TESTING A THEORY OF REPAIRING SEGMENTAL SPEECH ERRORS

Sieb Nooteboom & Hugo Quené

Utrecht institute of Linguistics OTS
Utrecht University, the Netherlands

ABSTRACT

We propose that in self-monitoring for segmental speech errors "repairs" stem from word forms that compete during speech preparation and self-monitoring with the selected form. Activation of these potential repairs decreases during the time lag between detection in internal and overt speech. Earlier it was demonstrated that repaired speech errors can be classified as detected in internal or in overt speech. A re-analysis of data obtained in two SLIP experiments shows that: (1) Error-to-interruption times are longer after non-elicited and multiple errors than after single elicited errors. (2) Single elicited errors are relatively more often detected in internal speech than other errors are. (3) The correct word form is the most frequent form used as repair, but less frequently after external than after internal detection. (4) Interruption-to-repair times are shorter for single elicited than for other errors, but less so after external detection. These findings support our theory of repairing.

Keywords: Speech errors, self-monitoring, internal speech, overt speech, repairs.

1. INTRODUCTION

This paper deals with self-repairing of segmental speech errors. The main question we try to answer is: "Where do, in self-monitoring, self-repairs of segmental speech errors come from?" We recently proposed the outline of a new theory of self-monitoring and repairing speech errors [14].We believe, with Nozari and colleagues [15, 16], that during speech preparation and self-monitoring, speech error detection is triggered by conflict between candidate word forms competing for the same slot in the utterance being prepared. We also are confident that, as claimed by Levelt and his associates [9, 10], speech errors are detected both before and after speech initiation, i.e. both in internal speech and in overt speech. Error detection triggers interruption of the speaking process and the start of a repair process. During this repair process, candidate repairs do not necessarily stem from re-compilation as was proposed in [9] and [10], but repairs can be highly activated competitor word forms. The assumption that error word forms and candidate repairs can be simultaneously active, competing for the same slot in the utterance being prepared, would explain that often segmental speech errors are articulatory blends between two competing segments [4, 6, 7, 11, 12, 20]. If the error form has the highest activation, a command to start articulation is issued when phonological encoding of this form is completed. We also assume that during speech preparation, activation of the correct word form is sustained from the lexical level, activation of error forms is not.

Levelt and his associates [9, 10] argued that speech errors can be detected at two stages, both before and after speaking the error form is initiated, i.e. both in internal and in external or overt speech. In principle, errors detected in internal speech could be repaired before speech initiation. If so, these would be unobservable. However, we have demonstrated [14] that such covert repairs, if they occur at all, are extremely rare: The distribution of error-to-interruption times is not clearly truncated, leaving little room for covert repairs. Most errors detected in internal speech surface as interrupted error forms as in "b..good beer", where the "b.." is an anticipation of the initial consonant of "beer". One may note that speech fragments in such interrupted error forms often are shorter than humanly possible reaction times. Such short speech fragments must reflect error detection in internal speech (cf. [1, 5, 8]). But, as we will see below, the empirical separation between internal and external detection on the basis of error-to-interruption times, is not determined by the shortest possible reaction times.

Until recently there was no way to distinguish empirically between repaired errors detected in internal speech and those detected in overt speech. This also led to the default assumption that detection and repair processes are the same in self-monitoring internal and overt speech. We have demonstrated [13] that there is a way to distinguish between repaired errors detected before and after speech initiation. As it happens, the distribution of error-to-interruption times is clearly bimodal, as one would predict from Levelt's assumption [9, 10] that there are two consecutive stages of self-monitoring. This bimodal distribution may be described with two underlying gaussian distributions with an intermediate separation threshold. If we assume that all error-to-interruption times shorter than the separation value belong to repaired errors detected in internal speech and all other error-to-interruption times belong to repaired errors detected in overt speech, then we can at least statistically separate between these two classes of repaired speech errors. Interestingly, the time lag between self-monitoring internal and overt speech appears to be in the order of 500 ms. This is considerably longer than previously supposed. (For example, the computational model proposed in [8] predicts that this time lag would be in the order of 200 ms). For all practical purposes we are now in a position to investigate differences between these two classes of repaired errors. This helps us in testing our theory of repairing speech errors.

We do this in two SLIP experiments. The main point of a SLIP experiment is that specific speech errors, here reversals between initial consonants of two CVC words, are elicited by the structure of precursor word pairs, priming segmental speech errors. This priming boosts the activation level of both the correct target forms and the elicited error forms, and thereby suppresses the activation of other competing error forms. Typically in such experiments we find correct responses, elicited error responses and non-elicited error responses. In testing our new theory of repairing speech errors we capitalize on the distinction between detection in internal and overt speech and on the distinction between elicited and other errors. Below we derive four predictions from the current theory of repairing segmental speech errors. These predictions are specific for SLIP experiments.

First, after single elicited errors there is mainly competition between a specific single error form and the correct target form. After other errors, potentially there is competition between more activated lexical items, providing more opportunities for repairing, but also taking more time to be resolved.

(1) Error-to-interruption times are longer after non-elicited and multiple errors (together "other" errors) than after single elicited errors.

Second, if detection takes more time for other errors than for elicited errors, there will be more cases in which the time available for internal detection has expired. In those cases, detection may take place in overt speech. This leads to our second prediction.

(2) Single elicited errors are relatively more often detected in internal speech than other errors.

Third, we assume that correct word forms are sustained from the lexical level during the process of self-monitoring, whereas other forms are not. But we also assume that during the delay between internal and external self-monitoring activation of the available candidates for repair decreases. Together these assumptions lead to the third prediction:

(3) The correct lexical form will be the most frequent form used as repair, but this effect will be weaker after external than after internal detection.

Finally, the theory assumes that after other errors there are more and less activated competing items than after elicited errors, and therefore selecting or re-activating a repair will take more time. However, we also assume that between detection in internal and and detection in overt speech, activation of potential repairs decreases. Therefore, more often a repair must be re-activated. This effect will be stronger for other errors than for elicited errors. Together these assumptions lead to our fourth prediction.

(4) Interruption-to-repair times are shorter for single elicited than for other errors. This effect will be stronger after internal than after external error detection.

Below these predictions will be tested.

2. METHOD

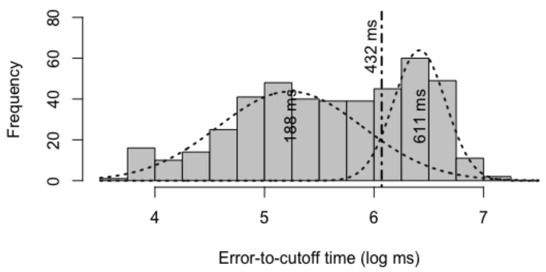
For testing our predictions, we employ data obtained in two SLIP experiments, described in [13]. The two experiments differed in the phonological contrasts between the interacting initial consonants. Differences in results between these two experiments are not relevant for our current purpose and will not be discussed in the current paper.

In both experiments anticipatory single segmental errors (exchanges and anticipations) were elicited by phonological priming of initial consonants in Dutch CVC CVC word pairs. Each stimulus word pair was preceded by five precursor word pairs the last three of which primed a reversal between the two initial consonants. For example, five precursor word pairs *bouw jool, lijf deed, koet pop, kuur poet, kas piet*, preceded the stimulus word pair *paf kiep*. The stimulus word pair was followed by a sequence of 6 question marks, serving as a cue to speak the last word pair seen. All word pairs and the speaking cue were presented on a screen during 900 ms followed by a blank during 100 ms. In each experiment there were two lists of test stimuli (32 in Experiment 1; 64 in Experiment 2) and filler stimuli (23 in Experiment 1; 46 in Experiment 2). The number of precursors for the fillers varied from 0 to 4. These precursors were not eliciting speech errors. Their sole function was to make the moment a word pair had to be pronounced unpredictable. In each experiment there were two parts: one part, employing one list of stimuli, with auditory feedback, and one part, employing the corresponding other list of stimuli, without auditory feedback. Unexpectedly, the effect of auditory feedback was found to be negligible [13]. Experiment 1 had as participants 106 native speakers of Dutch, Experiment 2 had as participants 124 native speakers of Dutch. For further details we refer to [13]. Whereas in [13] only correct and fluent and single elicited segmental error responses to the test stimuli were analysed, here we provide a first report and analysis also of other error responses. These are mainly non-elicited and multiple error responses. We will also compare single elicited error responses against these "other" error responses. Results of the two experiments will be collapsed below.

3. RESULTS

Because the contrast between internally and externally detected errors is important for testing our predictions, we begin with testing whether at least statistically we can separate between these two classes of repaired errors. We do this by using the same procedure as described in [13], but here applied to the entire set of relevant (elicited and other) error responses containing a repair. Log-transformed error-to-interruption times were modeled (without supervision) as a mixture of two gaussian distributions [2], [3], [18], visualized in figure 1.

Figure 1. Histogram of log-transformed durations of error-to-interruption intervals of repaired initial responses. Distributions plotted with dotted lines indicate the estimated distributions from an uninformed gaussian mixture model (see text). The vertical dashed line indicates the interpolated boundary value (6.07, corresponding to 432 ms) between the two distributions.



Based on this mixture model, we will from now on assume that repaired errors with error-to-interruption times shorter than 432 ms were detected "internally", and those with error-to-interruption times longer than 432 ms were detected "externally". Because the two estimated distributions overlap, some repaired errors are necessarily misclassified, causing some unavoidable statistical noise. Below we will test our four predictions.

Prediction 1 focuses on error-to-interruption times. These were analyzed by means of linear mixed-effects modeling (LMM), with subjects and stimuli as random intercepts.

(1) Error-to-interruption times are longer after other errors than after single elicited errors.

Log-transformed error-to-interruption times were indeed found to be considerably and significantly longer after "other" errors (6.146, or 467 ms) than after single elicited errors (5.239, or 188 ms), as confirmed by the LMM [β=+0.908, t=13.27, p<.0001; 95% CI (0.761, 1.047); masking noise yields no effect, β=-0.055, $|t|$<1, n.s.].

Prediction 2 focuses on the odds of elicited and other errors being detected internally or externally. The count frequencies were analyzed by means of loglinear (count-based) modeling with detection stage and type of error response as two predictors; due to the low frequencies per subject or per item, random effects had to be ignored here.

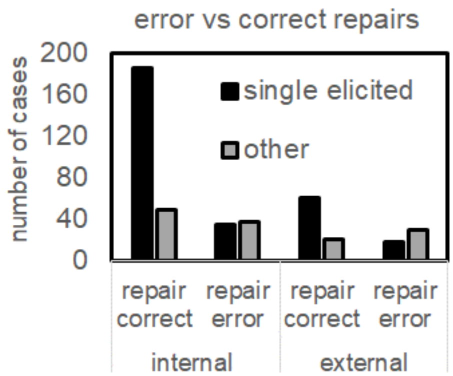
(2) Single elicited errors are relatively more often detected in internal speech than other errors.

For single elicited errors, the odds of being detected internally were 235:38. This is indeed far higher than for other errors, viz. 65:114. In other words, most of the internally detected errors were elicited ones (235:65), most of the externally detected errors were "other" error responses (38:114). The interaction was confirmed by the loglinear model (β=2.42, Z=10.2, p<.0001; adding masking noise did not improve the fit of the model, p=.2498).

Prediction 3 focuses on the odds of repairs of elicited and other errors being correct or incorrect.

(3) The correct lexical form will be the most frequent form used as repair, but this effect will be weaker after external than after internal detection.

Figure 2. Error forms and correct forms used as repairs separately for single elicited and other errors and for internal and external detection.



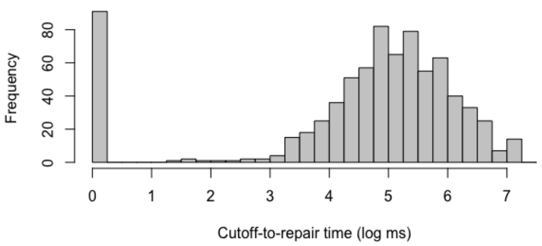
Again, these odds were analyzed by means of loglinear modeling, with the relevant factors as fixed predictors. Figure 2 shows that indeed correct forms are much more often used as repairs than error forms are (GLM, main effect of repair status, β=2.04, *Z*=8.19, *p*<.0001), and that this effect is weaker after external than after internal error detection (interaction β=-0.60, *Z*=-2.15, *p*=.0032). In addition, repairs are much more frequent after single elicited errors than after other errors (main effect of error status, β=0.95, *Z*=8.49, *p*<.0001), and this effect is also far weaker after external than after internal detection (interaction β=-2.25, *Z*=-10.68, *p*<.0001). Other effects were not significant.

Prediction 4 focuses on interruption-to-repair times:

(4) Interruption-to-repair times are shorter for single elicited than for other errors. This effect will be stronger after internal than after external error detection.

Figure 3 below is a histogram of interruption-to-repair times. The histogram includes the value 0 ms (converted to 1 ms before taking the logarithm). That this value stands apart in a single bin is an artefact of the logarithmic scale. But in the ms scale too, the value of 0 ms is overrepresented. The histogram of interruption-to-repair times (Figure 3) shows that the value of 0 ms (converted to 1 ms before log transformation) is overrepresented. These "zero" values correspond to *immediate* repairs.

Figure 3. Histogram of log interruption-to-repair times.



We might test our prediction (4) in two different ways. One is to test whether interruption-to-repair times are significantly longer for "other" errors than for single elicited errors. This effect was indeed found in an LMM including participants and stimuli as random factors [β=1.01, *t*=6.18, *p*<.0001]. However, the distributions of interruption-to-repair times deviate strongly from normal precisely because the overrepresentation of immediate repairs. Therefore, we also tested whether the incidence of immediate repairs is greater for single elicited than for other errors, using a GLMM [17]. With the detection stage included as a fixed predictor, we find that after internal detection the odds of immediate repairs are indeed higher after single elicited errors (35:198) than after other errors (3:52), although not significantly so (β=+1.04 logit units, *Z*=1.59, *p*=.112). After external detection the effect is much weaker (single elicited 2:36, other 4:109), but there are so few immediate repairs after external detection that the interaction is not significant (β=-0.89 logit units, *Z*=-0.79, *p*=.431; for the same reason the effect of masking noise could not be added to the GLMM).

4. DISCUSSION

We have tested four predictions derived from our proposal in [13] and [14] that in repairing segmental speech errors, repairs, both after internal and after external error detection, do not necessarily stem from re-compilation of the correct form [9, 10, 11], but that repairs may stem from active forms competing with the selected form. We found (1) that in SLIP experiments error-to-interruption times are longer after non-elicited and multiple segmental errors than after single elicited errors. This supports the idea that when competition is not limited to an elicited specific error form and a correct target form, error detection takes longer. Later interruption after other than after elicited errors possibly also reflects that interruption can be postponed, although not indefinitely, when no repair is readily available [19, 21]. The finding that error-to-interruption times are longer for other errors than for elicited errors is corroborated by the observation (2) that single elicited errors are relatively more often detected in internal speech than other errors: Obviously, when the time available for error detection in internal speech runs out, the error can be detected later in overt speech. We also found (3) that error forms are more often repaired with the correct target form than with other competing error forms. But crucially, this effect is much weaker after external than after internal error detection. This confirms both that correct forms are frequently used as repairs and that activation of potential repairs decreases during the 500 ms delay between internal and external detection. Finally, we found (4) that interruption-to-repair times are shorter for single elicited than for other errors, but more so after internal than after external detection. This confirms our proposal that repairing is more difficult for non-elicited and multiple errors than for single elicited errors, and that activation levels of potential repairs decrease in the interval between internal and external detection.

In sum, these findings support our proposal that at least in this experimental task, repairs stem from correct forms and error forms being simultaneously active and competing at the moment of speaking the error form.

5. REFERENCES

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

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

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

[4] 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.

[5] Gauvin, H.S. & Hartsuiker, R.J. 2018. Towards a new model of verbal monitoring. *Journal of Cognition* (accepted).

[6] 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.

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

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

[9] Levelt, W.J.M. 1989. *Speaking. From intention to articulation*. Cambridge, Massachusetts: The MIT Press.

[10] 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.

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

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

[13] 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.

[14] Nooteboom, S.G. & Quené, H. 2019. Temporal aspects of self-monitoring for speech errors. *Journal of Memory and Language*, 105, 43-59.

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

[16] 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.

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

[18] Scrucca L., Fop M., Murphy T. B. and Raftery A. E. 2017. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal* 8/1, pp. 205-233.

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

[20] 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.

[21] Tydgat, I., Stevens, M., Hartsuiker, R.J. & Pickering, M. J. 2011. Deciding where to stop speaking. *Journal of Memory and Language* 64, 359-380.