

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

1.1. INTRODUCTION AND OBJECTIVES

RC: 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.

AR: We thank the reviewer for their comment. In the revised manuscript we have tried to be consistent with how the terms are implemented in the phonetics/phonology, SLA, and psycholinguistic literatures. We reserve the term ‘comprehension’ to refer to one’s general understanding of linguistic input (see Esteve-Gibert et al., 2020 for examples). In the majority of cases we have simplified the revised manuscript by opting for ‘perception’/‘speech perception’, which we use to refer to the ability to perceive linguistic structure in the speech signal. We make a distinction between general ‘perception’ and ‘sentence processing’ when we refer to comprehension at the sentence-level and specifically to insight derived from online tasks (reaction times, eye-tracking gaze data, etc.) as opposed to response accuracy. We believe these distinctions have been made more explicit throughout the revised manuscript. We have opted not to explicitly operationalize these terms in the prose, as we believe they are generally understood by the readership of Applied Psycholinguistics, though we are happy to be more explicit at the discretion of the editor.

RC: 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).

AR: This observation was also pointed out by reviewer 2 and duly noted. We have systematized our use of these terms in the revised manuscript. When we refer to questions (generally), we now say interrogative utterances, and, more specifically, ‘yes/no questions’ and ‘wh- questions’ instead of ‘absolute interrogatives’ and ‘partial interrogatives’, respectively. When we refer to declarative utterances (generally) we now say ‘declarative statements’, and, more specifically, ‘broad focus statements’ and ‘narrow focus statements’. These changes are also reflected in all plots and tables, as well as the supplementary materials.

RC: 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?).

AR: We thank the reviewer for their suggestion. In the revised manuscript we have reduced the amount of times this point is emphasized to avoid repetition and in the referred text we have avoided describing the learning process as “difficult” or “challenging” because at this juncture of the introduction we are not interested in comparing the relative difficulty of different linguistic elements, nor explaining how different intonation patterns might cause more or less problems. Instead, our aim is illustrate that semantic meaning mapping is language-specific and that L2 learners typically have non-target-like outcomes due to cross-linguistic influence. We believe we have made this point more clear in the revised manuscript. We include the relevant text below for convenience.

This is, in part, because in everyday discourse speakers can use intonation to indicate syntactic structure, to signal whether an utterance is a question or a statement, to focus constituents, as well as to convey affective meaning. Notably, the manner in which intonation is mapped to meaning is often language-specific. As a consequence, L2 intonation is often produced in a non-target-like fashion due to cross-linguistic influence.

Intonation has a semantic function and through adequate cognitive decoding of the signal a listener can interpret the intended meaning of a given utterance. For example, an intonational contour can indicate to a listener whether the utterance of an interlocutor is a question or a statement. As touched upon above, a speaker can use prosody to signal numerous additional pragmatic functions as well. This rich variation in pragmatic uses makes the interpretation and decoding of intonational contours during speech comprehension a non-trivial task for the language learner. Moreover, the use of first language (L1) prosodic features when speaking the target language can result in misunderstandings because the same prosodic features can convey different linguistic and paralinguistic meaning in the target language (Chen, 2005; Cruz-Ferreira, 1987; Mennen, 2007; Pickering, 2001). As noted by Levis (2016), prosody is also “[...] critical for L2 pronunciation because it plays a major role in cementing social bonds as a key marker of social identity” (p. 154).

- RC:** 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).
- AR:** In the revised manuscript the topic of speaker variety is reported in the results section on pages 25-26 and we dedicate 4 paragraphs of the discussion section to possible explanations for our findings (approx. pp. 32-34). We hope that the reviewer finds this to be sufficient. Furthermore, at the request of this reviewer (see below) and reviewer 2, we provide more detail regarding self-reported familiarity with Spanish. This information is also included in the discussion section and the supplementary materials.
- RC:** 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.
- AR:** At the recommendation of all three reviewers we have included more information about the stimuli used in our study (please see the introduction section and the supplementary materials, as it is too much information to include here). The 4 sentence types, while not exhaustive, were selected for two reasons. First, they are the most commonly analyzed utterances in the intonational phonology literature. Thus, these tunes represent the intonational contours about which we know the most. Second, these are the sentence types (and the exact sentences) used in Brandl, González, and Bustin (2020), which our study replicates conceptually.
- RC:** 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.

AR: These are excellent questions. We appreciate the reviewer's critical eye in pointing out these missing elements in our motivation of empathy. We have included more information in the introduction to address these points. The basic idea proceeds as follows. From the perspective of the listener, empathy is likely critical because it allows one to understand the intentions of others, predict their behavior, and understand their emotions (Baron-Cohen & Wheelwright, 2004). Researchers that work on this construct have described two types of empathy that oftentimes might be difficult to distinguish. *Affective empathy* represents one's ability to be emotionally aligned with the interlocutor and *cognitive empathy* refers to recognizing and understanding the feelings and thoughts of an interlocutor. Thus, we believe empathy, to some degree, is a necessary element when an individual seeks to understand and interact with its interlocutors.

When we apply this to everyday communication, we find that speakers' expressions and feelings cannot be limited to only lexical or semantic cues. Speakers employ other strategies, such as using intonational cues, with the intention of producing a meaningful expressions that can include pragmatic connotations, and mental states, among other meaningful expressions. For instance, speakers can signal pragmatic information via pitch accents. A high tone at the end of a phrase can signal a yes/no question the same way as a final high tone preceded by a rising pitch accent can indicate unexpected information (in Spanish). The extant literature suggests that individual pragmatic skills modulate intonation processing (Bishop, 2016; Bishop, Chong, & Jun, 2015; Bishop & Kuo, 2016; Diehl, Bennetto, Watson, Gunlogson, & McDonough, 2008) The critical studies in the case that concerns us are Esteve-Gibert, Portes, Schafer, Hemforth, and D'Imperio (2016), Esteve-Gibert et al. (2020), and Orrico and D'Imperio (2020), which suggest that individuals with higher empathy (operationalized as a pragmatic skill) may be better at processing intonation-meaning associations.

Importantly, *why* and *how* particular acoustic cues lead to distinct processing outcomes on the part of the listener is still unclear. A priori, one would have reason to believe that the effects of empathy on processing intonation-meaning associations would be present across the board, i.e., for any utterance type. In our data, we find that higher empathy is moderately associated with higher accuracy (holding proficiency constant) for all utterance types, with the exception of yes/no questions. Importantly, this also happens to be the most difficult utterance condition in our data—and that of Brandl et al. (2020)—as well as the condition for which the listeners show the least sensitivity (See d' analysis in the supplementary materials). At this juncture, our focus is on expanding this line of research in two ways: (1) to individuals with different linguistic experience (specifically, second language learners), and (2) to different communicative situations (utterance types). We are not concerned (yet) with understanding *why* different pitch contours affect intonation perception, particularly with regard to the role of empathy, because there is inherent variability in how speakers realize their communicative intentions, at the variety-level and at the individual-level, within utterance types. This variability is present in our stimuli. We believe a fruitful avenue for future research is to explore *why* and *how* particular acoustic realizations of pitch within utterance types lead to distinct processing outcomes.

In the revised manuscript we have expanded on these issues in both the introduction and the discussion sections. We do not copy the revised text here due to excessive length, but we direct the reviewer's attention to approximately p. 10, and p. 37.

RC: **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.

AR: We thank the reviewer for this comment. Given that there is no research that we know of regarding empathy and L2 perceptual development, our motivation is based solely on findings from studies on monolinguals. The logic is the following.

Recent studies related to monolingual sentence processing finds that empathy influences speech perception (Esteve-Gibert et al., 2016, 2020; Orrico & D’Imperio, 2020). This line of work operationalizes the construct empathy as a pragmatic skill and has focused on it as a source of individual differences. To elaborate on just one example, Esteve-Gibert et al. (2020) shows that listeners with different levels of empathy interpret intonation and meaning in contexts in which a temporary lexical ambiguity could only be resolved through intonation. The higher empathy group was more sensitive to intonation cues in the process of forming sound-meaning associations. In other words they used intonation to resolve temporary lexical ambiguities. As mentioned, there is limited research in SLA involving empathy, and there are no studies, as far as we know, dealing with empathy and speech perception. That being said, if empathy plays a role in monolingual sentence processing, a valid question is to ask if this is also the case in second language sentence processing. We have tried to make this clear in the revised manuscript and include below a relevant section illustrating the logic explained above.

The particular body of work linking empathy with SLA has focused on speech production, or, more specifically, on what early scholars considered ‘authentic pronunciation’ and, more recently, ‘pronunciation aptitude’ (See Rota & Reiterer, 2009), though no strong associations have been found. To the best of our knowledge, no studies have explored the construct empathy as it pertains to L2 perceptual development. Thus, at this time we do not know if empathy plays a role in L2 sentence processing in a similar manner to monolingual sentence processing. The present project extends this research to the SLA context to determine if individual differences in this pragmatic skill affect the development of intonation in L2 perception and sentence comprehension.

With regard to research question 3, our hypothesis that learners will have more difficulty with the Cuban variety is grounded in our pilot data with monolingual Spanish speakers. We found that participants were much less accurate in the 2afc task when responding to the stimuli provided by the Cuban speaker (and to a lesser extent the Puerto Rican speaker). A priori we would have had no justification for this hypothesis had we not collected the pilot data. In hindsight, it would have made sense to pre-register a hypothesis stating that participants would be more accurate responding to speakers of varieties with which they were more familiar. However, we would not have been able to answer this question—at least not to the extent one would hope—as the listeners in the data we collected mainly stated being most familiar with U.S. Spanish (See pp. 32-33 for brief discussion). We now offer an additional plausible explanation based on familiarity with Spanish. Specifically, we analyzed the data from the participants who claimed to be most familiar with a Spanish variety that was included in our speaker varieties (Peninsular and Mexican Spanish). We coded the participants’ responses to familiar versus unfamiliar varieties and fit a Bayesian logistic regression model to the data (complete details provided in the manuscript). In short, we find that, marginalizing over proficiency and empathy, participants were indeed more accurate when responding to a familiar variety. This is true for all utterance types to some degree, but more clearly the case for questions (likely because responses to declarative utterances were near ceiling). Figure 1 and Table 1 below illustrate the familiarity effect and are now included in the discussion section and supplementary materials.

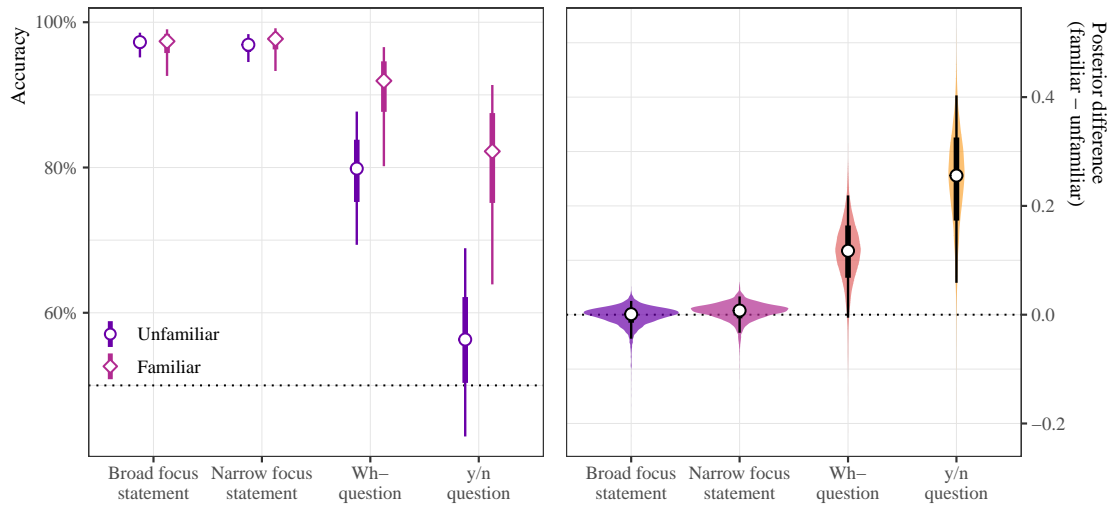


Figure 1: Response accuracy as a function of utterance type for unfamiliar and familiar Spanish varieties. Values represent posterior medians along with the 95% HDI for unfamiliar and familiar conditions (left panel), as well as the posterior difference (familiar - unfamiliar; right panel). The posterior predictive distribution is based on data from participants who claimed to be familiar with Mexican ($n = 47$) and Peninsular ($n = 44$) Spanish.

Table 1: Conditional effects of response accuracy as a function of sentence type and familiarity with the Spanish variety. Values represent posterior medians along with the 95% HDI for unfamiliar and familiar conditions, along with the posterior difference (familiar - unfamiliar). The posterior predictive distribution is based on data from participants who claimed to be familiar with Mexican ($n = 47$) and Peninsular ($n = 44$) Spanish.

Sentence type	Unfamiliar	Familiar	Difference
Broad focus statement	0.97 [0.95, 0.99]	0.97 [0.94, 0.99]	0.00 [-0.04, 0.03]
Narrow focus statement	0.97 [0.95, 0.99]	0.98 [0.95, 1.00]	0.01 [-0.03, 0.04]
Wh- question	0.80 [0.70, 0.88]	0.92 [0.83, 0.97]	0.12 [0.00, 0.22]
y/n question	0.56 [0.43, 0.69]	0.82 [0.67, 0.93]	0.26 [0.07, 0.41]

Finally, we have expanded on the motivation for our hypothesis in the revised manuscript and also included our exploratory analyses of the pilot data in the discussion section and the supplementary materials. This was the only motivation for this hypothesis and precisely what was preregistered prior to collecting any data. Copied below is the relevant change to the introduction.

Finally, with regard to RQ3, we hypothesize that, overall, L2 learners will have the most difficulty (lower accuracy, slower response time) with the Cuban variety. This hypothesis is grounded in exploratory analyses of pilot data collected from 113 monolingual Spanish speakers (See Supplementary Materials for more information).

1.2. STIMULI

RC: 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?

AR:

RC: Please provide the results of the 100 monolingual Spanish speakers rating the quality of items.

AR:

RC: 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.

AR:

1.3. RESULTS & DISCUSSION

RC: 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.

AR:

RC: Could you check whether individuals with higher empathy were not the ones with higher proficiency? How did these two factors correlate?

AR: r_eq_lt_bf r_eq_lt_nf r_eq_lt_yn r_eq_lt_wh

r_eq_lt_random_post

r_eq_lt_post

RC: 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).

AR: The use of proficiency as a continuous variable is not an innovation of our study. The reason continuous variables are discretized in other research likely results from the choice of analysis, e.g., analysis of variance (ANOVA), which requires a continuous outcome variable be predicted by factors (categorical variables).

While we do believe that the common practice of discretization of continuous variables is less than ideal, it is not our intention to speak to this topic in the manuscript. The description regarding how the variable is coded is included (a) for the sake of completeness and reproducibility, and (b) to highlight our departure from Brandl et al. (2020), which our study replicates conceptually.

RC: 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.

AR:

2. Reviewer #2

2.1. Comments to the Author

RC: 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.

AR:

2.2. Recommendations:

RC: 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.

AR:

RC: 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.

AR:

RC: 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?

AR:

RC: 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>>

AR:

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

AR:

RC: 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

AR:

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

AR:

RC: 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

AR:

RC: 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

AR:

RC: 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".

AR: Following this suggestion and that of reviewer 1, we have systematized how we refer to all utterance types throughout the revised manuscript. We now refer to absolute interrogatives as 'yes/no questions'.

RC: 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?

AR:

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

AR: Fixed.

RC: 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

AR:

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

AR: This issue was addressed previously. We include our response again here for the sake of completeness.

Following this suggestion and that of reviewer 1, we have systematized how we refer to all utterance types throughout the revised manuscript. We now refer to absolute interrogatives as ‘yes/no questions’.

RC: 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.

AR:

RC: 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”)?

AR:

RC: 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.”

AR:

RC: 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.

AR:

RC: 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.

AR:

RC: 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.

AR:

3. Reviewer #3

3.1. Comments to the Author

RC: 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.

AR:

RC: 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) [Link]

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) [Link]

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.

AR:

RC: 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.

AR:

RC: 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".

AR:

RC: 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.

AR:

RC: 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?

AR:

RC: 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.

AR:

RC: 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.

AR:

RC: 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.

AR:

RC: 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.

AR:

RC: 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).

AR:

RC: And then your Andalusian speaker has been innovative and has chosen for most of the cases the statement of the obvious pattern (realised as LH%) 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?)

AR:

RC: 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.

AR:

RC: 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.

AR:

RC: 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.

AR:

RC: 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.

AR:

From Hualde and Prieto (2015)

Andalusian - Sosa (1999) - Henriksen and García-Amaya (2012)

Argentine - Sosa (1999) - Colantoni and Gurlekian (2004) - Colantoni (2011) - Labastía (2006) - Labastía (2011)

Chilean - Ortiz-Lira and Cid-Urbe (2000) - Ortiz-Lira (2003)

Cuban - Sosa (1999) - Alvord (2006)

Madrialeño - Quilis (1981) - Quilis (1987) - Quilis (1993) - Sosa (1999) - Estebas-Vilaplana and Prieto (2010)

Mexican - Quilis (1981) - Quilis (1987) - Quilis (1993) - Sosa (1999) - Sosa (2003) - Butragueño (2003) - Butragueño (2004)

Peruvian - Sosa (1999) - O'Rourke (2005)

Puerto Rican - Quilis (1981) - Quilis (1987) - Quilis (1993) - Sosa (1999) - Sosa (2003) - Armstrong (2010) - Armstrong (2012)

4. References

- Alvord, S. M. (2006). *Spanish intonation in contact: The case of Miami Cuban bilinguals* (PhD thesis). University of Minnesota.
- Armstrong, M. E. (2010). Puerto Rican Spanish intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of intonation of the Spanish language* (pp. 155–189). Munich: Lincom Europa.
- Armstrong, M. E. (2012). *The development of yes-no question intonation in Puerto Rican Spanish* (PhD thesis). The Ohio State University.
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), 163–175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>
- Bishop, J. B. (2016). Individual differences in top-down and bottom-up prominence perception. *Proceedings of Speech Prosody, 2016*, 668–672.
- Bishop, J. B., Chong, A. J., & Jun, S.-A. (2015). Individual differences in prosodic strategies to sentence parsing. *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. London: International Phonetic Association.
- Bishop, J. B., & Kuo, G. (2016). Do “autistic-like” personality traits predict prosody perception. *Talk Presented at LabPhon15 Satellite Workshop on Personality in Speech Perception and Production, Ithaca, NY*.
- Brandl, A., González, C., & Bustin, A. (2020). The development of intonation in L2 Spanish: A perceptual study. In A. Morales-Front, M. J. Ferreira, R. P. Leow, & C. Sanz (Eds.), *Hispanic linguistics: Current issues and new directions* (pp. 12–31). John Benjamins Publishing Company. <https://doi.org/10.1075/ih11.26>
- Butragueño, P. M. (2003). Hacia una descripción prosódica de los marcadores discursivos. Datos del español de México. *La Tonía: Dimensiones Fonéticas y Fonológicas*, 375–402.
- Butragueño, P. M. (2004). Configuraciones circunflejas en la entonación del español mexicano. *Revista de Filología Española*, 84(2), 347–373.
- Chen, A. (2005). *On the universal and language-specific perception of paralinguistic intonational meaning*. Utrecht: LOT.
- Colantoni, L. (2011). *Broad-focus declaratives in Argentine Spanish contact and non-contact varieties*. John Benjamins.
- Colantoni, L., & Gurlekian, J. (2004). Convergence and intonation: Historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition*, 7(2), 107–119. <https://doi.org/10.1017/S1366728904001488>
- Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: An experimental study. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 103–120). Berlin: Mouton de Gruyter.
- Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and*

- Language*, 106(2), 144–152. <https://doi.org/10.1016/j.bandl.2008.04.002>
- Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of intonation of the Spanish language* (pp. 17–48). Munich: Lincom Europa.
- Esteve-Gibert, N., Portes, C., Schafer, A., Hemforth, B., & D’Imperio, M. (2016). *The role of individual empathic skills on the online processing of intonational meaning*. Bilbao, Spain: Basque Center on Cognition, Brain; Language. <https://doi.org/10.13140/RG.2.2.19401.13926>
- Esteve-Gibert, N., Schafer, A. J., Hemforth, B., Portes, C., Pozniak, C., & D’Imperio, M. (2020). Empathy influences how listeners interpret intonation and meaning when words are ambiguous. *Memory & Cognition*, 48, 566–580. <https://doi.org/10.3758/s13421-019-00990-w>
- Henriksen, N. C., & García-Amaya, L. J. (2012). Transcription of intonation of Jerezano Andalusian Spanish. *Estudios de Fonética Experimental*, 109–162.
- Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 350–391). Oxford University Press.
- Labastía, L. O. (2006). Prosodic prominence in Argentinian Spanish. *Journal of Pragmatics*, 38(10), 1677–1705. <https://doi.org/10.1016/j.pragma.2005.03.019>
- Labastía, L. O. (2011). Procedural encoding and tone choice in Buenos Aires Spanish. *Procedural Meaning: Problems and Perspectives*, *CRISPI*, 25, 383–413.
- Levis, J. (2016). Accent in second language pronunciation research and teaching. *Journal of Second Language Pronunciation*, 2(2), 153–159. <https://doi.org/10.1075/jslp.2.2.011ev>
- Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.), *Non-native prosody. Phonetic description and teaching practice* (pp. 53–76). Berlin: De Gruyter Mouton.
- O’Rourke, E. (2005). *Intonation and language contact: A case study of two varieties of Peruvian Spanish* (PhD thesis). University of Illinois at Urbana-Champaign.
- Orrico, R., & D’Imperio, M. (2020). Individual empathy levels affect gradual intonation-meaning mapping: The case of biased questions in Salerno Italian. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 11(1). <https://doi.org/10.5334/labphon.238>
- Ortiz-Lira, H. (2003). Los acentos tonales en un corpus de Español de Santiago de Chile: Su distribución y realización. *La Tonía: Dimensiones Fonéticas y Fonológicas*, 303–316.
- Ortiz-Lira, H., & Cid-Urbe, M. E. (2000). La prosodia de las preguntas indagativas y no-indagativas del Español culto de Santiago de Chile. *LEA: Lingüística Española Actual*, 22(1), 23–49.
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *Tesol Quarterly*, 35(2), 233–255. <https://doi.org/10.2307/3587647>
- Quilis, A. (1981). *Fonética acústica de la lengua española*. Madrid: Gredos.
- Quilis, A. (1987). Entonación dialectal hispánica. In H. López Morales & M. Vaquero (Eds.), *Actas del I Congreso internacional sobre el español de América. San Juan, Puerto rico: Academia puertorriqueña de la lengua española* (pp. 117–163).

- Quilis, A. (1993). *Tratado de fonología y fonética españolas*. Madrid: Gredos.
- Rota, G., & Reiterer, S. M. (2009). Cognitive aspects of pronunciation talent. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and brain activity* (pp. 67–96). Mouton de Gruyter. <https://doi.org/10.1515/9783110215496>
- Sosa, J. M. (1999). *La entonación del español*. Madrid: Cátedra.
- Sosa, J. M. (2003). Wh-questions in Spanish: Meanings and configuration variability. *Catalan Journal of Linguistics*, 2, 229–247. <https://doi.org/10.5565/rev/catjl.51>