-word and nonword errors for the interruptions differs significantly from the same distribution for undetected errors (Fisher's exact test: $p < .0001$). The distribution for interruptions can only be explained by assuming that a lexicality criterion has been operative in monitoring inner speech. The data also show that the distribution over real-word and nonword errors for the repaired completed errors does not differ from the same distribution for undetected errors (Fisher's exact test: $p = 1$). This can only be explained by assuming that in monitoring overt speech no lexicality criterion is applied. As predicted, early interruptions reflect a criterion of lexicality, repairs of completed exchanges do not.

3.1.2 *Speech errors in spontaneous speech*

To find out whether the pattern in experimentally elicited errors is supported by the pattern in speech errors made in spontaneous speech, a count was made in the Utrecht corpus of speech errors of how many segmental speech errors were interrupted before completion of the error form, separately for errors that, when completed, would have generated a real word and for errors that would have generated a nonword. It was also counted how often the error form was completed but the speaker had stopped after the error form for making a repair. In the collection there are 1100 segmental speech errors. After removal of all those cases that for a variety of reasons were difficult to classify, there remained 744 segmental errors.

Table 2. Numbers of unrepaired completed segmental speech errors, early interruptions, and repaired segmental speech errors, taken from a collection of speech errors made in spontaneous Dutch (see text)

	unrepaired completed exchanges	Interruptions	repaired completed exchanges
Real-word errors	163	52	121
nonword errors	158	127	123

The relevant classification of these errors, comparable to the classification in Table 1, can be seen in Table 2. Again, as with the elicited speech errors, unrepaired completed exchanges are interpreted as not being detected by the speaker, interruptions are interpreted as detected in inner speech, and repaired completed exchanges as errors detected in overt speech. It might be observed that the numbers of unrepaired completed exchanges do not seem to show much of a lexical bias. However, this is misleading. The probability of a real-word error to be made is roughly 50% for *CVC* words, but rapidly decreases with word length. In the corpus words of all possible lengths were included. Average length of the words in the corpus that were involved in segmental speech errors is roughly 6 phonemes. The lexical bias must be considerable if the numbers for real-word and nonword errors are about equal. Again, as earlier with the elicited errors, there may be uncertainty about the lexicality of interrupted errors. An English example might be *ba.. marvelous boat*. In each individual interruption the lexicality was judged from the combination of speech error and repair. In this example, the interrupted speech error is assumed to be *barvelous*, which is a nonword. It may be seen that the pattern of the data in Table 2 is very similar to the pattern found for elicited speech errors in Table 1. The distribution over real-word and nonword errors for the interruptions differs significantly from this distribution for the undetected completed speech errors (Fisher's exact test: $p < .0001$). The distribution for the interruptions suggests that in monitoring inner speech a criterion of lexicality is applied.

The distribution over real-word and nonword errors for the repaired completed speech errors is identical with this distribution for undetected completed speech errors (Fisher's exact test: $p = .93$). This suggests that in monitoring overt speech no criterion of lexicality is applied.

3.2 *Hypothesis 2. Overt repairs following interrupted speech errors have shorter offset-to-repair times than overt repairs following completed speech errors.*

As there are no speech wave forms available for the speech errors made in spontaneous speech, this hypothesis will only be tested with the corpus of elicited speech errors.

3.2.1 *Elicited speech errors*

The hypothesis was tested on repaired interruptions and repaired completed exchanges elicited in the experiment described in Nooteboom (2005b) and Experiment 1 as described in Nooteboom and Quené (2008). Although it is assumed that all interruptions have been detected in inner speech, it is not necessarily so that all experimentally elicited interruptions are also repaired, although most are. Probably due to the time pressure in SLIP experiments, of all early interruptions 13% are left without repair and 87% are repaired. Because the dependent measure of interest here is the offset-to-repair time, interruptions not followed by a repair are left out. Only those interruptions were included in which the spoken fragment retained at least the greater part of the vowel. This was done because the same set of cases was used to test the hypotheses on intensity difference and pitch difference between error and repair (see below). The greater part of the vowel was needed for this comparison in intensity and pitch. A "repair" was in most cases identical with the correct target, but in some cases it was not. It was required, however, that the vowels of the spoken error fragment and of the first syllable of the "repair" were identical. In some included cases the interrupted "error" was identical with the beginning of the correct target, and then followed by the correct target as in *kee...keel taart*. There remained 108 interruptions in the two experiments combined that were suitable for inclusion.

For the completed speech errors only those "repaired" errors were included in which the error consisted of a full *CVCCVC* form, which was either the correct target or the elicited consonant exchange or yet another *CVCCVC* form. It was required that "error" and "repair" had the same vowel in the first syllable. Quite a few cases were included in which the correct target was repeated. No indication was found that such repetitions differed in any way from real errors followed by the correct targets as repairs. Most of these repetition cases were preceded by an early interruption, as in *gfeit goud..feit goud*. There remained 90 repaired completed errors. The average offset-to-repair time for the 108 repaired interruptions is 182 ms, ranging from 0 to 944 ms, with a standard deviation of 199 ms, the average offset-to-repair time for the 90 repaired completed speech errors is 398 ms, ranging from 0 to 1371 ms, with a standard deviation of 285 ms. Both means, 182 ms for interruptions and 398 ms for completed errors, differ significantly on a Student's t test for independent means ($t = -6.25$ sd = 242. df =196; $p < .0001$). These distributions of offset-to-repair times confirm that on the average repairs following early interruptions come much faster than repairs following completed exchanges.

It should be noted that an offset-to-repair time of 0 ms in the case of a completed exchange such as *bood gear* instead of *good beer* corresponds to a much later reaction to the error than an offset-to-repair time of 0 ms in case of an early interruption such as *boo..good beer*, simply because of the time it takes to pronounce the completed exchange. For the *CVCCVC* word pairs used in this experiment the difference is virtually always in the order of 500 ms. The estimated average difference between the two classes of speech errors in the time needed to react to the overt error therefore is in the order of (398+500)−182≅716 ms, and the range of times available for the subject to detect the error and plan a repair runs from 500–1870 ms. Obviously in most completed errors the speakers had plenty of time to detect the error in overt speech and plan a repair. Even in the four cases in which the measured offset-to-repair time indeed was 0 ms, the speakers had some 500 ms during the pronunciation of the error for detecting the error in the first consonant and planning a repair. It has recently been shown that the pronunciation of the error form and planning a repair often go on in parallel (Hartsuiker, Catchpole, De Jong & Pickering 2008). Clearly, on the average repairs of errors detected in inner speech come fast, repairs of errors detected in overt speech come very much slower, probably not only because this time is needed for error detection, but also because the speaker needs this time to plan the repair.

3.3 *Hypothesis 3. Overt repairs following interrupted speech errors have more intensity and higher pitch than the errors themselves. Overt repairs following completed speech errors have less intensity and lower pitch than the errors themselves.*

As no speech wave forms are available for speech errors made in spontaneous Dutch, this hypothesis was only tested with elicited speech errors.

3.3.1 *Elicited speech errors*

This hypothesis was tested with the same set of cases used for measuring the offset-to-repair times above. In all interruption-plus-repair combinations and all completed-exchange-plus-repair combinations, using the standard intensity display in the PRAAT speech analysis software, the maximum of intensity both in the (first) vowel of the error and in the first vowel of the repair was assessed in decibels SPL, and then the latter was subtracted from the first. Similarly, the pitch in the (first) vowel of the error and the first vowel of the repair, in both cases at the moment of maximum intensity, was assessed in Hz by taking the inverse of the pitch period duration in seconds. Here also the value found in the repair was subtracted from the value found in the error. The hypothesis states that for interruptions both intensity and pitch have higher values in the repair than in the error, whereas for completed exchanges, both intensity and pitch have lower values in the repair than in the error. The pattern in the data is displayed in Figure 1.

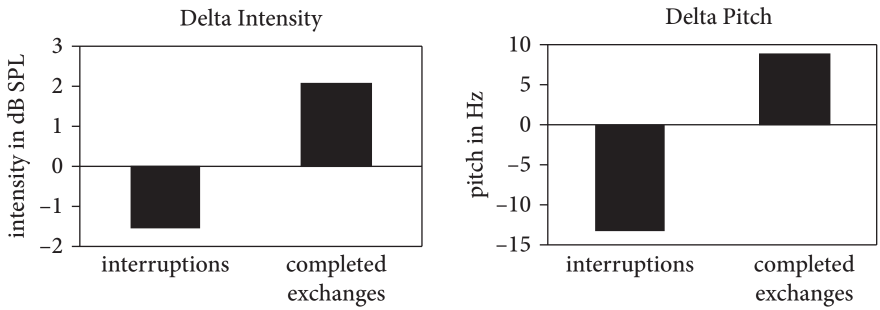

Figure 1. Average difference in intensity (left) and pitch (right) between error and following repair, separately for interruptions and completed exchanges

The average difference in intensity is −1.56 dB (sd = 3.5 dB) for interruptions and 2.04 dB (sd = 4.36 dB) for completed exchanges, suggesting that indeed for interruptions on the average intensity is higher in the repair than in the error, and for completed exchanges the average intensity is lower in the repair than in the error. The difference between −1.56 dB for interruptions and 2.04 dB for completed exchanges is significantly different on a student's t test for independent means (t = 6.46; sd = 3.92; df = 196; p <. 001), confirming that indeed the speakers, after having interrupted speech errors, mostly made repairs with more intensity than the errors, and after having completed exchanges they made repairs with less intensity than the errors.

The average difference in pitch at the moment of highest intensity was −12.93 Hz (sd = 1.713 Hz) for interruptions, and 8.62 Hz (sd = 2.07 Hz) for completed exchanges. The difference between −12.93 Hz for interruptions and 8.62 Hz for completed exchanges is significantly different on a student's t test for independent means (t = 8.09; sd = 18.7; df = 196; p <. 0001). Apparently, speakers generally speak the repair on a higher pitch than the error when they are repairing an interruption, and on a lower pitch than the error when they are repairing a completed speech error.

The classification of speech errors as repaired interruptions and repaired completed speech errors most likely does not fully match with the distinction between repairs planned in inner speech and repairs planned in overt speech. Possibly, repairs of interruptions with long offset-to-repair times are only planned during overt speech, and repairs of completed errors with very short offset-to-repair times were already planned during inner speech. If so, one expects that on the average the intensity difference and pitch difference is a function of the offset-to-repair time: the contribution of misclassified errors becomes bigger when offset-to-repair times get longer for interruptions and shorter for completed errors. Therefore differences in intensity and pitch were correlated with the offset-to-repair times, using Pearson product moment correlations. This was done separately for interruptions and completed speech errors. No significant correlation was found for the interruptions (intensity difference: $r = -.147$; $t = -1.53$; $df = 106$; $p >. 1$; pitch difference $r = .0999$; $t = 1.03$; $df = 106$; $p > .3$). However, for the repaired completed errors there are significant positive correlations for both intensity difference ($r = .286$; $t = 2.802$; $df = 88$; $p <. 007$) and pitch difference ($r = .3$; $t = 2.999$; $df = 88$; $p < .004$). These correlations suggest that both intensity and pitch in the repair tend to become lowered more relative to the error, as the offset-to-repair time increases. This suggests that repairs of completed speech errors with very short offset-to-repair times more often than repairs with long offset-to-repair times are reactions to inner speech and thus suffer from the tendency to have raised instead of lowered intensity and pitch.

3.4 *Hypothesis 4. Overt repairs following interrupted speech errors are much less often accompanied by editing expressions than overt repairs following completed speech errors.*

This hypothesis will first be tested with elicited speech errors and then with speech errors from spontaneous speech.

3.4.1 *Elicited speech errors*

As it happens, in the repairs of elicited speech errors editing expressions are on the whole not very frequent. In order to have as many cases as possible to test the hypothesis statistically, all repaired interrupted errors and all repaired completed *CVCCVC* speech errors taken from three experiments, viz. the experiment described in Nooteboom (2005b) and both experiments described in Nooteboom and Quené (2008), not only the many completed and interrupted exchanges in the test condition, but also the much less frequent completed and interrupted exchanges in the base-line condition, i.e. the condition where no segmental biasing was used to elicit exchanges. The relevant pattern in the data is given in Table 3.

Table 3. Repaired interruptions and repaired completed *CVCCVC* speech errors taken from three experiments (see text), separately for repairs accompanied by an editing expression and repairs not accompanied by an editing expression

	interruptions	**completed exchanges**
editing expression	15	15
no editing expression	230	32

In these three experiments combined and the test and base-line conditions combined there are 245 interrupted speech errors. Only 15 (6%) of these are accompanied by an editing expression. Editing expressions are *er, hhh, sorry, of zo iets (or something like it), no, mmm, hm, umnja, kweenie (dunno), geen idee (no idea)*. In the three experiments and the test and base-line conditions combined there are 221 completed exchanges. Of these only 47 are followed by a repair. Very likely, most of the other 174 exchanges were not detected by the speaker.

Of the 47 repairs of completed exchanges, 15 (32%) are accompanied by an editing expression. Editing expressions are *ja (yes), er, pfhum, mm, hi, andersom (the other way round), ahh..sorry, oh, hahaha, correctie (correction), hm*. The distribution of the numbers of detected speech errors with and without an editing expression is significantly different for interruptions and completed exchanges (Fisher's exact test: $p < .0001$). As predicted, editing expressions are rare for repairs of interruptions and relatively frequent for repairs of completed exchanges. This supports the idea that speakers have a tendency to draw the attention of their audience away from the observable consequences of speech errors detected in inner speech, but tend to direct the attention of their audience to the speech errors they have detected in their overt speech.

Following the same reasoning as applied earlier to the differences in intensity and pitch between error and repair, one would predict that the probability of a repair being accompanied by an editing expression increases with increasing offset-to-repair time. It turns out, however, that far too few editing expressions are made to test this hypothesis statistically.

3.4.2 *Speech errors in spontaneous speech*

Numbers of repaired segmental speech errors with and without an editing expression were counted in the Utrecht corpus of speech errors, separately for early interrupted speech errors and completed speech errors. Early interrupted here means that at most the initial CV of the error form was pronounced before interruption. Completed speech errors were all those cases were the error form was completed, and the speaker stopped only then for making a repair. It was found that in repairing segmental speech errors speakers hardly ever spoke more than the error form before stopping. They also included hardly ever more than the correct target form itself in the repair (See Nooteboom 2005a: 174–175). There are 122 repaired early interruptions in the collection. The relevant data are given in Table 4.

Table 4. Repaired interrupted and repaired completed segmental speech errors taken from a corpus of speech errors made in spontaneous Dutch (see text), separately for repairs accompanied by an editing expression and repairs not accompanied by an editing expression

	interruptions	completed exchanges
editing expression	5	38
no editing expression	117	210

Only 5 (4%) of the 122 repairs following interrupted speech errors are accompanied by an editing expression. Thirty-eight (15%) of the 248 repairs following completed speech errors are accompanied by an editing expression. The distribution of overt repairs with and overt repairs without an editing expression is significantly different for early interruptions and completed speech errors (Fisher's exact test: $p < .002$). Obviously repairs of interrupted errors are only rarely accompanied by editing expressions and repairs of completed errors are more often accompanied by editing expressions. This confirms that speakers strive towards making a repair as inconspicuous as possible for their audience when they have detected an error in inner speech, but have a tendency of making a repair conspicuous for the listeners when they have detected an error in their overt speech. It should be noted, though, that speech errors made in spontaneous speech seem even less often accompanied by editing expressions than experimentally elicited speech errors.

4. Discussion

In the introduction to this paper it was argued that monitoring for speech errors in inner speech potentially differs from monitoring for speech errors in overt speech, because these two forms of self-monitoring have different functions. When an error is detected in inner speech the speaker strives towards preventing this error from becoming public, preferably with as few perceivable consequences as possible. Because there is very little time to do this, monitoring for speech errors in inner speech is under time pressure, whereas monitoring for errors in overt speech is not. One consequence of this is, as has been confirmed in the present analysis, that, other things being equal, more nonword than real-word segmental errors are detected and then early interrupted, but that the probability of being detected and repaired is the same for nonword and real-word completed segmental errors. The reader may recall that this state of affairs was predicted in the introduction to this paper from the earlier published finding that under time pressure monitoring for speech errors detects more nonword than real-word errors, but when the time pressure is removed, this is not so (Nooteboom & Quené 2008), together with the above-mentioned assumption that monitoring inner speech operates under time pressure, but monitoring overt speech does not. It should be pointed out, however, that the current data have confirmed something that is not easily explained theoretically. Recently, Nozari and Dell (submitted) have gone a long way, both in arguing and in demonstrating experimentally, that a "lexical editor", being part of a monitor for speech errors, makes little sense, because (1) a lexical editor is too slow to be of much use, and (2) it adds little or nothing to an editor that compares each potential error with the correct target. These authors assume that the working of a "lexical editor" as part of the monitor is similar to a "lexical decision task". In a lexical decision task participants push one of two buttons, *word* or *nonword*. Nozari and Dell point out that, other things being equal, in a lexical decision task reaction times are much longer than in a task in which participants push one or two buttons, to indicate whether a stimulus word is the same as or different from a pre-specified word target (cf. Foss & Swinney 1973). The problem here is (1) that in theories about monitoring for speech errors, the assumption of a direct comparison between error and target is needed anyway, to explain the frequent and rapid detection of real-word segmental errors, and (2) that currently the lexical decision task is the only model we have of a monitor that detects more nonword than real-word errors. The earlier finding that more nonword than real-word errors are detected in inner speech under time pressure but not under relaxed conditions (Nooteboom & Quené 2008), and the current finding that more nonword than real-word errors are detected in inner speech but not in overt speech, for the time being seems theoretically awkward. It is as yet unclear what monitoring mechanism would detect both real-word errors and nonword errors, but under time pressure more nonword errors than real-word errors, and without time pressure equally many.

The idea underlying this paper is that there are two different "strategies" in monitoring for speech errors, one strategy for inner and one for overt speech. This terminology suggests that in both situations some measure of flexible attentional control is involved. The validity of this assumption is possibly not self-evident for all readers. The reader might perhaps have thought that, although possibly monitoring for speech errors in overt speech is semi-conscious and under flexible attentional control, monitoring for speech errors in inner speech surely is fast and automatic and cannot be under attentional control. However, there are quite many published experiments demonstrating that, although the participants may not be aware of this, some measure of attentional control is also exerted over monitoring for speech errors in inner speech. Examples are Baars et al. (1975) and Hartsuiker, Corley and Martensen (2005). Both papers show that pre-articulatory editing is sensitive to the lexicality of the context. Motley (1980) demonstrated that the probability of a consonant exchange increases considerably when the error is preceded by word pairs creating a semantic bias for the error, such as *bad mug* turning into *mad bug*, when preceding word pairs included *angry insec*t and *irate wasp*. Motley (1980) also showed that a conversational setting affected the probability of specific errors: *shad bock* turning into *bad shock* became more likely when the participants expected an electrical shock, and *goxi firl* turning into *foxy girl* was stimulated by an attractive and provocatively attired experimenter. Another example is provided by Motley et al. (1982), who found an effect of the taboo status on the probability of error detection in inner speech. Also the finding by Nooteboom and Quené (2008) that time pressure affects the relative probabilities of error detection suggests some measure of attentional control. All this is relevant, because the current attempt to show that there are different behavioural patterns in repairing errors detected in inner speech and repairing errors detected in overt speech, derives its logic precisely from the assumption that to a certain extent these different strategies are strategic, stemming from communicative needs. It should also be noted that the actual patterns found, seem to make sense in this context.

If indeed speakers strive to hide, or make less noticeable, for their listeners the perceivable consequences of errors detected in inner speech, they are well advised to keep the offset-to-repair times short. This is precisely what they often seem to do, much more often than when they repair errors detected in overt speech. Admittedly, there are also cases of repairs of interrupted speech errors with longish offset-to-repair times. But obviously, precisely in studying elicited errors of speech and their repairs, one is dealing with an attentional control system that is overburdened. The detection of an error in inner speech, and the decision to interrupt the error as soon as possible, compete under time pressure with the control needed to come up with the correct target as a repair. Interestingly, there are quite a few cases in which the speaker after an interruption hesitates, and then repeats the error that apparently still internally competes with the correct target.

If the speaker (or some subconscious part of the speaker's mind) wants to distract the attention of listeners from the error fragment that was just inadvertently spoken, it may be a good idea to speak the following repair not only fast, but also both louder and on a higher pitch than the error fragment itself was spoken. This is precisely what speakers do. This does not mean that the first syllable of the repair is getting a pitch accent, where the error fragment had not. Virtually all participants in the speech error elicitation experiments speak all regular responses with something that might best be described as double-stress. The impression is definitely that speaking the repair louder and higher-pitched than the error, holds for both syllables, and is not a function of making a pitch accent. In this context it may be relevant that Shattuck-Hufnagel and Cutler (1999) showed results on taped speech errors in spontaneous speech, suggesting that repairs of segmental errors are not marked with an (extra) pitch accent, whereas repairs of lexical errors, involving morphemes or whole words as displaced units, are more often marked with an extra pitch accent.

If the speaker has made a full error, that can hardly have been missed by the audience, he or she should not attempt to hide the fact that an error has been made, but instead should focus the attention of the audience on both the error and the repair. As the error has already been spoken, the only way to do this prosodically is by adapting the prosody of the repair. This can most easily be done by deviating from the often repeated sing song prosody of the regular responses, that of course was also given to the error (because when the error was made it was not yet detected as error). Speaking with somewhat less intensity and on a somewhat lower pitch is a most economical way to deviate from the regular pattern. And this is what the speakers generally do. The lowered pitch makes the impression of some form of finality, as if the repair is more of a final response than the error was. This seems communicatively effective under the circumstances.

Unfortunately, it must remain unclear here how speakers treat their repairs of interrupted and completed speech errors prosodically in spontaneous speech. Shattuck-Hufnagel and Cutler (1999), in investigating acoustically the accent patterns in repairs of segmental and lexical speech errors, did not make a distinction between interruptions and completed errors. This is something awaiting further research on a collection of taped speech errors in spontaneous speech. With respect to the use of editing expressions like *sorry, hahaha, wrong, no, er,* etc. speakers of both experimentally elicited speech errors and speech errors in spontaneous speech behave in a communicatively effective way. They add rarely editing expressions to repairs of errors detected in inner speech and more often to repairs of errors detected in overt speech. Levelt (1983) found, in a task where participants had to describe simple networks, that 62% of all error repairs (including not only repairs of segmental errors but also of errors involving lexical units) were accompanied by editing expressions. This is much more than found in the current analysis. Probably, in the SLIP experiments the use of editing expressions is constrained by the time pressure the participants are under. In the corpus of spontaneous speech editing expressions may be rare because they are not always observed or noted down by the collectors. Also in this respect a re-analysis of existing corpora of taped speech errors in spontaneous speech would be welcome. The scarcity of editing expressions in the current data also prohibits further analysis of the possible differences in the use of specific editing expressions, although it seems to be case that most editing expressions tend to follow the repair, but that *uhh* or *er* more frequently precedes the repair. This confirms that *uhh* or *er* first of all serves to hold the floor while a repair is being planned (cf. Levelt 1989: Chapter 12).

The current analysis supports an important aspect of Levelt's perceptual loop theory of monitoring, viz. that both inner and overt speech are being monitored for speech errors (see also Hartsuiker & Kolk 2001; Hartsuiker, Kolk & Martensen 2005). There is at least a delay of 200 or 250 ms between the two (Hartsuiker & Kolk 2001), and potentially much more, depending on how much material is buffered in inner speech. If an error is not detected in inner speech, detection of this error in overt speech is of course perfectly appropriate. However, it is unlikely that speakers can monitor inner and overt speech with equal attention at each moment in time. It has been assumed that monitoring for speech errors is under attentional control (Hartsuiker, Kolk & Martensen 2005) so that the speaker may either direct his or her attention more to inner speech or more to overt speech. In this respect it is revealing that the current analysis shows that in SLIP experiments there are far more interrupted errors (169) than repaired completed errors (29). Apparently, during these experiments the participants' attention is focused on inner speech. But the data on speech errors in spontaneous speech show that there are more repaired completed speech errors (244) than interrupted speech errors (179), showing that in normal spontaneous speech monitoring attention is certainly not less for overt speech than for inner speech. This makes sense for two reasons. First, in SLIP experiments participants have little time to detect and repair errors in overt speech.

Immediately after a response, the next series of biasing stimuli starts. Secondly, in SLIP experiments participants do not have to worry about detecting lexical errors, but in spontaneous speech this is an important aspect of monitoring. These lexical errors are probably much harder to detect in inner speech than in overt speech, because the phrasal context is needed for error detection (cf. Nooteboom 2005a).

The current findings complicate our view of self-monitoring. Whereas for a long time it was thought that self-monitoring operates in the same way whether it is directed at inner speech or at overt speech, using a few straightforward criteria for filtering out speech errors (cf. Levelt 1989), it must now be admitted that this view was too simple. The ways self-monitoring for speech errors operates are controlled by different goals in inner and in overt speech, and these different goals lead to different behavioural patterns. It seems indeed the case that speakers attempt to make speech errors detected in inner speech as little conspicuous as possible and to draw the listeners' attention to repairs of speech errors detected in overt speech. This explains the different behavioural patterns found in these two classes of overt repairs of detected speech errors.

References

Baars, B.J., Motley, M.T. & MacKay D. 1975. Output editing for lexical status from 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.

Del Viso, S., Igoa, J.M. & Garcia-Albea, J.E. 1991. On the autonomy of segmental encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research* 20: 161–185.

Dell, G.S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93: 283–321.

Dell, G.S. & Reich, P.A. 1981. Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 20: 611–629.

Foss, D. & Swinney, D. 1973. On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior* 12: 246–257.

Garrett, M.F. 1976. Syntactic process in sentence production. In *New Approaches to Language Mechanisms*, Wales, R.J. & Walker, E.C.T. (eds), 231–256. Amsterdam: North-Holland.

Hartsuiker, R.J. 2006. Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes* 21(7–8): 856–891.

Hartsuiker R.J., Antón-Méndez I., Roelstraete B. & Costa A. 2006. Spoonish spanerisms: A lexical bias effect in Spanish. *Journal of Experimental Psychology, Learning, Memory and Cognition* 32 (4): 949–53.

Hartsuiker, R.J., Catchpole, C.M., De Jong, N.H. & Pickering, M.J. 2008. Concurrent processing of words and their replacements during speech.

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

Hartsuiker, R.J. & Kolk, H.J. 2001. Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology* 42: 113–157.

Hartsuiker, R.J., Kolk, H.J. & Martensen, H. 2005. Division of labor between internal and external speech monitoring. In *Segmental Encoding and Monitoring in Normal and Pathological Speech,* R. Hartsuiker, Y. Bastiaanse, A. Postma & F. Wijnen (eds), 187–205. Hove: Psychology Press.

Kolk, H.J. 1995. A time-based approach to agrammatic production. *Brain and Language* 50: 282–303.

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

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation*. Cambridge MA: 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.

Motley, M.T. 1980. Verification of 'Freudian slips' and semantic prearticulatory editing via laboratory-induced spoonerisms. In *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*, V.A. Fromkin (ed.), 133–148. New York NY: Academic Press.

Motley, M.T., Camden, C.T. & Baars, B.J. 1982. Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 21: 578–594.

Nooteboom, S,G. 1980. Speaking and unspeaking: Detection and correction of segmental and lexical errors in spontaneous speech. In *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*, V.A. Fromkin (ed), 87–95. New York NY: Academic Press.

Nooteboom, S.G. 2005a. Listening to one-self: Monitoring speech production. In *Segmental Encoding and Monitoring in Normal and Pathological Speech,* R. Hartsuiker, Y. Bastiaanse, A. Postma & F. Wijnen (eds), 167–186. Hove: Psychology Press.

Nooteboom, S.G. 2005b. Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication* 47(1–2): 43–58.

Nooteboom, S.G. 2007. Alphabetics. From phonemes to letters or from letters to phonemes? *Written Language and Literacy* 10(2): 129–144.

Nooteboom, S.G. & Quené, H. 2008. Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in segmental speech errors. *Journal of Memory and Language* 58: 837–961.

Nozari, N. & Dell, G.S. 2008. More on lexical bias. How efficient can a "lexical editor" be? Submitted for publication.

Shattuck-Hufnagel, S. & Cutler, A. 1999. The prosody of speech error corrections revisited. Proceedings of the International Congress of Phonetic Sciences, August 1–6, 1999, 1483–1486, in San Francisco. San Francisco: Regents of the University of California. (on CD-ROM).

Stemberger, J.P. 1985. An interactive activation model of language production. In *Progress in the Psychology of Language,* Vol. 1, A.W. Ellis (ed), 143–186. Hillsdale NJ: Lawrence Erlbaum.