

## Author response to reviews of

Manuscript APS-Feb-22-0015

### Using intonation to disambiguate meaning: The role of empathy and proficiency in L2 perceptual development

Joseph V. Casillas

submitted to *Applied Psycholinguistics*

---

**RC: Reviewer Comment**   AR: Author Response   ☐ Manuscript text

Dear XXX,

Thank you for taking the time to consider our manuscript *Using intonation to disambiguate meaning: The role of empathy and proficiency in L2 perceptual development* (APS-Feb-22-0015) for publication in *Applied Psycholinguistics*. Our understanding is that we needed to consider the reviewers' comments and revise accordingly before the manuscript could be reevaluated for publication. We have considered thoroughly the detailed feedback provided by all three reviewers and resubmit what we believe to be a much improved version of the manuscript. In this letter we address the reviewers' concerns point-by-point. Where feasible, we quote all revised text in this document, otherwise, we refer the reader to the relevant lines in the revised manuscript. We thank you and the 3 anonymous reviewers for all comments and suggestions and again enthusiastically submit the revised manuscript for consideration in *Applied Psycholinguistics*.

Sincerely,

Joseph Casillas

(corresponding author)

## 1. Reviewer #1

Reviewer: 1

Comments to the Author INTRODUCTION AND OBJECTIVES

- The terms perception, processing and comprehension are used in the manuscript not very systematically. In some cases the authors use “perception and processing” together, and in some others they seem to summarize them both in the term “comprehension”. A more systematic use of these terms is needed.
- The authors used distinct terms to refer to the same sentence type: absolute interrogative, polar question, yes/no question, total interrogative... It also alternates between the term ‘statement’ and the term ‘declarative’. It recommends using one term in a systematic way and to describe and exemplify what these terms mean for readers that are not experts on intonation (or syntax or pragmatics). The authors now provide some general description of the intonational patterns in the introduction (e.g. ‘nuclear hat pattern’, ‘final rise’, but I doubt whether a reader that is not an expert on intonation will understand these descriptions).
- The introduction mentions several times that “intonation may result in comprehension and communication difficulties” and that “interpreting L2 intonation is challenging”, and it feels a bit repetitive. These statements should also be better motivated and better justified (intonation more than other linguistic aspects? which kind of difficulties? what does ‘challenging’ mean? why is it challenging? all intonation patterns work equally?).
- The factor “speaker dialectal variety” seems to appear and disappear from the manuscript. The introduction includes a paragraph on dialectal variance in Spanish intonation, and it is included as a third research question, but the authors should elaborate further on why dialectal variance may affect the processing of L2 intonation or even if it may affect learning. In the discussion and conclusion this factor is not mentioned when the authors summarize the aims of the study at the beginning of these sections. In sum, one feels that this research question is not well integrated into the study (or at least the manuscript).
- The study explores 4 specific sentence types that differ at the intonation level. The authors should motivate the selection of these sentence types, describe them at the intonational level for all varieties used, and elaborate on why certain sentence types are more difficult to process for L2 learners and, more specifically, for English learners of Spanish. Otherwise it is not possible to understand the nature of the differences that had to be perceived by the participants of the study and that, more generally, have to be learned in an L2.
- Also, why empathy should play a role in how these specific sentence types are perceived and processed? Why do the authors think that the listeners’ ability to feel/think what the other feel/think may impact in distinguishing questions from statements? In other words: what are the intonational cues in a declarative or in a yes/no question that may lead to distinct processing accuracy and speech by listeners with distinct empathy skills? Please motivate these aspects further.
- Please elaborate and motivate further why you think that empathy may lead to higher/faster intonational development in an L2 (expectation for RQ2). Likewise, please motivate further your expectations for RQ3, as they are now somehow disconnected from the literature reviewed in the introduction. The authors mention that statements might be more difficult to differentiate from questions in the Cuban

variety, but without information on why this must be the case (a distinct intonational pattern for certain questions, maybe?) it is impossible to evaluate whether this expectation makes sense.

## STIMULI

- Details on the stimuli used for the experimental task are much needed. Please describe the sentences used at the acoustic and phonological level, for all varieties, and provide a list of these sentences. Importantly, were they controlled at the syntactic and lexical levels to ensure that potential differences in processing speed were not due to syntactic or lexical factors? On a related note, and crucially to interpret your results, how can we be sure that listeners responded based on the intonational cues they perceived and not on potential syntactic and lexical cues in the sentences?
- Please provide the results of the 100 monolingual Spanish speakers rating the quality of items.
- Please provide details on which variety the listeners were familiar with, and put these data in relation to the participants' responses in the results in a more systematic way.

## RESULTS & DISCUSSION

- Results show that proficiency and empathy interact in wh-questions but not in yes/no questions or declaratives, and that empathy had an effect in all sentence types except for yes/no question. There is no discussion on how these findings relate to previous literature on sentence comprehension, or even on what may originate these differences across sentence types. Instead, the discussion and conclusion now seem to mask these differences across sentence types.
- Could you check whether individuals with higher empathy were not the ones with higher proficiency? How did these two factors correlate?
- In the discussion the authors highlight the fact that proficiency was treated as a continuous variable, and the reader would benefit from knowing whether this is actually an innovation of the present study or if, instead, other studies have done it before (now it seems that it is an innovation).
- The authors try to explain the variety-specific responses by providing three hypotheses. However, I think that none of these hypotheses are sufficiently motivated or explained. First, they state that "familiarity with the target variety may account for variety-specific response accuracy", although this cannot be derived from the results you obtained since most of your listeners were most familiar with US Spanish (which was not in the stimuli). Second, they state that they may derive from the fact that distinct varieties produce each sentence type using distinct intonational patterns, although it is impossible to evaluate if this hypothesis is plausible because details on the intonational features of the stimuli are missing and because results are never matched to the nature of the contour that was perceived. Finally, the authors propose an explanation linked to the distinct speech rate used in distinct varieties. The authors should provide a detailed description of the speech rate of the stimuli, by variety, in order for the reader to be able to evaluate if this explanation is indeed plausible. All in all, I recommend the authors to adjust the discussion of the effect of listener/speaker variety to the results that were actually obtained, and to provide enough details of the stimuli to ensure the reader can assess the validity of these explanations.

## 2. Reviewer #2

Reviewer: 2

Comments to the Author This is a very interesting study looking at the roles of proficiency and empathy in the development of L2 prosodic perception. The authors use sophisticated statistical methods to show that both proficiency and empathy seem to be at play in the development of L2 prosodic perception, as well as the specific type of prosody. The paper also considers the role of dialect, which is very often not dealt with in studies of L2 prosody. In paying attention to these aspects of prosodic acquisition, the paper is cutting edge. I think that the paper would benefit from a clear theoretical model with specific predictions based on that model, and perhaps the authors can investigate whether Mennen's LiLT model would help them do that or not. Below I present some general recommendations, followed by some specific comments.

Recommendations:

I would like to see more discussion about proficiency which as I understand was really about vocabulary. What do the authors think the relationship between vocabulary size and prosodic comprehension is? It would be interesting to see more discussion of this in the paper, since the authors made a decision to measure "proficiency" in this way. Is it that learners might be acquiring more "things" overall? More words but also more tunes? This could be explored further.

As mentioned in the detailed comments, I don't think it makes sense to conflate empathy and "pragmatic skills". I know the authors claim that they are "operationalizing" empathy as pragmatic skills, but I don't see a need to do this. While being highly empathetic might and probably does play a role in one's pragmatic skills, I don't think it makes sense to say they are the same thing. I think the authors can simply talk about empathy and leave it there, which is what they are doing in the analysis anyway.

This is also mentioned in the more detailed comments, but I would like to see some sort of description of the tunes (at least the nuclear configurations) for each variety and sentence type. At the very least a very phonetic description would do (e.g. a rise to a high tone in the nuclear stressed vowel, followed by a fall to a low boundary tone). But I think this would make for a better discussion of the findings. This might also help the authors to make sense of the finding about wh-questions and the role of proficiency and empathy for those. Why wh-questions?

While the paper points out the need for L2 models to account for levels further than the segment, the paper doesn't make any suggestions about how to do so and does not mention the most recent proposal of a model for L2 intonation. Please check the following for the model more generally:

Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussaire, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact* (pp. 171–188). Heidelberg, Germany: Springer

And specifically for the case of Spanish:

Sánchez Alvarado, Covadonga. 2020. The production and perception of subject focus prosody in L2 Spanish. University of Massachusetts Amherst dissertation.

Alvarado, Covadonga Sánchez and Armstrong, Meghan. "Prosodic Marking of Object Focus in L2 Spanish" *Studies in Hispanic and Lusophone Linguistics*, vol. 15, no. 1, 2022, pp. 211-250. <https://www.degruyter.com/document/doi/10.1515/shll-2022-2060/html>

p. 3, line 23-29 Please also include literature on information structure (focus, givenness, etc.) as well as

speaker belief states

p. 4, 47-50

I think it could be useful to include some concrete examples of the various meanings here, to demonstrate the plethora of meanings that can arise

p. 5, 23-29 Also reference the discussion regarding the instruction of prosody in Durwing & Monroe 2015

p. 6, paragraph beginning on line 47 - there should be some mention of Mennen's LiLT (L2 intonational learning theory) model of L2 intonation and the recent work using this model

p. 6 line 39 - both Caribbean Spanish and Argentine (porteño?) Spanish use final f0 falls for yes-no questions (absolute interrogatives), I'm not sure this description actually captures the differences. Also, there are differences in the actual falls in different Caribbean Spanishes. Varieties of DR Spanish use what is best described as H+L\* L% while Puerto Rican Spanish uses ;H\* L%. I think Hualde & Prieto also shows that question falls are actually rather common in Spanish, even if their pragmatic meanings are more restricted.

The author(s) might consider using the term polar or yes/no questions throughout, since these terms are more accessible outside the Hispanic Linguistics lit, which tends to use "absolute interrogatives".

p. 9 line 4 - has empathy been "operationalized as pragmatic skill" in the past? Is it really possible to say that they are the same thing, vs. one leading to the other?

p. 10 - line 45 - perceptual—> perceptual??

research question 2 - Again, it's hard for me to buy empathy AS a pragmatic skill itself. I would say it's an ability that could be helpful for pragmatic skills/pragmatic reasoning. But not the same thing. If the authors disagree I'd like to see this justified more. Given what the study does, I'm not sure you have to say anything about pragmatic skills at all, you can just call it empathy

p. 11, line 8, again, why not just call them yes-no questions?

p. 15, line 25 - I'm wondering what perceptual learning would mean here. We know that the encoding of narrow focus and question intonation differs across the varieties that are listed here. How would we know if these participants had experience with these varieties? It would be helpful to know what the tunes for each type of sentence were for each variety used in the stimuli as well.

p. 26, lines 53-6(on next page)- I'm not sure this makes sense" In practical terms, this implies that high proficiency, high empathy learners required more information to reach a decision and responded at a slower rate, particularly with regard to low empathy learners (top row), regardless of proficiency level."

Do they mean "compared" to low empathy learners (as opposed to "particularly with regard to low empathy learners")?

p. 29, line 28 - "This is taken as evidence suggesting that pragmatic skill can modulate the rate of development in L2 prosody." again, I don't see a reason to call this pragmatic skill. It's just empathy. better to say - "That is to say, higher empathy individuals may develop L2 prosody at an earlier stage than lower empathy individuals."

p. 31 - line 5 - The idea that there is one "US Spanish" is very controversial. US Spanishes still have the characteristics of the varieties of origin, so even though respondents are familiar with Spanish spoken in the U.S. they would be exposed to Mexican Spanish, Dominican Spanish, Puerto Rican Spanish, etc. as spoken in

the US. Looking at which region of the country respondents were from could be helpful in understanding this.

p. 31 - line 20 - The authors discuss the possibilities of why the different varieties might have been perceived differently. I think it would be helpful for the authors to share what the tunes in their stimuli actually were, which would help them to better surmise here. Puerto Rican Spanish (not mentioned in the discussion) and Cuban Spanish were the varieties people had the most difficulties with, but according to most descriptions PRS and CS use the same tune for ys questions (also again, NB that Dominican ynqs are not the same as PRS and CS ynqs, this description conflates all Caribbean Spanish questions, which are also not necessarily hat patterns). I would recommend an appendix with a description of the tunes in the stimuli for each variety, and for this to be brought into the discussion for research question #3.

p. 32, line 7 - again, the LiLT model needs to be mentioned, since this reads as if no theory of L2 intonation has been proposed.

### 3. Reviewer #3

Reviewer: 3

Comments to the Author The study analyses the influence of empathy in the perception and processing of intonation in questions and statements in L2 Spanish and shows that higher empathic individuals, in comparison with lower empathic individuals, appear to be more sensitive to intonation cues. In my opinion, the quality of the paper is very good and the conclusions make it a highly citable one and not only for the case of Spanish L2 but in general.

Below you can find a by-section review and some comments.

The goals and hypothesis of the paper are clear and well-defined.

As for the state of the art, page 29, line 6 states: “the current body of research is limited to studies on pronunciation accuracy.” And I think that is not accurate, there are studies that include empathy as an important factor for intonation perception (in L1 and mostly from psychology that is true) <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008759>

And, then, specifically for Spanish L2, a recent poster presented in New Sounds 2022 uses the Autism Spectrum Quotient (I don't know if they publish abstracts as short papers): [https://www.researchgate.net/publication/361371026\\_The\\_Role\\_of\\_Native\\_Language\\_Experience\\_and\\_Individual\\_Features\\_in\\_the\\_Cross-linguistic\\_Perception\\_of\\_Spanish\\_Intonation](https://www.researchgate.net/publication/361371026_The_Role_of_Native_Language_Experience_and_Individual_Features_in_the_Cross-linguistic_Perception_of_Spanish_Intonation)

Maybe you can add this info in the introduction stating that it is a very recent and promising field of research and that it helps us to understand individual variation and that way you won't need to change much in your text.

The methodology of the paper is outstanding, a great sample size (225 listeners), good materials (with several Spanish varieties included) and a state-of-the-art statistical analysis for fussy reviewers and an innovative one (at least for our discipline), that I don't fully understand, so maybe it is a good idea to organize a workshop on that in a future congress.

I have only have questions regarding those “64 critical items” that (I think) can be easily solved including examples or a description of the patterns the speakers used or better (and you will see why in the comments) how you elicited the data and saying something like “variation is wonderful, students must be exposed to a wide range of intonational patterns and we are not interested in their pragmatic meaning for this study”.

First, silly question, why do the listeners need intonation in wh- questions? They have the marker “qué, quién, cómo”... That is all they need, so I don't understand why this type of question is included in the study.

I know that the stimuli are in the supplementary materials, but I miss something in the text either an example (or two of them) or making explicit of the type of stress that you have included and why. The reason is that a lot of foreign speakers (that is the case for example for Dutch, Japanese and Chinese if I remember it correctly) find more difficult to distinguish between statements and questions when the last word of the sentence is stressed in the last syllable (aguda). And that has two main reasons, 1) compression and truncation of the tonal movements and 2) (and of special importance given the varieties included among your stimuli) most of Mexican, Peninsular Spanish, Chilean, Peruvian statements can be slightly rising. In pragmatic descriptions such as Prieto & Roseano (2010) that final contour is characterised as “statement of the obvious” but, in fact, that is the pattern that we find most of the times that we elicit statements that have the slightest hint of common ground information, for example, in narratives. Moreover, in some varieties of Mexico and

Chile, slightly rising (L\* !H%) is the standard neutral pattern. This has a direct implication in the perception of questions by native speakers that counterintuitively, need a higher F0 rise in order to say that something is a question than foreigners that speak usually speak languages where there are not rising statements (I would give you references but they are mine, so maybe when the paper is published, we can talk in a congress). Maybe you don't find that because English-speaking speakers are used to uptalk?

Before I start criticizing your stimuli in the follow paragraphs, I want you to know that I very much appreciate that you made them public and that if every researcher did this we will find problems in every dataset. So, this does not diminish your work at all. Now, some of the things that I am going to say are not solvable in a reviewed version and I am aware of that, and I do not expect you to solve them, the paper is publishable as long as you include the sentence that I said before saying that the sentences have several contours and several pragmatic meanings.

Firstly, eliciting real broad-focus sentences is not easy and not only because the tendency to the statement of the obvious, for example, your andalusian\_match\_declarative-broad-focus\_David-leia-el-libro has a slightly (but audible and present in the pitch contour) narrow-focus intonation (rising stressed syllable) and the only narrow-focus that I could find among all your stimuli. So, maybe we are calling narrow-focus to different things, let me know if we are.

As for the wh- questions, Spanish has loads of patterns, for example in Argentina wh- questions can be rising or falling depending on politeness, in Peninsular Spanish are usually falling but they can be rising, and usually is the same for any dialect. And, then we have all the biased intonation patterns, for example, your Andalusian speaker is using the “surprised one” L+HL%, *which happens to coincide with the pattern for focus, in all the wh- questions except in “andalusian\_match\_interrogative-partial-wh\_Por-que-ama-la-navidad”, where she uses the LH%*, the prototypical one for yes-no questions in standard Castilian Spanish.

For the rest, most of your speakers have chosen rising patterns for their wh-questions but that is not the most frequent pattern.

As for focus, all your speakers should be doing a rising stressed syllable and falling ending (L+H\* L%), that is pretty much panhispanic and crosslinguistically really frequent. But they are not, the case of the Peruvian speaker is salient because is using a monotonous tone, just the opposite of what she should be doing.

In the case of the Argentinian speaker if you look at the sentences side by side you will see that in the supposedly narrow sentence the peak is a bit earlier but it is still in the post-stressed syllable. We will expect here the Argentinian narrow-focus pattern, which is the most recognizable thing of their accent and the only case in which Spanish has a tritonal accent L+H\*+L (peak in the middle of the stressed syllable surrounded by low tones) .

And then your Andalusian speaker has been innovative and has chosen for most of the cases the statement of the obvious pattern (realised as L\*H%) therefore using the y-n question pattern. The context that Prieto and Roseano (2010) have for that is: A: ¿De quién es el hijo? B: ¡De quién va a ser! de Guillermo So for your data would be something like: A: María bebe el vino. (while she is thinking: obvio, ¿por qué me lo preguntas?)

In the rest of them I think that she is trying to put the focus on the verb, but the result is really strange. To do that, you would need something like “Mariano HAbLa | del tiempo (and does not another thing)” and you will have a small pause with compressed post-focal material etc. What she does is just weird.

And lastly, yes-no questions, in general rising (here the Andalusian speaker is using the Madrid standard Spanish), but falling in Cuba, Puerto Rico, all the Caribbean, North of Spain, Canary Islands. . . The good thing is that the falling patterns that you have are not easily mistaken with a statement (for example questions



of Medellín (Colombia) are) and the bad thing again is that you have more than one pattern.

Ok, so the message after this is that the choice of a pattern by the speakers has made that some of the stimuli are more difficult than others and that maybe you could solve that by adding a new variable “pattern used”, which for most of the cases will be the interaction between variety and utterance type but for those speakers that use different patterns it won’t. In fact, if you redo the analysis in order to see the effect of the contour, I would use this variable instead of utterance type because most of the supposedly narrow-focus sentences are broad-focus and will share the L\*L% contour. And again, I am not saying that you do that for the current paper, you can prepare a new experiment to focus on intonational patterns.

Results. Nothing to say, beautifully solved, I loved the inclusion of speech rate as a variable that could affect the results.

Congratulations on your work and I hope to see the reviewed version soon.

## **4. References**