Proceedings Paper:

Carter, P and Plug, L (2014) Prosodic marking, pitch and intensity in spontaneous phonological self-repair in Dutch. In: Proceedings of the 10th International Seminar on Speech Production. 10th International Seminar on Speech Production, 05-08 May 2014, Cologne, Germany. .

Prosodic marking, pitch and intensity in spontaneous phonological self-repair in Dutch

Paul Carter[1], Leendert Plug[2]

[1]*University of Sheffield, United Kingdom*
[2]*University of Leeds, United Kingdom*

paul.carter@sheffield.ac.uk, leendert.plug@leeds.ac.uk

Abstract

We report on a phonetic analysis of instances of spontaneous phonological error repair sampled from Dutch spontaneous speech. We investigate whether phonological error repairs can be 'prosodically marked' and what factors constrain repair prosody. Previous studies of 'prosodic marking' in self-repair have suggested that while lexical error repairs are regularly realized with marked prosody, phonological error repairs are generally unmarked. Moreover, it has been asserted that the temporal organization of a repair has no bearing on its pitch and intensity characteristics. Our findings suggest that in fact, phonological error repairs are realized with prosodic marking as frequently, and through the same acoustic correlates as lexical repairs. We also show that repair timing is a significant predictor of prosodic marking judgments and f0 and intensity measurements.

Keywords: self-repair, prosody, spontaneous speech

1. Introduction

This paper reports on a phonetic analysis of instances of spontaneous phonological error repair such as *sa ... fat soap*, in which a mispronunciation is corrected. We address two questions: first, whether phonological error repairs can be 'prosodically marked'; second, what factors constrain repair prosody. The concept of prosodic marking in repair was first introduced by Cutler (1983), who describes a collection of spontaneous speech error repairs. She describes an 'unmarked' repair as one in which the pitch, intensity and speaking rate of the repair component — in the case of phonological error repair, the correct pronunciation of the target word — are not noticeably different from those of the reparandum — in the case of phonological error repair, 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 (1983: 83) states that unlike lexical repairs, in which one lexical item is rejected in favor of another, phonological error repairs are, as a rule, unmarked: her data set contains no prosodically marked phonological error repairs. Levelt (1989: 495) reiterates the claim, and emphasizes that repair prosody is '*semantically* motivated'. According to Levelt, a speaker's decision to mark or not to mark a repair is made at the message level. The greater the perceived semantic contrast between two lexical items, the greater is the likelihood of a decision to assign 'contrastive prominence' to one of the items. Phonological error repairs do not involve semantic contrast at the message level, as they involve two attempts at a single target word — hence their consistently unmarked form.

While Levelt's reasoning is appealing, there is reason to doubt that Cutler's observations generalize beyond her data set. Shattuck-Hufnagel & Cutler (1999:1485) report that in their data set, prosodic marking is 'less likely' in phonological error repair than in lexical repair, but the difference does not reach statistical significance. This must mean their data set *does* contain prosodically marked phonological error repairs. Furthermore, Nooteboom (2010) reports a pattern of prosodic differentiation of phonological error repairs that suggests that at the very least, labeling them all 'prosodically unmarked' is a simplification.

With reference to the second question, Cutler (1983), Levelt & Cutler (1983) and Levelt (1989) have described semantic factors that constrain speakers' choices for or against prosodic marking in lexical repair. They explicitly describe the choice as 'orthogonal to the time course of error detection and correction' (Cutler 1983: 81), or the 'interruption-and-restart structure of the repair' (Levelt & Cutler 1983: 211). Still, in Nooteboom's (2010) pattern of prosodic differentiation among phonological error repairs, it is exactly repair timing that is the most significant factor: Nooteboom 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. Furthermore, Kapatsinski (2010) has shown that highly frequent words resist interruption of the reparandum item in lexical repair: in other words, a repair's timing is correlated to some extent with its lexical frequency contour.

These considerations warrant detailed analysis of the prosody of phonological error repair in spontaneous speech. In this paper we model the pitch and intensity characteristics of a collection of phonological error repairs sampled from the Spoken Dutch Corpus, complementing a previous study of lexical repairs drawn from the same corpus (Plug & Carter 2013). We derive the characteristics from auditory judgments of prosodic marking, following Cutler (1983) and Levelt & Cutler (1983), as well as acoustic measurements, following Nakatani & Hirschberg (1994), Nooteboom (2010) and others. We explore the relationship between the auditory judgments and measurements, and evaluate the role of temporal and frequency-related factors in accounting for both.

2. Data and method

2.1. Data

The data for this paper comprise 325 instances of phonological 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 repaired mispronunciations containing at least one consonant and one vowel with primary or secondary lexical stress. We left aside repairs occurring in utterance-initial and utterance-final positions. We included instances ambiguous between phonological and lexical repair if the immediate context contained a plausible trigger for phonological error. Representative examples include [b]*aarbij* – [w]*aarbij* 'with which', *vana*[l] *de* – *vana*[f] *de* 'from the' and *met*[e]*rol-* *met*[eo]*rologisch* 'meteorological'.

2.2. Acoustic analysis

We segmented all instances, placing boundaries at the start and end of the erroneous target word attempt and the start and end of the repair stretch. We included any lexical items following the target word attempt in the reparandum, for the purpose of calculating target-to-offset and target-to-repair durations. We delimited all vowel portions within the erroneous target word attempt and correct realization. We labeled the first vowel with primary or secondary stress separately.

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. We did this for the first stressed vowels in the erroneous target word attempt and its correction, and across all of the vowels in these repair components. In each case we calculated a delta value by subtracting the value derived from the erroneous target word attempt from the value derived from the correct production. This yields a measure of the prosodic difference between the crucial components of the repair, as well as introducing some speaker normalization.

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 pitch or loudness, or both, 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'; remaining cases of 'marked' *vs* 'unmarked' as 'possibly marked'. In what follows, we will refer to the marking classification by its variable name, *Prosodic marking*.

2.4. Temporal analysis

In order to assess whether repairs with an erroneous target word attempt that is interrupted early have different prosodic characteristics from repairs with a completed attempt, we classified each reparandum item as interrupted or completed prior to repair (*Completeness*). In addition, we explored the relevance of continuous measures, on the assumption that these might capture more fine-grained differences between 'early' and 'late' repairs. First, we measured the duration from the start of the erroneous target word attempt to the abandonment of speech prior to repair (*Target-to-offset duration*) and the duration from the start of the erroneous target word attempt to the onset of repair (*Target-to-repair duration*). Repairs with a low target-to-offset duration tend to have a low offset-to-repair duration too (Nooteboom 2010), so that *Target-to-repair duration* might show greater differentiation between 'early' and 'late' repairs. All other things being equal, the higher these measures, the later the repair. We also included a binary classification of whether the repair is preceded by an editing term — *uh* or *of* 'or' in our data set — or not (*Editing term).*

Second, we took a proportional measure of target word completeness (*Proportional completeness*). This is appropriate since our target words are not independently controlled for word length or speaking rate; as a result, raw duration measures can only partially capture repair timing. We divided the number of segments in the erroneous target word attempt by the number of segments in the correct realization. We ignored segment deletions: the crucial question was which target word segment was reached in the first attempt. The measure is bounded by 1, which corresponds to the level 'complete' of our binary variable *Completeness*. All other things being equal, the higher the value, the later the repair.

2.5. Lexical frequency and control variables

We took two measures of the frequency of the target word in our quantitative analysis: its word form frequency (*Word frequency*) and its lemma frequency (*Lemma frequency*) as represented in the CELEX lexical database.

In modeling our prosodic parameters, we considered several other potentially relevant variables. First, we included a classification of each repair as involving consonantal error, vowel error or both (*Error type*), on the expectation that if prosodic marking is attested in phonological error repair, the correction of a vowel error is more likely to be marked than that of a consonantal error. Second, we included several measures of similarity between the vowels of the erroneous target word attempt and correction, to control for any effects of the relative number or nature of the vowels compared through our prosodic delta values. None of these yielded significant effects, so we leave them aside here. Similarly, we included several speaker-related factors, which yielded no significant effects.

3. Results

3.1. Occurrence of prosodic marking

In our final coding for *Prosodic marking*, 63 instances (19%) are classified as marked, 49 (15%) as possibly marked and 213 (66%) as unmarked. These proportions are very similar to those we have reported for lexical repairs (20%, 11% and 69% respectively; Plug & Carter 2013). Therefore, our findings provide no support for Cutler's assertion that while lexical repairs are regularly prosodically marked, phonological error repairs are generally unmarked.

3.2. Acoustic measures and marking judgments

The pitch and intensity characteristics of our phonological repairs are very similar to those reported in Plug & Carter

(2013) for lexical repairs. We focus here on delta values derived from the first stressed vowels in the erroneous target word attempt and its correction, as these allow for the most direct comparison with Nooteboom's (2010) data.

Figure 1 shows corresponding f0 maximum and intensity maximum delta measures plotted against each other, with marked, possibly marked and unmarked instances labeled 'm', 'p' and 'u' respectively. (Equivalent plots for median and mean delta values show similar patterns.) If f0 and intensity are manipulated independently in the prosody of self-repair, we would expect data points to fall into discrete clouds. Moreover, if our acoustic parameters are among those on which the perception of prosodic marking is based, we would expect data points representing marked, possibly marked and unmarked instances to cover distinct subareas of the plot — following Cutler (1983), marked instances should cluster around the periphery of the plot, where data points represent instances with a large delta value for one or both parameters.

Figure 1 shows that most instances have delta values around 0 for both f0 and intensity. Moreover, the scatter shows what looks like a single cloud of data points with a positive correlation between the two dimensions, and only small numbers of instances around the peripheries of the plots (ρ=0.3817, p<0.0001). Most data points corresponding to prosodically marked instances occupy the top right quarter of the plot: these instances have a rise in f0 and intensity between the reparandum and repair. Instances with a fall in f0 and intensity do occur in our dataset; however, few of them are perceived as marked. On the whole, the distributions suggest that the higher the increase in f0 and intensity maximum between a reparandum and repair item, the greater the likelihood that the repair is perceived as prosodically marked. This is in line with Cutler's (1983: 80–81) observation that 'typically', a marked repair 'is uttered on a higher pitch and with greater intensity than the erroneous material'.

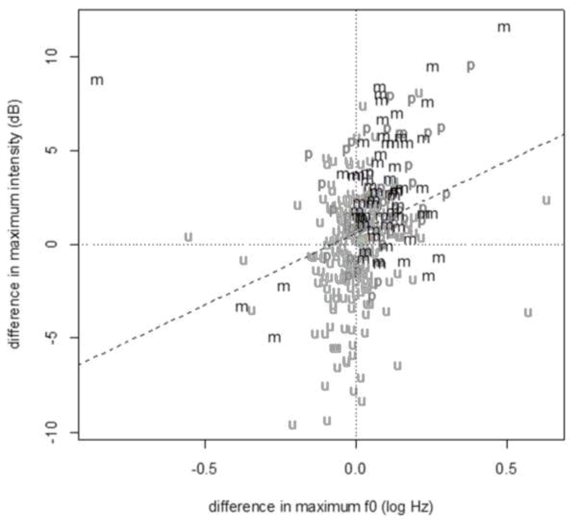

Figure 1: Scatterplot of f0 deltas by intensity deltas for maximum measures across the first stressed vowel in the erroneous target word attempt and correct realization. Data points are labeled for their auditory status: see text. The slope represents the outcome of a simple linear regression model.

These observations are confirmed by further analysis. Modeling of the marking judgments on the basis of our f0 and intensity maximum delta measures using conditional inference regression trees (Tagliamonte & Baayen 2012) reveals three homogeneous subsets of data: 185 instances with intensity delta values up to 2.76 and f0 delta values up to 0.05, of which less than 20% are perceived as possibly marked or marked; 73 instances with intensity delta values up to 2.76 and f0 delta values above 0.05, of which about 50% are perceived as such; and 67 instances with an intensity delta above 2.761, of which about 50% are perceived as marked.

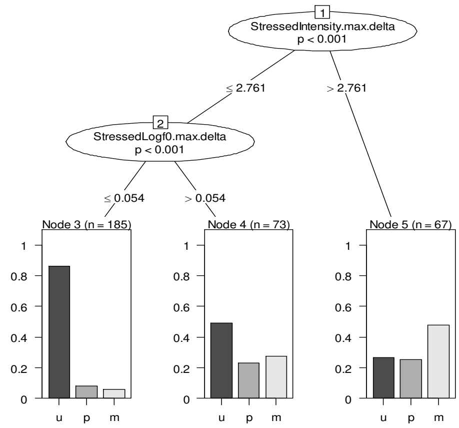

Figure 2: Conditional inference regression tree modeling the prosodic marking judgments on the basis of f0 and intensity maximum delta measures.

3.3. Modeling the prosodic parameters

In order to establish the predictive value of the factors described above, we modeled each of our acoustic parameters and our marking judgments, again using conditional inference regression trees. The modeling revealed consistent effects across dependent variables; we focus here on f0 maximum delta measured across the first stressed vowels in the erroneous target word attempt and its correction. In modeling this parameter, we included the corresponding intensity measure as a control variable, so that candidate predictor effects that are observed only in subareas of Figure 1 may emerge in the form of interactions between those predictors and the control prosodic variable.

Figure 3 shows the resulting regression tree. We see that the data is first split on *Stressed vowel intensity maximum delta*, such that instances with a high intensity maximum delta also have a high f0 maximum delta. This is in line with the positive correlation visible in Figure 1. More noteworthy is that in the subset of instances with a relatively low intensity maximum delta — and therefore a relatively low f0 maximum delta, at or below 0 on average — two further splits are possible, on *Completeness* and *Target-to-repair duration*. First, the 101 instances within this subset with an incomplete erroneous target word attempt have a higher f0 maximum delta (0 on average) than the remaining 37 instances with a completed target word attempt (below 0 on average). These 37 instances can be further split, such that instances with a higher target-to-repair duration have a lower (on average negative) f0 maximum delta.

Figure 3 illustrates that there is a systematic relationship between repair timing and repair prosody in our data. The relationship is not strong: instances with f0 deltas above 0 are mostly unconstrained by repair timing. Still, the direction of the relationship is consistent with the pattern of prosodic differentiation of early and late repairs described by Nooteboom (2010): early repairs are associated with higher delta values than late ones. None of our frequency-related variables or control variables feature in the analysis, including whether the mispronunciation concerns a vowel or consonant. However, it is worth noting that direct modeling of our timing variables (not shown in detail here) does reveal that higher frequency target words are more likely to be completed prior to repair than lower frequency target words. This provides some support for the notion that higher-frequency lexical items form more cohesive units in speech production (Kapatsinski 2010), although this interaction between timing and frequency variables does not appear to have a significant effect on repair prosody.

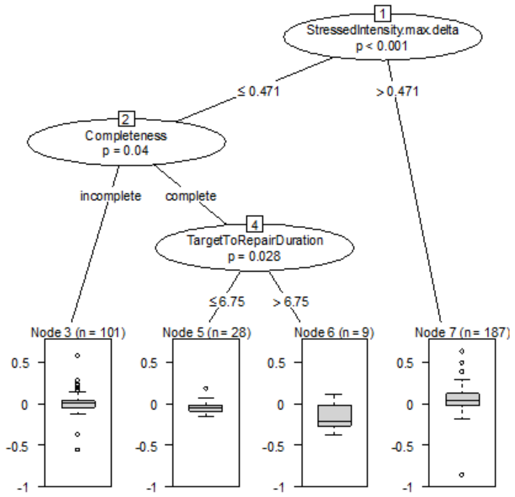

Figure 3: Conditional inference regression tree modeling the difference in f0 maximum between the first stressed vowels in the erroneous target word attempt and its correct production.

4. Discussion and conclusion

As indicated above, our findings provide no support for Cutler's (1983) assertion that while lexical repairs are regularly prosodically marked, phonological error repairs are, as a rule, unmarked. Comparing the findings with those of our previous study of lexical repairs (Plug & Carter 2013), we can only conclude that the prosody of phonological repairs is not as different from that of lexical repairs as Cutler suggests. This arguably has implications for our understanding of the function of prosodic marking in self-repair. In Levelt's (1989) model of speech production, a speaker's decision to mark or not to mark a repair prosodically is a semantic one, made at the message level. While phonological repairs do not, strictly speaking, involve semantic contrast at the message level, it seems that contrastive prominence can be assigned to the repair. Repair prosody is perhaps not exclusively 'semantically motivated' — or perhaps an incorrect and correct attempt at the same target word should be considered semantically distinct, at least in the context of self-repair.

It is worth highlighting that we found no prosodic differences between repairs of vowel and consonant mispronunciations. If we assume that segments can function as prosodic domains, attracting narrow focus (Van Heuven 1994), it would seem plausible that in phonological repair, contrastive prominence is assigned at the segment level, while in lexical repair, it is assigned to the entire repair item. If so, corrections of vowel errors should be more likely than corrections of consonant errors to be associated with pitch marking, simply because many consonants cannot be marked through pitch. The absence of this pattern in our data suggests that there is little qualitative difference between prosodic marking in lexical repair and prosodic marking in phonological repair: both are equally frequent, and both appear to be implemented through similar speech production processes, across similar domains.

Our findings are consistent with Cutler's (1983: 80–81) assertion that 'typically', a marked repair 'is uttered on a higher pitch and with greater intensity than the erroneous material', and suggest that the independence of pitch and intensity parameters implied by Levelt & Cutler's (1983: 206) definition of prosodic marking — 'a noticeable increase or decrease in pitch, in amplitude, or in relative duration' — should not be overestimated. Nakatani & Hirschberg (1994) have reported similar results.

Our findings are also consistent with those of Nooteboom (2010). We predicted that early repairs are associated with higher delta values for f0 and intensity than late ones: early repair should be associated with a mean rise in pitch and intensity between reparandum and repair, while late ones should be associated with a mean fall. We found some evidence that this is the case in our data. When modeling the three f0 delta measures between the first stressed vowels in the erroneous target word attempt and its correct production, several of our measures of repair timing yielded significant effects. The effects are all in the predicted direction: for example, incomplete erroneous target word attempt have a higher f0 maximum delta than complete ones in one subset of the data, and instances with a relatively high target-to-repair duration have a relatively low f0 maximum delta in another. The effects are weak, but similar to those we found in our study of lexical repair (Plug & Carter 2013). Therefore, we can conclude that it is at best a simplification to suggest that repair prosody is 'orthogonal to the time course of error detection and correction' (Cutler 1983: 81).

5. Acknowledgements

This work was supported by ESRC grant RES-061-25-0417. We thank Christina Englert for her contribution.

6. References

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.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, Mass.: MIT Press.

Shattuck-Hufnagel, S. & A. Cutler (1999). "The prosody of speech error corrections revisited". In: *Proceedings ICPhS 1999*, San Francisco.

Nooteboom, S. (2010). "Monitoring for speech errors has different functions in inner and overt speech". In: *The linguistic enterprise*, ed. M. Everaert et al., Amsterdam: Benjamins, pp. 213–233.

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

Plug, L. & P. Carter (2013). "Prosodic marking, pitch and intensity in lexical self-repair in Dutch". In: *Phonetica* 70, pp. 155–181.

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.

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.

Van Heuven, V. J. (1994). "What is the smallest prosodic domain?" In: *Phonological structure and phonetic form: Papers in laboratory phonology III*, ed. P. A. Keating, Cambridge: Cambridge University Press, pp. 76–98.