UNIVERSITY OF LEEDS

This is a repository copy of *Acoustic correlates of prosodic marking in spontaneous self-repair in Dutch*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/78903/

Version: WRRO with coversheet

Proceedings Paper:

Plug, L (2014) Acoustic correlates of prosodic marking in spontaneous self-repair in Dutch. In: Fuchs, S, Grice, M, Hermes, A, Lancia, L and Mucke, D, (eds.) Proceedings of the 10th International Seminar on Speech Productions. Proceedings of the 10th International Seminar on Speech Production, 05-08 May 2014, Cologne, Germany. .

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

promoting access to White Rose research papers

White Rose Research Online

Universities of Leeds, Sheffield and York
http://eprints.whiterose.ac.uk/

This is an author produced version of a paper published in **Proceedings of the 10th International Seminar on Speech Production.**

White Rose Research Online URL for this paper:

http://eprints.whiterose.ac.uk/78903/

Paper:
Plug, L (2014) *Acoustic correlates of prosodic marking in spontaneous self-repair in Dutch.* In: Fuchs, S, Grice, M, Hermes, A, Lancia, L and Mucke, D, (eds.) Proceedings of the 10th International Seminar on Speech Productions. Proceedings of the 10th International Seminar on Speech Production, 5 - 8 May 2014, Cologne, Germany.

White Rose Research Online
eprints@whiterose.ac.uk

Acoustic correlates of prosodic marking in spontaneous self-repair in Dutch

Leendert Plug

University of Leeds, United Kingdom

l.plug@leeds.ac.uk

Abstract

This paper reports on a phonetic analysis of instances of lexical and phonological self-repair sampled from Dutch spontaneous speech. The focus is on the relationship between f0, intensity and articulation rate characteristics of the repairs on the one hand, and their perceived status as 'prosodically marked' or 'unmarked' on the other. The concept of prosodic marking was first introduced in psycholinguistic work on self-repair, and its phonetic correlates have remained unclear, although available descriptions suggest they may be different from those of other forms of prominence marking, in particular 'hyper-articulation'. This paper confirms that prosodic marking in self-repair is a form of prominence marking whose phonetic features warrant its distinctive label.

Keywords: self-repair, prosody, hyper-articulation

1. Introduction

This paper reports on a phonetic analysis of instances of self-repair, including lexical repairs such as *Thursd- uh Friday* and phonological repairs such as *expela- explanation*, sampled from Dutch spontaneous speech. The focus of the analysis is on the relationship between f0, intensity and articulation rate characteristics of a repair on the one hand, and its perceived status as prosodically 'marked' or 'unmarked' on the other.

The concept of prosodic marking in self-repair was first introduced by Cutler (1983) and Levelt & Cutler (1983), as a terminological improvement on Goffman's (1981) distinction between 'strident' and 'flat' repairs. Cutler (1983) describes an 'unmarked' repair as one in which the pitch, intensity and speaking rate of the repair component — in the case of lexical repair, the second lexical item; in the case of phonological repair, the correct pronunciation of the target word — are not noticeably different from those of the reparandum — the first lexical item, or the erroneous target word attempt. A 'marked' repair, on the other hand, 'is distinguished by a quite different prosodic shape from that of the original utterance' (Cutler 1983: 81). By leaving a repair unmarked, the speaker 'minimises the disruptive effect of the error on the utterance as a whole', while marking assigns 'salience' to the correction (Cutler 1983: 80).

Cutler's (1983) description of marked repairs suggests that prosodic marking in self-repair can be achieved in a variety of ways. Levelt & Cutler (1983: 206) suggest the same, stating that it can be implemented through 'a noticeable increase or decrease in pitch, in amplitude, or in relative duration'. Interestingly, this is somewhat at odds with Goffman's description of 'strident' repairs: according to Goffman, these involve raised pitch, loudness and tempo — not the opposite. Unfortunately, none of these sources present results of systematic acoustic analysis.

The acoustic correlates of prosodic marking in self-repair are of interest because a pertinent question from the perspective of speech production modelling is whether prosodic marking is any different from other forms of prominence marking — or in other words, whether its distinctive label is justified. If Goffman (1981) is right, it probably does, as pitch and intensity raising generally go together with temporal expansion, not compression. If Levelt & Cutler (1983) are right and prosodic marking can be achieved through, for example, a noticeable drop in intensity alone, it probably does too. If, on the other hand, most instances perceived as prosodically marked have a repair component produced at a higher pitch and intensity and a lower tempo than the reparandum, 'prosodic marking' would seem a redundant synonym of 'hyper-articulation' (Lindblom 1996).

More recent studies into the phonetics of self-repair have shown that an increase in pitch and intensity between reparandum and repair is the norm (Howell & Young 1991, Nakatani & Hirschberg 1994, Cole et al. 2005), as is an increase in articulation rate (Plug 2011). However, these studies present results of acoustic analysis only, without considering which of the repairs in their data sets sound prosodically marked. The current paper represents a first attempt to establish how prosodic marking is implemented in spontaneously produced self-repairs.

2. Data and method

2.1. Data

The data for this paper comprize 580 instances of lexical (N=214) and phonological (N=366) error repair extracted from four sub-corpora of the Spoken Dutch Corpus containing spontaneous speech. We extracted instances of speech which were coded as mispronounced or interrupted and did a number of additional, unsystematic data trawls. We only included lexical repairs in which one word was retroactively replaced by another, and phonological repairs containing at least one consonant and one vowel with primary or secondary lexical stress. Representative examples include *met de au- met de bus* 'by ca- by bus', *een leuke k- een mooie keuken* 'a nice k- a beautiful kitchen', [b]*aarbij* – [w]*aarbij* 'with which', and *vana*[l] *de* – *vana*[f] *de* 'from the'. We left aside repairs occurring in utterance-initial and utterance-final positions in order to minimize the possible effect of prosodic boundary marking.

2.2. Acoustic analysis

We segmented all instances, placing boundaries at the start and end of the crucial lexical items in the reparandum and repair — the word that is subsequently replaced and its replacement in the case of lexical repair; the two attempts at the target word in the case of phonological repair. We delimited all vowel portions within these domains.

We measured f0 (in Hertz) and intensity (in decibels) at every millisecond across the segmented vowel portions, and log-transformed f0 values. We then calculated mean, median and maximum values. In each case we calculated a delta value by subtracting the value derived from the reparandum from that derived from the repair. This yields a measure of the prosodic difference between the crucial components of the repair, as well as introducing some speaker normalization. Analysis not reported in detail here (but see Plug & Carter 2013) revealed that the mean, median and maximum delta values are tightly correlated, with the maximum delta values most informative in subsequent modeling. In what follows, we therefore restrict our attention to the maximum delta values, which we will refer to by their variable names, *F0 max delta* and *Intensity max delta*.

In addition, 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 square-root-transformed rate values to normalize their distribution, and calculated a delta value for each instance by subtracting the (transformed) value derived from the reparandum from that derived from the repair.

2.3. Auditory analysis

Following Levelt & Cutler (1983), we classified all instances as prosodically marked or unmarked based on auditory analysis. The question in each case was whether the correct target word realization sounds particularly salient because of its prosody, relative to the erroneous attempt. We allowed for the intermediate classification of 'possibly marked' (see Plug & Carter 2013).

The classification was done by two raters: the second author and a Dutch discourse analyst with no particular knowledge of the phonetics of self-repair. The two raters classified all instances independently. They reached the same judgment in 250 cases (77%). Of the 75 instances for which the raters proposed a different classification, 24 involved one rater proposing 'possibly marked' and the other 'marked'. In order not to overestimate the proportion of prosodically marked repairs, we coded these instances as 'possibly marked'. The remaining 51 instances either involved 'possibly marked' *vs* 'unmarked' or 'marked' *vs* 'unmarked'. All of these instances were reconsidered independently by both raters. In nine cases, this resulted in straightforward agreement, while in 42, the raters confirmed their initial judgments. Remaining cases of 'possibly marked' *vs* 'unmarked' were coded as 'unmarked'; cases of 'marked' *vs* 'unmarked' as 'possibly marked'. In the final coding, 385 instances (66%) are 'unmarked', 81 (14%) 'possibly marked' and 114 (20%) 'marked'.

In what follows, we will refer to the marking classification by its variable name, *Prosodic marking*. For the purpose of the quantitative analysis reported here, this variable was transformed into a binary one. Exploratory modeling suggested that collapsing 'possibly marked' and 'marked' results in a better fit with the acoustic measurements than collapsing 'possibly marked' and 'unmarked'. We will therefore report on the former.

2.4. Quantitative analysis

We investigated the relationship between the prosodic marking judgments and the acoustic measurements using cluster analysis (run in SPSS) and mixed effects regression modeling (using the lme4 package in R). For the purpose of modeling, we transformed the delta values into Z-scores. We incorporated *Speaker* as a random effect where relevant. Explorative modeling showed that *Repair type* — lexical or phonological — has no explanatory value, so this variable is not further discussed here.

3. Results

3.1. Correlations among acoustic parameters

Figure 1 provides a visual impression of the relationship between the three acoustic parameters. *Intensity max delta* is significantly correlated with both *F0 max delta* (unstandardized: Spearman's ρ=0.37, p<0.001) and *Rate delta* (ρ=−0.10, p=0.02). *F0 max delta* and *Rate delta* are not significantly correlated (ρ=−0.04, p=0.38). The correlation between *Intensity max delta* and *F0 max delta* is positive, while that between *Intensity max delta* and *Rate delta* is negative: the higher the intensity of a repair compared with its reparandum, the higher its f0, but the lower its articulation rate. Delta values are more often positive than negative across the three parameters (*Intensity max delta* 65%>0, *F0 max delta* 67%>0, *Rate delta* 70%>0), and 32% of instances have positive delta values only. Instances with negative values only are rare (4%), and instances with positive and negative deltas tend to show intensity and f0 clustering together against rate.

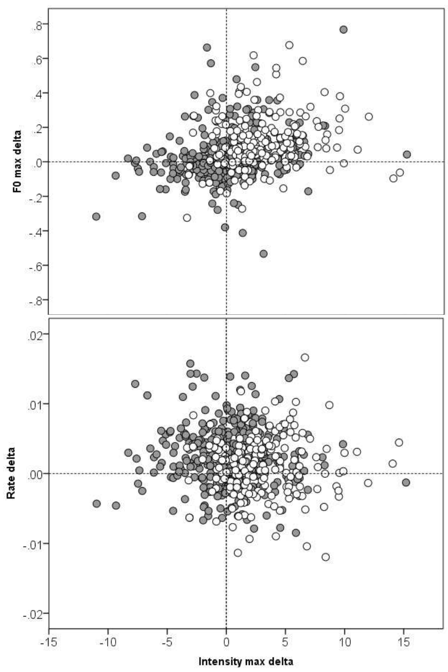

Figure 1: *Scattergrams of* Intensity max delta *against* F0 max delta *(top) and* Rate delta *(bottom). Dotted lines mark delta values of 0. Grey points represent unmarked instances; white points marked.*

3.2. Modeling the marking judgments

Looking now at the distribution of grey and white data points in the scattergrams in Figure 1, representing prosodically marked and unmarked repairs respectively, it seems clear that there are systematic relationships between the three acoustic parameters and the marking judgments. In the top scatter, white points predominantly cluster in the top right corner, while grey points have a wider spread. This means that prosodically marked instances are more strongly associated with positive values for *Intensity max delta* and *F0 max delta* than unmarked ones. In the bottom scatter, white data points cluster on the right accordingly, and appear to lean somewhat towards the bottom. The latter means that prosodically marked instances are associated with lower values for *Rate delta* than unmarked ones, although most are positive. These observations are confirmed by the boxplots in Figure 2 and associated unpaired comparisons: marked instances have significantly higher values for *Intensity max delta* (standardized: t(578)=9.40, p<0.001) and *F0 max delta* (t(578)=6.58, p<0.001), and significantly lower values for *Rate delta* (t(578)=−3.61, p<0.001).

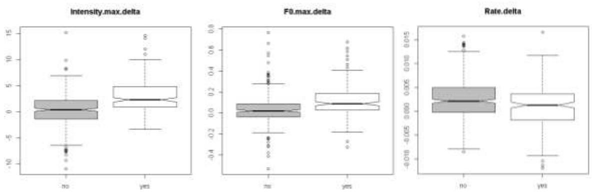

Figure 2: *Boxplots for* Intensity max delta*,* F0 max delta *and* Rate delta *with Prosodic marking as grouping variable ('no' meaning unmarked; 'yes' marked).*

To gain further insight into the predictive value of the three acoustic parameters, we built a mixed effects regression model with *Prosodic marking* as target variable, (standardized) *Intensity max delta*, *F0 max delta* and *Rate delta* as predictors and *Speaker* as a random factor. The model is shown in Table 1. The strongest single predictor was *Intensity max delta*. Given its significant correlations with the other two predictors, we orthogonalized the latter by replacing them with the residuals of simple linear models predicting their values on the basis of *Intensity max delta*. The model shows main effects consistent with the results of the paired comparisons cited above, and confirms that despite the significant correlations among the three acoustic parameters, each has predictive value of its own. In addition, a three-way interaction improves the model fit.

Table 1: *Mixed effects regression model with* Prosodic marking *as target variable and (standardized and orthogonalized)* Intensity max delta*,* F0 max delta *and* Rate delta *as predictors.*

Factor	Est.	SE	z	$p(>\lvert z\rvert)$
Intercept	−0.93	0.11	−8.33	<0.001
Intensity max delta	1.04	0.13	8.15	<0.001
F0 max delta	0.43	0.11	3.99	<0.001
Rate delta	−0.33	0.11	−3.08	0.002
Intensity~F0~Rate	−0.28	0.12	−2.38	0.017

The model is visualized in Figure 3 in the form of a conditional inference regression tree (Tagliamonte & Baayen 2012). The tree algorithm establishes which subdivisions in the data provide the most homogeneous groupings of observations with respect to a target variable — in this case *Prosodic marking*. Partitioning is recursive, so that predictors can feature more than once in the tree.

The tree in Figure 3 reflects the strength of *Intensity max delta* as a predictor: it yields three splits at the top of the tree (Nodes 1, 2 and 7), dividing the data into instances with substantial negative deltas, of which less than 10% are perceived as marked (Node 3), instances with high positive deltas, of which over 60% are marked (Node 11), and two sets of instances with intermediate delta values which are further split according to their values for *F0 max delta* (Node 4) and *Rate delta* (Node 8). Both *F0 max delta* and *Rate delta* split subsets of instances into two further subsets: one with around 20% and one with around 50% perceived as marked (Nodes 5 and 6, and 9 and 10). For *F0 max delta*, the subset with around 50% perceived as marked is associated with relatively high, positive values (Node 6); for *Rate delta* it is associated with relatively low ones, including negatives (Node 9). The asymmetry in the tree, with *F0 max delta* occurring in the left half only and *Rate delta* in the right half, is consistent with the significant three-way interaction in the linear model in Table 1.

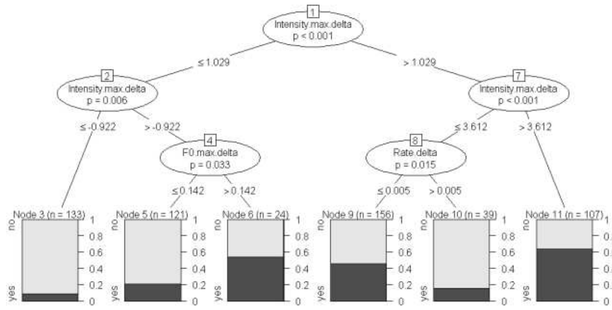

Figure 3: *Conditional inference regression tree with* Prosodic marking *as target variable and* Intensity max delta*,* F0 max delta *and* Rate delta *as predictors.*

3.3. Cluster analysis

The findings presented so far clearly confirm that intensity, f0 and articulation rate are systematically manipulated in the production of prosodic marking in self-repair. Evidence for acoustically distinct subtypes of marking, as suggested by Levelt & Cutler's (1983) definition, is less apparent: there is no obvious cloud separation in Figure 1, or evidence of multimodality in the distributions in Figure 2. Still, the regression tree in Figure 3 reflects some clustering among instances perceived as marked: these can have very high values for *Intensity max delta*, or intermediate values combined with relatively high values for *F0 max delta*, or low values for *Rate delta*. In order to assess the significance of this clustering, we carried out a cluster analysis on the standardized delta values. We first performed an unconstrained hierarchical cluster analysis using Ward's method. Inspection of the agglomeration schedule suggested that the most informative model should contain three to seven clusters. Subsequent constrained fitting was done through k-means clustering (Zellers & Ogden 2012). We report the model with five clusters, as it provides the best fit with the marking judgments. It also closely matches the regression tree in Figure 3. The model is summarized in Table 2.

Table 2: *Results of a k-means cluster analysis with (standardized) deltas as input, plus frequencies for* Prosodic marking*. Cluster centres are marked by distance from the mean for the relevant parameter: '++' >1 SD above, '…' <0.5 SD above or below, '−' >0.5 below, '−−' >1 SD below.*

	1	2	3	4	5
Intensity max delta	++	…	…	−−	…
F0 max delta	…	++	…	−	…
Rate delta	−	…	−	…	++
N	75	65	168	127	145
% marked	59%	52%	36%	12%	29%
% out of total marked	23%	17%	31%	8%	22%

Table 2 shows that in two of the five clusters (Clusters 1 and 2), more than 50% of instances are perceived as marked. Instances in these clusters have high values for either *Intensity max delta* or *F0 max delta*, and either low or average values for *Rate delta*. (Recall that on average, *Rate delta* is positive; in Cluster 1, its mean is just below zero.) Over a third of instances in Cluster 3 are perceived as marked, and this cluster accounts for the largest proportion of marked instances overall. The prosody of these instances shows little difference between reparandum and repair in intensity and f0, but a relatively low *Rate delta* average, just below 0. Cluster 5 is the same except that *Rate delta* is high: instances in this cluster are predominantly characterized by an increase in articulation rate between reparandum and repair. Nearly a third of instances in this cluster are perceived as marked. Cluster 4 has the lowest frequency of marked instances, and accounts for less than a tenth of marked instances overall. Instances in this cluster are characterized by a decrease in intensity and f0 between reparandum and repair, and little difference in articulation rate.

To assess the predictive value of cluster membership, we entered it as a predictor in a mixed effects model with random factor *Speaker*. This confirms that cluster membership is a significant predictor (Est=−0.44, SE=0.07, z=−5.99, $p(>|z|) <0.001$). However, log-likelihood comparison shows that its impact on model fit is considerably weaker than that of *Intensity max delta* alone, and not significantly stronger than that of *F0 max delta* or *Rate delta*. In other words, the observed data clustering is not strong enough to invalidate the linear model in Table 1, and *Intensity max delta* remains the strongest predictor of marking judgments.

4. Discussion

This paper has reported on an attempt to establish the acoustic correlates of prosodic marking in self-repair, using spontaneous repairs drawn from Dutch speech. Cutler's (1983) and Levelt & Cutler's (1983) descriptions suggest that in prosodic marking, speakers do not aim to produce a repair with a particular set of prosodic characteristics: rather, speakers can produce the repair in a number of ways, as long as it is noticeably different from the reparandum on any one or more prosodic parameters. Goffman's earlier description of 'strident' repairs suggests that these do have a recurrent overall prosodic shape, combining raised pitch, loudness and tempo.

The findings presented above are consistent to some extent with both descriptions. The implementation of prosodic marking in our data is less variable than Cutler's (1983) and Levelt & Cutler's (1983) descriptions might suggest: the prosodic marking judgments are best modeled linearly, with each of intensity, pitch and tempo contributing predictive value through predominantly unimodal relationships with prosodic marking. A clear majority of marked instances have a repair component with a higher maximum f0, a higher maximum intensity and a higher articulation rate than the reparandum. Still, while marking through a noticeable drop in maximum f0 or intensity is rare, it does occur, and so does marking through a noticeable drop in articulation rate.

In relation to Goffman's (1981) description, it is worth pointing out that tempo raising is not a feature that distinguishes marked from unmarked instances: in unmarked instances, too, the norm is for a repair to be produced at a higher articulation rate than its reparandum. In fact, the average tempo increase is *smaller* for prosodically marked repairs than for unmarked ones. This means that while prosodically marked repairs are mostly temporally compressed relative to their reparanda, they are temporally expanded relative to unmarked repairs. The latter makes prosodic marking similar to 'hyper-articulation' (Lindblom 1996), which combines pitch and intensity raising with temporal expansion. Still, if our findings are representative, it would be inappropriate to equate 'prosodic marking' to variation along a 'hypo-hyper continuum': a typical prosodically marked repair has a repair component that is higher in pitch, intensity and tempo than the reparandum — and is therefore neither locally hypo-articulated, nor locally hyper-articulated.

Why should 'prosodic marking' in self-repair be distinct from 'hyper-articulation' in this way? It is worth noting that a prevalence for speeding up at the repair tallies well with the observation in conversation-analytic work that speakers tend to initiate self-repair quickly, while delaying other-repair (see Plug 2011). Seen in this light, speeding up is consistent with a drive by speakers to get self-repair work done as soon as possible, and return to normal fluency. If a drive to allow for a soonest possible resumption of post-repair speech constrains the production of most instances of self-repair, including those that speakers do not aim to attach particular salience to, then it is arguably not surprising that the temporal characteristics of instances that *are* marked as salient are distinct from those of speech marked as salient outside of the context of self-repair. The upshot of this line of argument is that 'prosodic marking' in self-repair *can* be seen as a form of 'hyper-articulation' — but its result is not a 'hyper-articulated' repair as such, as the latter's characteristic of slow articulation is moderated by a drive to get the repair done as quickly as possible. Whether this account is on the right lines or not, it should be clear that prosodic marking in self-repair is an intriguing form of prominence marking whose phonetic correlates warrant its unique label.

5. Acknowledgements

This work was supported by ESRC grant RES-061-25-0417. I would like to thank Paul Carter and Christina Englert for their contributions to the research reported here.

6. References

Cole, J., M. Hasegawa-Johnson, C. Shih, et al. (2005). "Prosodic parallellism as a cue to repetition and error correction disfluency". In: *Proceedings of DiSS'05*, pp. 1–4.

Cutler, A. (1983). "Speakers' conceptions of the function of prosody". In: *Prosody: Models and measurements*, ed. A. Cutler & D.R. Ladd, Heidelberg: Springer, pp. 79–91.

Goffman. E. (1981). "Radio talk". In: *Forms of talk*, ed. E. Goffman, Oxford: Blackwell, pp. 37–46.

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

Levelt, W.J.M. & A. Cutler (1983). "Prosodic marking in speech repair". In: *Journal of Semantics* 2, pp. 205–217.

Lindblom (1996). "Role of articulation in speech perception: Clues from production". In: *Journal of the Acoustical Society of America* 99, pp. 1683–1692.

Nakatani C. H. & J. Hirschberg (1994). "A corpus-based study of repair cues in spontaneous speech". In: *Journal of the Acoustical Society of America* 95, pp. 1603–1616.

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

Tagliamonte, S. & R.H. Baayen (2012). "Models, forests and trees of York English: *Was/were* variation as a case study for statistical practice". In: *Language Variation and Change* 24, pp. 235–178.

Zellers, M. & R. Ogden (2013). "Exploring interactional features with prosodic patterns". In: *Language and Speech* online advance access, doi: 10.1177/002383091 3504568.