



**HAL**  
open science

# Visual channel influences the accuracy of listeners' comprehension of Brazilian Portuguese intonation of wh-questions and wh-exclamations

Luma Miranda, João Moraes, Albert Rilliard

► **To cite this version:**

Luma Miranda, João Moraes, Albert Rilliard. Visual channel influences the accuracy of listeners' comprehension of Brazilian Portuguese intonation of wh-questions and wh-exclamations. 4TH PHONETICS AND PHONOLOGY IN EUROPE, Jun 2021, Barcelona, Spain. hal-03512423

**HAL Id: hal-03512423**

**<https://hal.science/hal-03512423>**

Submitted on 5 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EcQs and InfoS. Phonetically, EcQEp, which have a hint of surprisal, differed from both EcQPer and InfoS. The results do now show support for a different semantics for EcQs and InfoSs.

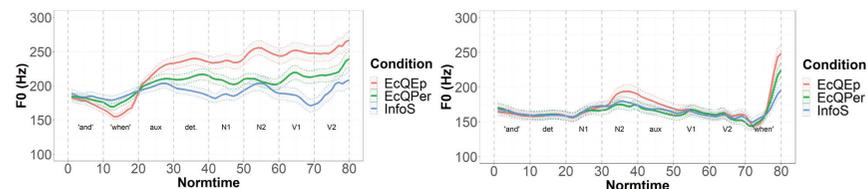


Figure 1. Averaged f0-contours for the two word orders, as derived from ProsodyPro ([8]), (a) fronted wh-word, (b) final wh-word. Grey lines indicate the 95% confidence interval of the mean.

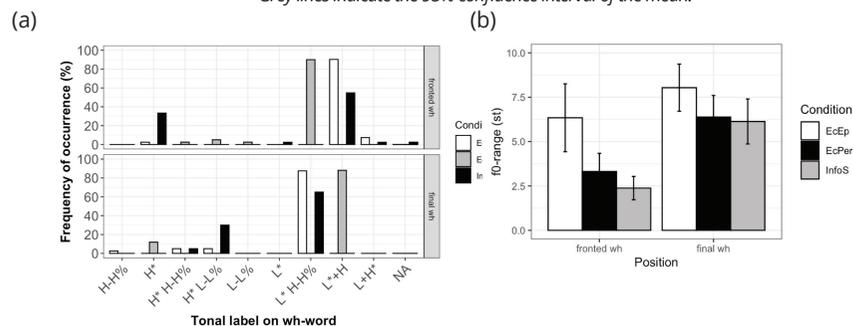


Figure 2. Frequency of occurrence of the different ToBI-labels for the wh-word (a) and average f0-range in semitones (b).

## References

- [1] Artstein, Ron. 2002. A focus semantics for echo questions. In Workshop on Information Structure in Context, ed. Agnes Bende-Farkas and Arndt Riester, 98–107. IMS Stuttgart.
- [2] Bartels, Christine. 1999. *The intonation of English statements and questions*. Garland Publishing.
- [3] Beck, Sigrid, and Marga Reis. 2018. On the form and interpretation of echo wh-questions. *Journal of Semantics* 1–40.
- [4] Krifka, Manfred. 2011. Questions. In von Stechow, Ralf, & Portner, Paul, *Semantics: An international handbook of natural language*, Vol. 2, 1742–1748.
- [5] Repp, Sophie, and Lena Rosin. 2015. The intonation of echo wh-questions. In Proceedings of the 16th Annual Conference of the International Speech Communication Association. Dresden, Germany: Interspeech.
- [6] Ginzburg, Jonathan, and Ivan Sag. 2001. *Interrogative investigations*. Stanford, CA: CSLI Publications.
- [7] Xu, Y. 2013. A tool for Large-Scale Systematic Prosody Analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), 7–10. Aix-en-Provence, France.
- [8] Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In Sun-Ah Jun (ed.), *Prosodic typology*, 9–54. Oxford: Oxford University Press.
- [9] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. <http://CRAN.R-project.org/package=lme4>.
- [10] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57, 289–300.

## Visual channel influences the accuracy of listeners' comprehension of Brazilian Portuguese intonation of wh-questions and wh-exclamations

Luma da Silva Miranda<sup>1</sup>, João Antônio de Moraes<sup>2,3</sup>, Albert Rilliard<sup>2,4</sup>

<sup>1</sup>Eötvös Loránd University, Hungary

<sup>2</sup>Federal University of Rio de Janeiro, Brazil; <sup>3</sup>CNPq, Brazil

<sup>4</sup>Université Paris-Saclay, CNRS, LISN, France

The multimodality of speech helps listeners to perceive prosodic functions such as sentence mode [1,2,3,4,5,6]. One of the protocols applied to evaluate the particular influence of the visual channel is based on McGurk & McDonald's work [7] that uses audio and visual modalities to convey a set of categories. The particular category presented through each modality may be the same (congruent) or not (incongruent presentations). This work aims at measuring the relative influence of each modality on a given function [8]. In a previous study [6] on the multimodal production and perception of Brazilian Portuguese (BP) wh-questions and wh-exclamations, it was shown that, acoustically, there is a steeper F0 fall in the nucleus for the former and a slighter movement for the latter, plus a higher initial F0 in wh-questions than in wh-exclamations. The intensity patterns of both contours are different. Visually, specific facial gestures for each speech act were described: lowered eyebrow plus the head turned right for wh-questions; lips corner raiser and up and down head movement for wh-exclamations. A perception test showed that BP listeners can identify each contour based on the auditory and visual cues, either separately or combined.

In the present study, the audio and the visual performances of BP wh-questions and wh-exclamations were mixed to create congruent (same speech act) and incongruent (different speech acts in the two modalities) stimuli, so as to evaluate the relative strength of these modalities in the identification of these speech acts. The stimuli were recorded by ten BP speakers (five females and five males) from Rio de Janeiro who repeated both speech acts ten times (two hundred utterances). The eighth and ninth repetitions of each act produced by each speaker were chosen so as to avoid a qualitative selection of the performances. The production of the stimuli was made with the Vegas Pro software [9] following these steps: for the congruent stimuli, two repetitions of the same speech act were used and the audio of one repetition was dubbed into the video of the other. For incongruent stimuli, one repetition of each speech act was used, dubbing the audio of one speech act into the video from the other speech act. The outcome of this procedure was a set of forty stimuli: ten congruent and ten incongruent stimuli of wh-questions and wh-exclamations. By using the PsyToolkit website [10,11], a perceptual identification protocol was designed and applied over the internet. Thirty-six BP listeners participated; they had to identify, for each stimulus, the mode of the perceived sentence (wh-question or wh-exclamation).

The answers to each stimulus were then considered as “hit” (1) or “miss” (0) whether the judge correctly identified the mode of the audio modality. This was done because the audio modality is supposed to dominate; we evaluate here the effect on the audio identification ratio introduced by the visual modality. It is hypothesized that an incongruent presentation will decrease identification, while a congruent one will increase it. The proportion of “hit” or “ratio of correct answers” was used as the dependent variable of a logistic regression, which used as independent variables the “speech act” (two levels: wh-question, PQ, and wh-exclamation, EX), the “congruence” between modalities (two levels: congruent, C, and incongruent, I), and the “speaker groups” who produced the stimuli (five levels of “speakers groups” showing coherent answers were obtained at the end of a simplification procedure;

see [12]). The model showed a main effect of factors “congruence” and “speaker groups” and significant interactions between “speaker groups” and “congruence”, as well as between “speaker groups” and “speech act”. The main effect of the “congruence” factor indicated that the incongruent stimuli significantly lowered the ratio of correct identification. Fig. 1 presents the mean identification ratio for the significant interactions. One can observe that the identification level for each speech act is mostly above 60%. This illustrates not only the primary role of the audio modality for the expression of the sentence mode, in line with previous studies [3,4,5,8], but also the speaker-dependent use of modalities for the expression of these speech acts in Brazilian Portuguese: some relying more than others on visual performances – but the audio modality being always dominant. It also shows that the QP expression has a higher identification ratio than EX, when produced by some speakers: this underlines the importance of individual performances.

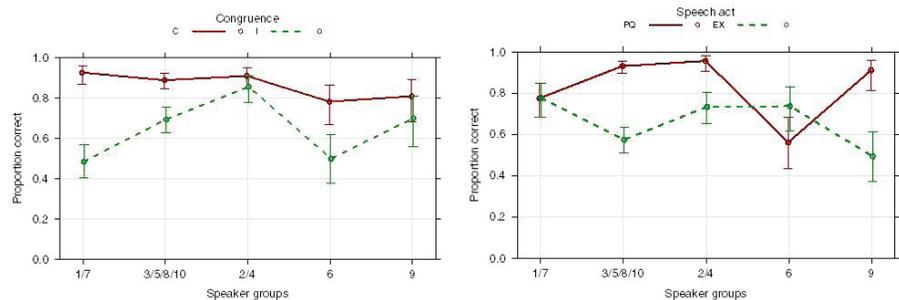


Figure 1: Means and confidence intervals estimated by the regression model for the interactions between factors “Congruence” (separate lines) and “Speaker groups” (x-axis) (left), and between “Speech act” (separate lines) and “Speaker groups” (x-axis) (right).

## References

- [1] House D. 2002. Intonational and visual cues in the perception of interrogative mode in Swedish. *Proceedings of the 7<sup>th</sup> International Conference on Spoken Language Processing*, Denver, Colorado, pp. 1957–1960.
- [2] Srinivasan RJ, Massaro DW. 2003. Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, 46 (1), pp. 1–22.
- [3] Borràs-Comes J, Prieto P. 2011. ‘Seeing tunes.’ The role of visual gestures in tune interpretation. *Laboratory Phonology*, 2 (2), pp. 355–380.
- [4] Cruz M, Swerts M, Frota S. 2017. The role of intonation and visual cues in the perception of sentence types: Evidence from European Portuguese varieties. *Laboratory Phonology*, 8 (1), 23, pp. 1–24.
- [5] Miranda L, Swerts M, Moraes J, Rilliard A. 2021. The role of the auditory and visual modalities in the perceptual identification of Brazilian Portuguese statements and echo questions. *Language and Speech*, 64 (1), pp. 3–23. DOI: <https://doi.org/10.1177/0023830919898886>
- [6] Miranda LS, Moraes J, Rilliard A. 2019. Audiovisual perception of wh-questions and wh-exclamations in Brazilian Portuguese. *Proceedings of the 19<sup>th</sup> International Congress of Phonetic Sciences*, August 5-9, Melbourne, Australia, pp. 2941–2945.
- [7] McGurk H, MacDonald J. 1976. Hearing lips and seeing voices. *Nature*, 264, pp. 746–748.
- [8] Swerts M, Kraemer E. 2004. Congruent and incongruent audiovisual cues to prominence. In: Bel B, Marlin I. (eds.), *Proceedings of 2<sup>nd</sup> International Conference on Speech Prosody*, Nara, Japan, pp. 69–72.

- [9] Vegas Pro software, Version 14 of Vegas Pro. 2016. Copyright © [2016] MAGIX. Software available at: <https://www.vegascreativesoftware.com/>.
- [10] Stoet G. 2010. PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42 (4), pp. 1096–1104.
- [11] Stoet G. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44 (1), pp. 24–31.
- [12] Crawley MJ. 2013. *The R book* (Second edition). United Kingdom: Wiley.