Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Investigating the causes of prosodic marking in self-repairs: an automatic process?

Morgane Peirolo[1], *Antje S. Meyer*[1,2,3], *Candice Frances*[1]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands
[3]Radboud University, Nijmegen, The Netherlands

morgane.peirolo@mpi.nl, antje.meyer@mpi.nl, candice.frances@mpi.nl

Natural speech involves repair. These repairs are often highlighted through prosodic marking (Levelt & Cutler, 1983). Prosodic marking usually entails an increase in pitch, loudness, and/or duration that draws attention to the corrected word. While it is established that natural self-repairs typically elicit prosodic marking, the exact cause of this is unclear. This study investigates whether producing a prosodic marking emerges from an automatic correction process, or if it arises for other purposes, e.g., a communicative purpose. In the current study, we elicit corrections to test whether all self-corrections elicit prosodic marking. Participants carried out a picture-naming task in which they described two images presented on-screen. To prompt self-correction, the second image was altered in some cases, requiring participants to abandon their initial utterance and correct their description to match the new image. This manipulation was compared to a control condition in which only the orientation of the object would change, eliciting no self-correction while still presenting a visual change. We found that the replacement of the item did not elicit a prosodic marking, regardless of the type of change. Theoretical implications and research directions are discussed, in particular concerning theories of prosodic planning.

Index Terms: *speech prosody, prosody planning, prosody marking, prominence, focus structure, repairs*

1. Introduction

Natural speech is not without imperfections. While traditional perspectives considered speech errors as mere outliers, recent research has highlighted their crucial role in spontaneous conversations. In everyday speech, we often find ourselves hesitating, repeating, correcting, and making repairs (Pouplier & Hardcastle, 2005). As an indicator, the latter occurs in around 10% of utterances (Nakatani & Hirschberg, 1994). Instead of being exceptions, these errors have proven to be valuable in unveiling the hidden mechanisms of speech production. For instance, the study of prosodic marking offers insight into how speech errors are typically repaired (Postma, 2000). According to Levelt (1983), the repair process involves self-monitoring, where the speaker becomes aware of the error and produces a correction, often marked by pitch accenting. This accenting emphasizes a specific part of speech acoustically, creating a prosodic marking that corrects (e.g., was it a big pear? No, it was a big *apple*) or clarifies information (I am going to take the bike. Which one? The *blue* bike). Existing literature, such as studies by Plug (2011, 2014), has demonstrated that natural corrections result in prosodic marking. However, it remains unclear whether prosodic marking arises from communicative purposes, or if it is an automatic process triggered by abandoning an utterance and resuming fluent speech. Previous work from Arnold (2008) highlights the role of the addressee in modulating our referential expressions. When in the presence of a listener, we will make our speech easier to interpret (e.g., offering a more detailed description of the object, acoustically emphasizing parts of the utterance). However, this listener-oriented adaptation is not present when producing internal speech. Another line of work from Nooteboom (2004) points at the distinctive strategies used in inner and overt speech during error monitoring. While overt speech repairs aim at drawing the attention of the listener to the corrected utterance, inner speech monitoring aims to minimize the salience of the repair, thus preventing the error from being made public. This kind of repair arises during speech planning, placing it under great time pressure. While showcasing the importance of communicative purposes in self-initiated repairs, no conclusion can be drawn for externally cued corrections.

The current study investigates whether prosodic marking is a generalized automatic process inherent to corrections. Specifically, we explore whether prosodic marking results from recognizing and repairing errors or from a change in planning.

To address this question, we induced repairs in sentences without contrastive information using an online picture-naming task. Participants were presented with two images on the screen, and in some cases, the image on the right was replaced, requiring a repair. If prosodic marking is not solely tied to a communicative purpose but also stems from automatic processing, we anticipate observing the prosodic markers of repairs. This might manifest as increased intensity, lengthening of vowel production on the stressed syllable, and/or a shift in fundamental frequency (F0) toward a later and higher peak (Grice & Kügler, 2021). These acoustic parameters, including average intensity, fundamental frequency, and duration, will be compared across noun phrases (Calhoun, 2010).

In summary, the fundamental distinction between the two hypotheses lies in whether acoustic markers are present in induced repairs. If prosodic marking is evoked automatically by the repair, these markers should be evident; otherwise, if it is driven by other (for example communicative) purposes, no such markers should be found, irrespective of the presence of induced repairs.

 10.21437/SpeechProsody.2024-218

2. Methods

2.1. Participants

Participants were 15 native Dutch speakers (M=24,555 years; SD=3, 2398; [20;32]; 13 females, 2 males). All participants were recruited through the internal database of the Max Planck Institute for Psycholinguistics and compensated 7 euros. Two participants were removed for not correctly following directions. The final set of participants was 13. All participants reported no hearing, speech, or neurological disorders, and had normal or corrected-to-normal vision. The participants gave written consent before taking part in the experiment, which was approved by the Ethics Committee Faculty of Social Sciences of the Radboud University, approval number ECSW-2019-019.

2.2. Apparatus

The experiment was built in Frinex (Whiters, 2016), the internal experimental platform of Max Planck Institute for Psycholinguistics. The experiment was run online on the participants' personal computers. The audio files were recorded through the microphone of each computer.

2.3. Stimuli

Stimuli consisted of 60 line-drawn images, each of them in one of six colors. The complete dataset consisted of 360 images.

2.3.1. Image and colour selection

The 60 images were selected from (Severens et al., 2005), on the basis of their visual and linguistic characteristics (Perret & Bonin, 2019). All the selected images had a name agreement equal to or higher than 80%. Additionally, we removed compound names, as well as names with more than 8 letters or a frequency lower than 6 words per million. The final dataset was composed of names consisting of one or two syllables. The final set of words and images is accessible in the Supplementary material repository.

Each item was modified using the recolour function of PowerPoint so that the lines were in one of six colours. The colour tones were selected according to the hex codes of primary colours (see Supplementary material repository). Additionally, all colors started with a consonant and were composed of a single syllable in Dutch. The final color set was red, purple, blue, yellow, brown, and green ("rood", "paars", "blauw", "bruin", "geel", and "groen").

2.3.2. Experimental conditions

Two variables of interest were manipulated. The first variable was Change (two levels: change; no change), which refers to whether the right image would change for a new one or stay the same. Nested into the change condition was the second variable Type of change (two levels: orientation change, object change). In the orientation change condition, the new contrastive object was a mirror image of the previous one. In the object change condition, the new contrastive object was different both in terms of shape and color from the old objects.

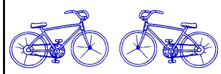
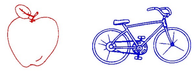
Experimental condition	**Target noun phrase**	**Stimuli**
No change	*Dit is een blauwe fiets.*	
Orientation change	*Dit is een blauwe fiets.*	
Object change	*Dit is een blauwe fiets.*	

Table 1: Summary of experimental conditions

2.4. Task

The task consisted of the presentation of two stimuli displayed on the left and right sides of the screen. In each condition, the trial started with a central fixation cross for 500 ms, followed by a first stimulus screen composed of an image on the left side of the screen, present for 3000 ms. After that delay, a second stimulus screen displayed an image on the right side of the screen. Condition 1 (*No-change condition*) was composed of these two subsequent screens. In condition 2 (*Orientation change condition*), the second screen was presented for 1000 ms then followed by a third one for 2000 ms, showing a mirror image of the right item. In condition 3 (*Object change condition*), the second screen was presented for 1000 ms and was followed by a third one for 2000 ms, picturing an image different both in shape and colour from the two old ones.

The experiment consisted of 3 blocks of 60 trials, for a total of 180 trials: thirty trials of each experimental condition, with the remaining 90 trials being fillers. Items were manually randomized within the experiment, such that the first three trials were fillers, no more than two consecutive trials were either condition of interest or filler, and no consecutive trials were of the same experimental condition.

2.5. Procedure

The experiment took around 40 minutes to be completed. At the beginning of the experiment, a recording test allowed the participant to test their microphone. At the end of each block, the participants were allowed to take a self-paced break.

Participants were instructed to describe the objects appearing on the screen, using a standardized sentence structure: "dit is een [adjective] [noun]." The participants were instructed to respond as quickly as possible. In every picture, the adjective referred to the colour, and the noun to the shape (e.g., "dit is een blauwe fiets"; *this is a blue bike*). Participants were instructed to first describe the image on the left of the screen, then move on to the right side of the screen and describe the second image. In condition 3, where the right item was replaced by a new one, participants were instructed to ignore the old item and describe the new item out loud as quickly as possible. If they had started to describe it, they were instructed to abandon the utterance and restart describing the new image.

Measures

2.5.1. Annotations

The utterances of the target images from the COI trials were annotated in Praat, with the help of the pitch and the intensity drawings on the spectrogram. For condition 3 (Object change condition), only the utterance of the replacement image was annotated. The target noun phrases (adjective plus noun) were boundary-delimited, and then a script extracted average and peak values for:

1) **Intensity (dB),** defined as the peak intensity over the target noun phrase (e.g., "blue bike), and expressed as the loudest point over the target noun phrase.

2) **Fundamental Frequency (Hz),** defined as the peak F0 over the target noun phrase, which is expressed as the highest value of the fundamental frequency over the target noun phrase.

3) **Duration (ms),** defined as the length of the target noun phrase.

For each of these measures, outliers more than 2.5 SD from the mean were removed before analysis (17,64 % of the data).

Trials with an incorrect answer, interruptions, and hesitations between the adjective and the noun were removed. In total, 41 trials were removed, leaving a dataset of 995 trials for analysis.

2.5.2. Statistical analyses

Analyses were conducted in RStudio (version 4.3.1). The data was analysed with Bayesian logistic mixed effects regression, with the package *brms* (Bürkner, 2017). The rationale behind using Bayesian statistics is that it allowed us to investigate the evidence for the null hypothesis, in other words validating that our results yielded equal values (Dienes, 2014; Nicenboim et al., 2018; Vasishth et al., 2018). Following (Stefan et al., 2019), we used an uninformed prior. The crucial contrasts to assess our hypotheses were 1) Change (change vs no change), allowing us to see whether a new item replacing a previous one would make the participants produce focus; 2) Type of change (orientation vs object change), allowing us to see, in the case of prominence production, whether the change of item or the change itself was leading the potential effects. To determine if the object change had an effect, we analysed the three acoustic parameters (pitch, intensity, and duration) across all participants, by experimental condition. Two model comparisons were run to assess the effect of Change and Type of change. Each model was respectively compared to an intercept only model, allowing to compute Bayesian Factors. Data was normalized using a scale function. Participant and Stimulus were fitted as random intercepts. Participants mean accuracy on the task was overall of 96,6%. The results are reported in terms of the 95% Credible intervals (95%-CrI) and Bayes factors (BF_{10}) for the contrasts of interest. The details of the models can be found in the Supplementary material repository.

2.6. Fundamental frequency, peak Hz

The Bayesian analysis provided strong evidence for the absence of effect of type of change, with $BF_{10} = 0.04652$, and 95% credible interval (CrI) [-18.04, 8.99]. Furthermore, the analysis provided strong evidence for the absence of effect of change, with $BF_{10} = 0.04837$, and 95% credible interval (CrI) [-14.78, 16.55]. This suggests that changing the stimuli didn't lead the participants to treat self-repair as a case of prosodic marking, regardless of the type of change.

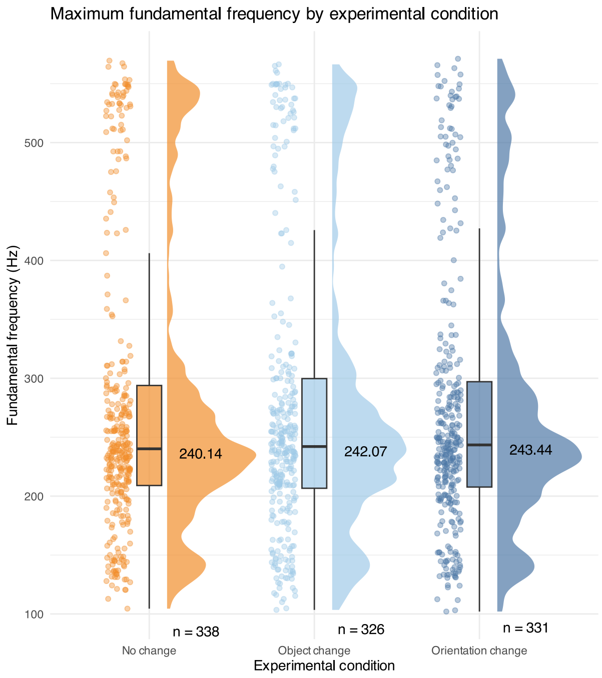


Figure 1: maximum fundamental frequency over the target noun phrase by experimental condition.

N=X represents the number of datapoints per condition; the values next to the boxplot are the median of each condition.

2.7. Intensity, peak dB

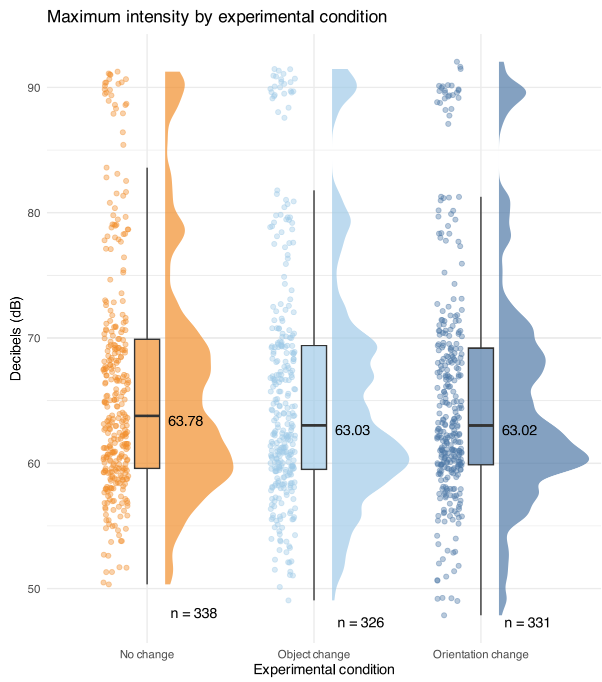


Figure 2: maximum intensity over the target noun phrase by experimental condition.

N=X represents the number of datapoints per condition; the values next to the boxplot are the median of each condition.

The Bayesian analysis provided moderate evidence for the absence of effect of type of change, with $BF_{10} = 0.25498$, and 95% credible interval (CrI) [-0.73, -0.02]. Furthermore, the analysis provided anecdotal evidence for absence of effect of change, with $BF_{10} = 0.77225$, and 95% credible interval (CrI) [-0.13, 0.69]. This suggests that changing the stimuli did lead the participants to treat self-repair as a case of prosodic marking, but that effect was not modulated by the type of change.

2.8. Average duration

The Bayesian analysis provided provided strong evidence for the absence of effect of type of change, with $BF_{10} = 0.04237$, and 95% credible interval (CrI) [-9.76, 19.70]. Furthermore, the analysis provided provided strong evidence for the absence of effect of change, with $BF_{10} = 0.03743$, and 95% credible interval (CrI) [-14.89, 22.16]. This suggests that changing the stimuli didn't lead the participants to treat self-repair as a case of prosodic marking, regardless of the type of change.

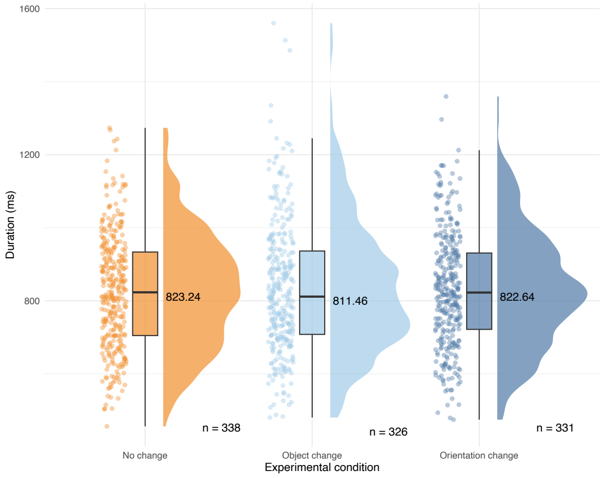


Figure 3: average duration over the target noun phrase by experimental condition.

N=X represents the number of datapoints per condition; the values next to the boxplot are the median of each condition.

3. Discussion

The current study explored the impact of induced self-repairs on prosodic marking. One hypothesis posited that prosodic marking arises from an inherent, automatic cognitive process. Alternatively, a second hypothesis suggested that prosodic marking serves a communicative purpose. In the former case, this would imply that prosodic marking results from a shift in planning; the latter case suggests that the marking is associated with communicating a repair.

To address this question, we prompted repairs in sentences lacking contrastive information through an online picture-naming task. Participants were presented with two images on the screen, and in certain trials, the right image was replaced, necessitating a repair. If prosodic marking is not solely linked to a communicative purpose but also originates from automatic processing, we hypothesized observing the prosodic markers of repairs, such as increased intensity, lengthening of the word, and/or a shift in fundamental frequency.

A Bayesian analysis found that, overall, no change in the three acoustic parameters was present, apart from the intensity of the target, which was significantly different between the change and no-change conditions. Given that this was not dependent on the repair per se, being present even in the change condition in which no repair was necessary, this effect is more likely related to the attentional shift to the visual change.

One possible future direction is to investigate whether the prosodic marking is affected by the presence of an interruption. In the present study, the delay between the first and second screen was long enough to allow the participants to finish describing the first image without interruption. A design eliciting interruptions would provide clarity on whether and how these affect the production of cued corrections. This could further be done by comparing a dataset of elicited interruptions and corrections. Whereas not taking from the conclusions of the current study, this analysis would allow us to have a better understanding of speech production and monitoring, in particular whether the interruption itself and its inherent repair induces prosodic marking.

Overall, our data strongly support the hypothesis that prosodic marking stems from processes other than the repair itself, such as for communicative purposes. A similar pattern of behaviours emerges from our study and the ones from Arnold (2008) and Nooteboom (2004). Despite differences in the origins of corrections (externally cued corrections vs production internal processes), it seems that speakers accurately modulate their speech based on contextual—be it internal or external—cues.

Moreover, aside from a difference in intensity between the change and no-change trials, the 13 participants maintained fundamental frequency and duration constant. The variation in intensity could potentially be attributed to a surprisal effect triggered by the image change. The key result is that the two change conditions—the one requiring repair and the other only showing a visual change—showed the same acoustic characteristics. The equality of values for the two other acoustic measures strengthen our conclusion that repairs themselves are not enough to modulate prosodic marking and that this phenomenon is due to other processes such as internal monitoring or communication which cannot be evoked by induced repairs.

4. Conclusions

In the present study, we found that prosodic marking is not an automatic process derived from repairing or quickly modifying an utterance. This suggests that prosodic marking in repairs may be the result of other processes or even serve communicative purposes rather than being an automatic process. Although further studies investigating the time course and characteristics of repairs are needed, especially in relation to self-interruptions, these results provide a solid foundation for the understanding of prosodic marking in repairs.

5. Data availability statement and supplementary material

The stimuli and datasets used in this study, along with any relevant supplementary material (such as scripts and full model outputs) can be found on the online OSF repository https://osf.io/nxvp7/?view_only=efada9e9f2bb44ed87646244296925f1.

6. References

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527. https://doi.org/10.1080/01690960801920099

Bögels, S., Schriefers, H., Vonk, W., & Chwilla, D. J. (2011). Pitch accents in context: How listeners process accentuation in referential communication. *Neuropsychologia*, *49*(7), 2022–2036. https://doi.org/10.1016/j.neuropsychologia.2011.03.032

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*(7–9), 1099–1140. https://doi.org/10.1080/01690965.2010.491682

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00781

Grice, M., & Kügler, F. (2021). Prosodic Prominence – A Cross-Linguistic Perspective. *Language and Speech*, *64*(2), 253–260. https://doi.org/10.1177/00238309211015768

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M., & Cutler, A. (1983). Prosodic Marking in Speech Repair. *Journal of Semantics*, *2*(2), 205–218. https://doi.org/10.1093/semant/2.2.205

Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *The Journal of the Acoustical Society of America*, *95*(3), 1603–1616. https://doi.org/10.1121/1.408547

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and Confirmatory Analyses in Sentence Processing: A Case Study of Number Interference in German. *Cognitive Science*, *42*(S4), 1075–1100. https://doi.org/10.1111/cogs.12589

Nooteboom, S. (2004). *SELF-MONITORING OF INNER SPEECH AND OVERT SPEECH*.

Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, *51*(6), 2533–2545. https://doi.org/10.3758/s13428-018-1100-1

Plug, L. (2011). Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. *Journal of Phonetics*, *39*(3), 289–297. https://doi.org/10.1016/j.wocn.2010.08.001

Plug, L. (2014). On (or not on) the 'upgrading-downgrading continuum': The case of 'prosodic marking' in self-repair. *Prosody and Phonetics in Interaction*, 70–86.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*(2), 97–132. https://doi.org/10.1016/S0010-0277(00)00090-1

Pouplier, M., & Hardcastle, W. (2005). A Re-Evaluation of the Nature of Speech Errors in Normal and Disordered Speakers. *Phonetica*, *62*(2–4), 227–243. https://doi.org/10.1159/000090100

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*(2), 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042–1058. https://doi.org/10.3758/s13428-018-01189-8

Vasishth, S., Nicenboim, B., Beckman, M., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161. https://doi.org/10.1016/j.wocn.2018.07.008

Whiters, P. (2016). *Frinex: Framework for Interactive Experiments | Max Planck Institute*. https://www.mpi.nl/publications/item2353700/frinex-framework-interactive-experiments