Journal of Memory and Language 105 (2019) 43–59

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Temporal aspects of self-monitoring for speech errors

Sieb Nooteboom*, Hugo Quené

Utrecht institute of Linguistics OTS, Utrecht University, the Netherlands

ARTICLE INFO

Keywords:
Speech errors
Self-monitoring
Timing
Selective attention
Tongue twisters

ABSTRACT

We report two four-word tongue twister experiments eliciting consonantal errors and their repairs, in word initial and medial positions, testing some predictions relating to temporal aspects of self-monitoring. Main findings: (1) After internal error detection interrupting the speaking process takes more time than speech initiation of the error form. This implies that “covert repairs” are rare. (2) Word onset-to-cutoff times are longer for medial than for initial errors. This implies that scanning internal word forms for errors takes time. (3) Cutoff-to-repair times of 0 ms are overrepresented. This shows that often repairs are available at interruption. (4) Cutoff-to-repair times are longer for medial than initial consonants. This shows that repairing takes more time for medial than for initial errors. (5) Detection rate decreases from early to late within word forms. Temporal aspects of self-monitoring suggest time-consuming scanning of internal word forms, strategic postponement of interruption, and variations of selective attention.

Introduction

This paper is about self-monitoring for speech errors. Observed frequencies of sound form errors such as *bood geer* instead of *good beer* may be distorted by processes of detection and repair during self-monitoring. Distortion would be serious if under some conditions more errors would be detected and repaired covertly by the speaker, i.e. before speech initiation, than under other conditions. The possible frequent occurrence of such unobservable repaired speech errors, called “covert repairs“ by some (Kolk & Postma, 1997; Levelt, 1989; Postma & Kolk, 1993) and “prepairs” by others (Schlenck, Huber & Wilmes, 1987), is suggested by two aspects of a computational implementation by Hartsuiker and Kolk (2001, from now on called the H&K model) of the dual loop perceptual theory of self-monitoring by Levelt (1989) and Levelt, Roelofs and Meyer (1999). First, the H&K model with reasonable parameter settings, predicts that after detection of speech errors in internal speech, that is before speech initiation, the moment of interruption of the speaking process would come some 200 ms later than the moment of initiating articulation of the error. However, the variance in error-to-cutoff times is such that the distribution of error-to-cutoff times would probably be truncated at 0 ms. Of course, all cases in which the virtual error-to-cutoff time is shorter than 0 ms represent covert, unobservable errors. Second the H&K model assumes that planning pronunciation is faster than pronunciation itself. Therefore, in multi-word

* Corresponding author at: Cor Ruyslaan 20, 3584 gd UTRECHT, the Netherlands.
E-mail address: s.g.nooteboom@uu.nl (S. Nooteboom).

https://doi.org/10.1016/j.jml.2018.11.002
Received 5 January 2018; Received in revised form 4 November 2018
Available online 04 December 2018

utterances, later words are buffered for a longer period of time than earlier words, leading to more covert repairs for later than for earlier words. A table with the most important parameters of the H&K model is given in Footnote.[1]

Below we have confronted the H&K model with the results of later research. This has led to some proposed changes in and additions to the model and to some new predictions on temporal aspects of detecting and repairing speech errors in self-monitoring some of which differ from predictions that can be derived from the H&K model. Our predictions (developed below) were tested in two experiments eliciting sound form errors and their repairs in four-word tongue twisters. Based on recent publications on self-monitoring for speech errors and the current results we then, in the general discussion, give an account of how we think detecting and repairing sound form errors of speech in self-monitoring works. This account in some respects differs from current theories of self-monitoring. It capitalizes on competition between word form candidates all the way during selection of word form candidates, planning of pronunciation, articulation and beyond, and continuing into detecting and repairing sound form errors.

Some major assumptions of the Hartsuiker and Kolk (2001) implementation of the dual perceptual loop theory of self-monitoring in the light of later research, and some predictions

(1) *During self-monitoring speakers inspect both their internal speech, before speech initiation, and their overt speech, after speech initiation.* This basic tenet of the perceptual loop theory of self-monitoring (Levelt, 1989; Levelt, Roelofs & Meyer, 1999) was verified by Hartsuiker and Kolk (2001) by demonstrating that distributions of experimentally observed error-to-cutoff times and cutoff-to-repair times as reported by Oomen and Postma (2001) could only be simulated convincingly with the computational implementation of the dual perceptual loop theory if the implementation took both stages of self-monitoring into account. From this aspect of the H&K model Nooteboom and Quené (2017) predicted and confirmed, that distributions of error-to-cutoff times for sound form errors are bimodal, with two underlying gaussian distributions reflecting the two stages of self-monitoring. The time delay between internal and external self-monitoring was found to be c. 500 ms. This is considerably longer than one would predict from the H&K model. There is little doubt that there are indeed two stages of self-monitoring, one directed at internal and one directed at overt speech. This implies that error-to-cutoff times have a wide and non-normal distribution.

(2) *Both internal speech and overt speech are fed into the same speech comprehension system that is also used in perceiving other-produced speech: internal speech from a buffer with a representation of speech being prepared for articulation, overt speech via audition and the resulting auditory representation.* This prominent role of the speech comprehension system has come under attack in different ways. Nozari, Dell and Schwartz (2011) reported that in aphasics successful self-monitoring is correlated with production measures and not with perception measures. Nooteboom and Quené (2017) found that the detection of sound form errors is independent of auditory feedback, both in detecting errors in internal speech and, more spectacularly, in detecting errors in external speech. It was suggested that detection of sound form errors after speech initiation employs somatosensory and proprioceptive feedback from the articulators (cf. Hickok, 2012; Lackner, 1974; Pickering and Garrod, 2013). For discussion of other relevant literature with respect to the role of the speech comprehension system see Hartsuiker and Kolk (2001) and Nozari et al. (2011). One implication of these findings is that with respect to internal error detection, those temporal parameters in the H&K model that are related to speech comprehension should be interpreted with caution. These parameters are mainly related with how errors are detected. Temporal implications of the H&K model with respect to what happens after error detection might still be valid.

(3) *The output of the speech comprehension system is fed into a centrally located monitor, looking for speech errors (or other deviations from the speech plan).* A problem with this assumption is that it remains unclear what form the output of the speech comprehension would have, and with what it would be compared in a centrally located monitor where word forms have not yet been compiled. It seems a natural assumption that for example sound form errors are detected in comparing a word form candidate containing an error with a correct version of the word form candidate. In this respect a convincing and interesting production-based proposal has been made by Nozari et al. (2011) who implemented a domain-general conflict-based model of self-monitoring directed at conflict between simultaneously activated and competing responses. In the case of speech, these responses would be competing word candidates at the levels of both lexical selection and sound form selection (phonological encoding). This proposal capitalizes on the fact that models of word production generally assume that at various levels word candidates are competing for being selected (e.g. Levelt et al., 1999; Roelofs, 2003; Roelofs, 2005). Interestingly, competition between word candidates is often sustained during speech preparation, articulation and repair (Hartsuiker, Pickering & De Jong, 2005; Nozari, Freund, Breining, Rapp & Gordon, 2016; Nooteboom & Quené, 2017).

(4) *In case an error is detected by the monitor, immediately a command is issued to interrupt speech.* This assumption in the H&K model deviates from Levelt (1989) who assumed that a command to repair is only issued at the moment of interruption. The H&K model predicts that, apart from the considerable effect of noise in the system, after internal error detection, the speaking process is interrupted some 200 ms later than the moment of speech initiation. Given the wide distribution of error-to-cutoff times, this would explain that every now and then error-to-cutoff times can be very short. Nooteboom and Quené (2017) found that after internal error detection the estimated distribution of error-to-cutoff times runs from close to 0 ms to over 1000 ms. The very long error-to-cutoff times, not only after external but also after internal error detection, are in in line with results reported by Tydgat, Stevens, Hartsuiker and Pickering (2011) who found that, although speakers can stop very quickly after error detection, they do not always do this. Speakers can postpone interruption for strategic reasons, for example because no repair is available (see also Seyfeddinipur, Kita & Indefrey, 2008, for a similar proposal). This suggests that the interruption time predicted by the H&K model at best predicts a minimum value, not an average value. This has important consequences for our purpose.

[1] Basic durations of each time interval in speech generation and self-monitoring according to H&K. In the "Stage" column we added some terms between brackets for the sake of clarification. σ stands for syllable, ω for lexical item c.q. lexical form.

Stage	Symbol	Duration (ms)	Per unit
Phonological encoding	T_{phon}	110	σ
Selection (of action plan)	T_{sel}	100	ω
Command (to execute action plan)	T_{com}	100	σ
Audition (in case of external monitoring)	T_{sel}	50	ω
Parsing (either internal or perceived lexical form)	T_{pars}	100	ω
Comparing (encoded or perceived form with correct target)	T_{comp}	50	ω
Interrupting (execution of action plan or overt speech)	T_{int}	150	ω
Restart planning (of a repair)	$T_{restart}$	50	ω

Note (by H&K). "Restart planning" is a parameter that represents the duration of repeated execution of selection processes before phonological encoding minus the time benefit from priming the to-be-selected units.

If indeed speakers can postpone interruption because a candidate repair is not sufficiently activated, this means that the distribution of error-to-cutoff times is shifted toward longer durations, and error-to-cutoff times close to 0 ms would be rare (**Prediction 1**). Covert repairs would also be rare or would not exist at all. If so, the prediction by the H&K model that the relative number of covert repairs increases from earlier to later words in multi-word utterances, would remain untestable.

(5) *H&K are silent about the time course of error detection within lexical forms.* This is only reasonable, because their data set did not allow distinguishing between different positions in the word. Wheeldon and Levelt (1995) demonstrated in three phoneme and syllable detection experiments with Dutch participants reacting to silent Dutch translations of auditorily presented English words, that scanning internal word forms takes time. For our purpose it is relevant that, although scanning of internal word forms appeared to be faster than speaking aloud, at least the necessary time in scanning word initial syllables is in the same order of magnitude as speaking time for these same syllables: Reactions to syllable initial consonants were 150 ms later for second than for initial syllables in the case of initial stress and 95 ms later in the case of final stress. Scanning appeared to go faster for later syllables. Also, although the order of phonemes was reflected in reaction times, the actual durations of phonemes in their spoken forms were not. It seems not too far-fetched to assume that scanning internal word forms is comparable in detecting phonemes and in detecting sound errors. This, of course, has consequences for temporal aspects of self-monitoring. Particularly, under the assumption that interrupting the speaking process is initiated at the moment of internal error detection, we predict that, whereas *word onset*-to-cutoff times would be significantly longer for word-medial than for word-initial errors, *error*-to-cutoff times (measured from the onset of the error segment to the moment of interruption) would be roughly equal for both error types (**Prediction 2**).

(6) *The H&K model predicts that repairs are often available at the moment of interruption.* This prediction is reinforced by the demonstration by Tydgat et al. (2011) that speakers often postpone interruption, for example because there is no available repair. From this one would expect that the distribution of cutoff-to-repair times (measured from the moment of overt interruption to the onset of the repair) is strongly "censored" at 0 ms (**Prediction 3**). If so, this would mean that there are relatively many cases with effectively (and latently) negative values, corresponding to cases where a repair has come available to the mind of the speaker before speech interruption, and that all these unknown values are censored at the boundary value of 0 ms. The value of 0 ms should be overrepresented. However, we also expect that the amount of selective attention (see below under (7)) available for self-monitoring decreases from earlier to later during preparation for speaking of a word form. (Selective attention refers to the capacity for a process of reacting to certain stimuli selectively when several occur simultaneously; selective attention can be likened to the manner by which a bottleneck restricts the flow rate of a fluid). Therefore, we expect cutoff-to-repair times to be longer for word-medial than for word-initial errors (**Prediction 4**).

(7) *H&K are silent about the possible role of selective attention in self-monitoring.* Selective attention in self-monitoring potentially affects both rate of error detection and speed of processing (e.g. Bates & Stough, 1997; see also Roelofs, 2003, for a thorough discussion of the role of attentional control in explaining 50 years of results in the so-called "Stroop task", and Piai, Roelofs and Schriefers, 2012, for a discussion of the role of selective attention in competition between candidate responses in Picture Word Interference tasks). Nooteboom (2011) demonstrated that segmental speech errors are detected less frequently in phrasal lexical items (e.g. proverbs and cliché's) than in novel phrases. This was explained by assuming that amount of selective attention in self-monitoring for speech errors is inversely related to the predictability of the items to be monitored. Within word forms predictability of the next speech sound rapidly increases from early to late (Marslen-Wilson & Welsh, 1978) so that less and less selective attention is required. Therefore, we expect that, in line with the longer cutoff-to-repair times for word medial than for word initial errors, error detection rate rapidly decreases from word onset errors to errors later in the word (**Prediction 5**).

A similar argument could be made for error detection in multi-word utterances: Because in normal utterances generally predictability increases from earlier to later words, one would perhaps expect that error detection rate decreases from earlier to later words. This does not appear to be the case: Levelt (1983, 1989, Figure 12.2) found in an experiment involving pattern descriptions that frequency of repairs of misspoken color names rapidly increased from early to late within constituents. He explained this assuming that early in the constituent more selective attention is needed for message planning. Later in the constituent more selective attention is available for self-monitoring. These arguments are not necessarily applicable to the kind of tongue twisters used in the two experiments reported below, because there is no syntactic and semantic structure. In such tongue twisters many speech errors are made, often more than one per utterance. Therefore, one expects selective attention to become more and more occupied from earlier to later words. This would possibly lead to a decrease in detected errors from earlier to later words (**Prediction 6**).

In sum, we have made the following six predictions for the experiments reported below:

Prediction 1: Error-to-cutoff times have a rather wide distribution, for two reasons. One is the 500 ms delay between internal and external error detection, the other is the strategic postponement of interruption. Error-to-cutoff times run from close to 0 ms when a repair is available before or at speech initiation, to much longer values when interruption is postponed. Missing values below 0 ms are rare.

Prediction 2: Word onset-to-cutoff times are significantly longer for word-medial than for word-initial sound form errors and error-to-cutoff times are roughly equal for word-initial and word-medial sound form errors.

Prediction 3: The distribution of cutoff-to-repair times is censored at 0 ms, making values of 0 ms overrepresented.

Prediction 4: Cutoff-to-repair times are significantly longer for word-medial than for word-initial errors.

Prediction 5: Rate of error detection is lower for word-medial than for word-initial consonant errors.

Prediction 6: Rate of error detection decreases from earlier to later words.

Below we will report two experiments in which the above predictions are put to the test.

Experiment 1

Experiment 1 was originally not designed to test the above predictions. It was designed to investigate the rate of single segment interactional speech errors as a function of segment position in the word and segment position relative to stress (cf. Shattuck-Hufnagel, 1992). However, it appeared to us that the results of such an experiment are not easy to interpret when the effects of self-monitoring on observable error frequencies are unknown. As it happens, the design of the experiment is suitable to investigate effects of the time course of self-monitoring and variations of selective attention available for self-monitoring on observed error rates. This we set out to do. The reader may note that most of the predictions were made before we knew the relevant structure of the data. In this sense it is not the case that our "predictions" are rather interpretations of earlier findings.

Method of Experiment 1

The experiment consists of two parts. One part is a replication of Experiment 2 reported in Shattuck-Hufnagel (1992), but this time in Dutch. The other part is a modification of Shattuck-Hufnagel's experiment by employing two-syllable Dutch words only. The basic idea of the entire experiment is to elicit segmental errors by having participants rapidly and repeatedly speak word sequences that have properties of tongue twisters, and then compare error frequencies between conditions that differ in what properties the potentially interacting consonants share or not share, in particular word onset position and pre-stress position. In one half of the experiment all sequences of four words had an initial and final monosyllabic word and two intermediate disyllabic words, replicating Shattuck-Hufnagel's (1992) Experiment 2. We will refer to these stimuli as the "1 + 2 + 2 + 1" stimuli. In the other half of the experiment all sequences of four words had disyllabic words only. We will refer to these stimuli as the "2 + 2 + 2 + 2" stimuli. This set-up makes it possible to compare a situation in which consonants sharing pre-stress position were in different positions within the word, viz. word initial and word medial, with a situation in which the consonants sharing pre-stress position were in the same position in the word, which was either word initial or word medial. It also makes it possible to compare segmental interactions between competing words similar or dissimilar in stress pattern, and words in four different positions in the utterance.

Stimuli

A basic unit in constructing the stimuli for the experiment was a quartet of stimuli for the four sharing conditions B, W, S, N, as exemplified for Dutch in the following two quartets, one for the "1 + 2 + 2 + 1" stimuli, and one for the "2 + 2 + 2 + 2" stimuli. In both quartets the potentially interacting consonants are **w** and **r**. The third consonant used by Shattuck-Hufnagel for eliciting unexpected errors, we do not indicate here because this gets confusing in the "2 + 2 + 2 + 2" stimuli. For the sake of clarity, the two potentially interacting consonants are given in bold face here, and the stressed vowels are marked *diacritically* (as in á). This was of course not done in the actual visual stimuli. The meaning of the four conditions B, W, S, and N is as follows:

Type B: The two consonants share both word onset position and pre-stress position.
Type W: The two consonants share word onset position but not pre-stress position.
Type S: The two consonants share pre-stress position but not word onset position.
Type N: The two consonants share neither position.

Here follows an example of a set of four stimuli, one for each of the conditions B, W, S, N, for 1 + 2 + 2 + 1 and the 2 + 2 + 2 + 2 stimuli separately (Table 2.1).

Table 2.1
Examples of two corresponding sets of four stimuli, with target consonants in bold face. Stimulus type "1 + 2 + 2 + 1" represents a sequence of "a one-syllable word + a two-syllable word + a two-syllable word + a one-syllable word", stimulus type "2 + 2 + 2 + 2" represents a sequence of "four two syllable words". For condition B, W, S, N see text.

Condition	Stimulus type	
	1 + 2 + 2 + 1	2 + 2 + 2 + 2
B	**w**ok **r**ápper **r**óeper **w**al	**w**áter **r**ápper **r**óeper **w**állen
W	**w**ad **r**appórt **r**apíer **w**ol	**w**óeker **r**appórt **r**apíer **w**íkkel
S	**w**in pa**r**íjs po**r**éus **w**el	be**w**íjs pa**r**íjs po**r**éus ju**w**éel
N	**w**it píe**r**en pá**r**el **w**as	la**w**áai píe**r**en pá**r**el ge**w**ín

As exemplified here, stimulus word pairs of the "2 + 2 + 2 + 2" type were derived from those of the "1 + 2 + 2 + 1" type. We have decided that members of such related quartets should not be presented to the same participant because this might be confusing. Therefore we created two lists of stimuli each with 12 quartets of the "1 + 2 + 2 + 1" type and 12 quartets of the "2 + 2 + 2 + 2" type, in such a way that for each quartet of the "1 + 2 + 2 + 1" type the corresponding quartet of the "2 + 2 + 2 + 2" type was in the other list and vice versa. Thus each list had 24 quartets and therefore 96 sequences of four words, 12 quartets containing 1 + 2 + 2 + 1 stimuli and 12 quartets containing 2 + 2 + 2 + 2 stimuli. The pairs of potentially interacting consonants were: 1: w/ʁ; 2: w/ʁ; 3: n/m; 4: n/m; 5: b/v; 6: v/b; 7: p/k; 8: k/p; 9: l/ʁ; 10: l/ʁ; 11: j/l; 12: j/l for one half of each list and 1: d/j; 2: z/d; 3: k/χ; 4: χ/k; 5: t/d; 6: t/d; 7: p/t; 8: d/z; 9: ʃ/s; 10: s/ʃ; 11: v/z; 12: z/v for the other half of each list. The complete lists of stimulus word pairs, organized in quartets, are given in Appendix A.

Participants

There were 28 participants, 20 females and 8 males, all students at Utrecht University. Their age ranged from 18 to 26. Data from one participant (female, even-numbered) were lost due to technical malfunction. The analysis reported below is based on the remaining 27 participants.

Procedure

Participants were tested individually, in a sound-treated booth, seated in front of a PC screen. The session started with an instruction appearing on the screen. This instruction, translated into English, ran as follows:

> "Dear participant,
> Thank you for participating in this experiment. Shortly you will see a sequence of four words on the screen. Read these four words aloud as fast as you can. You are to repeat the whole sequence of four words three times. Then you should push the blue button. As a result, the words will disappear from the screen. You are to speak the sequence of four words once again, this time from memory. Repeat the sequence once again three times. Thereafter, push the blue button once again. The next sequence of four words will appear on the screen. In total there will be 96 sequences of four words. On the screen, bottom right, you can see how far you have come in the experiment. We start with a set of 10 practice items. Push the blue button to start the experiment".

There were 10 practice items specifically constructed for the purpose. In the test phase, the 96 word sequences were presented in random order to each odd-numbered participant. Each even-numbered participant got the same order of presentation as the immediately preceding odd-numbered participant, but then from List 2 instead of list 1 (cf. Appendix A). All speech produced by each participant in the test phase was recorded with a Sennheiser ME 50 microphone, and digitally stored on disk with a sampling frequency of 48,000 Hz. For each participant two separate audio files were recorded for each stimulus sequence of four words, one recording for the phase in which the words were visible on screen and one for the phase where the words were invisible, and they had to be spoken from memory. Thus, for each participant 192 audio files were created.

Scoring

All speech of each audio file of each participant was transcribed by the first author, with the help of an audiovisual display in PRAAT (Boersma and Weenink, 2009) in normal orthography or in phonetic transcription when necessary. For each response we recorded the number of the participant, the number of the trial, the stimulus identity, whether the stimulus was of the "1 + 2 + 2 + 1" type or of the

"2 + 2 + 2 + 2" type, the condition B, W, S, or N, the list number, the two consonants for which interaction was expected, and whether the stimulus word sequence was visible or invisible. Also, we categorized speech errors as "targeted" or not "targeted". "Targeted" were those single consonant substitutions that the condition was intended to elicit, "not targeted" were other single consonant substitutions, involving initial or medial consonants. As valid responses we counted (a) fluent and correct responses (**w**áter **r**ápper **r**óeper **w**állen ≫ **w**áter **r**ápper **r**óeper **w**állen), (b) completed exchanges involving initial or medial consonants (**w**áter **r**ápper **r**óeper **w**állen ≫ **r**áter **w**ápper **r**óeper **w**állen or **w**áter **r**ápper **r**óeler **w**áppen), (c) interrupted speech errors against single initial or medial consonants (**w**áter **r**ápper **r**óeper **w**állen ≫ **w**áter **r**ápper **w**óe…**r**óeper **w**állen or **w**áter **r**ápper **r**óel..**r**óeper **w**állen), (d) anticipations involving single initial or medial consonants (**w**áter **r**ápper **r**óeper **w**állen ≫ **r**áter **r**ápper **r**óeper **w**állen or **w**áter **r**ápper **r**óeler **w**állen), and (e) perseverations involving single initial or medial consonants (**w**áter **r**ápper **r**óeper **w**állen ≫ **w**áter **r**ápper **r**óeper **r**állen or **w**áter **r**ápper **r**óeper **w**áppen). All other error types and errors involving other segmental positions were coded as invalid.

If a response contained more than a single valid speech error, these speech errors were categorized separately. When a participant repeated a stimulus word sequence more than 3 times either in the visible phase or in the invisible phase, the response utterances beyond the third response utterance were discarded. When a participant produced fewer than three response utterances either in the visible or in the invisible phase, the lacking utterances were counted as omissions, thereby becoming invalid responses. We also coded as invalid all those responses that did deviate from the intended stress pattern or from the intended segment pronunciation, because these responses did not accord with the experimental variables. An example is the spoken response "bot vázal vizier bit" to the stimulus "bot vazal vizier bit", where the participant erroneously stressed the first syllable of "vazal".

Unfortunately, there appeared to be quite some hysteresis in the responses in the sense that when a participant made a particular speech error in response to a stimulus, quite often that speech error was repeated unchanged during the six responses to that stimulus. This, of course, violated the required independence of the successive errors made in response to that stimulus. For this reason, we regarded as invalid all repetitions by the same speaker of a specific speech error to a certain stimulus. The analysis took each four-word sequence as a stimulus and considered all valid single consonant substitutions in initial or medial consonant position to this stimulus together as one "super response", separately for the visible phase and the invisible phase (but see below). The numbers of these valid interactional single consonant substitutions in initial or medial position were counted for each "super response". This formed a main dependent variable. Each substitutional error was coded as to the word in which it occurred from word 1 to word 4. Complete exchanges were in this respect coded as to the word in which the anticipatory part of the error occurred.

Due to the strict constraints on the stimuli, the phonotactic opportunities for targeted errors in initial and medial position were exactly equal, at least in the "2 + 2 + 2 + 2" stimuli. Each valid speech error was coded as "unrepaired" or "repaired". For all repaired valid speech errors, we measured word onset-to-cutoff times (from the onset of the word containing the error to the moment of interruption) and error-to-cutoff times (from the onset of the error segment to the moment of interruption). Of course, word onset-to-cutoff times and error-to-cutoff times are identical for errors in initial consonants. All repaired errors were classified as "interrupted" when the word containing the error was not completed or "not interrupted" when the error word was completed. We also measured cutoff-to-repair times in all valid errors.

Results of Experiment 1

A survey and some relevant breakdowns of the errors observed in Experiment 1 are given in Table C.1 in Appendix C. For testing our predictions 1 and 2, we focus on the error-to-cutoff times of the

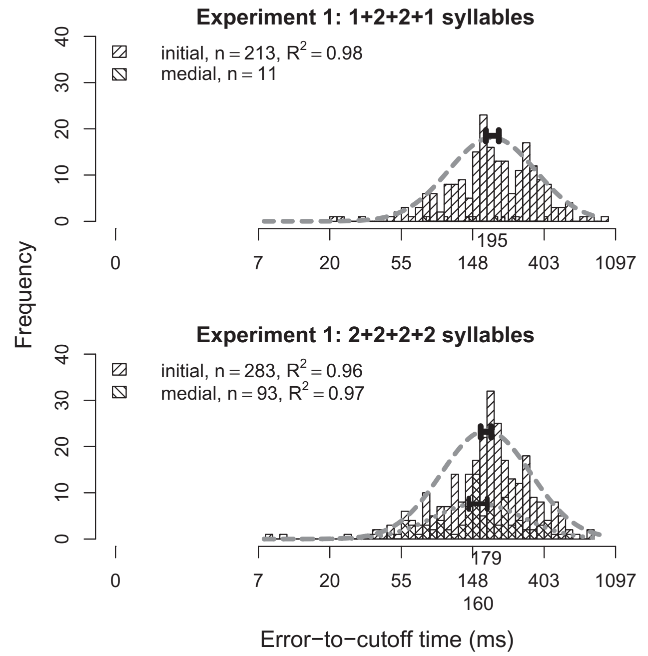

Fig. 2.1. Histograms of observed error-to-cutoff times, broken down by syllable structure (upper panel 1 + 2 + 2 + 1 syllables, lower panel 2 + 2 + 2 + 2 syllables) and by position of the error, with lognormal density distributions fitted by a tobit regression model (initial errors: dashed, medial errors: dotted). R^2 indicates the proportion of observed variance (after log transformation) captured by the fitted lognormal distribution. The horizontal error bars near the peak of a distribution indicate the bootstrapped 95% confidence interval of the location of that peak (over 500 replications).

repaired errors. One positive outlier value exceeding 1000 ms was discarded (1 out of 601 observations). Fig. 2.1 shows the observed (histograms) and fitted (curves) distributions of lognormal error-to-cutoff times, as well as bootstrapped 95% confidence intervals (Efron & Tibshirami, 1993) for the means of the fitted distributions. The error-to-cutoff times were analyzed by means of tobit regressions for censored data in R (Tobin, 1958; Kleiber & Zeileis, 2008, R Core Team, 2017), using a lognormal distribution with position as the only predictor (more complex models including visibility condition, centered trial number, and their interactions, did not yield greater likelihoods; hence the simpler tobit model including only position is summarized here). Although the error-to-cutoff times were all positive and above zero, tobit regression was used here because the same technique was used in the analysis of cutoff-to-repair times (which were censored), reported below.

Our first prediction (Prediction 1) was that error-to-cutoff times have a rather broad distribution running from approximately zero to much longer values, and that missing values below zero ms would be rare. In fact, error-to-cutoff times run from close to zero to over 1000 ms, and the distribution is nearly complete. As predicted, few, if any, errors seem to be missing due to truncation at 0 ms. This means that it rarely is the case that the speaking process is stopped before speech initiation. Note that when the error-to-cutoff time is close to 0 ms, the error must have been detected in internal speech (cf. Blackmer & Mitton, 1991). Errors detected in overt speech are relatively few and have error-to-cutoff times of many hundreds of ms (Nooteboom & Quené, 2017). That the peak of the distribution of error-to-cutoff times lies far above 0 ms (but below 200 ms), can only mean that interruption after internal error detection often comes much later than speech initiation.

Our second prediction (2) was that *word onset*-to-cutoff times are much longer for word-medial than for word-initial errors and that *error*-to-cutoff times are roughly equal for both error types. In the

1 + 2 + 2 + 1 set we find average word onset-to-cutoff times of 195 ms in word initial errors, and of 324 ms for medial errors (beta = +0.5076 on lognormal scale, Z = +2.71, p = .0067). In the 2 + 2 + 2 + 2 set, the word-onset-to-cutoff times are 180 ms and 337 ms respectively (beta = +0.630, Z = +9.19, p < .0001). Thus, in both sets of data the difference is significant. The reader may note that in the case of word medial errors such as *koene* for *koele* it is possible that the word was interrupted before the error was spoken, as in *koe…koele*. If there would have been many such cases, this would have upset our comparison between initial and medial errors. There are two ways to demonstrate that this was not so. One is that if this was so, then the distribution of error-to-cutoff times (cf. Fig. 2.1) would have been truncated at much higher values for medial than for initial errors. This was not the case. A second is to look at the relative numbers of hesitations (all such cases as *koe..koele* were labeled as hesitations). If relatively many medial errors were interrupted before being spoken, we would find a significantly greater relative number of hesitations in the conditions eliciting medial errors than in the conditions eliciting initial errors. This was not found. Apparently, both the moment of internal error detection and the moment of interruption are shifted from early to later when we compare medial with initial consonant errors. Although this shift is in the same order of magnitude as the distance in speaking time between initial and medial consonants, we also see in Fig. 2.1 that error-to-cutoff times are somewhat shorter (160 ms) for medial than for initial errors (179 ms). This difference is not significant (beta = −0.111 in lognormal scale, Z = −1.48, p = .139), but it is in the direction predicted from the results in Wheeldon and Levelt (1995) that scanning the initial syllable of internal word forms is somewhat faster than speaking these same syllables.

For testing our predictions 3 and 4, we focus on the *cutoff-to-repair times* of the repaired errors only. Outlier values exceeding 1500 ms were discarded (3 out of 601 observations). The cutoff-to-repair times were again analyzed by means of tobit regression for censored data (Tobin, 1958; Kleiber & Zeileis, 2008), separately for 1 + 2 + 2 + 1 and 2 + 2 + 2 + 2 stimuli, using the lognormal distribution and with position as the single predictor. (More complex tobit models, including visibility condition, centered trial number, and their interactions, suggest a marginally significant main effect of the visibility condition, with cutoff-to-repair times marginally slower if the stimulus is not visible, beta = +0.198, Z = +1.784, p = .0744; we will report the tobit model including only main effects of position and of visibility condition). Thus, the cutoff-to-repair times scored as 0 ms are still included in the model and contribute to the resulting estimated lognormal distribution. Fig. 2.2 shows the observed and fitted lognormal distributions, as well as the bootstrapped 95% confidence intervals for the cell means of the fitted distributions.

Our Prediction 3 was that the distribution of cutoff-to-repair times is censored at 0 ms, and that cutoff-to-repair times of 0 ms are overrepresented (note that the clear separation of 0 ms from the rest of the distribution in Fig. 2.2 is an artefact of the lognormal distribution; see below for further discussion). Obviously, the three distributions with enough observations in Fig. 2.2 are indeed strongly censored at 0 ms, suggesting that a considerable number of cases have an observed cutoff-to-repair time of 0 ms. These potentially would have had a negative value if only we could have assessed the actual moment that the repair came available to the mind of the speaker.

Our Prediction 4 was that cutoff-to-repair times are significantly longer for word-medial than for word-initial errors. For the 1 + 2 + 2 + 1 stimuli, there are too few repaired errors to test this prediction. For the 2 + 2 + 2 + 2 stimuli (Fig. 2.2, lower panel), the tobit regression analysis yielded a significant effect of consonant position (beta = +0.324 on log scale, Z = +2.54, p = .0112), however, this position effect was only weakly supported by bootstrap validations (over 500 replications) of the model, as illustrated by the overlapping bootstrapped 95% confidence intervals of cell means in Fig. 2.2 (lower panel).

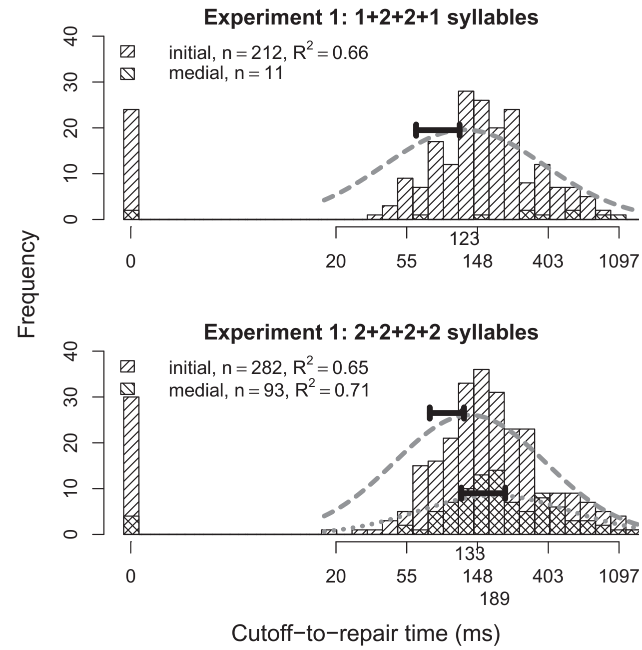

Fig. 2.2. Histograms of observed cutoff-to-repair times, broken down by syllable structure (upper panel 1 + 2 + 2 + 1 syllables, lower panel 2 + 2 + 2 + 2 syllables) and by position of the error, with lognormal density distributions fitted by a tobit regression model (initial errors: dashed, medial errors: dotted). R^2 indicates the proportion of observed variance (after log transformation) captured by the fitted lognormal distribution. The horizontal error bars near the peak of a distribution indicate the bootstrapped 95% confidence interval of the location of that peak (over 500 replications).

A tobit regression analysis, as well as similar methods for censored or truncated data, assumes that all observations are generated by a single process, here fitted by a lognormal distribution. This assumption is questionable here, however, for two reasons. From an empirical perspective, the estimated lognormal distributions fit somewhat poorly to the observed cutoff-to-repair times (see Fig. 2.2); this fit did hardly improve when we attempted to fit the data to gaussian or weibull rather than lognormal distributions. For example, in Fig. 2.2 (lower panel), the higher number of censored (0 ms) observations for initial consonants exert a stronger leftward pull on the centre of the estimated lognormal distribution, relative to the medial consonants, even though the nonzero distributions seem to overlap at first glance. Secondly, from a theoretical perspective, there is no reason for us to assume that a single process, corresponding to a unimodal lognormal distribution, has generated the observed cutoff-to-repair times. Very likely the distribution captures both internally and externally detected repaired errors, which have very different temporal properties (cf. Nooteboom & Quené, 2017). For these reasons, we also inspected the odds of a repair being immediate (i.e., having a censored cutoff-to-repair time of 0 ms) vs non-immediate (cutoff-to-repair time longer than 0 ms), by means of a Generalized Linear Mixed Model (GLMM; Quené & Van den Bergh, 2008; Bates, Maechler, Bolker & Walker, 2016; R Development Core Team, 2017; participants and item sets were used as random intercepts). Models including Stimulus Type and Position as fixed predictors did not provide a better fit than the intercept-only model (which had beta = −2.5743, Z = −8.49, p < .0001), despite the different log odds for initial (−2.46) and medial (−3.21) consonants. Thus, neither consonant position nor stimulus type did affect the odds of a repair being immediate, if the random variation between participants and between item sets in these odds was taken into account. This may well have been due to a power problem, as the number of repaired errors per participant and per item have been too low to assess effects of interest.

Because our prediction is about the detection of errors in internal speech, we might also focus on cutoff-to-repair times of which we can be reasonably certain that they correspond to internally detected errors. The data in this experiment do not allow an estimation of the underlying gaussian distributions as done by Nooteboom and Quené (2017). Nooteboom and Quené (2017) found that the time delay between internal and external error detection is c. 500 ms. To be on the safe side we now limited our analysis to cutoff-to-repair times of those repaired errors that have an error-to-cutoff time lower than 350 ms (note that this selection is based on short error-to-cutoff times, not on cutoff-to-repair times; the value of 350 ms is rather arbitrary, but is so short that repaired speech errors detected in overt speech are virtually excluded). We re-ran the tobit regression analyses on these selected lower-censored responses again separately for the two stimulus types, and again with position in the word (initial vs medial) as a fixed factor. For the 1 + 2 + 2 + 1 stimuli, there are again too few repaired errors for modelling. For the 2 + 2 + 2 + 2 stimuli, the tobit regression analysis on these selected responses yielded a significant effect of consonant position (beta = +0.407 on lognormal scale, Z = +2.992, p = .0028). If real, this effect implies that, conform to our Prediction 4, repairing speech errors in medial position is slower than repairing speech errors in initial position.

For testing our prediction 5 relating to a possible effect of variation in selective attention, we focus on the *odds of detection* (detection rate) of the valid errors. Detected errors were coded as hits, and undetected errors as misses; these binomial responses were analysed by means of a single mixed-effects Generalized Linear Model (GLMM; Quené & Van den Bergh, 2008), with position (initial vs medial), word number in stimulus (1–4) and stimulus structure (1 + 2 + 2 + 1 vs 2 + 2 + 2 + 2) as three fixed predictors. Participants and item sets (matching stimuli) were included as random intercepts, and position and word number were also included as random slopes at the participant level. The log odds of detection are summarized in Fig. 2.3, broken down by the three fixed predictors in the GLMM.

Our Prediction 5 was that within spoken lexical forms, rate of detection is higher for word-initial errors than for non-initial errors (here in word-medial position). This predicted difference is indeed clearly visible in both panels of Fig. 2.3 and is confirmed by the main effect of position in the GLMM (beta = −1.286, Z = −3.61, p = .0003). The position effect did not interact with other predictors in the GLMM. Hence, speech errors in word-initial position have a much higher probability to be detected in self-monitoring than speech errors in word-medial position. This points at a difference between initial and medial consonants in amount of selective attention available for self-monitoring.

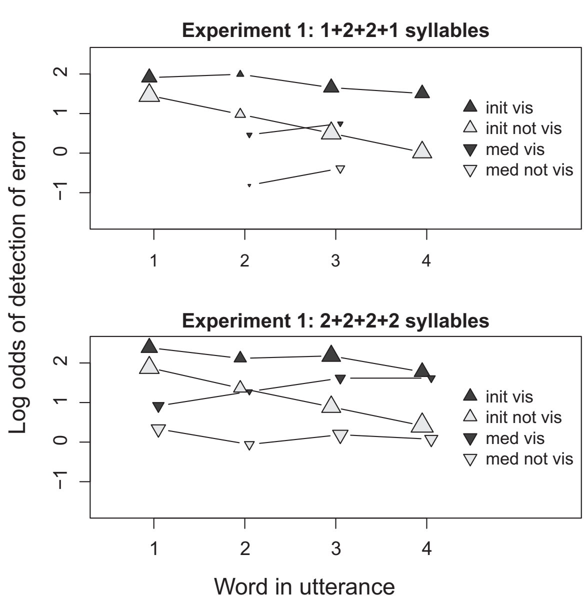

Fig. 2.3. Estimated log odds of detection of valid errors, broken down by syllable structure (upper panel 1 + 2 + 2 + 1 syllables, lower panel 2 + 2 + 2 + 2 syllables), by position of the error (initial: upward triangles, medial: downward triangles), by word number in the response utterance (1–4, along horizontal axis), and visibility condition (darker symbols: visible, lighter symbols: not visible. Symbol sizes correspond with the numbers of detectable valid errors in each cell.

Our Prediction 6 was that within multi-word utterances (not longer than a single intonational unit), the odds of error detection decrease from earlier to later words in the utterance. This predicted effect of word number is visible in both panels of Fig. 2.3, most clearly for initial-consonant errors in the condition with stimuli being not visible. Indeed, the GLMM contains an insignificant main effect of word position (beta = −0.120, Z = −0.79, p = .4300), combined with a marginally significant interaction effect of word position and visibility condition (beta = −0.325, Z = −1.90, p = .0570), with large variability among speakers (s = 0.71). Although the patterns in Fig. 2.3 may suggest an interaction effect between position and word number, this was not significant in the GLMM (beta = +0.289, Z = 1.58, p = .113), most likely due to the low numbers of valid, detectable errors for word-medial consonants. Other interactions in the GLMM were not significant. [The reader may also note that the total numbers of errors per word position, coded in the symbol sizes in Fig. 2.3, does not increase from earlier to later as one would expect from the results reported by Choe & Redford (2012)].

Discussion of Experiment 1

In the current experiment we have set out to test some predictions of effects of timing and selective attention in self-monitoring on rates of detection of segmental speech errors in different positions. We will briefly discuss the results in terms of our six predictions:

Prediction 1: *Error-to-cutoff times have a rather wide distribution, running from close to 0 ms when a repair is available before or at speech initiation, to much longer values when interruption is postponed. Missing values below 0 ms are rare.* This prediction was confirmed. The distribution of error-to-cutoff times indeed runs from close to zero ms to values far above 1000 ms. Of course, long error-to-cutoff times are not only caused by postponement of interruption after internal error detection, but also because errors in many cases are detected only after speech initiation. According to Nooteboom and Quené (2017) the time delay between internal and external error detection is in the order of 500 ms.

Possibly the distribution is truncated at zero ms, but if so this is in the far lower tail of the distribution, suggesting that missing values below zero are rare. This demonstrates that, conform the H&K computational model, after internal error detection generally interruption of the speaking process takes more time than initiation of speaking the error form. This also implies that so-called "covert repairs", i.e. repairs silently made before speech initiation, are rare or do not occur at all (unless of course by an entirely different process). As this has consequences for the interpretation of a number of earlier publications, we will return to this point in the general discussion.

Prediction 2: *Word onset-to-cutoff times are significantly longer for word-medial than for word-initial sound form errors and error-to-cutoff times are roughly equal for word-initial and word-medial sound form errors.* Word onset-to-cutoff times were indeed found to be significantly longer for word-medial than for word-initial errors, on average in the order of 150 ms. This suggests that word-medial errors are internally detected later than word-initial errors, as predicted from the assumption that internally errors are detected by scanning internal word forms from early to late in a time-consuming way. Error-to-cutoff times did not differ significantly, although these were somewhat shorter for medial than for initial errors, as one would expect from the finding by Wheeldon and Levelt (1995) that scanning internal word forms is faster than speaking.

Prediction 3: *The distribution of cutoff-to-repair times is censored at 0 ms, making values of 0 ms overrepresented.* This was confirmed by the data. This means that often repairs are available to the mind of the speaker before speech interruption. If a repair is available before speech interruption this may be because the repair was available very fast, or that interruption was postponed in order to gain time for re-activating the repair. Because cutoff-to-repair times in by far most cases are longer (and often much longer) than zero ms, it is obvious that strategic postponement of interruption is limited (for errors detected internally the postponement probably has a maximum somewhat above 500 ms, cf. Nooteboom & Quené, 2017; most longer values probably correspond to errors detected after speech initiation).

Prediction 4: *Cutoff-to-repair times are significantly longer for word-medial than for word-initial errors.* This prediction is borne out by the data. This suggests that repairing word-medial errors takes more time than repairing word-initial errors. Under the assumption that selective attention affects speed of processing, this also suggests that selective attention is higher for word-initial than for word-medial consonants.

Prediction 5: *Rate of error detection is lower for word-medial than for word-initial consonant errors.* We have assumed that amount of selective attention during speech planning and self-monitoring internal word forms is inversely proportional to predictability of speech segments. Because predictability very rapidly increases from word onset to later segments (Marslen-Wilson & Welsh, 1978), variation in amount of selective attention is supposed to be considerable. This ties in with our results: There is a very big and significant difference in error detection rate between word-initial and word-medial consonant errors.

Prediction 6: *Within spoken utterances (not longer than a single intonational unit), rate of error detection decreases from earlier to later.* This predicted effect was not found to be significant in the current experiment. We will come back to this in describing Experiment 2.

Experiment 2

Method of Experiment 2

Experiment 1 has shown that the 1 + 2 + 2 + 1 stimuli, derived from the original Experiment 2 in Shattuck-Hufnagel (1992), are not very efficient in eliciting segmental interactions. In our Experiment 2 we will refrain from using these 1 + 2 + 2 + 1 stimuli. Instead we have opted for a setup that makes it possible to investigate the contribution of targeting specific consonant positions for interaction, by repeating or not repeating consonants in these positions, as in (2a), where interaction between /w/ and /r/ is elicited in initial position and (2b), where no such interaction is elicited by consonant repetition in initial position:

(2a) water rapper roeper wallen
(2b) water roeper lommer bikkel

We will refer to these two groups of stimuli as "eliciting" versus "not eliciting". Note that these terms apply to a specific segmental position. We reserve the terms "targeted" versus "not targeted" for distinguishing between the specific position in which interaction is or is not elicited, such as the initial position in the above example, and other positions. In the examples above the initial position is "targeted" for interaction, but interaction is only elicited in the (2a), not in (2b). As in Experiment 1, there are four conditions B, W, S and N. B means that the consonants in the targeted positions are both word initial and followed by a stressed vowel, W means the targeted positions are all word initial but not followed by a stressed vowel, S that the targeted positions are all followed by a stressed vowel but are not all word initial and N means they share neither word onset positions nor stress position.

Table 3.1
Examples of two corresponding sets of four stimuli. "Eliciting" = interaction provoked by consonant repetition in the targeted position; "not eliciting" = no interaction provoked by consonant repetition in the targeted position. For the four conditions see text.

Condition	Eliciting	Not eliciting
B	**w**áter **r**ápper **r**óeper **w**állen	**w**áter **r**óeper **l**ómmer **b**íkkel
W	**w**óeker **r**appórt **r**apíer **w**íkkel	**w**íjzer **p**aríjs **d**oríen **g**ózer
S	be**w**íjs pa**r**íjs po**r**éus ju**w**éel	be**w**íjs pa**r**íjs lo**k**áal ge**n**óot
N	la**w**áai píe**r**en pá**r**el ge**w**ín	la**w**áai píe**r**en bá**kk**en ge**s**óp

Stimuli

Whereas in Experiment 1 we had created quartets of stimuli targeting the same consonants for interaction, due to limitations in the Dutch vocabulary we did not always succeed in doing this for Experiment 2. Here follows an example of two sets of four stimuli, one stimulus for each of the conditions B, W, S, N, for eliciting and not eliciting stimuli separately (see Table 3.1).

We again created two lists of stimuli each with 12 quartets of the "eliciting" type and 12 quartets of the "not eliciting" type, in such a way that for each quartet of the "eliciting" type the corresponding quartet of the "not eliciting" type was in the other list, and vice versa. Thus, each list had 24 quartets and therefore 96 sequences of four words. The complete lists of stimulus word pairs, organized in quartets, are given in Appendix A.

Participants

There were 30 participants, 25 females and 5 males, all students at Utrecht University. Their age ranged from 18 to 53, with an average of 24.8 years. All participants reported having no hearing, speech or vision problems. They were paid for their participation.

Procedure

The procedure was the same as in Experiment 1.

Scoring

Scoring was the same as in Experiment 1, except that the labels "1 + 2 + 2 + 1" and "2 + 2 + 2 + 2" were replaced by the labels "eliciting" and "not eliciting".

Results of Experiment 2

A survey and relevant breakdowns of the errors observed in Experiment 2 are given in Appendix D. For testing our predictions 1 and 2, we focus again, as we did in Experiment 1, on the error-to-cutoff times of the repaired errors. No observations exceeded the outlier criterion value of 1000 ms (all 558 valid observations remaining). Error-to-cutoff times were again analysed by means of tobit regressions for censored data in R (Tobin, 1958; Kleiber & Zeileis, 2008; R Core Team, 2017), using a lognormal distribution with position as the only predictor. (More complex models including visibility conditions, centered trial number, the contrast between fixed-stress [B, S] vs varying-stress conditions [W, N]. and their interactions, did not yield greater likelihoods; hence the simpler tobit model including only position is summarized here.)

Fig. 3.1 shows the observed (histograms) and fitted (curves) distributions of lognormal error-to-cutoff times, as well as bootstrapped 95% confidence intervals (Efron & Tibshirami, 1993) for the mean of the fitted distribution. The histograms in Fig. 3.1 indicate that the error-to-cutoff times are again, as predicted, nearly complete: If the distributions are truncated at all, they are so only in the far lower end tails.

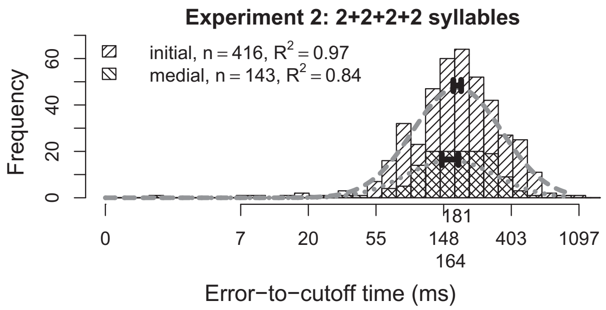

Fig. 3.1. Histograms of observed error-to-cutoff times, broken down by position of the error, with lognormal density distributions fitted by a tobit regression model (initial errors: dashed, medial errors: dotted). R^2 indicates the proportion of observed variance (after log transformation) captured by the fitted lognormal distribution. The horizontal error bars near the peak of a distribution indicate the bootstrapped 95% confidence interval of the location of that peak (over 500 replications).

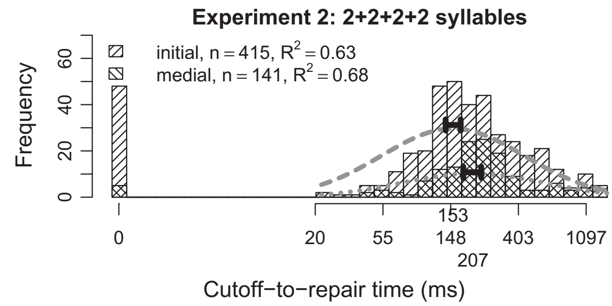

Fig. 3.2. Histograms of observed cutoff-to-repair times, broken down by position of the error, with lognormal density distributions fitted by a tobit regression model (initial errors: dashed, medial errors: dotted). R^2 indicates the proportion of observed variance (after log transformation) captured by the fitted lognormal distribution. The horizontal error bars near the peak of a distribution indicate the bootstrapped 95% confidence interval of the location of that peak (over 500 replications).

This implies that, conform the H&K computational model, the average moment of interruption is considerably later than the average moment of speech initiation. The difference is in the order of 180 ms for initial and in the order of 160 ms for medial consonant errors. This basically confirms what we found in Experiment 1. It supports our Prediction 1 that, although the distribution of error-to-cutoff times is truncated close to 0 ms, generally speech interruption not only has a broad distribution but also follows speech initiation. This again entails that there are very few missing values below zero ms, if any. This renders the prediction from the H&K model that in the distributions of error-to-cutoff times the number of covert repairs, and therefore the number of missing values due to truncation, increases with position of the word, untestable.

Our second prediction (2) was that *word onset*-to-cutoff times are significantly longer for word-medial than for word-initial sound form errors and that *error*-to-cutoff times are roughly equal for word-initial and word-medial sound form errors. We find in this experiment that the average word onset-to-cutoff interval is 181 ms for word initial and 345 ms for word medial errors. This difference is significant according to a tobit model using a lognormal distribution (beta = +0.642 on log scale, Z = +11.25, $p < .0001$). As in Experiment 1, here too there are very few observations missing due to truncation at zero ms, for both error positions (cf. Fig. 3.1). This leaves little room for the possibility that the average word onset-to-cutoff interval is biased because of truncation. As predicted, the error-to-cutoff times are roughly equal for word-initial and word-medial errors (see Fig. 3.1). The tobit regression analysis did not yield a significant effect of consonant position (beta = −0.098 on log scale, Z = −1.541, p = .123). This suggests again that interruption requires about the same amount of time after initial and after medial errors: later error detection in medial than in initial position (in absolute time) is compensated by later interruption in medial than in initial position (in absolute time). The estimated (back-transformed) averages of error-to-cutoff times are 181 ms and 164 ms for initial and medial error respectively. Although not significant, this difference is again, as in Experiment 1, in the direction predicted by the finding of Wheeldon and Levelt (1995) that speed of scanning of internal word forms is in the first syllable somewhat faster than speaking time for these same syllables.

For testing our predictions 3 and 4, we focus on the *cutoff-to-repair times* of the repaired errors only. Outlier values exceeding 1500 ms were discarded (3 out of 559 observations). The cutoff-to-repair times were again analysed by means of tobit regression for censored data (Tobin, 1958; Kleiber & Zeileis, 2008; R Core Team, 2017), separately for initial and medial position, using lognormal distribution. Thus, the cutoff-to-repair times scored as 0 ms are still included in the model, and contribute to the resulting estimated lognormal distribution, representing possibly virtual negative values. (More complex tobit models, including visibility condition, centered trial number, the contrast between fixed-stress [B, S] vs varying stress conditions [W, N], and their interactions, did not yield greater likelihoods than the simpler tobit model including only the position effect; hence that simpler model is summarized here.) Fig. 3.2 shows the observed and fitted lognormal distributions, as well as the bootstrapped 95% confidence intervals for the means of the fitted distributions.

Our third prediction was that the distribution of cutoff-to-repair times is censored at 0 ms (this means that values of 0 ms are overrepresented). Obviously, the distributions in Fig. 3.2 are indeed censored at 0 ms, which suggests once more that a number of cases showing immediate repairs correspond to *negative* time intervals between moment of interruption and the moment a repair comes available to the mind of the speaker. Of course, the censored distribution deviates from lognormal because of the overrepresentation of intervals of 0 ms.

Our prediction 4 was that cutoff-to-repair times are longer for word-medial than for word-initial repaired error. As in Experiment 1, the tobit regression analysis yielded a significant effect of error position (beta = +0.299 on log scale, Z = +2.905, p = .0037), and in this experiment, this position effect was indeed supported by the bootstrapped 95% confidence intervals of the peaks of the distributions [138, 172] and [179, 233]. Again, we see that the tobit-modeled distributions fit somewhat poorly to the observed distributions of log-transformed error-to-cutoff times. For these reasons, we again inspected the odds of a repair being immediate (i.e., having a censored cutoff-to-repair time of 0 ms) vs non-immediate (cutoff-to-repair time longer than 0 ms), by means of a Generalized Linear Mixed Model (GLMM; Quené & Van den Bergh, 2008; Bates et al., 2016; R Core Team, 2017; participants and item sets were used as random intercepts). A GLMM including position as fixed predictor performed significantly better than the intercept-only model [Likelihood Ratio Test, $\chi^2(1) = 8.579$, p = .0034; Pinheiro & Bates, 2002]. For initial consonant errors, the odds of a repair being immediate were 0.079 (or 8%), whereas for medial consonants the odds were significantly and very much lower at 0.023 (or 2%) [beta = −1.29, Z = −2.59, p = .01]. This supports our prediction 4 that cutoff-to-repair times are longer for non-initial than for initial errors.

As in Experiment 1, we re-ran the tobit regression analyses on the repaired errors with error-to-cutoff times lower than 350 ms ($n = 386$ responses that most probably corresponded to internally detected errors; the value of 350 ms is arbitrary but so low that errors detected in overt speech are virtually excluded) to avoid contamination of the analysis by the long error-to-cutoff times of externally detected repair errors, again with position in the word (initial vs medial) as a fixed factor. And again, the tobit regression analysis with lower and upper censoring yielded a significant effect of consonant position (beta =

+0.326 on lognormal scale, Z = +2.678, p = .0074). This confirms the preliminary finding of Experiment 1 that repairing speech errors in medial position is slower than repairing speech errors in initial position, in accordance with prediction 4. The exact difference in cutoff-to-repair times between initial and medial consonant errors is difficult to assess, given the non-gaussian distributions, but Fig. 3.2 suggests this difference to be in the order of 50 ms.

For testing our predictions 5 and 6, relating to possible effects of variation in selective attention, we focus again on the *odds of detection* (detection rate) of the valid errors, modeled by a single mixed-effects Generalized Linear Model (GLMM; Quené & Van den Bergh, 2008), with position (initial vs medial), word number in stimulus (1–4), visibility condition (visible vs not visibly) and elicitation status (true: eliciting vs false: not eliciting) as four fixed predictors. Participants and matching item sets (of matching stimuli) were included as random intercepts. Models including elicitation status or including this main effect plus its interactions did not perform better than models without these terms, according to Likelihood Ratio Tests [$\chi^2(1) = 0.1422$ and $\chi^2(3) = 0.1077$, respectively, both n.s.], so these terms were dropped from the GLMM. (As suggested by one reviewer, the centered trial number was also added as predictor, but this did not improve the GLMM, and computational problems arose when interactions of centered trial number and other predictors were also added; hence we report the optimal GLMM without centered trial number). The log odds of detection are summarized in Fig. 3.3, broken down by the two remaining fixed predictors in the GLMM.

Prediction 5 was that within spoken lexical forms, rate of error detection is higher for word-initial segments than for later segments (here in word-medial position). This predicted difference is indeed clearly visible in Fig. 3.3, and it is confirmed by the main effect of position in the GLMM (beta = −0.900, Z = −3.56, p = .0004). As in Experiment 1, speech errors in word-initial position have a much higher probability to be detected in self-monitoring than speech errors in word-medial position. Note that also the total numbers of detectable errors, as coded in the symbol sizes, are systematically lower in medial than in initial position.

Prediction 6 was that within multi-word utterances (not longer than a single intonational unit), the odds of error detection decrease from earlier to later words. This predicted effect of word number is visible in Fig. 3.3, most clearly for the conditions with stimuli being not visible. Nevertheless, neither the main effect of word number (beta = +0.0395. Z = −0.277, n.s.) nor its interaction with visibility condition (beta = −0.261, Z = −1.719, p = .0855) is significant, perhaps due to the large variability in odds of detection among speakers (s = 0.61) and among stimuli (s = 0.42), as in Experiment 1. The interaction effect between position and word number was also not significant in the GLMM (beta = −0.205, Z = −1.517, p = .1293). Note that the total numbers of detectable errors as coded in Fig. 3.3 in the symbol sizes do not increase from early to late as one would expect from the results reported by Choe & Redford (2012). There is a clear alternating pattern.

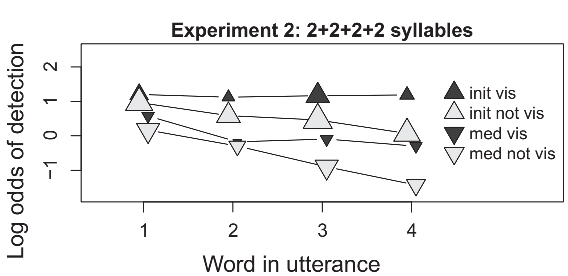

Fig. 3.3. Estimated log odds of detection of valid errors, broken down by position of the error (initial: upward triangles, medial: downward triangles), by word number in the response utterance (1–4, along horizontal axis) and visibility condition (darker symbols: visible, lighter symbols: not visible). Symbol sizes correspond with the numbers of detectable valid errors in each cell.

This pattern appears to be due to the structure of the tongue twisters, cf. Croot, Au and Harper (2010), Goldrick, Keshet, Gustafson, Heller and Needle (2016) and Wilshire (1998).

Discussion of Experiment 2

In Experiment 2 we have set out to see whether some results obtained in Experiment 1 would stand further testing. We will shortly discuss the results of Experiment 2 in terms of our six predictions. Before we do that, we wish to point out that our data also show some unpredicted and unexpected results: In Experiment 1 the data (see Appendix C) suggested that eliciting versus not eliciting interaction between two consonantal segments by repetition of a consonant in a specific position has a rather strong effect in initial position, but not in medial position. This unexpected finding was confirmed in Experiment 2, in a much more convincing test, because now the “eliciting” and “not eliciting” stimuli were in all other respects comparable (see Appendix D). We also found that the error rate is much higher in initial than in medial position.[2]

Prediction 1: *Error-to-cutoff times have a rather broad distribution, running from close to 0 ms when a repair is available before or at speech initiation, to much longer values when interruption is postponed. Missing values below 0 ms are rare.* Error-to-cutoff times run from close to zero ms to nearly 1000 ms, showing indeed a wide distribution. Of course, this is not only caused by postponement of interruption after internal error detection, but also because errors in many cases are detected only after speech initiation. According to Nooteboom and Quené (2017) the time delay between internal and external error detection is in the order of 500 ms. The distribution of error-to-cutoff times is nearly complete: very few cases are missing from the lower tail of the distribution. This suggests that generally the speaking process is stopped after speech initiation, rarely before speech initiation. This leaves little room for covert repairs. The prediction from the H&K model that the relative number of covert repairs increases from early to late within utterances is, at least in these experiments, untestable.

Prediction 2: *Word onset-to-cutoff times are significantly longer for word-medial than for word-initial sound form errors and error-to-cutoff times are roughly equal for word-initial and word-medial sound form errors.* Word onset-to-cutoff times were again found to be significantly longer for word-medial than for word-initial errors, on average in the order of 150 ms. This suggests that word-medial errors are internally detected later than word-initial errors, as predicted from the assumption that internally errors are detected by scanning internal word forms from early to late in a time-consuming way. Error-to-cutoff times did not differ significantly, although these

[2] There is one aspect of our results that requires specific mention: we found that all other things being equal, word-initial consonant errors are far more frequent than word-medial consonant errors (see Appendices C and D). Nooteboom and Quené (2015) demonstrated that a strong word-initial effect in a collection of speech errors in spontaneous Dutch could be explained statistically on the basis of the assumption that error rate is a function of the number of opportunities each segment has for interaction with segments in the immediate environment in similar positions. However, in the current Experiment 2, number of opportunities for interaction were by design kept equal for word-initial and medial-consonants. The predominance of word-initial errors and detection rate in word-initial errors in these experiments is completely in line with our assumptions that both error rate and amount of selective attention are a function of predictability of the speech sound. Obviously, we need another explanation for the finding by Nooteboom and Quené (2015). Post hoc we observe that in spontaneous Dutch the overwhelming majority of word tokens are monosyllabic. Statistically, medial consonants hardly count. Unfortunately, the conclusions by Nooteboom and Quené (2015) cannot be generalized to utterances with more polysyllabic words.

were somewhat shorter for medial than for initial errors, as one would expect from the finding by Wheeldon and Levelt (1995) that scanning internal word forms is faster than speaking.

Prediction 3: *The distribution of cutoff-to-repair times is censored at 0 ms.* Inspecting the distributions of cutoff-to-repair times separately for the "eliciting" and the "non-eliciting" conditions and for initial and medial positions has shown that these distributions indeed are censored at 0 ms. This suggests that, as in Experiment 1, the columns containing cases of 0 ms hide quite a few cases where the actual moment a repair came available to the mind of the speaker fell a varying amount of time before the moment of interruption. This result demonstrates that after internal error detection often repairs are available before interruption. We also find confirmed that cutoff-to-repair times may be very long, even in the order of 1000 ms. Long cutoff-to-repair times probably are caused by no repair being sufficiently activated to be spoken, even after postponement of interruption.

Prediction 4: *Cutoff-to-repair times are significantly longer for word-medial than for word-initial errors.* This prediction was borne out. This suggests that repairing medial consonant errors takes longer than repairing initial consonant errors. This in turn suggests that during self-monitoring selective attention decreases from early to late within word forms.

Prediction 5: *Rate of error detection is lower for word-medial than for word-initial consonant errors.* This was also predicted from the assumption that selective attention decreases from early to late with word forms. Because we have assumed that selective attention is inversely proportional to predictability of segments within a word form, and we also know that this predictability rapidly increases from word onset to later segments, its effect on rate of detection is supposed to be big. This is precisely what we found with percentages of repaired speech errors decreasing over the four words in the tongue twisters from 73% to 59% for word initial positions and from 56% to 26% for word medial positions.

Prediction 6: *Rate of error detection decreases from earlier to later words.* This predicted effect, although clearly visible in Fig. 3.3, was not significant. Of course, the effect found by Choe and Redford (2012) that the total number of detectable errors increases from earlier to later in utterances, was clearly absent from out data. Instead, the structure of our tongue twisters seems to have induced an alternating pattern, as also found by Croot, Au and Harper (2010), Goldrick, Keshet, Gustafson, Heller and Needle (2016) and Wilshire (1999).

Experiment 2 has confirmed the main findings of Experiment 1. We have found strong support for the proposal that interruption of the speaking process after internal error detection, although sometimes very fast, often is following and rarely preceding speech initiation. Also, the proposal that internal sound form error detection employs scanning the internal word forms from early to late is supported. The results again demonstrate that selective attention is a major determinant of detection and repair of speech errors in self-monitoring.

General discussion

In this paper we have focused on temporal aspects of detecting and repairing sound form errors in self-monitoring. The motivation of the investigation was to find out how observed frequencies of sound form errors of speech are affected by temporal aspects of self-monitoring for speech errors. We first confronted a computational implementation by Hartsuiker and Kolk (H&K model, 2001) of the dual perceptual loop theory by Levelt (1989) and Levelt et al. (1999) with later investigations in the literature. This led us to make some testable predictions that not always are conform to what one would have predicted from the H&K model. Based on the literature discussed in the introduction to this paper and our own results obtained in the current two experiments, testing various predictions on self-monitoring, below we will provide a somewhat different account of various processes controlling temporal aspects of detecting and repairing sound form errors of speech in self-monitoring.

We capitalize on the assumption that during speech preparation there frequently is competition between word form candidates. Such competition is often assumed in theories of word production. In non-cascading theories (Levelt, 1989; Levelt et al., 1999; Roelofs, 2003) the outcome of a selection process, for example phonological encoding or word form selection, is with only few exceptions, a single candidate. This would entail that in the articulatory buffer that is supposedly the output of phonological encoding, there is seldom competition between word form candidates for the same slot in the utterance. We propose, on the contrary, that competition between word form candidates is the norm rather than the exception. That nevertheless speech errors are relatively rare (Levelt, 1989: p. 199; Garnham, Shilcock, Brown, Mill & Cutler, 1982) does not mean that "goal referenced selection" (Roelofs, 2003) or "cognitive control" (Nozari et al., 2016) de-activate all candidates but one, but rather that generally the correct candidate gets higher activation than other candidates.

Cascading theories of word production and persistence of competition during articulation are supported by the frequent occurrence of articulatory blending in experiments eliciting speech errors (Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein et al., 2007; McMillan & Corley, 2010; Mowrey & MacKay,1990; Slis & Van Lieshout, 2016). We propose that competition is also sustained during repairing. This explains why, also in the current findings, so often a repair is immediately available at the moment of interruption, even with very short error-to-cutoff intervals.

Nooteboom and Quené (2017) found that error detection not only before but also after speech initiation does not depend on auditory feedback. This made us assume that competition between word candidates leads to error detection via conflict-based monitoring as proposed by Nozari et al. (2011). Conflict-based monitoring can be directed at internal speech, i.e. the contents of the articulatory buffer before speech initiation. We propose that it can also be directed at somatosensory and proprioceptive feedback from articulatory gestures after speech initiation. This is supported by the unexpected finding in Nooteboom and Quené (2017) that so-called externally detected errors, with a distribution of error-to-cutoff times that is 500 ms delayed with respect to the distribution of so-called internally detected errors, are detected with the same frequency with and without auditory feedback.

It has been assumed in various publications that speech errors detected in internal speech can be repaired silently, i.e. before the error has become overt. These repairs are called "covert repairs" (Kolk & Postma, 1997; Levelt, 1989; Postma & Kolk, 1993) or "prepairs" (Schlenk et al., 1987). Our results show that, at least in our experiments, interruption of the speaking process after internal error detection rarely occurs before speech initiation. This suggests that when people actually speak, covert repairs of sound form errors either are very rare or do not occur at all. Of course, covert repairs remain fully possible when people prepare speech internally for later usage. Also, our results do not exclude the possibility that covert repairs are much more frequent after syntactic, semantic or appropriateness errors than after sound form errors. However, we propose that after internal detection of sound form errors during speaking, interruption of the speaking process is generally so slow that covert repairs are virtually excluded. This implies that the prediction from the H&K model that the number of covert repairs increases from earlier to later words in an utterance is untestable for sound form errors. It should be noted that testing the H&K computational model by Hartsuiker and Kolk (2001) involved a set of speech errors including both sound form errors and higher order errors.

An interesting property of self-monitoring for sound form errors that is suggested by our results is that these errors are detected in internal speech by scanning word forms from early to late, conform to the results on phoneme detection reported by Wheeldon and Levelt (1995). However, we note that it cannot be excluded that both the Wheeldon and Levelt and our own results can be explained from variations in selective attention. If indeed, as we assume, selective attention is inversely proportional to predictability, then one would expect that in polysyllabic words selective attention decreases from the first speech segment in a word form to the so-called uniqueness point, i.e. the point in the word form after which the word is uniquely distinguished from all other words in the lexicon (Marslen-Wilson & Welsh, 1978). The role of the uniqueness point in self-monitoring was investigated by Özdemir, Roelofs and Levelt (2007). These authors found a strong uniqueness point effect in a silent phoneme monitoring task. They argued that such an effect is specific for perception, and therefore is an argument in favor of perception-based monitoring. However, both in word selection and in production-based self-monitoring similarities and differences between word candidates as well as early-to-late processing are involved. These are the two ingredients necessary for uniqueness point effects. There seems little reason to believe that uniqueness point effects are specific for perception. After the uniqueness point the required selective attention is minimal because all remaining segments are predictable. This would explain that Wheeldon and Levelt (1995) found that in detecting phonemes in silent internal word forms, scanning speed increased rapidly after the first syllable.

Time intervals from error to interruption (error-to-cutoff times) in our experiments as in earlier experiments show a wide distribution, running from 0 ms to more than 1000 ms. This suggests that, although speakers can and sometimes do interrupt the speaking process very quickly, they do not always do this. We propose, following Tydgat et al. (2011), that speakers often postpone interruption "for strategic reasons". In our experiments, and also in self-monitoring for speech errors in spontaneous speech, such "strategic reasons" very probably have to do with the non-availability of a repair. In our view of self-monitoring for speech errors a repair is a (mostly correct) word form candidate that is competing with the error form that is being produced and abandoned after error detection. When activation of the competing form is relatively high, it is rapidly available as repair, when activation is lower, reactivating it will consume more time. This view of repairing leads to testable predictions. We give two examples: (1) In experiments eliciting sound form errors by phonological priming activation of both the correct word form candidate and the elicited error form candidate is boosted with respect to other possible candidates. Therefore cutoff-to-repair times will be shorter in the eliciting condition than in a control condition without phonological priming. This is still to be tested. (2) Because activation of the repair candidate decreases during the time delay between internal and external error detection, cutoff-to-repair times will be significantly longer after external than after internal error detection. This was indeed found by Nooteboom and Quené (2017), the average difference being in the order of 200 ms.

We think that self-monitoring for sound form errors is affected by variations in selective attention. It has been suggested to us that the assumption that self-monitoring is semi-conscious and (partly) controlled by selective attention, for example by Levelt (1989), is coupled to the hypothesized prominent role of the speech comprehension system. However, we see no reason why conflict-based monitoring of speech production would not be under attentional control. We have formulated some predictions making this assumption, and we find these predictions corroborated by the data. Notably we find that word-medial sound form errors are repaired more slowly than word-initial sound form errors. It has been pointed out to us, that there is an alternative explanation of this aspect of our data. Tydgat, Diependaele, Hartsuiker and Pickering (2012) demonstrated in picture naming experiments in which the picture was changed unexpectedly after the naming was initiated, that a phonological similarity between the onset of the interrupted word and the onset of the repair, as in *pawij…parijs*, may delay the production of the repair due to interference between the two similar word forms. However, our interpretation in terms of decreasing selective attention is supported by the considerable difference in detection rate between word initial and word medial errors. We are not aware that such a difference in detection rate between word initial and word medial errors, although a considerable effect, has been reported earlier. This may be related to the fact that controlled experiments eliciting segmental errors virtually always have been limited to word initial errors. Nooteboom (2011), comparing error detection rate in novel phrases with error detection rate in phrasal lexical items, proposed that selective attention for self-monitoring is high when errors are more frequent (novel phrases) and low when errors are less frequent (phrasal lexical items). In the same vein the much lower error detection rate in word medial than in word initial position can be explained from the often-observed high error rate in word initial and low error rate in word medial position (a.o. Dell, 1986; Nooteboom & Quené, 2015; Shattuck-Hufnagel, 1992). We conclude that self-monitoring for speech errors is under firm control of selective attention.

Finally, we wish to point out that the question that originally motivated this investigation, viz. how are relative numbers of observed speech errors distorted by processes of detection and repair in self-monitoring, has a re-assuring answer: At least for sound form errors such distortion is minimal, because covert repairs are either very rare or do not occur at all.

Conclusions

Based on earlier results in the literature and of results obtained in two experiments reported in this paper, we propose an account of detecting and repairing sound form errors of speech in self-monitoring that in some respects deviates from earlier proposals. This theory assumes that in word production there is competition between activated word form candidates. This competition persists during the preparation of articulation, during articulation, and even after abandonment of erroneous word forms. Word form errors are detected on the basis of conflict between competing word candidates, either before speech initiation (internal error detection) or after speech initiation (external error detection). Error detection after speech initiation is based in conflict between articulatory gestures as perceived by somatosensory and proprioceptive feedback from the articulators. Repairs stem from active or re-activated word candidates competing with the abandoned error forms. Activation of potential repairs decreases during the time delay of 500 ms between internal and external error detection. Efficiency and temporal aspects of self-monitoring are controlled by variations in selective attention.

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. We are grateful to the technical staff of the institute. Alex Manus helped us with setting up and running both experiments, and Dr. Iris Mulders was extremely helpful in finding enough speakers. We are also grateful to two anonymous reviewers and Rob Hartsuiker, whose comments helped very much to improve this paper. The raw data of the two experiments in both .csv and .xlsx files can be found through the following web page: https://osf.io/6wfS3.

Appendix A. Stimuli Experiment 1

	List 1				List 2			
nr	1vs2 syll	cons	cond	stimulus	2vs2 syll	cons	cond	stimulus
01	1vs2 syll	w/r	B	wok rapper roeper wal	2vs2 syll	w/r	B	water rapper roeper wallen
02	1vs2 syll	w/r	W	wad rapport rapier wol	2vs2 syll	w/r	W	woeker rapport rapier wikkel
03	1vs2 syll	w/r	S	win parijs poreus wel	2vs2 syll	w/r	S	bewijs parijs poreus juweel
04	1vs2 syll	w/r	N	wit pieren parel was	2vs2 syll	w/r	N	lawaai pieren parel gewin
05	1vs2 syll	w/r	B	wig radar ridder weg	2vs2 syll	w/r	B	waggel radar ridder wegen
06	1vs2 syll	w/r	W	won radauw radijs wiel	2vs2 syll	w/r	W	woning radauw radijs wielen
07	1vs2 syll	w/r	S	web direct dorien wak	2vs2 syll	w/r	S	gewoon direct dorien bewaar
08	1vs2 syll	w/r	N	wis dieren duren wil	2vs2 syll	w/r	N	gewis dieren duren beween
09	1vs2 syll	n/m	B	nap molen maken nok	2vs2 syll	n/m	B	neder molen maken nodig
10	1vs2 syll	n/m	W	nog maleis meloen nuf	2vs2 syll	n/m	W	noten maleis meloen nuffig
11	1vs2 syll	n/m	S	nut lamel limiet nies	2vs2 syll	n/m	S	benut lamel limiet genies
12	1vs2 syll	n/m	N	nuk lemmet lommer net	2vs2 syll	n/m	N	geniet lemmet lommer benul
13	1vs2 syll	n/m	B	nep morren mieren nis	2vs2 syll	n/m	B	nepper morren mieren nissan
14	1vs2 syll	n/m	W	nat merijn marien nek	2vs2 syll	n/m	W	nader merijn marien nimmer
15	1vs2 syll	n/m	S	niet remous romein nul	2vs2 syll	n/m	S	teniet remous romein genoot
16	1vs2 syll	n/m	N	neut raming rommel nies	2vs2 syll	n/m	N	genot raming rommel genies
17	1vs2 syll	b/v	B	bak vazen vezel bel	2vs2 syll	b/v	B	bakker vazen vezel boling
18	1vs2 syll	b/v	W	bot vazal vizier bit	2vs2 syll	b/v	W	botter vazal vizier bitter
19	1vs2 syll	b/v	S	bies zovéél zovér boek	2vs2 syll	b/v	S	debut zovéél zovér tabak
20	1vs2 syll	b/v	N	bon zuivel zever bed	2vs2 syll	b/v	N	gebod zuivel zever gebed
21	1vs2 syll	v/b	B	vak bellen balen vet	2vs2 syll	v/b	B	vader bellen builen veter
22	1vs2 syll	v/b	W	voet ballon balein vis	2vs2 syll	v/b	W	voeder ballon balein visser
23	1vs2 syll	v/b	S	vod labiel libel vies	2vs2 syll	v/b	S	gevat labiel libel devies
24	1vs2 syll	v/b	N	vin lebber lobben val	2vs2 syll	v/b	N	ravijn lebber lobben revier
25	1vs2 syll	p/k	B	pet kamer kommer pop	2vs2 syll	p/k	B	pieter kamer kommer pooier
26	1vs2 syll	p/k	W	pof kameel komeet pin	2vs2 syll	p/k	W	poging kameel komeet peiling
27	1vs2 syll	p/k	S	poet mekaar makaak poch	2vs2 syll	p/k	S	tepas mekaar makaak gepoch
28	1vs2 syll	p/k	N	poes makker mocca peen	2vs2 syll	p/k	N	kaping makker mocca keper
29	1vs2 syll	k/p	B	kas pater peter kil	2vs2 syll	k/p	B	kajak pater peter ketter
30	1vs2 syll	k/p	W	kom patat potent kus	2vs2 syll	k/p	W	koter patat potent kussen
31	1vs2 syll	k/p	S	kop tapijt topaas keel	2vs2 syll	k/p	S	tekoop tapijt topaas bekeer
32	1vs2 syll	k/p	N	kor tepel tapir kiep	2vs2 syll	k/p	N	bekom tepel tapir tekijk
33	1vs2 syll	l/r	B	lies rakker rekel lach	2vs2 syll	l/r	B	liever rakker rekel ladder
34	1vs2 syll	l/r	W	log raket rekest los	2vs2 syll	l/r	W	logger raket rekest lover
35	1vs2 syll	l/r	S	lik karos koraal lef	2vs2 syll	l/r	S	gelik karos koraal beleg
36	1vs2 syll	l/r	N	lak kerel karig lam	2vs2 syll	l/r	N	belet kerel karig meloen
37	1vs2 syll	l/r	B	lis raven rover lied	2vs2 syll	l/r	B	lekker raven rover loeder
38	1vs2 syll	l/r	W	lot ravijn rivier lus	2vs2 syll	l/r	W	lokker ravijn rivier leiding
39	1vs2 syll	l/r	S	les varaan viriel loep	2vs2 syll	l/r	S	gelijk varaan viriel beloop
40	1vs2 syll	l/r	N	lol virus varen lor	2vs2 syll	l/r	N	geloof virus varen zeloot
41	1vs2 syll	j/l	B	jas later loting jek	2vs2 syll	j/l	B	jekker later loting jopper
42	1vs2 syll	j/l	W	juf latijn letaal jap	2vs2 syll	j/l	W	juffer latijn letaal jarig
43	1vs2 syll	j/l	S	jacht talent teloor joch	2vs2 syll	j/l	S	gejacht talent teloor gejuich
44	1vs2 syll	j/l	N	joop teler tema juich	2vs2 syll	j/l	N	gejok teler tema gejaagd
45	1vs2 syll	j/l	B	jacht loper liepen jol	2vs2 syll	j/l	B	jager loper liepen jokker
46	1vs2 syll	j/l	W	jas lapel lipoom jek	2vs2 syll	j/l	W	jammer lapel lipoom joker
47	1vs2 syll	j/l	S	jool paleis poliep jucht	2vs2 syll	j/l	S	gejoel paleis poliep bejuicht
48	1vs2 syll	j/l	N	jaar paling peiler juf	2vs2 syll	j/l	N	bejaard paling peiler gejouw
49	2vs2 syll	d/j	B	dolen jura jarig duiker	1vs2 syll	d/j	B	doof jura jarig duit
50	2vs2 syll	d/j	W	duvel jorien jeroen doper	1vs2 syll	d/j	W	dut jorien jeroen doek
51	2vs2 syll	d/j	S	gedoe radijs redoute gedaas	1vs2 syll	d/j	S	dos radijs redoute dar
52	2vs2 syll	d/j	N	gedut redding roedel bedompt	1vs2 syll	d/j	N	dun redding roedel dom
53	2vs2 syll	z/d	B	zaling dame duimel zakker	1vs2 syll	z/d	B	zaal dame duimel zak
54	2vs2 syll	z/d	W	zuiger domein damast zaling	1vs2 syll	z/d	W	zuig domein damast zaal
55	2vs2 syll	z/d	S	bezaan modern madam seizoen	1vs2 syll	z/d	S	zaak modern madam zoen
56	2vs2 syll	z/d	N	gezag moeder modder bezet	1vs2 syll	z/d	N	zag moeder modder zet
57	2vs2 syll	k/g	B	kennis gekkie gele kater	1vs2 syll	k/g	B	ken golem gele kaai
58	2vs2 syll	k/g	W	kotter geluk geloof kapper	1vs2 syll	k/g	W	kot geluk geloof kap
59	2vs2 syll	k/g	S	bekijk legaat legaal bekort	1vs2 syll	k/g	S	kijk legaat legaal kort
60	2vs2 syll	k/g	N	bekoor lachen lichaam bekoel	1vs2 syll	k/g	N	koor lachen lichaam koel
61	2vs2 syll	g/k	B	gaper kale koele gene	1vs2 syll	g/k	B	gas kale koele geen
62	2vs2 syll	g/k	W	gokker kalot Colijn gister	1vs2 syll	g/k	W	gok kalot Colijn gist
63	2vs2 syll	g/k	S	begoot lakei loket begeef	1vs2 syll	g/k	S	gor lakei loket geef
64	2vs2 syll	g/k	N	begaf lakken lekken tegek	1vs2 syll	g/k	N	gaf lokken lekken gek
65	2vs2 syll	t/d	B	tijdig deken duiker topper	1vs2 syll	t/d	B	tijd deken duiker top
66	2vs2 syll	t/d	W	tukker decaan ducaat tinnef	1vs2 syll	t/d	W	tuk decaan ducaat tin
67	2vs2 syll	t/d	S	getij kadet kado beton	1vs2 syll	t/d	S	teil kadet kado ton
68	2vs2 syll	t/d	N	getik kader koddig getal	1vs2 syll	t/d	N	tik kader koddig tal
69	2vs2 syll	t/d	B	tuigen dapper doping togen	1vs2 syll	t/d	B	tuin dapper doping toog
70	2vs2 syll	t/d	W	tering depot depêche toner	1vs2 syll	t/d	W	teer depot depêche toon
71	2vs2 syll	t/d	S	getik pedant pedaal getob	1vs2 syll	t/d	S	teek pedant pedaal top
72	2vs2 syll	t/d	N	getal peddel padden baton	1vs2 syll	t/d	N	taal peddel padden tok
73	2vs2 syll	p/t	B	peter tekkel tikker poker	1vs2 syll	p/t	B	peet tekkel tikker pook
74	2vs2 syll	p/t	W	pieter tekeer tekort pover	1vs2 syll	p/t	W	pet tekeer tekort pof

75	2vs2 syll	p/t	S	gepees katoen katijf bepoot	1vs2 syll	p/t	S	pees katoen katijf poes
76	2vs2 syll	p/t	N	bepakt ketel kater gepeins	1vs2 syll	p/t	N	pad ketel kater peins
77	2vs2 syll	d/z	B	dader zuilen zaling daler	1vs2 syll	d/z	B	daad zuilen zaling dal
78	2vs2 syll	d/z	W	dekking zeloot zolang duiten	1vs2 syll	d/z	W	dek zeloot zolang duit
79	2vs2 syll	d/z	S	gedaan lazuur lysol gedoopt	1vs2 syll	d/z	S	daan lazuur lysol dop
80	2vs2 syll	d/z	N	bedot lozing lezer gedimd	1vs2 syll	d/z	N	dog lozing lezer dim
81	2vs2 syll	sj/s	B	sjalen sufferd saffie sjezen	1vs2 syll	sj/s	B	sjaal sufferd saffie sjees
82	2vs2 syll	sj/s	W	sjieker sofie saffier sjacher	1vs2 syll	sj/s	W	sjiek sofie saffier sjah
83	2vs2 syll	sj/s	S	gesjouw facet fossiel gesjok	1vs2 syll	sj/s	S	sjouw facet fossiel sjaak
84	2vs2 syll	sj/s	N	gesjor fasces facie gesjans	1vs2 syll	sj/s	N	sjoerd fasces facie sjans
85	2vs2 syll	s/sj	B	sijpel sjoemel sjekkie sollen	1vs2 syll	s/sj	B	sip sjoemel sjekkie sof
86	2vs2 syll	s/sj	W	soppen sjamaan chauffeur cello	1vs2 syll	s/sj	W	sop sjamaan chauffeur cel
87	2vs2 syll	s/sj	S	gesip machien michel gesol	1vs2 syll	s/sj	S	sik machien michel sol
88	2vs2 syll	s/sj	N	besef misje muisje gesim	1vs2 syll	s/sj	N	sein misje muisje sim
89	2vs2 syll	v/z	B	vitter zone zanik voeder	1vs2 syll	v/z	B	vit zone zanik vos
90	2vs2 syll	v/z	W	vielen zonee zonaal vodden	1vs2 syll	v/z	W	vief zonee zonaal vod
91	2vs2 syll	v/z	S	schavuit nasaal nazist tevol	1vs2 syll	v/z	S	vuist nasaal nazist vol
92	2vs2 syll	v/z	N	geval nazi neuzen gevang	1vs2 syll	v/z	N	val nazi neuzen vang
93	2vs2 syll	z/v	B	zessen vutter vette zullen	1vs2 syll	z/v	B	zes vutter vette zal
94	2vs2 syll	z/v	W	zeilen votief vitaal zagen	1vs2 syll	z/v	W	zeil votief vitaal zaag
95	2vs2 syll	z/v	S	bazin tevoet gevat gazon	1vs2 syll	z/v	S	zin tevoet gevat zon
96	2vs2 syll	z/v	N	dozijn tover toeval gezang	1vs2 syll	z/v	N	zijn tover toeval zang

Appendix B. Stimuli Experiment 2

stim nr	not elic	cons	cond	stimulus	elic	cons	cond	stimulus
01	not elic		B	water rapper lommer bikkel	elic	w/r	B	water rapper roeper wallen
02	not elic		W	nuttig lamel poreus niessen	elic	w/r	W	woeker rapport rapier wikkel
03	not elic		S	noteer maleis rapier varaan	elic	w/r	S	bewijs parijs poreus juweel
04	not elic		N	geniet lemming weiland gevit	elic	w/r	N	lawaai pieren parel gewin
05	not elic		B	nepper morren ridder wegen	elic	w/r	B	waggel radar ridder wegen
06	not elic		W	moter remous zover takken	elic	w/r	W	woning radauw radijs wielen
07	not elic		S	papier merijn vizier bedot	elic	w/r	S	gewoon direct dorien bewaar
08	not elic		N	genot raming lodder bezit	elic	w/r	N	gewis dieren duren beween
09	not elic		B	bakker vazen maken nodig	elic	n/m	B	neder molen maken nodig
10	not elic		W	doeken zoveel romein goten	elic	n/m	W	noten maleis meloen nuffig
11	not elic		S	beman vazal radijs gewoel	elic	n/m	S	benut lamel limiet genies
12	not elic		N	geloof virus tafel tekijk	elic	n/m	N	geniet lemmet lommer benul
13	not elic		B	neder molen roeper wallen	elic	n/m	B	nepper morren mieren nissan
14	not elic		W	wijzer parijs dorien gozer	elic	n/m	W	nader merijn marien nimmer
15	not elic		S	rozijn patat rekest manier	elic	n/m	S	teniet remous romein genoot
16	not elic		N	gebod zuiver lommer rivier	elic	n/m	N	genot raming rommel genies
17	not elic		B	vader bellen kommer pooier	elic	b/v	B	bakker vazen vezel boling
18	not elic		W	vatten labiel topaas doelen	elic	b/v	W	botter vazal vizier bitter
19	not elic		S	tomaat kanon boleet gevat	elic	b/v	S	debuut zovéél zovér tabak
20	not elic		N	ravijn lening zeker gebed	elic	b/v	N	gebod zuivel zever gebed
21	not elic		B	pieter kamer vezel bowling	elic	v/b	B	vader bellen builen veter
22	not elic		W	passen mekaar libel buiging	elic	v/b	W	voeder ballon balein visser
23	not elic		S	rabbijn kameel balein tonijn	elic	v/b	S	gevat labiel libel devies
24	not elic		N	tapijt makker wever gebed	elic	v/b	N	ravijn lebber lobben revier
25	not elic		B	koper pater rekel ladder	elic	p/k	B	pieter kamer kommer pooier
26	not elic		W	kochten tapijt makaak pochen	elic	p/k	W	poging kameel komeet peiling
27	not elic		S	kaneel rapport meloen figuur	elic	p/k	S	tepas mekaar makaak gepoch
28	not elic		N	bekom tepel karig malloot	elic	p/k	N	tapijt matter motte kapot
29	not elic		B	liever rakker peter ketter	elic	k/p	B	kajak pater peter ketter
30	not elic		W	likken karos topaas duiker	elic	k/p	W	koter patat potent kussen
31	not elic		S	balein raket potent kozijn	elic	k/p	S	tekoop tapijt topaas bekeer
32	not elic		N	belet kerel mocca tapuit	elic	k/p	N	bekom tepel tapir tekijk
33	not elic		B	lekker raven builen veter	elic	l/r	B	liever rakker rekel ladder
34	not elic		W	lijken varaan koraal zeggen	elic	l/r	W	logger raket rekest lover
35	not elic		S	loket ravijn banaan pedant	elic	l/r	S	gelik karos koraal beleg
36	not elic		N	lawaai pieren bakken gesop	elic	l/r	N	belet kerel karig meloen
37	not elic		B	jekker later doping koppig	elic	l/r	B	lekker raven rover loeder
38	not elic		W	jager talent viriel zoeker	elic	l/r	W	lokker ravijn rivier leiding
39	not elic		S	meneer latijn rivier moreel	elic	l/r	S	gelijk varaan viriel beloop
40	not elic		N	gejok teler varen bezoek	elic	l/r	N	geloof virus varen zeloot
41	not elic		B	jager loper rover moedig	elic	j/l	B	jekker later loting jopper
42	not elic		W	joelen paleis roman moker	elic	j/l	W	juffer latijn letaal jarig
43	not elic		S	kopij kajuit metaal rozijn	elic	j/l	S	gejacht talent teloor gejuich
44	not elic		N	bejaard paling deren gemaal	elic	j/l	N	gejok teler palet gejaagd
45	not elic		B	waggel radar mieren nissan	elic	j/l	B	jager loper liepen jokker
46	not elic		W	duvel jorien jeroen doper	elic	j/l	W	jammer lapel lipoom joker
47	not elic		S	konijn direct limiet fineer	elic	j/l	S	gejoel paleis poliep bejuicht
48	not elic		N	rebel bader mare bewaar	elic	j/l	N	bejaard paling peiler gejouw
49	elic	d/j	B	dolen jura jarig duiker	not elic		B	dolen jura kater gokken
50	elic	t/d	W	tover dozijn defect toeval	not elic		W	garen radijs legaal beker
51	elic	d/b	S	gedoe robijn gebal gedaas	not elic		S	katoen poliep piraat robijn
52	elic	t/d	N	getut redding roedel beton	not elic		N	genot redding mijlen berijk

53	elic	z/d	B	zaling dame duimel zakker	not elic		B	zaling dame gokken kater
54	elic	z/d	W	zuiger domein damast zaling	not elic		W	bazig tevoet gekat vallen
55	elic	z/d	S	bezaan modern madam seizoen	not elic		S	sigaar domein baron ballon
56	elic	z/d	N	gezag moeder modder bezet	not elic		N	gezag moeder roebel konijn
57	elic	k/g	B	kennis gekkie gele kater	not elic		B	kennis gekkie dadel poeder
58	elic	k/g	W	kotter geluk geloof kapper	not elic		W	boter legaat madam deining
59	elic	k/g	S	bekijk legaat legaal bekort	not elic		S	getob geluk damast venijn
60	elic	k/g	N	bekoor lachen lichaam bekoel	not elic		N	bekoor lachen modder bezet
61	elic	g/k	B	gaper kale koele gene	not elic		B	pater kale gokker later
62	elic	g/k	W	gokker kalot Colijn gister	not elic		W	gotisch pineut legaal teken
63	elic	g/k	S	begoot lakei loket begeef	not elic		S	gemaal kalot majoor rekest
64	elic	g/k	N	begaf lakken lekken tegek	not elic		N	begaf lekken modder facet
65	elic	t/d	B	tijdig deken duiker topper	not elic		B	tijdig deken roken bodem
66	elic	t/d	W	tukker decaan ducaat tinnef	not elic		W	geiten kadet pedaal boete
67	elic	t/d	S	getij kadet kado beton	not elic		S	bevel decaan pineut gemok
68	elic	t/d	N	getik kader koddig getal	not elic		N	gelik kapper padden beton
69	elic	t/d	B	tuigen dapper doping togen	not elic		B	tuigen dapper kikker poker
70	elic	t/d	W	tering depot depêche toner	not elic		W	balen pedant latei kippig
71	elic	t/d	S	getik pedant pedaal getob	not elic		S	teleen depôt metaal gemeen
72	elic	t/d	N	getal peddel vodden baton	not elic		N	getal peddel kamer balein
73	elic	p/t	B	peter tekkel tikker poker	not elic		B	peter tekkel duivel rapper
74	elic	p/t	W	pieter tekeer tekort pover	not elic		W	benig katoen lysol poker
75	elic	p/t	S	gepees katoen katijf bepoot	not elic		S	terrein tekeer zolang roman
76	elic	p/t	N	bepakt ketel kater gepeins	not elic		N	bepakt netel lekken piloot
77	elic	d/z	B	dader zuilen zaling daler	not elic		B	vader zuilen doping toegang
78	elic	d/z	W	dekking zeloot zolang duiten	not elic		W	danig lazuur fossiel kapper
79	elic	d/z	S	gedaan lazuur lysol gedoopt	not elic		S	ducaat zeloot gedimd lazuur
80	elic	d/z	N	gewis dieren bezig palet	not elic		N	bedot lozing jager gepeins
81	elic	sj/s	B	sjalen sufferd saffie sjezen	not elic		B	sjalen sufferd tikker pooier
82	elic	sj/s	W	sjieker sofie saffier sjacher	not elic		W	sjouwer facet katijf bijten
83	elic	sj/s	S	gesjouw facet fossiel gesjok	not elic		S	tevoet sofie zolang bemind
84	elic	sj/s	N	gesjor rozig lezen gesjouw	not elic		N	gesjor fakkel muizen begin
85	elic	s/sj	B	sijpel sjoemel sjekkie sollen	not elic		B	sijpel sjoemel zanik voeder
86	elic	s/sj	W	soppen sjamaan chauffeur cello	not elic		W	sippen machien bevel pater
87	elic	s/sj	S	gesip machien michel gesol	not elic		S	makaak sjamaan banaal kapoen
88	elic	s/sj	N	besef misje muisje gesim	not elic		N	besef misje ketting gevang
89	elic	v/z	B	vitter zône zanik voeder	not elic		B	vitter zône koker gading
90	elic	k/p	W	kamer piloot palet koter	not elic	v/z	W	jochie nasaal gebed kapen
91	elic	v/z	S	schavuit nasaal bezoek tevol	not elic	v/z	S	verrot chinees gezag vitaal
92	elic	v/z	N	geval razen neuzen gevang	not elic	v/z	N	geval nozem fakir gesjans
93	elic	z/v	B	zessen vutter vette zullen	not elic	z/v	B	zessen vutter bellen roedel
94	elic	z/v	W	zeilen votief vitaal zagen	not elic	z/v	W	koter tevoet gekat ratten
95	elic	z/v	S	bazin tevoet gevat gazon	not elic	z/v	S	zeloot votief japon kaneel
96	elic	d/z	N	bedot lozing lezer gedimd	not elic	z/v	N	dozijn tover toekan kapel

Appendix C.

Table C1 gives a first breakdown of the responses we obtained in this experiment.

In our further analysis we mainly focus on the category of valid errors. From Table C1 it is clear that the 2 + 2 + 2 + 2 stimuli are much more successful than the 1 + 2 + 2 + 1 stimuli in eliciting interactional segmental errors. This was to be expected, because, given that initial and medial segments rarely interact with each other, segments in the 1 + 2 + 2 + 1 stimuli simply have fewer opportunities for interaction than segments in the 2 + 2 + 2 + 2 stimuli (for the strong effect of shared position on error frequencies see Nooteboom & Quené, 2015a). Also the lack of prosodic similarity between one-syllable and two-syllable words may be involved (Nooteboom & Quené, 2015b).

Table C2 gives a further breakdown of the valid errors by the positions of the intended error being elicited by the tongue twister and of the realized valid error. For example, in the stimulus "nut lamel limiet nies" the initial consonant position is targeted for interaction, and the elicited error (if any) is therefore in initial position. This applies to all 1 + 2 + 2 + 1 stimuli. The intended position of the error corresponds with the condition: B and W conditions are meant to elicit errors in word initial position, and the S and N positions to elicit errors in medial positions (cf Table 2.1).

Table C2 shows that for the 1 + 2 + 2 + 1 stimuli, by far the most errors are indeed elicited in initial position, in accordance with the position targeted for interaction. Note that in the 1 + 2 + 2 + 1 stimuli medial consonants cannot be successfully targeted for interaction because they do not share this position in the four words. This is because initial consonants tend to interact with initial consonants and medial consonants with medial

Table C1
Numbers of responses obtained in Experiment 1. Invalid errors comprise all errors other than non-repeated single segment interactional substitutions in initial and medial position. "1 + 2 + 2 + 1" stands for a sequence of one syllable + two syllables + two syllables + one syllable. "2 + 2 + 2 + 2" stands for a sequence of four two-syllable words.

	1 + 2 + 2 + 1	2 + 2 + 2 + 2	Total
Fluent correct	6244	5436	11,680
Valid errors	315	509	824
Invalid errors	1509	2441	3950
Total	8068*	8386*	16,454

* *Note:* If only a single error could have been made per spoken tongue twister, these numbers would have been 27 × 48 × 6 = 7776. The surplus stems from multiple errors per response utterance.

Table C2
Numbers of valid errors, broken down by numbers of syllables in the stimulus, by position of the intended error (rows) and by position of the realized error (columns).

		Position of realized error	
		Initial	Medial
1 + 2 + 2 + 1			
	Initial (B, W)	201	8
	Medial (S, N)	94	12
2 + 2 + 2 + 2			
	Initial (B, W)	215	70
	Medial (S, N)	142	82

Table C3
Numbers of unrepaired and repaired single segmental interactional substitutions separately for 1 + 2 + 2 + 1 and 2 + 2 + 2 + 2 stimuli and for word initial and medial positions.

	1 + 2 + 2 + 1			2 + 2 + 2 + 2		
	Initial	Medial	Total	Initial	Medial	Total
Unrepaired	82	9	91	73	59	132
Repaired	213	11	224	284	93	377
Tot	295	20	315	357	152	509

consonants (Nooteboom & Quené, 2015). In this sense, in the 2 + 2 + 2 + 2 stimuli both positions are equal. Yet, we see that also in the 2 + 2 + 2 + 2 stimuli eliciting errors in initial position is more successful than eliciting errors in medial position. This was investigated by means of a Generalized Linear Mixed Model (GLMM, Quené & Van den Bergh, 2008) on the 2 + 2 + 2 + 2 stimuli only, with intended position as the only fixed predictor. Participants and item sets (matching stimuli) were included as random intercepts, and intended position was included as a random slope at the participant level. Results confirm the lower prevalence of realized errors when elicited in medial position (beta = –0.2883, Z = –2.729, p = .0064) as compared to initial position (baseline, beta = –0.314). Interestingly, targeting errors in medial position yields most errors in the non-targeted, i.e. initial position (142/224 = 63%), similar to the pattern observed for errors elicited in initial position (215/285 = 75%): By far the most of the valid errors were made in initial position, irrespective of which position was targeted for interaction (see Table C3).

Appendix D.

Table D1 gives a first breakdown of the responses we obtained in this experiment 2.

In our further analysis we mainly focus on the category of valid errors.

Table D2 gives a further breakdown of the valid errors as to whether the error was "targeted" or not. By "targeted" we mean that the error was elicited in the "eliciting" condition (but not in the "not eliciting" condition) by the structure of the tongue twister. In a stimulus such as "vader bellen builen veter", in the eliciting condition the initial position is targeted for interaction by consonant repetition in that position. In the corresponding

Table D1
Numbers of responses obtained in Experiment 2. Invalid errors comprise all errors other than non-repeated single segment interactional substitutions in initial and medial position.

	Eliciting	Not eliciting	Total
Fluent correct	6139	6752	12,891
Valid Errors	509	467	976
invalid errors	2720	1985	4705
Total	9368*	9204*	18,572

* *Note:* If only a single error could have been made per response utterance, these numbers would have been 30 × 48 × 6 = 8640. The surplus stems from multiple errors per response utterance.

Table D2
Numbers of valid errors, broken down by eliciting versus not eliciting, by position of the intended error (rows) and by position of the realized error (columns).

		Position of realized error		
		Initial	Medial	Total
Eliciting				
	Initial (B, W)	206	76	282
	Medial (S, N)	120	107	227
Not eliciting				
	Initial (B, W)	149	70	219
	Medial (S, N)	152	96	248

stimulus "vader bellen kommer pooier" in the "non eliciting" condition, although no interaction is elicited, we still refer to the initial position as "targeted", for reasons of comparison. This enables us to assess the effect of eliciting versus not eliciting interaction in a specific position under otherwise comparable conditions.

The effect of the position of elicitation on the number of errors in the two positions of realized errors together was investigated by a GLMM with the intended position as the only fixed predictor, with participants and item sets (matching stimuli) as random intercepts, and with intended position as random slope at the participant level. Results indicate that the overall error rates are approximately equal for conditions where errors were **intended** (elicited or not elicited) to be made in initial position (282 + 219, 6.6%) and in medial position (227 + 248, 6.2%) (beta = −0.076, Z = −1.06, p = .289). Thus both conditions of intended error position were equally effective in generating errors. The results in Table D2 also show, however, that valid errors actually **realized** were more frequently made in initial position than in medial position, irrespective of the conditions of intended error position: even if errors were intended to be made in medial position, more errors were in fact realized in initial position (120 + 152) than in medial position (107 + 96).

In Table D3, we provide a breakdown of all valid single interaction errors as initial versus medial and not repaired versus repaired.

Table D3
Valid single interactional substitutions separately for initial and medial position and separately for not repaired and repaired errors.

	Initial	Medial	Total
Unrepaired	211	206	417
Repaired	416	142	558
Total	627	348	975

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bates, T., & Stough, C. (1997). Processing speed, attention, and intelligence: Effects of spatial attention on decision time in high and low *IQ* subjects. *Science Direct, 23*(5), 861–868.

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. (2009). Praat: Doing phonetics by computer (Version 5.1.05) [Computer program]. < http://www.praat.org/ > .

Choe, W. K., & Redford, M. A. (2012). The distribution of speech errors in multi-word prosodic units. *Laboratory Phonology, 3*(1), 5–26.

Croot, K., Au, C., & Harper, A. (2010). Prosodic structure and tongue twister errors. In C. Fougeron, B. Kuehnert, M. d'Imperio, & N. Vallée (Eds.). *Papers in laboratory phonology 10: Variation, phonetic detail and phonological representation* (pp. 433–459). Berlin: De Gruyter Mouton.

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

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

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.

Garnham, E., Shillcock, R. S., Brown, G. D., Mill, A. I. D., & Cutler, A. (1982). Slips of the tongue in the London-Lund corpus of spontaneous conversations. In A. Cutler (Ed.). *Slips of the Tongue and Language Production* (pp. 261–263). Berlin: Mouton.

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.

Goldrick, M., Keshet, J., Gustafson, E., Heller, J., & Needle, J. (2016). Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition, 149*, 31–39.

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., & 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., 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*(5), 921–932.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nat. Rev. Neurosci. 13*, 135–145.

Kleiber, C., & Zeileis, A. (2008). *Applied Econometrics with R*. New York: Springer-Verlag.

Kolk, H. H. J. & Postma, A. (1997). Stuttering as a covert-repair phenonemon. In: R. Corlee & G. Siegel (Eds.), Nature and Treatment of Stuttering: New Directions (pp. 182–203), Boston: Allyn & Bacon.

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

Levelt, W. J. M. (1989). *Speaking. From Intention to Articulation.* Harvard, MA: 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.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions during word-recognition in continuous speech. *Cognitive Psychology, 10*, 29–63.

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. (2011). Self-monitoring for speech errors in novel phrases and phrasal lexical items. *Yearbook of Phraseology, 2*, 1–16.

Nooteboom, S. G., & Quené, H. (2015). Word onsets and speech errors. Explaining relative frequencies of segmental substitutions. *Journal of Memory and Language, 78*, 33–46.

Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language, 95*, 19–35.

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

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience, 31*(7), 886–903.

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.

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

Piai, V., Roelofs, A., & Schriefers, H. (2012). Distractor strength and selective attention in picture-naming performance. *Memory and Cognition, 40*(6), 614–627.

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

Pinheiro, J. C., & Bates, D. M. (2002). *Mixed Effect Models in S and S-plus*. New York: Springer Verlag.

Postma, A., & Kolk, H. H. J. (1993). The covert repair hypothesis. Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research, 36*(3), 472–487.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*, 413–425.

R Development Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: http://www.R-project.org.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modelling attentional control in the Stroop task. *Psychological Review, 110*(1), 88–125.

Roelofs, A. (2005). Spoken word planning, comprehending, and self-monitoring: Evaluation of WEAVER++. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.). *Phonological encoding and monitoring in normal and pathological speech* (pp. 42–63). Hove, UK: Psychology Press.

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*, 837–842.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition, 42*, 213–259.

Slis, A., & Van Lieshout, P. (2016). The effect of phonetic context on the dynamics of intrusions and reductions. *Journal of Phonetics, 57*(3), 430–445.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26*, 24–36.

Tydgat, I., Diependaele, K., Hartsuiker, R. J., & Pickering, M. J. (2012). How lingering representations of anbandoned context words affect speech production. *Acta Psychologica, 140*, 218–229.

Tydgat, I., Stevens, M., Hartsuiker, R. J., & Pickering, M. J. (2011). Deciding where to stop speaking. *Journal of Memory and Language, 64*, 359–380.

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

Wilshire, C. E. (1998). Serial order in phonological encoding: an exploration of the word onset effect using laboratory-induced errors. *Cognition, 68*, 143–166.