

Response to revision

2019/05/22

The following paragraphs report the reviewers' texts and the author's responses. Text in italics is from the manuscript prior to revision. The author's responses with details on how the recommended changes have been implemented are given in light purple. I would like to thank the anonymous reviewers for their helpful feedback.

Reviewer A

Specific line-by-line comments on details of the paper. Begin each comment with a page number, example number or paragraph/line from top or bottom, as appropriate.: Be careful with the use of 'such' in English. There are many instances in the paper where it should be followed by an indefinite article: p2, line 40; p4, line 143, p19 line 452; p 21 line 494, p26 line 604, p30 line 749;

Also, in describing closure duration results in lines 431-436, I think the author mixed up voiced and voiceless. Finally, p 22 line 554, change "As it can be seen" to "As can be seen"

The typos and errors have now been corrected. References to voiceless and voiced closure durations in lines 431-436 were the other way around, and have been now reversed to the correct order.

Reviewer B

Most of the concerns raised in my previous review have been adequately addressed with an exception of an important one: the paper draws a too strong conclusion in favor of temporal adjustment hypothesis given the data while dismissing other possibilities with subjective language such as "seems more natural" and "what it is claimed here". Two paragraphs in the discussion section need revision due to this subjective language and speculative hypotheses.

Moreover, I would like to add that, if misplacement of the V-C boundary is due to random error (which is a neutral assumption to make), the measured displacement from the 'actual' boundary will approximately follow a normal distribution with mean 0, so that negative and positive misplacements would cancel each other out.

This is not true. Even though the misplacement is due to random error, the negative and positive misplacements would not cancel each other out, because the vowel vs. closure measurements are not independent. If only one annotator annotates the data, the measurements are dependent and the results would not cancel each other out, but would result in negative correlation.

The reviewer is correct, even with normally distributed 'misplacements' the fact that the boundary is shared would result in a negative correlation. This sentence has been removed and the following statement has been added in lines 451-453: 'Note, moreover, that the negative correlation found here could be a consequence of annotation bias, since the vowel and closure share a boundary.'

However, what it is claimed here is that the stability of the interval is a logical antecedent, while the different CV of the release to release duration is 0.203, while that of the vowel onset to release duration is 0.232. The CQV is 0.127 for the release to release and 0.136 for the vowel onset to release. Lower values mean less dispersion/more cohesion.

And "This interpretation 635 seems more natural in light of accounts of gestural phasing like the one 636 proposed"

It is not relevant what authors claim. And what "seems more natural". The authors should avoid such subjective judgments and stick to empirical facts. It is true that the results are compatible with the compensatory temporal adjustment, and the body of the paper makes only this claim. The discussion section,

however, is too strong in its conclusions. The results are also compatible with the effects of the laryngeal features. A possible articulatory explanation (Tilsen 2013) is not evidence in favor of the proposal, because the laryngeal feature effect has also a possible articulatory explanation which the authors do not mention.

I agree that the wording is too strong. I have rephrased the ideas presented in section 4.2 (Compensatory temporal adjustment, especially the second part) to emphasise that other accounts cannot be ruled out, and that independent explanations are needed for the timing of the vowel offset/closure onset. I have removed subjective and ambiguous language. Possible articulatory mechanisms behind the laryngeal features effect are also now mentioned.

Even when multicollinearity between predictors is minimal, statistical significance of multiple terms cannot unequivocally inform us on the actual contribution of those terms, since it is possible that unknown relations between terms mask underlying mechanisms (for a discussion see McElreath 2015). For example, it is possible that, through time, different phonation categories (like ejective, voiceless, and voiced stops) can develop different closure duration sub-distributions (Sóskuthy 2013). This can result in the ability for the phonation predictor to capture variance in vowel durations, even if the original mechanism directly involved closure durations only

This is true, but presence of significant predictors is still more informative than the absence of any empirical evidence.

This problem is not an argument against the laryngeal features effect either. The same problem can work in the opposite direction. Why can't the same argument go in the opposite direction, where initially the "causal" effect is only that of laryngeal features? This is in fact what many scholars believe, where secondary durational differences arise due to perception. This paragraph is based on pure speculation without any empirical evidence. If the authors have any empirical evidence, please state so.

But why in the expected direction? This is not very elaborate (beyond being a possibility).

I agree that the statistical argument goes in both directions. I'd like to thank the reviewer to have pointed this out. Lines 661-686 now state that diachronic patterns can be masked by synchronic statistical patterns, and that this does not preclude the existence of a separate laryngeal features effect. It is true that the present study does not exclude the possibility of a separate laryngeal features effect, and articulatory mechanisms proposed to be behind the laryngeal features effect, like the ones summarised in Beguš 2017, would still be required to explain the differential timing of the VC boundary within the release to release interval, as I now explain in lines 638-657. In line with this more neutral and less strong view, lines 565-583 have been reformulated to propose a compensatory account in which the timing of the VC boundary within the release to release interval is the focus, rather than the duration of stop closure per se. Perceptual mechanisms that can produce secondary effects via enhancement are now referred to in lines 654-658. Section 4.2 concludes by reiterating that even in the present data a laryngeal effect could coexist with a compensatory mechanism.

Abstract:

While other factors (like perceptual biases and laryngeal effects) could also play a role in the development of the voicing effect, the data discussed here sheds new light on a possible production account of voicing-related differences in vowel durations.

Laryngeal effect is also compatible with a production account, not only the temporal compensation approach.

I reworded the abstract to clarify this point. The abstract now says 'The durational difference of the first vowel and the stop closure would then follow from differences in timing of the VC boundary within this interval. While other aspects, like production mechanisms related to laryngeal features effects and perceptual biases cannot be ruled out, the data discussed here are compatible with a production account based on compensatory mechanisms.'

Conclusion

The conclusion is adequate.

Such pluralist view has already been proposed, for example, for incomplete neutralisation in Winter and

Rottger (for a review of explanatory pluralism in the cognitive sciences, see Dale et al. 2009 and references therein).

Pluralist views are already proposed for the very problem discussed, not only for incomplete neutralization or in general in cognitive science.

I have now clarified this point by adding a mention to Beguš 2017 and Sanker 2018, who among others argue for a multi-factorial account of the voicing effect.