

## Overview

This paper has improved substantially and the authors have done an excellent job of integrating the comments and suggestions of the reviewers.

The majority of my suggestions are minor, but I still think discussion can be developed a bit to better explain the lack of effect for yes-no questions. I understand that the authors did not have the goal of explaining “understanding why different pitch contours affect intonation perception”, however, I think it would be a more satisfying discussion of the results that is already hinted at in the paper and could be developed more fully since the relevant information is available.

## Page 5

line 32 “to signal whether an utterance is a question or a statement” - can this be broader since intonation goes beyond this? - there are other sentence types that can be signaled by intonation

line 37 is often language-specific —> is language-specific

line 54 - please cite reference for Chilean Spanish questions - might also be worth citing Hualde & Prieto (2015) for discussion of pragmatic nuances in Spanish intonation for questions in varieties of Spanish

## Page 7

line 10

Please add following references for Spanish intonation applying the LiLT model:

Sánchez-Alvarado, 2022

<https://www.jbe-platform.com/content/journals/10.1075/jslp.20041.san>

Sánchez-Alvarado & Armstrong 2022

<https://www.degruyter.com/document/doi/10.1515/shll-2022-2060/html>

## Page 8

Review is missing Sánchez Alvarado 2022 and Sánchez Alvarado & Armstrong 2022 (Also see Sánchez Alvarado’s 2020 dissertation)

The paper talks about concepts such as pitch accents and boundary tones, which assume a specific framework. There should be a short section devoted to the Autosegmental Metrical Framework and the ToBI system of labeling intonation for Spanish specifically where the basic concepts (e.g. pitch accent) are explained. This is standard for papers that use these terms, which involve theoretical assumptions.

line 24 - missing reference for varieties with falling intonation - Gabriel et al. 2010 for Argentine Spanish, Willis 2010 for Dominican and Armstrong 2010 for Puerto Rican Spanish

line 28 - note that in fact most varieties of Spanish have some sort of question fall

## **Page 9**

Please cite Sánchez Alvarado's perception work:

[https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=2906&context=dissertations\\_2](https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=2906&context=dissertations_2)

Marasco's 2020 dissertation also looks at perception (not sure if she has published any of this):

<https://tspace.library.utoronto.ca/handle/1807/103553>

## **Page 31**

line 9 - perceptual—> perceptual?

For discussion on wh- questions, "A specific intonation contour is obligatory to force a question interpretation" - can you discuss the differences between wh intonation and declarative intonation in your stimuli?

Line 57 – note that there are prosodic differences between Que bebe María and Qué beba María (e.g. absence vs. presence of a pitch accent on [ke]) which in theory would help guide the listener to the intended interpretation.

## **Page 32**

“Perhaps for this reason yes/no questions require more effort and attention to intonation in order to distinguish them from statements in our task.” So here the authors are indeed intending to make sense of why it might take more effort to distinguish YNQs from statements. But I think this discussion oversimplifies what we see in the stimuli. The authors have rich information in their ToBI transcriptions and it would be helpful to dig into the melodic differences between varieties and how those differ from American English in order to explain the lack of an effect for YNQs. I understand that the intention was not to get into specific melodies and explain results in this way, but I think at least a cursory attempt is necessary to try to explain the results. More on this below.

## **Supplementary materials:**

Sp\_ToBI labelling

Note that the Sp\_ToBI labels used do not reflect the most current labeling conventions (see Hualde & Prieto 2015) - for example HH% is now !H% if the authors aren't using the most updated conventions, they should cite which version of Sp\_ToBI they are using. Note label change for PRS YNQ below.

## Suggestions for discussion re: YNQs

In the stimuli we see lots of variation in terms of yn questions, and less variation for declaratives (seems like  $H+L^* L\%$  or  $L^* L\%$ ) - analysis should take this into account. Not only is there a ton of variation for the YNQs, three varieties (Argentine, Cuban and PRS) use tunes that differ greatly from the Mainstream American English in the Northeast, so we might expect for these to be harder to identify.

Final rises in red, final falls in green

Madrid:  $L^* HH\%$  ( $L^* !H\%$ )

Andalusian:  $L+H^* HH\%$  ( $L+H^* !H\%$ )

Argentine:  $L+!H^* HL\%$

Chilean:  $H+L^* HH\%$  ( $H+L^* !H\%$ )

Cuban:  $L+H^* L\%$  (I would label this  $L+;H^* L\%$ )

Mexican:  $L^* HH\%$  ( $L^* !H\%$ )

Peruvian:  $L^* HH\%$  ( $L^* !H\%$ )

Puerto Rican:  $H+L^* L\%$  (this is actually not  $H+L^*$ , should be  $;H^* L\%$ , following Armstrong 2017: <https://www.degruyter.com/document/doi/10.1515/probus-2014-0016/html>)

I wonder what would happen if the authors either removed these varieties from the analysis or did an additional analysis that divided the YNQs between rises and falls (falls being Arg, Cub, PR), though if we look at Figure 10 and Table 8 it looks like it might just make sense to do this for Cuban and Puerto Rican). Perhaps responses for rising YNQs would pattern more with the other sentence types. I realize this goes beyond the scope of the paper though. But I think the possible reasons for the finding are readily available in the supplemental data, so why not have a stronger justification for the results?

In sum it would be great to see the authors discuss the variation and the challenges the YNQ variation presents for learners. We see way less variation for the other sentence types in terms of general directionality of the contour (rise vs. fall). So while again, I know the point of the paper is not to get into tunes, it would really help to at least superficially acknowledge these differences in the stimuli to make better sense of the results. The YNQs included here present different challenges compared to the other sentence types. I do see this mentioned in the variety-specific portion of the analysis, but should also be incorporated into the discussion of the lack of an effect for empathy for YNQS.