PROSODIC CHARACTERISTICS OF INTERRUPTED VERSUS COMPLETED SPEECH ERROR REPAIRS IN SPONTANEOUS DUTCH SPEECH

Paul Carter & Leendert Plug

University of Leeds, UK
p.g.carter@leeds.ac.uk; l.plug@leeds.ac.uk

ABSTRACT

We present an analysis of speech error repairs aimed at establishing whether prosodic differences between 'interrupted' and 'completed' error repairs found in elicited data [8] are observed in a corpus of Dutch spontaneous speech. It has been assumed that the interrupted–completed distinction is associated with the coordination of pre- and post-articulatory self-monitoring processes; we explore the use of more fine-grained measures of reparandum completeness than considered so far. We find no evidence in intensity and f0 measures to support a distinction between 'interrupted' and 'completed'. Results of regression analyses differ in detail from the findings in [8] and provide only weak support for prosodic marking of early repairs. We suggest that analysis of interactional factors absent from experimental data may prove fruitful.

Keywords: prosodic marking, self-repair, self-monitoring, Dutch

1. INTRODUCTION

Speech errors and their repairs have long been studied by psycholinguists [1, 5, 7] to inform debate about speech processing. One important question concerns the extent to which processes involved in self-monitoring of pre-articulatory and articulated speech overlap. While some maintain that all self-monitoring is performed by the speech comprehension system applying identical processes to 'inner' and 'overt' input [3, 6], others argue that pre-articulatory and post-articulatory monitoring are functionally different, and should be expected to be associated with different behavioural patterns [8]. The phonetic details of speech error repairs provide a useful test case.

Since the function of self-monitoring inner speech is to 'prevent errors … from becoming public' [8] p.215, the process is under considerable time pressure. Once the erroneous form has been produced this pressure disappears and 'the speaker should take his or her time to make clear to the listener that an error has been made' [8] p.216. Therefore, while repairs following error detections in inner speech should 'make the error as little noticeable as possible', repairs following detections in overt speech should 'stand out by a prosody that is markedly different from both the error and the regular correct responses' [8] p.217.

Consistent with this argument, differences are reported [8] between speech error repairs in which the error form is interrupted early (*ba- dark boat*), and instances in which it is completed before being repaired (*bark bo- dark boat*). In the former, the onset of repair is too close to the onset of mispronunciation for the error to have been detected in overt speech: it must have been detected *before* the start of articulation. In the latter, it is more likely that the error was detected through monitoring of overt speech [4, 7]. In experimentally elicited 'interrupted' repairs, the repair component is associated with higher intensity and pitch peaks than the reparandum, whereas in 'completed' repairs, it is associated with lower intensity and pitch peaks.

The main question addressed in the present study is whether the prosodic patterns described above are observed as clearly in spontaneous speech error repairs as in experimental data. Moreover, while [8] classifies any incomplete reparandum as 'interrupted' regardless of its duration or proportional completeness, the present study explores the implementation of continuous measures, on the assumption that 'the later the interruption comes after the error has been made the greater the probability that the error was detected in overt speech' [8] p.217.

2. METHOD

2.1. Data selection

All instances involve the mispronunciation of a single word, followed by a repair that includes the correct pronunciation; they were collected from the *spontaneous conversations (face-to-face)* and *interviews/discussions/debates (broadcast)* sub-corpora of the Spoken Dutch Corpus [9] through searches for utterances coded as incomplete or mispronounced. Lexical error repairs, which involve the erroneous selection of an item which is itself accurately produced, were not considered. Since ambiguous lexical/sound errors may pattern with sound errors [11], they were included if a source of the supposed mispronunciation could be found in the immediate context. Instances with poor audio quality (e.g. background noise) were excluded in order to facilitate acoustic analysis. Since analysis focused on prosodic details of stressed vowels, instances in which the reparandum contains only consonants were also excluded. To improve consistency with the data in [8], only instances with phonologically identical first stressed vowels in the reparandum and the repair were considered. In total 170 instances were analysed. Examples are given below.

(1) *baarbij* ~ *waarbij* ('with which')
(2) *vanal* ~ *vanaf* ('from')
(3) *vruch-* ~ *vluchtmiddel* ('means of escape')
(4) *boelima* ~ *boelimia* ('bulimia')

2.2. Segmentation

Following [10], all instances were segmented into the components *reparandum* and *repair*. For consistency with [8], the first lexically stressed vowel in each component was labelled.

2.3. Prosodic analysis

Pitch tracks were computed in Praat [2] using a ceiling of 500 Hz and a floor of 50 Hz, except in a few cases which required a lower floor. Each track was manually checked with corrections confirmed by comparison with the original audio. Values for f0 were then log-transformed. Intensity contours were also calculated. Maximum, minimum and mean values for intensity and f0 were extracted both for the first lexically stressed vowels and for the reparandum and repair as a whole.

2.4. Structural-temporal analysis

We examined two measures of reparandum completeness: (*a*) a binary classification, following [8], and (*b*) a proportional value. For each instance, a broad phonetic transcription was prepared, and the number of segments realised in the reparandum was expressed as a proportion of the number of segments in the repair. Diphthongs and long vowels were counted as single segments for this purpose. Instances with a proportional value of 1, e.g. (1) and (2) above, are 'completed': the speaker produces an attempt at the entire target form before correcting. Instances with a value below 1, e.g. (3), are 'interrupted'. The relevant domain of comparison was the phonological word: in (3), *vruch-* [vrʏx] was compared with *vlucht* [vlʏxt]. Instances in which the speaker reached the end of the target form, but deleted one or more segments, as in (4), were considered completed.

In addition, we examined the relevance of three continuous duration measures: (*c*) duration of the reparandum, (*d*) offset-to-repair duration (from the end of the reparandum to the start of the repair), and (*e*) onset-to-repair duration (from the start of the reparandum to the start of the repair). Two of these (*c* and *e*) are particularly appealing since the onset of the reparandum may provide the closest accessible approximation to the point at which an inner speech monitor identifies that a repair is required: later than the onset of the reparandum it is increasingly likely that the error has been picked up by a monitor of overt speech. Offset-to-repair durations (*d*) were examined in [8] on the grounds that short durations might reflect a strategy for deflecting the listener's attention from the error.

2.5. Statistical analysis

After replicating the t-tests in [8] using the binary classification (*a*), we fitted regression models (linear or logistic, as appropriate) for each of measures (*a*) to (*e*), using the difference in intensity and f0 peaks in stressed syllables between reparandum and repair as predictors. All duration measures were log-transformed. Of the continuous measures (*c*) to (*e*), only onset-to-repair duration (*e*) approximated a normal distribution. Shapiro-Wilk and K-S tests suggested that offset-to-repair duration (*d*) violated assumptions of normality, while reparandum duration (*c*) was clearly bimodal. Further modelling with the full range of prosodic measures outlined above was therefore based on onset-to-repair duration (*e*).

3. RESULTS

Figure 1 shows differences in maximum intensity and f0 between the first lexically-stressed vowels. Reparandum measurements were subtracted from repair measurements so that positive values indicate an increase in intensity or f0 in the repair. Mean values with t-tests (with the Welch approximation for unequal variances) are given in Table 1. Unlike [8] we have no evidence for relatively high intensity/f0 after interruptions or low intensity/f0 after completed reparanda.

Figure 1: Box-and-whisker plots for differences in maximum intensity (left panel) and f0 (right panel) in stressed vowels, split by whether the reparandum is completed or interrupted. See text for further details.

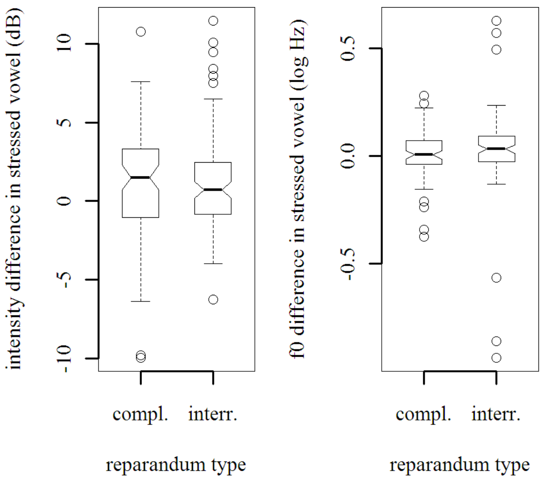


Table 1: Comparison of mean values for differences in maximum intensity and f0 in stressed vowels, split by whether the reparandum is completed or interrupted.

	compl.	interr.	t	df	*p*
intensity (dB)	0.9568	1.2236	0.4860	138.6	0.6277
f0 (log Hz)	0.0100	0.0247	0.6127	157.0	0.5410

Table 2 outlines the regression models, in none of which did the predictors have a significant effect, although intensity in model (*c*) approaches significance. Measure (*b*) includes several instances of 100% completeness in the reparandum and measure (*d*) includes several instances of zero duration between reparandum and repair which skew the distributions; the performance of the models is not improved noticeably by excluding these extreme values.

Table 2: Effects of models using the differences in intensity and f0 peaks in stressed vowels as predictors. Models: (*a*) logistic regression on binary completeness classification; linear regressions on (*b*) proportional completeness value, (*c*) reparandum duration, (*d*) offset-to-repair duration, (*e*) onset-to-repair duration.

	intensity (dB)		**f0 (log Hz)**	
	coeff.	*p*	coeff.	*p*
(*a*)	0.0173	0.7095	0.4625	0.6362
(*b*)	−0.4508	0.4172	5.2653	0.6525
(*c*)	−0.0173	0.0686	−0.1057	0.5967
(*d*)	0.0322	0.4216	−0.1374	0.8705
(*e*)	−0.0116	0.2436	−0.1015	0.6262

We then examined onset-to-repair duration (*e*) in more detail by broadening the range of prosodic measures considered. We used a log-likelihood-based stepwise model selection algorithm (AIC) to select candidate predictors from a variety of measures of the differences in intensity and f0 between reparandum and repair. We included the difference between the maximum, the minimum and the mean measurements and the difference between the ranges. In each case we supplied measurements both from the stressed vowel only and from the entire extent of the reparandum and repair, giving a total of 16 candidate predictors. This meant that the dependent variables reported in Table 2 (the difference in maximum values in the first lexically-stressed syllables) were still included as candidates in the algorithm. Only one variable was included in the model selected by the algorithm: the difference in intensity ranges over the whole reparandum or repair.

A linear regression model was then fitted to the data using this difference in intensity ranges as an independent variable. Inspection of the residuals suggested a number of data points were high-leverage observations; 8 observations (4.7%) were then removed and the model refitted, providing a better fit to the data (Table 3). Figure 2 plots the revised fitted model over the data points. A 1000-run bootstrap procedure confirmed that Δ intensity range should be retained in the model, though r^2 should be reduced to 0.0813.

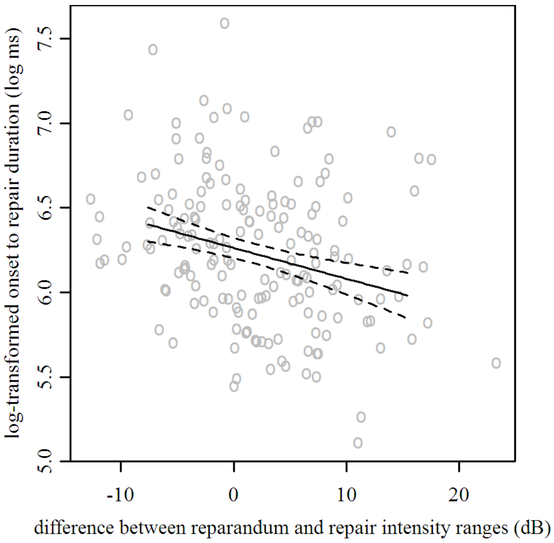
Table 3: Revised linear model fitted to the difference in intensity ranges between reparandum and repair.

	coeff.	SE	t	*p*
intercept	6.2631	0.0302	207.384	<0.0001
Δ intensity range	−0.0183	0.0045	−4.066	<0.0001
RSE: 0.3721 on 160 df				
r^2=0.0937; adjusted r^2=0.0880				

4. DISCUSSION

We have no evidence from our data to suggest that a binary distinction between interrupted and completed reparanda has an effect on the intensity or f0 peaks in the repair. If we assume an inner speech monitor is constrained by the time involved in articulating overt speech, however, there may not be reason to believe that there should be such a binary distinction; instead it seems reasonable simply to assume that reparanda of shorter duration are more likely than reparanda of longer duration to have been corrected by an inner speech monitor.

Figure 2: Scatter plot of the log-transformed onset-to-repair duration and the difference in intensity ranges between reparandum and repair. The solid line represents the effect of the predictor described in the text; dashed lines represent 95% confidence intervals.



However, we do not find evidence of prosodic differences dependent solely on reparandum duration either. The difference in intensity peaks in the first stressed vowels approaches a significant association with reparandum duration (Table 2) but the effect is masked once offset-to-repair duration is added in. The lack of significance in other measures in Table 2 means the more likely explanation for this near-significance is simply noise in the data, and the result should at this stage be counted as not significant.

Once we consider a wider range of prosodic measures, and onset-to-repair duration as a measure of how early repairs are initiated, we find that earlier repairs are associated with an increase in intensity range, whereas later repairs are associated with a reduction in intensity range. To the extent that increased intensity range could be seen as a phonetic exponent of emphasis, our findings are consistent with the idea that early repairs are prosodically marked to downplay the error, as suggested in [8] – although the details of this marking are different from those found in elicited instances.

Still, it is clear from the statistical models that measures of intensity differences between reparandum and repair are only associated to a small extent with the onset-to-repair duration: our final model accounts for only about 8% of the variance in the data. It is even possible that this increase in intensity range might reflect to an extent the fact that short, interrupted reparanda are less likely than their repairs to include low intensity items such as plosives. This is unlikely to be the entire explanation, though, since difference in intensity minima was not identified as a useful predictor of onset-to-repair duration.

Perhaps our most notable finding is the large amount of variance unaccounted for. We strongly suspect there may be interactional factors (by definition absent from experimentally elicited speech) having an effect on the prosody of spontaneous conversational data which we have not yet modelled; in particular, differing interactional structure may shed light on the fact that some interrupted reparanda are repaired with a large drop in f0 and some with a large increase (see Figure 1). This is clearly an area for future study.

5. REFERENCES

[1] Baars, B.J., Motley, M.T., MacKay, D. 1975. Output editing for lexical status from artificially elicited slips of the tongue. *J. Verbal Learn. Verbal Behav.* 14, 382-391.

[2] Boersma, P., Weenink, D. 2010. Praat: doing phonetics by computer [Computer program], Version 5.2. *http://www.praat.org/*

[3] Hartsuiker, R.J. 2006. Are speech error patterns affected by a monitoring bias? *Lang. Cog. Proc.* 21, 856-891.

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

[5] Levelt, W.J.M. 1983. Monitoring and self-repair in speech. *Cognition* 14, 41-104.

[6] Levelt, W.J.M., Roelofs, A., Meyer, A.S. 1999. A theory of lexical access in speech production. *Behav. Brain S.* 22, 1-75.

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

[8] Nooteboom, S. 2010. Monitoring for speech errors has different functions in inner and overt speech. In Everaert, M., Lentz, T., De Mulder, H., Nilsen, Ø. (eds.), *The Linguistic Enterprise*. Amsterdam: John Benjamins, 213-233.

[9] Oostdijk, N.J.H. 2002. The design of the Spoken Dutch Corpus. In Peters, P., Collins, P., Smith, A. (eds.), *New Frontiers of Corpus Research*. Amsterdam: Rodopi, 105-112.

[10] Plug, L. 2010. Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. *J. Phon.*, doi:10.1016/j.wocn.2010.08.001.

[11] Shattuck-Hufnagel, S., Cutler, A. 1999. The prosody of speech error corrections revisited. *Proc. 14th ICPhS* San Francisco. Vol. 2, 1483-1486.