Dear Editors,

We've now revised the manuscript. Thanks for all the insightful feedback on the previous version (both from the reviewers and the editors). We hope that the narrative is now much clearer, especially in the discussion. We give detailed responses to the comments below. Please let us know if you'd like to see any other changes made!

All the best,

Michael (Wagner) & Michael (McAuliffe)

Content

Highlights: Please reconsider the choice of highlights (and reformulate them). Neither the study of speech acts (first bullet point) nor interrogative rises (third bullet point) seem to be the main aspects of your paper.

> *We've revised them*

p.4, Fig.2: Explain the figure a bit more (incl. the abbreviations – some readers may not be familiar with them) or delete it.

> *We got rid of the abbreviations and explained the figure a bit more.*

p.11, second para: Give concrete examples for boosting and reduction of prosodic parameters.

> *We are more explicit now on what exactly is boosted/reduced.*

p.18, l.976ff. or p.19, l.1017: Also mention that phrase accents may be secondarily associated with postnuclear prominences/stressed syllables but do not constitute fully-fledged pitch accents. It would also be interesting to see how you would represent phrase accents (in their prominence-lending function) in your metrical model/grid.

> *We mention this now, and refer to the final discussion, where we also added some discussion of this.*

p.20, l.1045: Which language are you referring to (citing Feldhausen 2016)?

> *Right! We now mention that this paper is about Spanish.*

p.20, l.1085 ff.: …is used less -> in comparison with what?

We clarified this

p.24ff.: It seems to be problematic to talk about "accuracy" between the annotations and the intended bracketing in cases where the speakers simply produced the stimuli in a different way. The wording suggests that there is one correct way of expressing a given structure. It would probably be better to talk about "matching" productions instead (also e.g. Table 4).

> *Yes, we use a different wording here, but now in the later tables, since we make it clear then whether they guessed the original accurately (we don't mean to imply that the prosody is accurate, just that the condition was guessed correctly).*

p.46, l.2528f.: However, in the random forest analysis for declaratives, the results for wide and third focus are only slightly above chance (Table 4) – interestingly, though, the difference is much bigger for the human annotators (0.14 vs. 0.66). Can you comment on this result?

> *Yes, we comment on it now. It seems that both humans and RF are often unable to tell the difference between third focus and wide focus, but humans have a bias to interpret both types of utterances as narrow focus on the third constituent.*

p.48: The part on overlay models remains too vague – given the way in which these models were introduced above. Some more information is needed here (another option would be to shorten the discussion on overlay models).

> *We clarified this a bit, but also relegated part of the discussion to a footnote.*

p.49, l.2731: Probably just a formal mistake but if not, it is content-related: Do you really compare 15-b with 16-b, not with 15-a? 'Lauren' is the focus of 16-b and thus more prominent than 'Lauren' in 15-b.

> *Yes, that was just an error*

p.50, discussion of example 16; p.57, l.3177 f. 'the intensity cues to focus prominence and phrasing-related prominence are very different': It seems that (high) intensity goes hand in hand with accentuation, for which F0 is the main cue, while phrasing is expressed by (increased) duration. In other words: Pitch movement and duration are the most important cues to prominence perception; accents (focus prominence) comprise *both* cues, final lengthening (phrasal prominence) is marked by duration only (would also explain the high numbers for third focus in Table 6, p.58: accent plus final lengthening on constituent C).
Couldn't these major cues be represented as levels in a grid, here in an example of right-branching initial focus?:

```
(MEGAN) (and Lauren or Morgan)
(  x                          )     pitch accent (F0 + intensity)
(  x    ) (               x   )     duration
( x    ) (      x    ) (     x   )
```

The second level marks final lengthening, the third adds pitch movement (and accompanying intensity). An extra level may be necessary to differentiate between prenuclear and nuclear pitch accents (and postnuclear prominences = phrase accents?).

> *Thanks for suggesting this possibility. The revised discussion discusses this possibility (and attributes the suggestion to the editors), and points out why this would leave some crucial questions open.*

p.51, ex. (17) and subsequent grids (18-20): The representations go back to the word level, conflating the effects of initial and final syllables again. Can you think of an elegant way to keep this information?

> *We now use metrical representations that include both syllables. Although Féry didn't represent this (her representation only went down to the word level), we now added a line for feet in the grid, but we explain that this is our addition.*

p.57, l.3164 ff.: too vague

> *We rewrote this part of the discussion.*

p.58, l.3200: differ from those of prominence -> accentual prominence?

> *fixed*

<u>Structure</u>

p.2, l.86-140: The two paragraphs could be merged, since they basically state the same thing.

> *done*

p.6, l.294ff.: This chapter ends too abruptly. Guide the reader through the structure of the whole paper.

> *done*

p.6, l.301: Change title to "Prosodic cues to focus (prominence)", in analogy to chapter 3. Or merge both chapters into one background chapter.

> *done*

p.11, l.600: Mention phrase accents (along the lines of **Grice**, Ladd and Arvaniti, 2000) here already?

> *done*

p.20, end of ch.3: There is no transition to ch.4.

> *fixed*

p.37, l.2027: passive effects during articulation -> physiological effects (?)

> *We used the term 'passive' since we saw it used on the literature on the correlation between intensity and pitch, but it's probably better in this context to use 'indirect' or 'secondary', so we replaced it.*

p.39, ch.6: This chapter has to be introduced earlier/better motivated – it feels like an add-on after the main part (i.e. the experiment) is over.

> *There is now more of a segue*

p.40: The introduction to random forests can be shortened.

> *We considered shortening it, but we think that it might be useful given the audience to spend some time on this (see also the positive comment by reviewer 4 about this discussion).*

Formal aspects

p.6, l.312: get rid of title for subsection (see remark on structure above)

> *We'd like to keep the subsection headings if possible since they give the discussion some more structure. The earlier version had a confusing structure because a subsection was accidentally formatted as a section.*

p.7, l.382: delete (5), and change numbers of subsequent examples

> *done*

p.21: change numbering to 4.1., 4.1.1. etc. in the whole chapter (i.e. also changing 4.1. to 4.2. etc.)

> *done*

p.26, Fig.3: increase the size of the figures, the text is not readable; adjust the caption

We increased the size, but this might be improvable in the final editing process

p.27, Fig.4 (also Figs. 5 & 6): use the same scalings for the initial and the final syllable

done

p.29, Table 1 (accordingly for intensity and F0 in Tables 2 & 3): Add to caption sth. like “values in ms, with standard deviations in brackets”

done

p.50, l.2770 ff.: missing and wrong references to figures and tables

done

p.51, ex. (17): 1. Refer to the two grids as (a) and (b) (same in subsequent examples). 2. Left- and right-branching mixed up? 3. Are you using “branching” and “bracketing” interchangeably?

fixed

p.51, l.2840: …more prominent than Megan in (?) because -> which example are you referring to?

fixed

p.53, ex. (19): Again, left- and right-branching mixed up? Where is the highest prominence in the grid? Repeat the test words in the example again, for ease of understanding.

Fixed (we added which words A, B, and C correspond to in the text)

p.53, l.2954 f.: …Lauren should be more prominent than Megan… -> the other way around?

Fixed (the grids were just swapped, as you noted above)

p.55, ex. (20): Once more, left- and right-branching mixed up? Beat on second level of B wrong?

fixed

In general, would you mind using “broad” instead of “wide” focus throughout the text? This term is commonly used in our Special Issue.

done

Note that the titles of the cross-references mentioned in the text have been updated. Please change to:

Jason Bishop, Grace Kuo, Boram Kim (this volume). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*.
Jennifer Cole, José I. Hualde, Caroline L. Smith, Christopher Eager, Timothy Mahrt, Ricardo Napoleão de Souza (this volume). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics.*
Language

done

Please check the whole paper thoroughly for typos (especially for mistakes in number agreement), including omissions and doublings.

We tried our best to find these

Additional cases:

Thanks for these—we fixed all of these issues!

-**Reviewer 1**

- The authors considered all points raised by my review thoroughly and reported their changes in great detail. In my view, the final version is a great contribution to the problem of the different dimensions of prosodic prominence and I will be very happy to see it published. No further comments/suggestions from my side.

-**Reviewer 4**

-
Review of "The Effect of Focus Prominence on Phrasing
This is a revised version of a manuscript that I reviewed before, yet the title has changed. The study investigates the factors speech act, focus and syntactic constituent structure on the prosodic phrasing of English utterances, in particular the effects of focus in the post-focal domain. The authors argue for a theory that keeps prominence and phrasing separate. Two of my major concerns were successfully revised. The issue of emphasis has become more clear in looking at focal prominence. The issue of measurements has been redone on the level of the syllable and the data presentation becomes much clearer showing the acoustic effects on the accented syllable and those on the phrase-final syllable as each factor contributes (partly differently) to the acoustic cues.

For the structure of the paper (my third concern), I think that the authors did a good job, which however could be improved on. Sections 2 and 3 give a comprehensive and clear discussion of phonological and phonetic cues to focus and phrasing. I feel that Section 2 and 3 could be renumbered as 1.1 and 1.2 under the heading “Introduction”, or as 2.1 and 2.2 under a heading “Background”. Similar, the discussion could be one major heading with subheadings.

In this revised version I have another concern that relates to the discussion. In the introduction, two theories are presented that would make (partly) contradicting predictions, AM-theory and overlay models. In the discussion the authors only briefly discuss their results in relation to overlay models (l. 2647ff). The authors state that “This finding is compatible and maybe even expected under overlay models”. However, as no real predictions were spelled out in the introduction it is hard to follow what the compatibility is between models, which kind of measures / cues have been shown to adhere to the models. The authors are more in favor of AM models, which is fine and it is also necessary to explore their findings in relation to AM theory. So I would recommend to revise the discussion that the relation to overlay models becomes clearer.
Second, the discussion about phrasing and abstract prominence is far too long and not really on the point, to my view. In line 3032ff there is basically no difference between (20b), right-branching initial focus by Wagner 2005, and (18b) right-branching initial focus by Féry 2013. As far as I can get the discussion around the two versions, they are alike. So the lengthy text should be reduced to a clear point that the results of the study would support. A similar discussion about abstract grid prominence and the post-focal prosodic effects in German was given in Kügler & Féry 2017 that the authors also (partly) discuss at different places. A point in Kügler & Féry’s paper was that abstract prominence would not predict the prosodic effects alone as pre-focal and post-focal grids are similar, and the pitch register differs as a function of focus position; the pitch register difference is not directly derived from grid prominences. To some extent, the finding for English and those of Kügler & Féry on German should be related, and the new proposal here, the difference between focal prominence and phrasal prominence, should be more on point.

We have substantially changed and clarified the discussion now.

In general, I have to say that I found it partly disturbing that the manuscript contained a number of typos, missing or wrong cross-references, wrong citation formats (I guess that is due to wrong latex \cite commands). The manuscript looks like if resubmitted in a rush without careful reading.

We tried our best to fix this in the new version!

A list of minor issues by line number:
68ff The variability of sentence prosody in relation to factors like speech act, constituent structure and focus may not hold for all languages. Especially, languages like English appear to allow variation in constituent structure, other languages like Akan for instance do not allow such variation.

100 argue -> argues

fixed

100 the term “grouping” should be briefly explained here for all readers not familiar with Xu’s terminology (may be in parenthesis), e.g. move parenthesis from l. 139f up.

99ff and Figure 1 I wonder whether it is really necessary to illustrate the Penta model (for reasons of space, one could eliminate Figure 1.). Rather name Penta model in the text, i.e. “and Xu (2005) argues in his Penta model that focus and grouping …” The lacking prediction of focus location and phrasing is good!

We’d like to leave it in since many people are not familiar with this kind of model, and we found it helpful when we started thinking about this.

103 is -> are

fixed

103ff The sentence is unclear, what is the compatibility of overlay and interaction between functions? You would rather want to say something like: Given that different factors interact in shaping sentence prosody, overly models do not predict these interactions but simply show additive aggregate effects.
Yet, I am not sure if Xu would subscribe to this view. Of course, the different factors in the Penta model are separated yet they are not simply additive…

We clarified this part of the discussion now

148 remove one ‘)’ after Ladd (2008)
fixed

Fig 2 As with Fig 1 I wonder whether it could be removed for reasons of space. The prosodic hierarchy could then simply be explained in a sentence.

We decided to keep it, since we think it helps clarify the idea behind the model.

245 Move the punctuation character ‘.’ to the previous line

fixed

263 “focus-related emphasis” -> focus-related prominence?

fixed

295 Up to this part, the interaction of focus and constituent structure is motivated such that the investigation of phrasing and prominence of post-focal constituents can be done. Yet, no motivation for the factor *speech act* is given, which comes in l. 296. So before this paragraph, it might be necessary to contrast (1) with a corresponding declarative to illustrate different expectations around speech act.

> *We decided to keep this, but updated the figure in order to better anticipate the later discussion.*

378 (2) must be (4)

> *done*

382 remove (5) here

> *done*

403 / 405 (5) should be (4)

> *done*

413 (6) then is (5) – all cross-references need to be checked!

> *done*

518 add Baumann (2016) on second occurrence focus as a reference here as well.

> *done*

1049 Remove one 'the'

> *done*

1090 check sentence 'if' -> is ? 'consider' -> considered?

> *done*

1190 check sentence

> *done*

1305 exclusion of stimuli evenly distributed – make reference to some table of distribution here (it was mentioned in the response letter that such a table exists in supplementary materials)

> *This information was already in Table 7 in the appendix (which we assume will be handled as supplementary material)!*

1371 ‘using through Praat’s F0 analysis’ – delete ‘using’

> *done*

Fig 3 the individual plots need to be larger, use the space of the whole line width for each figure, otherwise the figures are very hard to read

> *done*

1492 ‘closes tracks’ ?

> *fixed*

1532 “compared to the final syllable of the final syllable of the last name” ? final syllable twice?

> *fixed*

1528ff The striking pattern of lesser final lengthening of the last name is not that striking if considering the phrasing: the final name is not phrase final, is it? It follows a verb, so why should the final syllable be lengthened?

> *Right, we dropped this paragraph now, it didn’t make much sense as it was.*

1552 ‘seems’ -> seem?

> *fixed*

Table 1 The abbreviations ‘First, late, second, late, …’ need clarification.
Could you use the coding as used before, i.e. (AB)C / A(BC), Decl / Interr, Focus: wide, first, second, third?
Also, which numbers are given?
A straightforward way would be to report the complete numbers of the model output.

> *We give estimates and standard errors (which is fairly standard), and make this explicit now.*

1723f delete one ‘better’

> *fixed*

1771 ‘phrase’ -> phrased

fixed

1771f delete one ‘compared’

fixed

1889 ‘It would seem like …’ I don’t get this sentence

fixed

1934 The sentence “The intensity of both” lacks a verb

fixed

2097 ‘They suggest that focus and phrasing are, for the most part are encoded independently of each other’ this seems to be evidence in favor of overlay models instead of phonological AM-models?

fixed

2159 ‘being are’ delete are?

fixed

2166 ‘this methods’ – this method

fixed

2225 Reference, check citing form (check through the whole document, there are many incorrect citation forms

fixed

Although I never run a random forest analysis I feel that the authors introduced the relevant parameters well in order to follow their procedure. The motivation to use that method is clear. I leave it to the editors to judge the quality of the analysis.

2320ff Add the abbreviations in the Table caption, i.e. e.g. ‘declarative data (-Dec)’ and mention what R.F. means.

done

Classification of focus: The random forest analysis reveals similar focus identification rates as the human annotators, earlier focus is better recognized than later focus. This is an interesting finding which should be discussed. For instance, focus identification in perception studies also revealed better focus identification for earlier focus than for a sentence-final focus (e.g. Botinis et al. (1999), Xu et al. (2012), Vainio & Järvikivi (2006)). In sentence-final position the prosodic cues for focus seem to be reduced or cancelled out.

> *We added two of these references to the Gussenhoven reference we had already given in the context of discussing this finding in the discussion section. The Vainio & Järvikivi (2006) article is also interesting, but doesn't seem to make the point at hand.*

Fig. 7 I would recommend to reduce the scale of the x-axis. If keeping it comparable with following figs, may be a level of 0.8 would do?

> *done*

2373 I miss a presentation of the interrogative data here. It is only said that F0-ini and duration ini + final are the best predictors. This is only true for declaratives, not so for interrogatives – if I interpret Fig 7, right, correctly. This difference needs some presentation and later on discussion.

> *We now added a paragraph.*

2401 'annotator' – annotators

> *fixed*

2416 'The last data set' This is misleading, rather say "The latter data set"

2431 This paragraph needs revision. First, the human annotators did overall better than random forests (0.7 vs. 0.6). Second, given this distinction the comparisons between focus conditions were comparable, also for initial focus.

> *We added discussion of the human annotators.*

2450 'for focus for' delete for

> *fixed*

2450ff Discussion: the results for focus in questions may be to some extent due to the fact the F0-Max is considered. Phonologically, the authors are aware of the fact that question are realized with rising pitch accents where the low part may be the accentual tone (L*). In that case, it might well be that focus affects this part of the accent and that F0-max as a cue for the

H trailing tone is not affected. The question is whether listeners did well identifying focus in questions? If that identification would be better than random forests then there might be some information in the signal that is not captured by these measures.

> *True, but the human annotators were not (much) better than the random forests, which leads us to believe that we didn't simply use the wrong acoustic cues. We make this clearer now.*

2496ff relate the discussion to other perception studies on early vs. late focus (see comment above)

> *done*

2531ff move this paragraph up to the previous one to l. 2495, for coherence to continue the discussion on interrogatives.

> *done*

2568 Remembr -> Remember

> *fixed*

2564 'phrasing information remains intact.' Here, add the evidence for this claim, relate this to the random forest variable in which figure this can be seen.

> *done*

2599f 'This model actually classified 86% of the utterances correctly for phrasing for the out-of-bag predictions.' The same addition needs to be added here, based on which variable were the 86% utterances classified correctly?

> *We made this clearer, and added a plot with the importance of the various sacoustic predictors for the additional model in the appendix.*

2609 The conclusion given here that phrasing is still detectable post-focally should be related to the other studies mentioned in the introduction (e.g. Norcliffe & Jaeger, Ishihara, Sugahara, Jun & Fougeron, Kügler & Féry).

> *We cite Norcliffe & Jaeger here, and then the other articles just a paragraph later.*

2652 Even here, add these references as this finding is not completely new (given the earlier studies on different languages). It is new for English.

> *done*

2675 The discussion on which model captures the data is interesting and relevant. See the discussion in Kügler & Féry 2017 who also assume a phonological mediation of phonetic effects.

> *The discussion has been reorganized a bit, but we cite this article now also in this context.*

2770, 2791, 2809 and elsewhere: missing cross-references

> *fixed*

2826 Represenation -> Representation

> *fixed*

2840 in – missing cross-reference

> *fixed*

References

Botinis, Antonis, Marios Fourakis & Barbara Gawronska. 1999. Focus identification in English, Greek and Swedish. In John J. Ohala, Y. Hasegawa, Manjari Ohala, D. Granville & A. C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, 1557–1560. San Francisco: University of California.

Vainio, Martti & Juhani Järvikivi. 2006. Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics* 34(3). 319–342.

Xu, Yi, Szu-Wei Chen & Bei Wang. 2012. Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review* 29. 131–147.