Paper:
Plug, L and Carter, P (2014) *Timing and tempo in spontaneous phonological error repair.* Journal of Phonetics, 45. 52 - 63.

http://dx.doi.org/10.1016/j.wocn.2014.03.007

Timing and tempo in spontaneous phonological error repair

Leendert Plug & Paul Carter
University of Leeds

Published in Journal of Phonetics 45: 52-63
http://dx.doi.org/10.1016/j.wocn.2014.03.007

Abstract

This paper reports on a study of the temporal characteristics of phonological error repair in spontaneous Dutch speech, with a focus on how the articulation rate of the correct target word production — the repair — compares to that of the preceding erroneous target word attempt — the reparandum. The study is motivated by two findings from recent independent studies: first, that self-repair is generally associated with relative temporal compression — that is, a local increase in articulation rate — following the repair initiation; second, that the timing of the repair initiation relative to the error is consequential for the prosody of the repair component. This study investigates to what extent these findings generalise to a collection of spontaneous phonological error repairs sampled from the Spoken Dutch Corpus. The study also asks how best to quantify repair timing, considers whether timing is consequential for the duration of the 'offset-to-repair interval', and tests for effects of lexical frequency. The results confirm that temporal compression following the repair initiation is more common than temporal expansion, and that repair timing has a significant effect on both offset-to-repair duration and repair tempo — at least in a subset of the data. A frequency effect is also observed. The results suggest that proportional measures of target word completeness provide the most informative quantifications of repair timing in modelling the overall temporal organisation of phonological error repairs.

Key words
Speech error, Repair, Self-monitoring, Articulation rate, Duration, Dutch

1 Introduction

Speech errors and infelicities and their repairs have long been of interest to phoneticians and psycholinguists alike (e.g. Fromkin 1973, Baars et al. 1975, Nooteboom 1980, Levelt 1983, Levelt & Cutler 1983, Blackmer & Mitton 1991, Postma & Kolk 1993, Shattuck-Hufnagel & Cutler 1999, Shriberg 2001, Jasperson 2002, Nooteboom 2005a, Hartsuiker 2006, Seyfeddinipur et al. 2008, Nooteboom 2010, Plug 2011, Tydgat et al. 2011). For phoneticians, the interest lies primarily in how speakers deal with the disfluency associated with repair, while for psycholinguists repair 'may reveal principles of organization of the speech production process that would be hard to discover on the basis of laboratory data alone' (Levelt 1984: 105) — in particular with reference to self-monitoring.

In fact, insights into the organisation of speech production processes can be gained exactly from a careful consideration of the phonetic details of repair, as shown by studies of prosodic 'marking' (Cutler 1983, Levelt & Cutler 1983, Howell & Young 1991) and the temporal organisation of repair (Blackmer & Mitton 1991, Oomen & Postma 2001, 2002, Seyfeddinipur et al. 2008, Nooteboom 2010, Plug 2011). For example, Blackmer & Mitton (1991) report that a substantial proportion of error corrections involve no delay between the abandonment of the erroneous lexical item and the onset of the repair item. This is inconsistent with Levelt's (1989) proposal that the repair is planned during this 'offset-to-repair' interval: at least in a proportion of instances, planning must precede the abandonment of the erroneous lexical item, and therefore take place simultaneous with its ongoing articulation. Oomen & Postma (2001) show that offset-to-repair durations are lower, and the proportion of zero durations higher, in faster speech. This is inconsistent with the idea that fast speech allows for less extensive look-ahead in the speech production process than slow speech (cf. Blackmer & Mitton 1991); rather, increasing speaking tempo appears to speed up self-monitoring for errors as well. More recently, Seyfeddinipur et al. (2008) have used measurements of offset-to-repair duration in comparing competing claims regarding error detection.

In this paper we report on an investigation of the temporal characteristics of phonological error — or mispronunciation — repair.[1] The primary focus of the investigation is on the speaking tempo during the repair, although we will also consider the duration of offset-to-repair interval. The investigation builds on that of Plug (2011), who analysed instances of self-repair in terms of their tempo following the repair initiation. Plug reports a predominance of temporal compression — that

[1] Among the various other terms for the type of error whose repair is considered here are 'sound-form error' (Levelt 1989), 'sound error' (Meyer 1992), 'phonetic error' (Brédart 1991), 'sub-lexical error' (Frisch & Wright 2002), 'phonological speech error' (Frisch & Wright 2002, Nooteboom 2005a) and 'speech error' (Nooteboom 2010). We will use 'phonological error' throughout, following Levelt (1983), Postma (2000), Oomen & Postma (2001, 2002) and others.

is, a relative speeding up after the repair initiation, and argues that this is unexpected on the basis of the 'H&H' theory of speech production (Lindblom 1990, 2000), which predicts a predominance of hyper-articulation in the context of repair. While informative, Plug's study has several limitations. First, it is based on a relatively small data set comprising various distinct types of repair: in addition to phonological and lexical repairs there are syntactic error repairs and 'different' repairs (see Levelt 1983). It is possible that distinct repair types are subject to differing temporal constraints; however, Plug does not address this possibility. Second, Plug's statistical modelling of repair tempo incorporates a small number of candidate predictors only, with a focus on the relevance of the distinction between error repairs, in which a factual or linguistic error is corrected, and appropriateness repairs, in which a correct but infelicitous phrasing is replaced (see Levelt & Cutler 1983). Recent findings by Nooteboom (2010) suggest that a factor not considered by Plug (2011) may be highly relevant to repair tempo, at least for phonological error repairs. This factor is the timing of the repair.

In a study of elicited errors and their repairs, such as sa ... fat soap or sat soap ... fat soap, Nooteboom (2010) observes that instances in which the repair comes in very early, as in sa ... fat soap, tend to have a repair component with a high pitch and intensity prominence on the first vowel compared with the reparandum. Instances in which the mispronounced word is completed before the onset of repair, as in sat soap ... fat soap, tend to have a repair component with a low pitch and intensity prominence on the first vowel compared with the reparandum. Nooteboom does not consider the tempo of the repairs, but if we associate high pitch and intensity with relative emphasis, or hyper-articulation (Lindblom 1996, Smiljanić & Bradlow 2009, Niebuhr 2010), we would predict that 'early' repairs are more likely to be associated with temporal expansion following repair initiation.

In fact, Nooteboom's account of the prosodic differentiation of phonological error repairs suggests exactly the opposite. Nooteboom's motivation for comparing 'early' and 'late' repairs is the idea that they result from different coordinations of two self-monitoring mechanisms: 'inner speech' monitoring, which monitors the compilation of a pre-articulatory speech plan, and 'overt speech' monitoring, which monitors the output of articulation (Levelt 1989, Levelt et al. 1999, Nooteboom 2005b, Hartsuiker et al. 2005a, b).[2] When an error is detected in self-monitoring, the monitor sends a signal to abandon ongoing speech and plan a repair. Several studies have suggested a delay of

---

[2] Several other terms are used to refer to these two types of self-monitoring (see e.g. Oomen & Postma 2001, 2002): inner speech monitoring is also called 'internal channel', 'inner loop' and 'pre-articulatory' monitoring, while overt speech monitoring is also called 'external channel', 'auditory loop' and 'post-articulatory' monitoring. We follow Nooteboom (2010) in this paper.

approximately 150–250 ms between error detection and the abandonment of an ongoing utterance (Marslen-Wilson & Tyler 1981, Levelt 1989, Blackmer & Mitton 1991, Hartsuiker & Kolk 2001, Slevc & Ferreira 2006) — a delay during which erroneous production may continue (Hartsuiker et al. 2008, Tydgat et al. 2011). This means that repairs with a reparandum duration below 150–250 ms — as is the case for most 'early' repair instances such as *sa … fat* soap — are likely to be occasioned by an error detection that precedes the overt onset of the mispronunciation: in other words, an error detection in inner speech monitoring (Levelt 1989, Blackmer & Mitton 1991, Nooteboom 2005a, b). Repairs with a reparandum duration well above 250 ms — as is the case for most 'late' instances such as sat soap … fat soap — are likely to be occasioned by an error detection in overt speech monitoring instead (Nooteboom 2005a, b, Hartsuiker et al. 2005b).[3]

Nooteboom (2010) argues that inner and overt speech monitoring serve different purposes in the speech production system, and are therefore subject to different constraints; these different constraints can in turn explain the different prosodic characteristics of 'early' versus 'late' repairs. Among other things, he suggests that the purpose of inner speech monitoring is to 'prevent errors … from becoming public' (Nooteboom 2010: 215); therefore, a major characteristic of the process is that it operates under considerable time pressure, and aims to minimise disfluency in production. Once the erroneous form has been produced, on the other hand, it is clear that fluency will need to be sacrificed and 'the speaker should take his or her time to make clear to the listener that an error has been made' (Nooteboom 2010: 216). Consistent with this account, Nooteboom reports that 'early' repairs have significantly lower offset-to-repair durations than 'late' ones (Nooteboom 2010: 223–224; see also Seyfeddinipur et al. 2008).[4] If Nooteboom is correct in concluding that repairs of errors detected by the inner speech monitor 'come fast' and are produced under time pressure, while repairs of errors detected by the overt speech monitor 'come very much slower' and allow the speaker to 'take his or her time' to produce the repair (Nooteboom 2010: 116, 224), a reasonable prediction is that we will observe an effect of repair timing on the speaking rate of the repair component. Concretely, if Nooteboom's theoretical account stands up we would predict that 'early' repairs are more likely than 'late' ones to be associated with temporal compression following the repair onset.

The study presented in this paper had two main aims: first, to assess whether the predominance of temporal compression in spontaneous self-repair reported by Plug (2011) is

[3] Note that this argument rests on the assumption, formalised by Levelt (1989) as the 'Main Interruption Rule', that an error is repaired as soon as it is detected. This may well be incorrect, as suggested by Hartsuiker et al. (2008), Seyfeddinipur et al. (2008), Tydgat et al. (2011). We will return to this issue in Section 4.

[4] This in turn tallies well with the notion that in the case of 'early' repairs, repair planning may have started before the error is produced (Hartsuiker & Kolk 2001, Tydgat et al. 2011).

observed in a larger and more homogeneous data set — in this case one containing speech error repairs only; and second, to assess whether repair timing has any effects on repair tempo in spontaneous, unelicited speech error repair. In addressing the latter question, we also assess whether repair timing has an effect on offset-to-repair duration, and we compare multiple methods for quantifying repair timing, in order to improve our understanding of the units of execution involved in speech error repair. The results of the study are presented in Section 3, followed by discussion in Section 4 and conclusions in Section 5. In Section 2, we first describe our data set and analysis methods.

2 Materials and method

2.1 Data selection

The data for this paper comprises 368 instances of speech error repair extracted from four sub-corpora of the Spoken Dutch Corpus (Oostdijk 2002), containing spontaneous face-to-face conversations, interviews with teachers of Dutch, broadcast interviews, discussions and debates, and non-broadcast meetings, discussions and debates. We selected these subcorpora on the grounds that the recordings are predominantly of unscripted speech. We extracted instances of speech which were coded as mispronounced or interrupted — a small subset of which contain subsequent repair — and did a number of additional, unsystematic data trawls.[5] We discarded many potential instances because of poor audio quality or overlapping speech, and only included instances containing at least one consonant and one vowel. We included instances ambiguous between phonological and lexical repair if the immediate phonological context contained a plausible trigger for phonological error; Shattuck-Hufnagel & Cutler's (1999) tar ... *c*ar talk would be an example of this. (1) contains representative examples from our data set.

(1) a. [b]aarbij // [w]aarbij 'with which'
b. v[r]uch- // v[l]uchtmiddel 'means of escape'
c. vana[l] // vana[f] 'from'
d. een b[a]nkj- // een b[ɑ]nkje 'a bench'
e. [w]is ik z- // [m]is ik zeker 'miss I certainly'
f. zel[s]- // zel[f]werkzaamheid 'ability to work independently'

[5] Incomplete erroneous target word attempts are coded as both mispronounced and incomplete in the corpus annotation scheme, so the two corpus searches partly yielded the same results.

2.2 Segmentation and temporal measures

We segmented all instances of repair in PRAAT (Boersma and Weenink 2010). We placed boundaries at the start and end of the target word attempt and the start and end of the repair stretch, following the segmentation criteria set out by Rietveld & Van Heuven (1997). In a small subset of instances (N=36), the repair includes the repeat of one or more words preceding the target word attempt: see een 'a' in (1d) above. We segmented these as part of the reparandum, although separately from the target word attempt. Similarly, we included any lexical items following the target word attempt, as in (1e), in the reparandum as well as the repair stretch. Finally, in addition to the start of the erroneous target word attempt, we marked the start of the first erroneous segment within it — for example, in (1b) we marked the start of [r] in v[r]uch-; in (1c) the start of [l] in vana[l]; and in (1d) the start of the first vowel in b[e]d[e]-. In instances such as (1a) and (1e), this point of course coincides with the start of the target word attempt. Figure 1 illustrates our segmentation method, with reference to the repair in (1d).

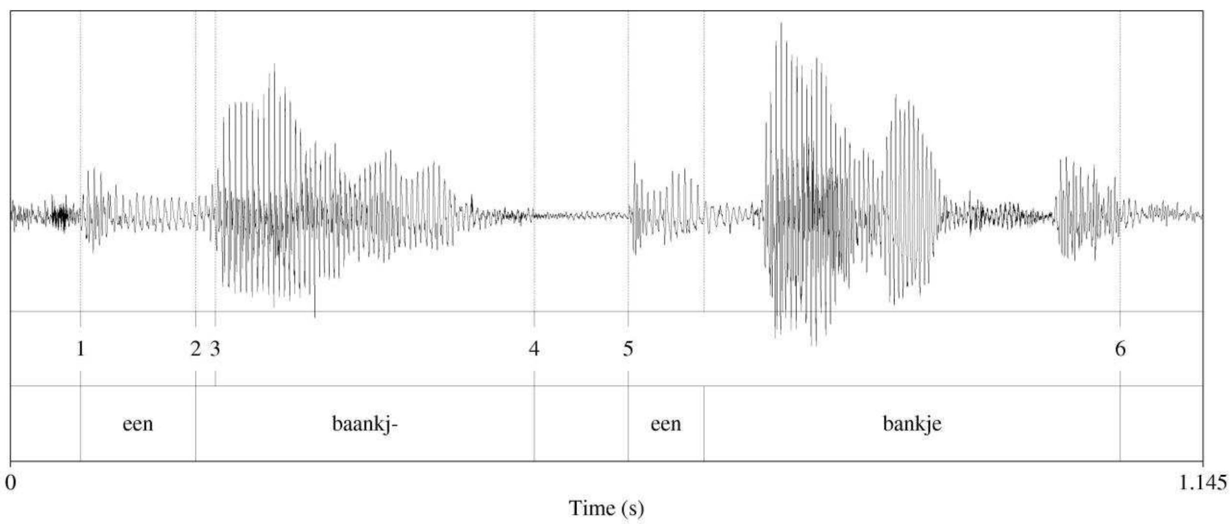


Figure 1. Illustrative segmentation of the repair een b[a]nkj- // een b[ɑ]nkje 'a bench' (1d). Boundary 1 marks the start of the reparandum; boundary 2 that of the target word and boundary 3 that of the error portion, at the onset of the erroneous vowel. Boundary 4 marks the reparandum, target and error offset. The portion between boundaries 4 and 5 constitutes the offset-to-repair interval, and the repair follows between boundaries 5 and 6.

We calculated the articulation rate for each segmented portion by dividing the number of surface segments articulated during the portion by its raw duration. We also used segment numbers and duration measurements independently in subsequent analysis, as measures of repair timing.

2.3 Quantifying repair timing

Of crucial importance to our study is the question of what constitutes an ‘early’ or ‘late’ repair. In Nooteboom’s (2010) study, speech errors and repairs are elicited using the SLIP technique (Baars et al. 1975). Subjects are asked to produce a CVC#CVC word pair, such as barn door, after being exposed to a number of different CVC#CVC sequences which prime an exchange of the word-initial consonants: in this case sequences such as dove ball, deer back and dark bone (Nooteboom 2010: 219). Nooteboom’s ‘early’ repairs are those in which the erroneous target production is abandoned after the first CV sequence, as in da ... barn door; and his ‘late’ repairs those in which a full CVC#CVC sequence is produced before repair, as in darn bore ... barn door. He does not discuss ‘intermediate’ instances: for example, instances in which the first word of the target two-word sequence is completed erroneously, then repaired — as in darn ... barn door — or instances in which the second word is interrupted — as in darn bo- ... barn door.

In the case of spontaneous phonological error repairs, it is often difficult to identify target word pairs. There are of course instances analogous to elicited (partial) spoonerisms, in which the error appears to be occasioned by the segmental make-up of a preceding or following word. However, there are also instances of apparently random segment substitution, omission or insertion, or erroneous gestural coordination (see Frisch & Wright 2002 for a review). In all such cases, it is possible to identify a target word, but not necessarily a target word pair. This means that Nooteboom’s comparison between ‘early’ CV repairs and ‘late’ CVC#CVC ones cannot be replicated exactly unless sampling is extremely restrictive. We can, however, compare instances in which the erroneous production of the target word is interrupted — as in (1b, d, f) above — with instances in which the target word is completed before being repaired — as in (1a, c, e). This is the approach taken in this study. If repair timing has an impact on repair tempo, we would expect to find a difference between these ‘interrupted error’ and ‘completed error’ repairs. Note that for the purpose of this study, all morphologically complex words, including compounds, were treated as single words: therefore, zels- in (1f) is treated as an interrupted error repair, even though the speaker has reached the final segment of the target morpheme zelf, which could be a word on its own in a different grammatical context.

Given that this study includes a consideration of a range of instances of repair which in Nooteboom's (2010) terminology can be called neither 'early' nor 'late', it would seem appropriate to implement a continuous measure of repair timing in addition to classifying instances as either 'interrupted error' or 'completed error' repair. This seems particularly worthwhile since in spontaneous error repairs, target word lengths vary. A repair of a three-syllable target form whose erroneous attempt is interrupted just before the final consonant can be considered less 'early' — and less clearly due to an error detection that precedes error production — than a repair of the same target in which the interruption comes after the first vowel. In fact, Nooteboom (2005a) reports a bimodal distribution of reparandum length in segments in his collection of elicited repairs, and uses this as supporting evidence for a qualitative distinction between 'early' and 'late' repairs.[6] One relevant question is, therefore, whether such a bimodal distribution is observed in our collection of spontaneous error repairs.

Moreover, if repair timing has an impact on repair tempo, we might discover patterns using a continuous measure of repair timing that would remain hidden if we implemented only a binary classification following Nooteboom (2010). For example, on the basis of the experimental evidence reviewed above, we might expect that repairs with a reparandum duration above 150–250 ms have different temporal characteristics than those with a duration below 150–250 ms, irrespective of whether they involve completed or interrupted target word attempts. Oomen & Postma (2001: 166, 2002: 168) and Hartsuiker et al. (2005b: 190) have questioned the validity of postulating a reparandum duration threshold in classifying error repairs as resulting either from inner or overt speech monitoring (see e.g. Liss 1998, who chooses a threshold of 500 ms in a study of repair by apraxic speakers). Implementing a continuous measure of repair timing should allow us to assess whether the data themselves provide evidence for such a threshold.

The matter is complicated by the fact that in spontaneous repairs of the type considered in our study, the location of the error in the target word varies. In Nooteboom's (2010) data, the error is always in the initial consonant: in other words, the error onset is simultaneous with the target word onset.[7] When this is the case, a completed error repair clearly involves either a later detection of the error or a later abandonment of speech after detection than an interrupted error repair. On the other

[6] Nooteboom's finding tallies well with Schegloff's (1979: 275) observation that in English spontaneous talk-in-interaction, repair is commonly initiated 'after the first sound of a word or just before its last sound'. Fox et al. (2009) show that this tendency can be observed to some extent across a number of genetically unrelated languages.

[7] Oomen & Postma (2001: 173) similarly operationalise what they call the 'error to cutoff time' as the duration from the start of the lexical item that is subject to repair to its end — even though in their data, too, the point of observable error in phonological error repairs must vary. Oomen & Postma (2002: 172), on the other hand, start measuring at the 'onset of the first erroneous phoneme'.

hand, when the error onset is not simultaneous with the target word onset, it is possible for a completed error repair to involve an earlier abandonment of speech relative to the point of observable error than an interrupted error repair. In terms of the self-monitoring mechanisms involved, the point of observable error constitutes the first point at which the overt speech monitor can detect the error, irrespective of how much of the target word has been produced already. If the involvement versus non-involvement of the overt speech monitor in the detection of a given error has an impact on the prosodic implementation of its repair, the earliest occasion for possible overt speech monitor detection would seem a highly relevant reference point for the implementation of a continuous measure of repair timing. Therefore, a relevant question in our study is whether continuous measures of repair timing which take the start of the first erroneous segment as the reference point perform differently from measures that refer to the start of the erroneous word in predicting the tempo of the subsequent repair.

On the basis of the above considerations, we took various measures of repair timing to enter into our statistical analyses. First, we implemented a binary dichotomy between repairs with an interrupted and a completed target word attempt. We will refer to this variable as Completeness. Second, we took a number of duration measurements and segment counts. These included the duration from the start of the reparandum (including any repeated lexical items) up to the abandonment of speech (Onset-to-offset duration) and the number of segments in this interval (Onset-to-offset segments); the duration and number of segments from the start of the target word to the abandonment of speech (Target-to-offset duration and Target-to-offset segments), and the duration and number of segments from the first erroneous segment to the abandonment of speech (Error-to-offset duration and Error-to-offset segments). We also measured the offset-to-repair interval (Offset-to-repair duration), which allowed us to enter the additional variables Onset-to-repair duration, Target-to-repair duration and Error-to-repair duration. Third, we implemented two proportional measures of target word completeness, by dividing the number of segments in the target word attempt by the number of segments in the subsequent complete realisation of the target word (Proportional target segments), and by dividing the number of segments in the target word attempt from the first erroneous segment onwards by the number of segments in the complete realisation of the target word, counting again from the first erroneous segment (Proportional error segments).

2.4 Statistical modelling

Our general method in modelling the articulation rate of the repair was to construct a series of linear mixed effects regression models with and without individual candidate predictors from the set described above, and use likelihood ratio tests to assess whether the inclusion of the relevant predictor contributed significantly to the model fit (see Baayen 2008). We also constructed a model on the basis of conditional inference regression trees, for reasons we will go into in Section 3.3. In Section 3.2, we will report on an analysis of offset-to-repair duration along similar lines.

In addition to the repair timing variables described above, which can be taken as our crucial candidate predictors, we included a number of other variables which might have some effect on the articulation rate of the repair. First, the articulation rate of the reparandum is expected to have a strong effect: this is a crucial control variable to include in the model. Second, we included the speaker's identity (Speaker), the subcorpus from which each instance was sampled (Subcorpus), and the language variety spoken (Netherlands versus Flemish Dutch, Variety) as random factors, and the speaker's gender (Gender) as a related control variable.[8] Third, we included two measures of the frequency of the target word: its word form frequency (Word frequency) and its lemma frequency (Lemma frequency) as represented in the CELEX lexical database (Baayen et al. 1995). Kapatsinski (2010) has shown that in lexical repair, more frequent words are less likely to be interrupted prior to repair than less frequent ones; at the same time, it seems reasonable to expect a higher-frequency word to allow for a faster repair than a low-frequency word.

Prior to statistical modelling, we transformed temporal measurement values in order to make their distributions as close to normal as possible. We took the square roots of all articulation rate values, and log-transformed raw duration values. We also log-transformed frequency values.

3 Results

Before turning to repair tempo in Section 3.3, we will first describe the general temporal make-up of the repairs in our dataset up to the repair component. In particular, we can compare our instances with those of Nooteboom (2005a, 2010) in terms of two temporal characteristics discussed above. First, Nooteboom (2005a) reports a bimodal distribution of reparandum duration and length in segments in his collection of elicited repairs, and takes this as evidence for a qualitative difference between 'early' and 'late' repairs. Second, Nooteboom (2005a, 2010) reports a significant difference

[8] We can expect the influence of Speaker to be minimal: the instances in our dataset are produced by 248 speakers, 174 (70%) of whom contribute one repair only, 65 (26%) contribute two or three repairs, and 9 (4%) contribute four to six.

in offset-to-repair duration between ‘early’ and ‘late’ repairs. We address these points in turn in Sections 3.1 and 3.2.

3.1 Reparandum duration

Figure 2 shows the distributions associated with durations, segment counts and proportional segment counts in the erroneous target word attempt, both from the start of the target word attempt and from the first erroneous segment. Looking first at durations and segment counts, we see that a majority of instances have Target-to-offset duration values between 200 and 400 ms, which roughly correspond to segment counts between two and five — although durations up to 1500 ms and 13 segments are attested. In the latter, speakers either produce an attempt at a very long target word, or complete a shorter target word and continue with following lexical material before initiating repair. Counting from the first erroneous segment, a majority of instances have Error-to-offset duration values between 100 and 300 ms, and segment counts of one to three: that is, an abandonment of the target word attempt immediately or very shortly after the erroneous segment. All of these distributions show evidence of positive skew, and Shapiro-Wilk tests on the (log-transformed) duration values confirm their non-normality (Target-to-offset duration: W=0.9896, p=0.0100; Error-to-offset duration: W=0.9840, p=0.0004). However, they show little evidence of bimodality. Hartigan’s dip tests on the duration values yields no significant evidence of multimodality, whether performed on the raw values (Target-to-offset duration: D=0.0174, p=0.6650; Error-to-offset duration: D=0.0134, p=0.9626) or on the log-transformed ones (Target-to-offset duration: D=0.0174, p=0.6650; Error-to-offset duration: D=0.0202, p=0.3928).

For the purpose of assessing whether evidence of bimodality emerges when target word length is controlled for, we normalised Proportional target segments and Proportional error segments by dividing each proportional segment count by the number of errors in the data set which could give rise to that proportion. This is to take account of the fact that some proportional segment counts require relatively long — and therefore relatively rare — target word lengths: for example, we could only arrive at a value in the 0.9–1.0 range for a target word that has at least 10 segments.[9] We divided proportional counts of 1 by the total number of errors in the data set, since words of any length could be produced entirely before a repair.

[9] We are very grateful to an anonymous reviewer for suggesting this method.

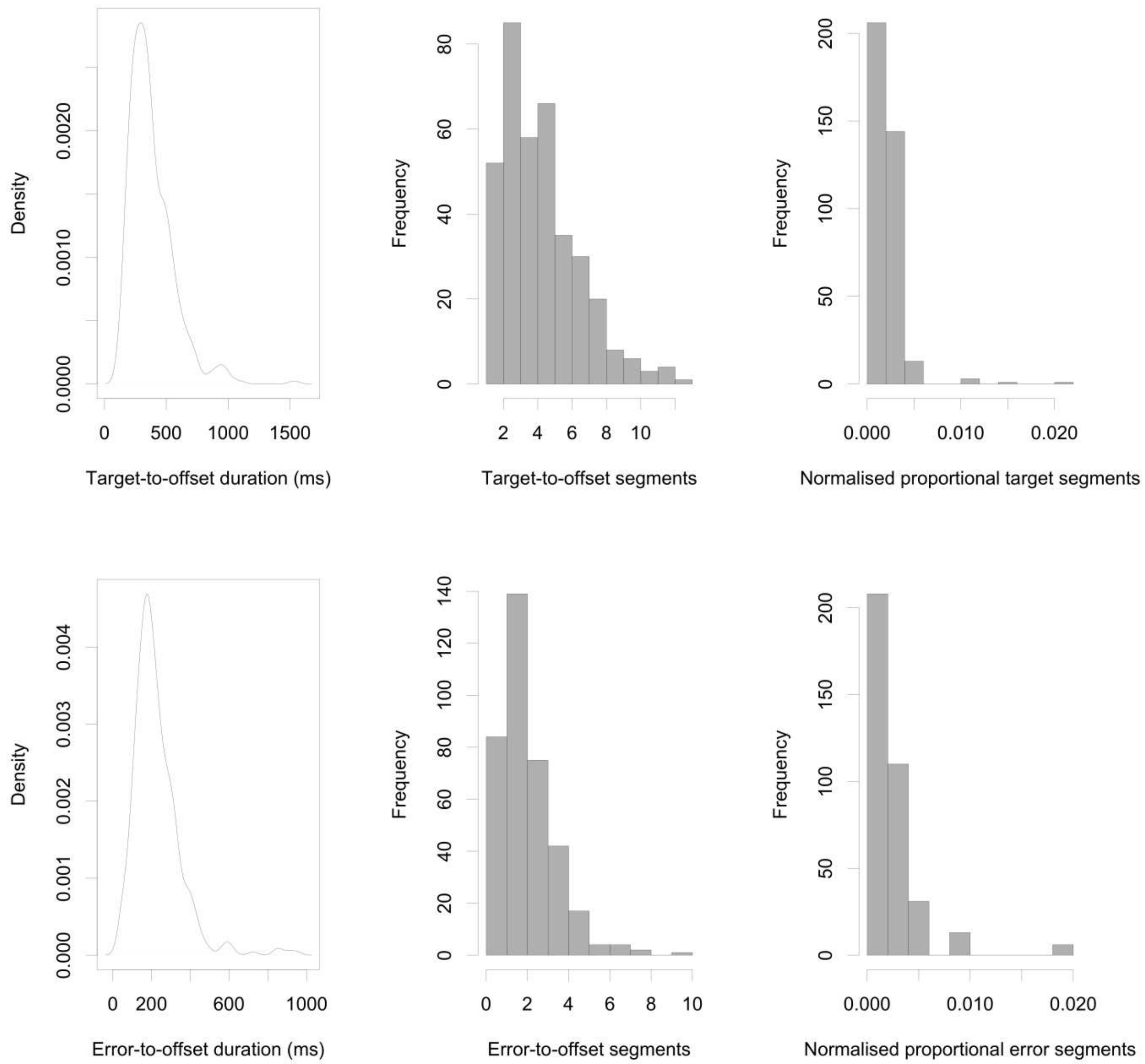


Figure 2. Density kernel plots for raw durations (leftmost), bar charts for raw and proportional segment counts (centre) and bar charts for proportional segment counts normalised for the distribution of target word lengths (rightmost) for the complete target word attempt (upper panels) and the error portion only (lower panels).

As can be seen in Figure 2, the distributions for Normalised proportional target segments and Normalised proportional error segments show no evidence of bimodality. This means we cannot be confident that our data are similar to Nooteboom's (2005a, 2010) with respect to the distribution of reparandum lengths, and we cannot claim our data are strongly suggestive of a qualitative difference between ‘early’ and ‘late’ repairs. Furthermore, it is worth noting that a great majority of our instances involve a repair initiation around 350 ms into the incorrect target word attempt, and around 200 ms following the onset of the first erroneous segment.[10] In the light of the estimated 150–250 ms detection-to-interruption latency discussed above, these values can be considered neither very low nor very high, and choosing a reparandum duration threshold to classify the repairs as resulting from inner or overt speech monitoring, as done by Liss (1998), would result in fairly arbitrary binning of a sizeable proportion of instances.

3.2 Offset-to-repair duration

With reference to offset-to-repair duration, we can ask two questions: first, whether the timing of the repair initiation allows us to predict the duration of the offset-to-repair interval, as found by Nooteboom (2005a, 2010); and second, which measure of timing discussed above has the greatest predictive value. In order to address these questions, we built linear mixed-effects regression models with the duration of the offset-to-repair interval as dependent variable.[11] As shown in Figure 3, the distribution of the variable is not normal (Shapiro-Wilk: $W=0.7307$, $p<0.0001$), as a large proportion of offset-to-repair intervals ($N=56$, or 16%) is zero. Log-transforming the raw values does not result in normality ($W=0.7668$, $p<0.0001$); however, excluding the zero durations and log-transforming the rest does ($W=0.9944$, $p=0.3548$). Therefore, we built two models: one to predict whether the raw interval duration is zero or not, and one to predict the log-transformed interval duration excluding raw zero values.[12] In both cases, we started with a base model containing only the random factor Speaker,[13] and assessed the performance of a subset of the candidate predictors listed in Section 2.2: Completeness, Proportional target segments, Proportional error segments, Target-to-offset duration and Error-to-offset duration.

[10] These values are comparable with those reported by Oomen & Postma (2001), who manipulate the time pressure under which subjects perform a verbal task: they report mean reparandum durations of 311 ms (normal condition) and 453 ms (fast condition). On the other hand, our error duration values are considerably lower than those reported by Oomen & Postma (2002), who add a distractor task to a verbal task: they report mean error-to-interruption durations of 419 ms (verbal task only) and 313 ms (verbal and distractor task). We do not have an explanation for the latter difference, although we can note that Oomen & Postma include various repair types in addition to phonological error repair.

[11] For the purpose of this analysis, we excluded a small subset of instances (5%) with a filled pause (containing several combinations of uh and of ‘or’), reducing the dataset to N=349.

[12] By contrast, Nooteboom (2005a) uses arbitrary binning of interval durations, with a lowest bin of 0–100, while Nooteboom (2010) presents results of t-tests only, with no indication as to the normality of the interval distribution.

[13] Exploratory modelling suggested that adding Subcorpus and Variety to Speaker in the base model does not improve its prediction and is of no consequence to the performance of our candidate predictors — even though Subcorpus and Variety subsequently came out of our random forest modelling as stronger predictors than Speaker, as shown in Figure 6 below.

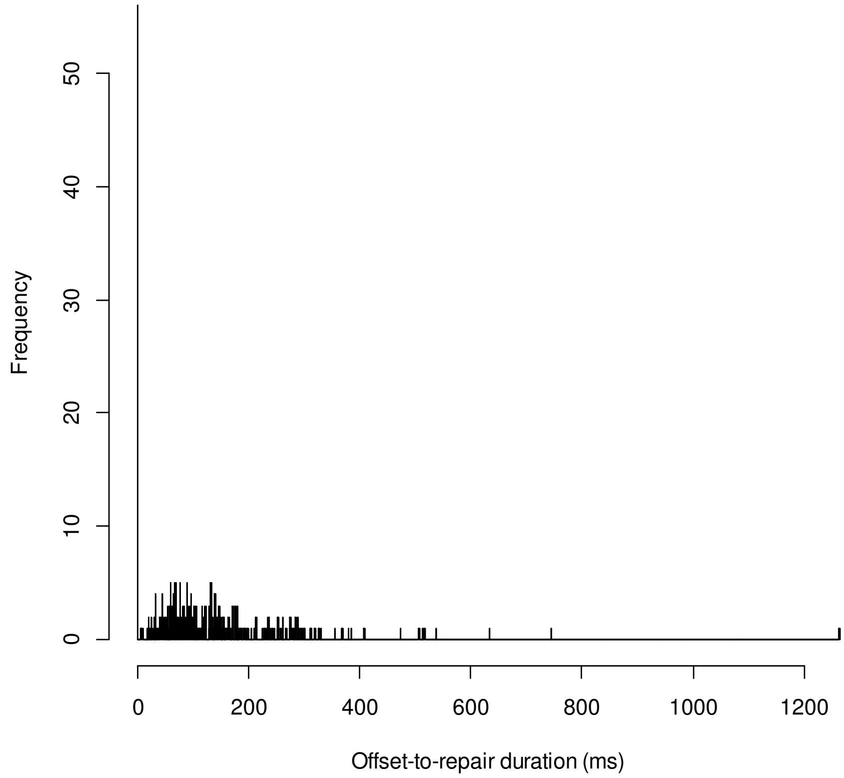


Figure 3. Histogram of raw offset-to-repair durations, excluding instances with filled pauses.

Table 1 lists the predictors whose addition to the base models results in a significant or near-significant improvement of fit — Table 1(a) for predicting whether the raw interval duration is zero or not, and Table 1(b) for predicting the log-transformed interval duration excluding raw zero values. We can see that the list is the same in both cases: Completeness, Proportional target segments and Target-to-offset duration. In both cases it appears that Completeness is the weakest predictor of the three; in fact, in the first model the addition of Completeness only results in a near-significant improvement of fit. For predicting whether the raw interval duration is zero or not, Proportional target segments outperforms Target-to-offset duration; for predicting the log-transformed interval duration excluding raw zero values it is the other way around. The effects are in the expected direction: for example, the lower the proportional segment count, the greater the likelihood of a zero offset-to-repair interval; and the greater the target-to-offset duration, the greater the offset-to-repair duration.

(a)	Predictor	Estimate	SE	z	C	p
	Completeness (incomplete)	−0.7509	0.4315	−1.740	0.9685	0.0539
	Proportional target segments	1.8344	0.7583	2.419	0.9734	0.0097
	Target-to-offset duration	0.8979	0.4594	1.955	0.9640	0.0349
(b)	Predictor	Estimate	SE	t	r^2	p
	Completeness (incomplete)	−0.1985	0.0879	−2.26	0.0172	0.0241
	Proportional target segments	0.3904	0.1643	2.38	0.0189	0.0180
	Target-to-offset duration	0.3314	0.0961	3.448	0.0390	0.0006

Table 1. Constructing linear mixed-effects regression models to predict (a) whether offset-to-repair duration is zero or not and (b) the log-transformed offset-to-repair durations excluding raw zero values: candidate predictors whose addition to a base model containing Speaker only results in a (near-)significant improvement of fit. Values for p refer to likelihood ratio tests of each two-factor model against the base model. Binary predictors are given with the value to which the estimate refers in parentheses.

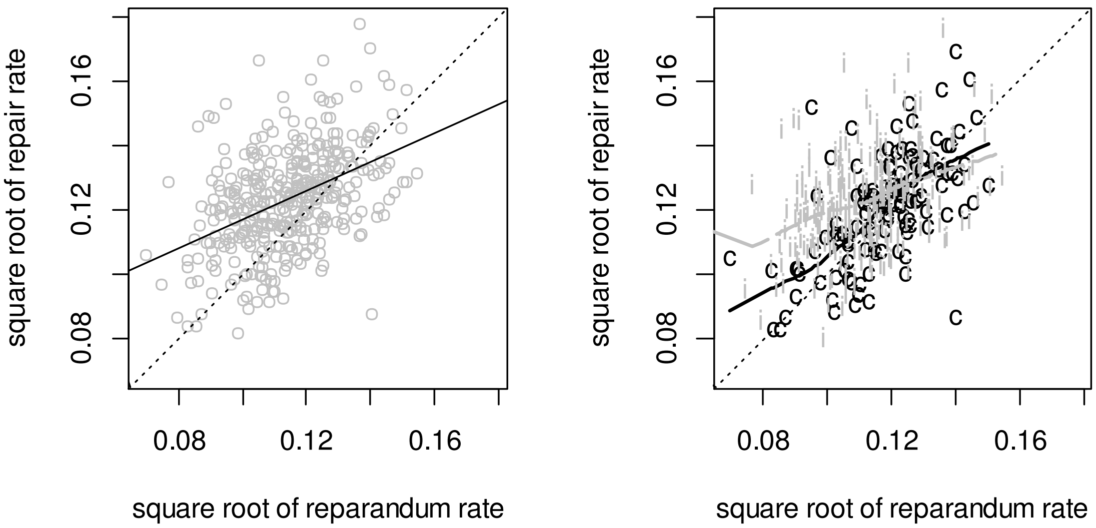
In sum, our dataset is comparable to that of Nooteboom (2010) in that repair timing has a significant effect on offset-to-repair duration: the earlier the repair, the shorter the offset-to-repair interval. The fact that this effect is captured by Target-to-offset duration and two measures of target word completeness, but not by Error-to-offset duration or Proportional error segments suggests that for understanding the general temporal make-up of our repairs up to the repair component, quantifying repair timing is best done with reference to the start of the target word attempt, rather than the start of the first erroneous segment. We now turn to the predictive value of these measures of timing for modelling repair tempo.

3.3 Repair tempo: Linear modelling

In modelling the articulation rate of the repair stretch (including any repeated lexical items),[14] we initially followed the same procedure as that described above for modelling offset-to-repair duration. In this case, we started with a base model containing the articulation rate of the reparandum

[14] Exploratory analysis not reported here showed that modelling repair articulation rate excluding repeated items leads to very similar results to those reported below; there are 56 instances of repairs (15% of the data set) which include repeated items in addition to the target word.

(including any repeated lexical items) and the random variable Speaker.[15] As might be expected, this reveals a significant correlation between reparandum and repair articulation rates (p_{mcmc}=0.0001), which accounts for approximately 20% of the variance in repair rate (r^2=0.1964). The correlation is illustrated in Figure 4(a). The figure further illustrates that consistent with Plug's (2011) results, we find that in most instances (72%), the articulation rate of the repair stretch is above that of the corresponding reparandum: in other words, temporal compression following the repair initiation is considerably more common than temporal expansion.



(a) (b)

Figure 4. Repair rate plotted against reparandum rate. The dotted line indicates values where the two rates are identical. In (a), the solid line indicates the prediction of a base model including only the additional factor Speaker. In (b), the data are split by the predictor Completeness: instances with a completed target word attempt are labelled "c"; those with an interrupted target word attempt "i". The black line is a locally-weighted polynomial smoother for the subset of data labelled "c"; the grey line the same for "i".

[15] Exploratory analysis not reported here showed that whether reparandum articulation rate is measured including or excluding lexical items that are repeated in the repair has no significant effect on the predictive value of the factor Reparandum rate in modelling repair tempo. This is consistent with the finding by Plug (2011) that in the context of self-repair, repetition is not consistently associated with temporal compression — in other words, whether a repair includes lexical repetition is itself not a significant predictor of the temporal organisation of the repair.

We then followed a stepwise algorithm to expand the model. We added each of the candidate predictors and control variables listed in Sections 2.2 and 2.3 to the base model to assess whether the addition resulted in a significant improvement of the model prediction. We expanded the model with the candidate predictor that resulted in the greatest improvement, and then assessed the impact of the remaining candidate predictors, and any interactions between them, on the expanded model. We continued expanding the model until no remaining candidate predictor yielded a significant improvement of model prediction. At the final stage, we trimmed 5 outliers (1.4% of the data set) to arrive at our best model, which is summarised in Table 2.[16] The model has an r^2 of 0.3558; that is, adding the extra predictors to our original base model allows it to explain an additional 16% of the variance in the articulation rate of the repair stretch.[17]

The model in Table 2 confirms that repair timing has a significant effect on the articulation rate of the repair. The measure with most predictive power is Proportional target segments. Its model estimates suggest that the higher the proportional completeness, the lower the articulation rate of the repair. In binary terms, the mean repair rate is higher following interrupted target word attempts than it is following completed ones, as illustrated in Figure 4(b). However, Figure 4(b) also shows that two separate states of affairs exist dependent on how fast the reparandum is: for slower reparanda, interrupted target word attempts are followed by a greater speeding up in the repair relative to completed ones than for faster reparanda. In other words, the overall effect of Proportional target segments is mainly due to an effect at relatively low local speaking rates, while the effect is attenuated at higher rates. The inclusion in our final model of an interaction between Reparandum rate and Proportional target segments reflects this observation. The model in Table 2 also includes Lemma frequency and Target-to-offset segments. The effect of Lemma frequency is in the expected direction: the higher the frequency, the higher the articulation rate of the repair. We return to the inclusion of Target-to-offset segments below.

[16] We present our best model with random intercepts. Additional analysis, not reported in detail here, shows that including random slopes does little to improve the model. Including Proportional target segments as a random slope by Speaker increases the value of r^2, but results in the same effect structure. No other inclusions of random slopes increases the goodness of fit or alters the effect structure.

[17] Variance due to Speaker is considerably below 0.0001. Note that the effect of Reparandum rate is not significant in the final model. This may be due to the interaction between Reparandum rate and Proportional target segments capturing most of the correlation between reparandum and repair rates. Likelihood ratio tests show that the model still has a higher r^2 with Reparandum rate included. Lemma frequency and Reparandum rate are not significantly correlated.

Factor	Estimate	SE	t	p_{mcmc}
Intercept	0.1052	0.0125	8.386	<0.0001
Reparandum rate	0.1179	0.1138	1.036	0.3097
Proportional target segments	−0.0836	0.0181	−4.629	<0.0001
Lemma frequency	0.0012	0.0002	5.477	<0.0001
Target-to-offset segments	0.0103	0.0021	5.025	<0.0001
Reparandum rate * Proportional target segments	−0.4896	0.1566	3.127	0.0022

Table 2. Summary of fixed effects in a linear mixed-effects model predicting the articulation rate of the repair. Binary variables are given with the value to which the estimate refers in parentheses. Values for p are estimated on the basis of 100,000 MCMC samples.

3.4 Repair tempo: Additional modelling

One problem in interpreting the linear model summarised in Table 2 is that it contains a certain degree of collinearity between variables.[18] In particular, Proportional target segments shows significant correlations with Target-to-offset segments ($r=0.5321$, $p<0.0001$) as well as Lemma frequency ($r=0.2918$, $p<0.0001$).[19] In order to assess the robustness of the model, we turned to an analysis based on conditional inference regression trees. As pointed out by Strobl et al. (2009) and Tagliamonte and Baayen (2012), this type of analysis, particularly when extended in random forest modelling, allows for a direct comparison between multiple, possibly correlated variables without the robustness of the final model being weakened by collinearity. Figure 5 plots the output of a conditional inference regression tree algorithm using all of the candidate predictors discussed above. The tree algorithm establishes which subdivisions in the data provide the most homogeneous groupings of observations with respect to the response variable — in this case, the articulation rate of the repair stretch.

[18] Together, the predictor variables in this model have a condition number, κ, of 23, indicating that collinearity may be approaching a harmful level.

[19] A t-test with Completeness and Lemma frequency confirms that the more frequent the reparandum item, the more likely it is to be completed prior to repair (Welch's $t(225)=4.0272$, $p<0.0001$). This is consistent with Kapatsinski's (2010) findings on lexical replacement repair in English.

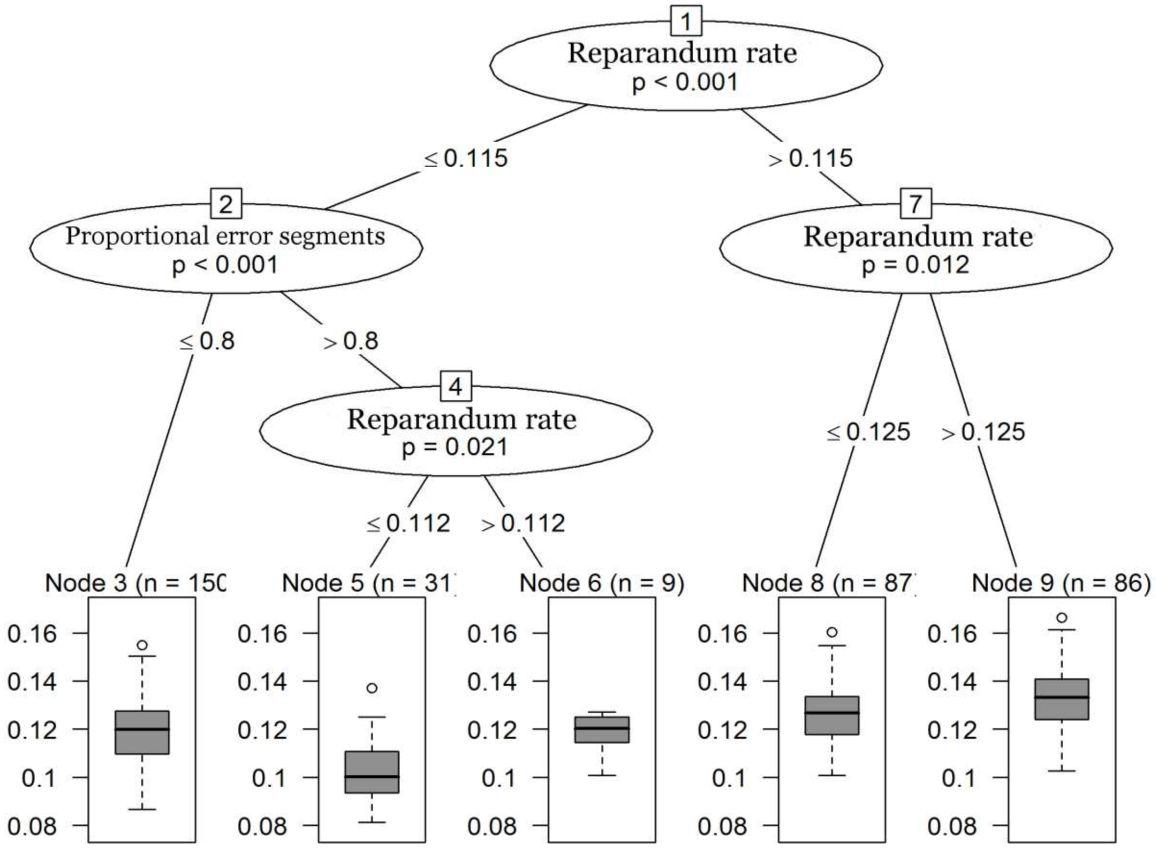


Figure 5. Conditional inference regression tree predicting Repair rate built on all candidate predictors and the data set excluding five outliers.

The regression tree analysis is consistent with our findings so far: it shows that the most important predictors of the articulation rate of the repair are the rate of articulation of the reparandum, a measure of the completeness of the erroneous target word attempt, and an interaction between the two. The measure of completeness it selects as having the greatest predictive value is Proportional error segments rather than Proportional target segments. The algorithm first splits the data into two subsets based on Reparandum rate: a relatively slow subset (left) and a relatively fast one (right). In the slow subset, a further split is made based on Proportional error segments, with values above 0.8 associated with higher repair articulation rates than values at or below 0.8. All other splits are based on the control variable Reparandum rate and point simply to the fact that the articulation rate of the repair increases as the articulation rate of the reparandum does. No further subdivisions result in an improvement of homogeneity. Lemma frequency and Target-to-offset segments do not feature in the analysis.

Finally, we constructed a random forest model to predict the repair rate on the basis of all our candidate predictors. Random forests (Breiman 2001) construct multiple conditional inference regression trees using subsets of the data and subsets of the predictor variables in order to provide test and training sets; the resulting predictions can be tested against the observed data and the relative importance of variables calculated. Our model grew 500 trees, each with five randomly-sampled input variables (approximately the square root of the total number of variables), and fits the observed data with $r^2=0.5691$. To calculate the relative importance of variables, we used a conditional variable importance measure (Strobl et al. 2008); this version of variable importance avoids a bias towards correlated variables. Our main interest was in assessing the robustness of our linear model in Table 2 and the regression tree in Figure 5.

The results are plotted in Figure 6, where bars which extend further to the right are indicative of the greater importance of the variable in question. Strobl et al. (2008) point out that the random permutations inherent in the algorithm mean that some unimportant variables will end up with a negative value on the importance scale. They suggest that variables with a positive value that does not exceed the absolute equivalent of the most extreme negative value can be disregarded in subsequent analysis. The relevant range of 'close-to-zero' measures is delimited in Figure 6 by means of vertical dashed lines. Figure 6 confirms that Reparandum rate is the most important predictor of the articulation rate of the repair, followed by a proportional measure of target word completeness. Like the regression tree analysis, the random forest model points towards Proportional error segments rather than Proportional target segments. Looking further down the list, we see a series of variables whose importance is rather similar, and substantially below that of Proportional error segments. Lemma frequency and Proportional target segments are the most important of these, followed closely by a set of segment counts which includes Target-to-offset segments.

In sum, the outcome of the regression tree and random forest analyses suggest we can have confidence in the robustness of our linear mixed-effects model, despite the possible collinearity between variables: all reveal an effect of repair timing on repair tempo, and all suggest that the best measure of timing is a proportional measure of target word completeness, with the possible addition of a measure based on simple segment counts. The linear and random forest models further suggest that lemma frequency has predictive value, such that higher-frequency words are repaired at a higher tempo than lower-frequency words.

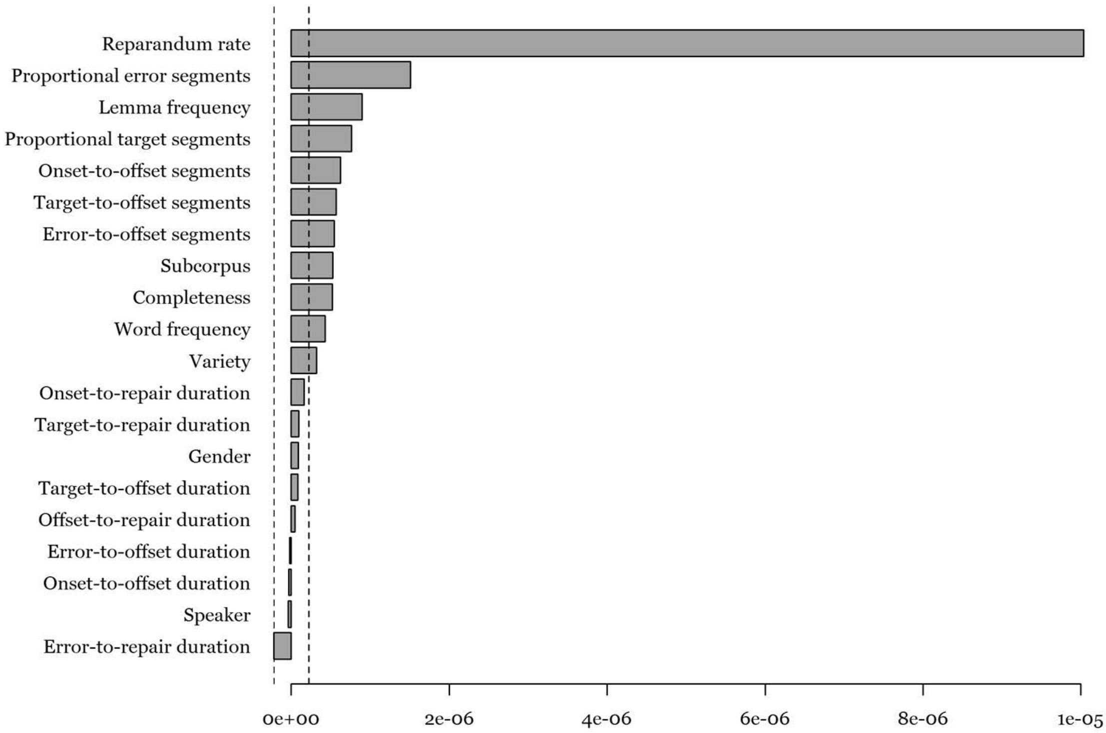


Figure 6. Barplot representing the conditional variable importance of all variables, according to a random forests model. Vertical dashed lines indicate the level of the most extreme negative importance and the positive importance with the same absolute value.

4 Discussion

Our study of the temporal characteristics of phonological error repair had two main aims: first, to assess whether the predominance of temporal compression in self-repair reported by Plug (2011) is observed in a larger, more homogeneous data set; and second, to assess whether repair timing has any effect on repair tempo, as suggested by Nooteboom's (2010) findings on a collection of elicited error repairs. In order to address the second question fully, we have also investigated the distribution of reparandum durations and the effect of repair timing on offset-to-repair duration. Moreover, we have asked what is the most relevant measure of repair timing, with choices between a binary classification of 'early' vs 'late', like the one implemented by Nooteboom, and a continuous measure; and between measures from the start of the erroneous target word attempt, and from the start of the first erroneous segment. We address the first two aims and related questions in Sections 4.1 and 4.2, and discuss the issue of how best to quantify repair timing in Section 4.3.

4.1 The extent of temporal compression

Our findings are consistent with those of Plug (2011): in our data set too, the articulation rate of the repair stretch is above that of the corresponding reparandum. In other words, temporal compression following the repair initiation is considerably more common than temporal expansion, and most repairs cannot be described as hyper-articulated on the grounds of their temporal organisation. In the case of phonological error repair, this may be due to the fact that this type of repair involves two attempts at the same lexical item. This means that the repair stretch has a relatively high level of informational redundancy, which would allow for some degree of articulatory reduction (see Aylett & Turk 2004, 2006, Pluymaekers et al. 2005). Still, it seems intuitively plausible that following a mispronunciation, at least the segments replacing erroneous ones will be produced with hyper-articulation in the repair stretch. Our analysis does not provide direct evidence to support or refute this intuition — but it does suggest that even if hyper-articulation can be observed on individual segments, it is either observed in a small minority of instances only, or its effect is too weak to counteract the general temporal compression across the repair stretch in the majority.

4.2 The influence of repair timing

Our findings are consistent in several respects with those of Nooteboom (2005a, 2010). First, like Nooteboom (2005a, 2010) we observe a significant effect of repair timing on offset-to-repair durations, such that early repairs are associated with shorter offset-to-repair intervals than late ones. Second, as predicted on the basis of Nooteboom's (2010) interpretation of observed prosodic differences between early and late repairs, we find a significant effect of repair timing on the articulation rate of the repair stretch, such that early repairs are on average produced with a higher repair tempo than late ones. Together, these findings provide support for Nooteboom's contention that a systematic relationship exists between the timing of the abandonment of an erroneous utterance for the purpose of self-repair and the phonetics of the immediately following stretch of speech — in this case, both the time between the abandonment and the start of the repair, and the tempo of the repair are to some extent correlated with repair timing.

Before we consider Nooteboom's account for the phonetic differentiation between early and late repairs in more detail, we should emphasise that the effect is observed in a subset of instances only: both our linear and tree-based models reveal a significant interaction between the articulation rate of the reparandum and repair timing, such that the higher the former, the less significant the effect of the latter. We cannot at present offer a convincing explanation for this. Notwithstanding this complication, it is worth highlighting that the observed effect of repair timing not only confirms Nooteboom's (2010) contention that repair timing and the phonetics of the immediately following stretch of speech are systematically related, but also bears out the most plausible prediction we can make regarding repair tempo on the basis of Nooteboom's theoretical account of this relationship. As pointed out in Section 1, Nooteboom (2010) argues that the different prosodic characteristics of early versus late repairs can be explained with reference to the distinct functions of inner and overt speech monitoring processes — to prevent errors from becoming public for the former, and to address errors that have already become public for the latter. The function of preventing errors from becoming public can plausibly be associated with execution under time pressure, which does not seem as plausible an attribute of addressing errors that are already public. This offers a straightforward explanation for Nooteboom's (2010) observation that early repairs have significantly lower offset-to-repair durations than late ones (see also Seyfeddinipur et al. 2008). As shown in Section 3.2, this is also the case in our data. Our additional finding that early repairs are on average produced with a higher repair tempo than late ones is consistent with the idea that early repairs are not only initiated, but also completed under a greater degree of time pressure than late ones.

Nooteboom's account rests on the hypothesis that early repairs are most likely due to an error detection by inner speech monitoring processes, and late repairs due to error detection in overt speech (see also Postma 2000). This makes sense if we assume that earlier repairs necessarily follow earlier error detection than later repairs. This assumption forms the basis of the Main Interruption Rule, or MIR (Nooteboom 1980, Levelt 1983, 1989, Brédart 1991), which dictates that speakers '[s]top the flow of speech immediately upon detecting trouble' (Levelt 1989: 478). However, the Main Interruption Rule has been challenged on the basis of experimental findings (Hartsuiker et al. 2008, Tydgat et al. 2011) as well as corpus data (Hartsuiker & Kolk 2001, Seyfeddinipur et al. 2008), which suggest that speakers can decide whether to initiate repair immediately after detection or to postpone initiation, and this decision is based partly on whether they prioritise local accuracy or fluency: interrupting an erroneous word maximises accuracy at the expense of fluency, while postponing repair initiation until the next word boundary maintains fluency at the expense of accuracy. Furthermore, if Kapatsinski's (2010) finding of an effect of frequency on reparandum item completeness, confirmed in this study, can be attributed to a high degree of automatization of the production of frequent words, we might wonder whether frequency constrains the speed of repair initiation following detection, irrespective of fluency considerations.[20]

[20] We are grateful to an anonymous reviewer for this suggestion.

Our findings do not provide direct evidence for or against Nooteboom's (2010) theoretical account or the Main Interruption Rule,[21] but the absence of evidence in our data of a bimodal distribution of reparandum lengths is relevant. Nooteboom (2005a, 2010) takes evidence of a bimodal distribution of reparandum lengths, which has also been observed in conversation-analytic studies of repair (Schegloff 1979, Fox et al. 2009), as support for a qualitative difference between early and late repairs. It is certainly difficult to see why, for example, 10% or 90% complete target word attempts should be attested more frequently than 50% complete target word attempts if repair is initiated immediately upon error detection by a single monitoring process. Nooteboom's suggestion that two distinct monitoring processes are involved is strengthened by the common understanding that inner speech monitoring is faster in detecting errors than overt speech monitoring (Postma 2000). This offers a straightforward explanation of bimodal distributions in reparandum length: as the slower second monitor takes over, there is a temporary lag in the detection of errors not picked up by the faster first monitor. Our analysis provides no empirical support for this line of reasoning.

4.3 How to quantify repair timing

With respect to the question of how best to quantify repair timing, we have seen that proportional measures of target word completeness are most informative in modelling the temporal organisation of our repairs. In modelling offset-to-repair duration, Proportional target segments performs similarly to a binary classification (Completeness) and a measurement of target word duration (Target-to-offset duration); in modelling repair tempo, Proportional target segments outperforms all other measures.

The fact that proportional measures of target word completeness outperform a binary classification confirms our hypothesis that not all interrupted reparanda constitute an equally early repair initiation: repairs initiated after 10% of the target word has been completed are not directly comparable to repairs initiated after 60% has been completed. The fact that proportional measures of target word completeness generally outperform measures of target word duration confounds our

[21] Seyfeddinipur et al. (2008: 838–839) use the observation of repairs after interrupted target word attempts having lower offset-to-repair durations than repairs after completed attempts, which our study has confirmed, as evidence against the Main Interruption Rule. Their reasoning rests on an exception to the MIR, formulated by Levelt (1989: 481), which states that speakers can postpone repair initiation until the end of a word if the word that is produced when trouble is detected is not itself erroneous: for example, in the case of lexical 'appropriateness' repair. If this is the case, one might predict that repairs initiated following the complete production of a non-erroneous word have shorter offset-to-repair durations than repairs initiated following an interrupted word — as the completion of the word allows the speaker some extra time for planning the repair. Seyfeddinipur et al. (2008) show that this prediction is not borne out. As our data set only contains error repairs, our findings do not provide clear support for Seyfeddinipur et al.'s (2008) argument: in all instances, the assumption based on the MIR must be that repair is initiated immediately upon error detection.

hypothesis that repairs with a reparandum duration above 150–250 ms may have different temporal characteristics than those with a duration below 150–250 ms, irrespective of whether they involve completed or interrupted target word attempts. This hypothesis is based on the argument that repairs with a reparandum duration below 150–250 ms are almost certainly motivated by an error detection in inner speech monitoring, while those with a higher reparandum duration are most likely motivated by a detection in overt speech monitoring. Our analysis confirms that cautions against postulating a reparandum duration threshold in classifying error repairs as resulting either from inner or overt speech monitoring (Oomen & Postma 2001, 2002, Hartsuiker et al. 2005b) are justified: we see no split in the data at a particular target word duration or segment count.[22]

We noted in Section 2.3 that unlike in Nooteboom's (2010) data, in ours the location of the error varies from target word to target word. We suggested that if the involvement versus non-involvement of the overt speech monitor in the detection of a given error has an impact on the prosodic implementation of its repair, the earliest occasion for possible overt speech monitor detection — that is, the point at which the first erroneous segment becomes overt — should be a highly relevant reference point for the implementation of a continuous measure of repair timing. Our statistical analyses are contradictory on the relative performance of repair timing variables that refer to the first erroneous segment and variables that refer to the start of the erroneous target word attempt: linear regression modelling suggests that the latter perform best, while modelling using conditional inference regression trees and random forests highlights the significance of the former.

In sum, while there is clearly something to be gained from implementing a measure of proportional target word completeness in investigating the phonetics of phonological error repair, a measure that refers to the start of the target word is likely to capture similar effects to one that refers to the precise error location. Our results can be taken as highlighting the importance of the word as a unit in speech production and self-monitoring. We do not find strong evidence to support a model of self-monitoring in which the difference between early and late repairs of overt errors is in the duration that elapses, or the number of segments that are produced, after the first erroneous segment has been articulated. Instead, what appears to matter is how much of the target word is articulated before the repair is initiated.

[22] In constructing our linear regression model, we tested for the significance of a range of thresholds between 150 and 500 ms by including corresponding binary variables (above vs below 150, above vs below 200 and so on) among our candidate predictors. None of these variables resulted in a significant improvement of the fit of our base model. Note that conditional inference regression modelling is particularly appropriate for identifying thresholds along a continuous parameter: the threshold value or range will give rise to a partition in the tree. As seen in Section 3.4, our duration variables do not give rise to significant splits in the data.

5 Conclusion

In this paper we have reported on a study of the temporal characteristics of phonological error repair in spontaneous Dutch speech. The study has confirmed Plug's (2011) finding of a predominance of temporal compression — that is, a local increase in articulation rate — following the repair initiation. The study has also confirmed Nooteboom's (2010) finding that the timing of repair initiation covaries with aspects of the prosody of the repair that follows, and confirmed that target word frequency is a relevant factor, in line with results reported by Kapatsinski (2010). With reference to repair timing, the study has revealed that proportional measures of target word completeness provide the most informative quantifications in modelling the overall temporal organisation of the repairs.

Acknowledgements

This work was supported by ESRC grant RES-061-25-0417 'Prosodic marking revisited: The phonetics of self-initiated self-repair in Dutch'. We thank Christina Englert for her contribution to the research reported here, and the associate editor of Journal of Phonetics, one anonymous reviewer and Sieb Nooteboom for helpful comments on the first submission of this paper.

References

Aylett, M. & A.Turk 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. Language and Speech 47: 31–56.

Aylett, M. & A.Turk 2006. Language Redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. Journal of the Acoustical Society of America 119: 3048–3058.

Baars, B.J., M.T. Motley & D.G. MacKay 1975. Output editing for lexical status in artificially elicited slips of the tongue. Journal of Verbal Learning and Verbal Behaviour 14: 382–391.

Baayen, R.H. 2008. Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press.

Baayen, R.H., R. Piepenbrock & L. Gulikers 1995. The CELEX lexical database. Release 2 [CD-ROM]. Philadelphia: Linguistics Data Consortium (Distributor).

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

Boersma, P. & D. Weenink 2010. Praat: doing phonetics by computer (version 5.1.34). http://www.praat.org/.

Brédart, S. 1991. Word interruption in self-repairing. Journal of Psycholinguistic Research 20: 123–137.

Breiman, L. 2001. Random forests. Machine Learning 45: 5–32.

Fox, B., F. Wouk, M. Hayashi, S. Fincke, L. Tao, M.-L. Sorjonen, M. Laakso & W. F. Hernandez 2009. A cross-linguistic investigation of the site of initiation in same-turn self-repair. In Conversation analysis: Comparative perspectives, ed. J. Sidnell. Cambridge: Cambridge University Press. 60–103.

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

Fromkin, V.A. 1973. The non-anomalous nature of anomalous utterances. In Speech errors as linguistic evidence, ed. V.A. Fromkin. The Hague: Mouton. 215–242.

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

Hartsuiker, R.J. & H.H.J. Kolk 2001. Error monitoring in speech production: A computational test of the Perceptual Loop Theory. Cognitive Psychology 42: 113–157.

Hartsuiker, R.J., R. Bastiaanse, A. Postma & F. Wijnen 2005a. Phonological encoding and monitoring in normal and pathological speech. In Phonological encoding and monitoring in normal and pathological speech, ed. R.J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen. Hove: Psychology Press. 1–14.

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

Hartsuiker, R.J., C.M. Catchpole, N.H. De Jong & M.J. Pickering 2008. Concurrent processing of words and their replacements during speech. Cognition 108: 601–607.

Howell, P. & K. Young 1991. The use of prosody in highlighting alterations in repairs from unrestricted speech. The Quarterly Journal of Experimental Psychology 43A: 733–758.

Jasperson, R. 2002. Some linguistic aspects of closure cut-off. In The language of turn and sequence, ed. C.E. Ford, B.A. Fox & S.A. Thompson. Oxford: Oxford University Press. 257–286.

Kapatsinski, V. 2010. Frequency of use leads to automaticity of production: Evidence from repair in conversation. Language and Speech 53: 71–105.

Levelt, W.J.M. & A. Cutler 1983. Prosodic marking in speech repair. Journal of Semantics 2: 205–217.

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

Levelt, W.J.M. 1984. Spontaneous self-repairs in speech: Processes and representations. In Proceedings of the Tenth International Conference of Phonetic Sciences, ed. M.P.R. Van den Broecke & Cohen. Dordrecht: Foris. 105–117.

Levelt, W.J.M. 1989. Speaking: From intention to articulation. Cambridge, Mass.: The MIT Press.

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

Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In Speech production and speech modeling, ed. W.J. Hardcastle & A. Marchal. Dordrecht: Kluwer. 403–439.

Lindblom, B. 1996. Role of articulation in speech perception: Clues from production. Journal of the Acoustical Society of America 99: 1683–1692.

Lindblom, B. 2000. Developmental origins of adult phonology: The interplay between phonetic emergents and evolutionary adaptations. Phonetica 57: 297–314.

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

Marslen-Wilson, W. & L. Tyler 1981. Central processes in speech understanding. Philosophical Transactions of the Royal Society London B259: 317–332.

Meyer, A.S. 1992. Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. Cognition 42: 181–211.

Niebuhr, O. 2010. On the phonetics of intensifying emphasis in German. Phonetica 67: 170–198.

Nooteboom, S.G. 1980. Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In Errors in linguistic performance, ed. V. Fromkin. New York: Academic Press. 87–95.

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

Nooteboom, S.G. 2005b. Listening to oneself: Monitoring speech production. In Phonological encoding and monitoring in normal and pathological speech, ed. R.J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen. Hove: Psychology Press. 167–186.

Nooteboom, S.G. 2010. Monitoring for speech errors has different functions in inner and overt speech. In The linguistic enterprise: From knowledge of language to knowledge in linguistics, ed. M.B.H. Everaert, T. Lentz, H. De Mulder, Ø. Nilsen & A. Zondervan. Amsterdam: John Benjamins. 213–233.

Oomen, C.C.E. & A. Postma 2001. Effects of time pressure on mechanisms of speech production and self-monitoring. Journal of Psycholinguistic Research 30: 163–184.

Oomen, C.C.E. & A. Postma 2002. Limitations in processing resources and speech monitoring. Language and Cognitive Processes 17: 163–184.

Oostdijk, N. 2002. The design of the Spoken Dutch Corpus. In New frontiers of corpus research, ed. P. Peters, P. Collins & A. Smith. Amsterdam: Rodopi. 105–113.

Plug, L. 2011. Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. Journal of Phonetics 39: 289–297.

Pluymaekers, M., M. Ernestus & R.H. Baayen 2005. Articulatory planning is continuous and sensitive to informational redundancy. Phonetica 62: 146–159.

Postma, A. 2000. Detection of errors during speech production: A review of speech monitoring models. Cognition 77: 97–131.

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

Rietveld, A.C.M. & V.J. van Heuven 1997. Algemene fonetiek. Bussum: Coutinho.

Schegloff, E.A. 1979. The relevance of repair to syntax-for-conversation. Syntax and Semantics 12: 261–286.

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

Shattuck-Hufnagel, S. & A. Cutler 1999. The prosody of speech error corrections revisited. In Proceedings of the Fourteenth International Congress of Phonetic Sciences (ICPhS 1999). Vol. 2. San Francisco. 1483–1486.

Shriberg, E. 2001. To 'errr' is human: Ecology and acoustics of speech disfluencies. Journal of the International Phonetic Association 31: 153–169.

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

Smiljanić, R. & A.R. Bradlow 2009. Speaking and hearing clearly: Talker and listener factors in speaking style changes. Language and Linguistics Compass 3: 236–264.

Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin & A. Zeileis 2008. Conditional variable importance for random forests. BMC Bioinformatics 9: 307.

Strobl, C., J. Malley & G. Tutz 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods 14: 323–348.

Tagliamonte, S. & R.H. Baayen 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. Language Variation and Change 24: 135–178.

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