



HAL
open science

Focus types in Brazilian Portuguese: Multimodal production and perception

Manuella Carnaval, João Antônio De Moraes, Albert Rilliard

► **To cite this version:**

Manuella Carnaval, João Antônio De Moraes, Albert Rilliard. Focus types in Brazilian Portuguese: Multimodal production and perception. *DELTA : Documentação de Estudos em Lingüística Teórica e Aplicada*, 2022, 38 (3), pp.1-34. 10.1590/1678-460X202258944 . hal-03836231

HAL Id: hal-03836231

<https://hal.science/hal-03836231>

Submitted on 2 Nov 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.

Articles

Focus types in Brazilian Portuguese: Multimodal production and perception

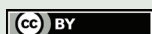
Tipos de foco no português do Brasil: Produção e percepção multimodal

Manuella Carnaval^{1,2}
João Antônio de Moraes^{3,4}
Albert Rilliard^{5,6}

ABSTRACT

This paper aims at describing prosodic focalization as a multimodal phenomenon in Brazilian Portuguese, evaluating the role of two modalities in focus production and perception: audio (A), visual (V), and their combined audiovisual presentation (AV). Five focus types are considered, according to their semantic-pragmatic values: (a) in declarative sentences: (i) IF - informational focus (answer to a previous question,

1. Federal University of Rio de Janeiro – UFRJ. Rio de Janeiro – Brazil. <https://orcid.org/0000-0002-4321-5859>. E-mail: manuellacarnaval@gmail.com.
2. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES.
3. Federal University of Rio de Janeiro – UFRJ. Rio de Janeiro – Brazil. <https://orcid.org/0000-0002-8894-166X>. E-mail: jamoraes3@gmail.com.
4. Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.
5. Université Paris Saclay, CNRS, LISN, Orsay – França. <https://orcid.org/0000-0001-6490-2386>. E-mail: albert.rilliard@limsi.fr.
6. Federal University of Rio de Janeiro. Rio de Janeiro – Brazil. <https://orcid.org/0000-0001-6490-2386>. E-mail: albert.rilliard@limsi.fr.



This content is licensed under a Creative Commons Attribution License, which permits unrestricted use and distribution, provided the original author and source are credited.

conveying new information), (ii) CF- contrastive (strong) focus (correction of information considered wrong); (iii) ATF - attenuated (weak) focus (proposition of an alternative solution to previous information); (b) in interrogative sentences: (i) INTF - interrogative focus (a new information is requested in the question); (ii) SF - surprise focus (one casts doubt on a previous information). Also, structural factors were evaluated, as focus extension and position in the sentence. After running a multimodal perceptual experiment and developing an acoustic and visual analysis on focus production, results show that multimodality plays a relevant role in focus production and perception. Different acoustic and visual parameters, or configuration of parameters, contribute to conveying distinct meanings, according to each focus type.

Keywords: *focalization; audiovisual prosody; Brazilian Portuguese.*

RESUMO

Este artigo descreve a focalização prosódica como um fenômeno multimodal no português do Brasil, investigando a relevância de duas modalidades na produção e percepção do foco: auditiva (A), visual (V), além de sua apresentação audiovisual (AV). Cinco tipos de foco são considerados, de acordo com seus valores semântico-pragmáticos: (a) em enunciados assertivos: (i) FI – foco informacional (uma resposta a uma pergunta prévia, constituindo informação nova), (ii) FC – foco contrastivo (correção de uma informação errada), (iii) FAT – foco atenuado (proposição de uma solução alternativa para uma afirmação prévia); (b) em enunciados interrogativos: (i) FINT – foco interrogativo (uma nova informação requisitada em uma pergunta); (ii) FE – foco com estranheza (uma informação prévia é colocada em dúvida). Também foram investigadas a extensão e a posição do foco na sentença. Após a aplicação de um experimento perceptivo e de uma análise multimodal, os resultados mostraram que a multimodalidade apresenta relevância na produção e percepção do foco. No entanto, diferentes parâmetros acústicos e visuais, ou a sua combinação, contribuem para a transmissão de diferentes significados, segundo cada tipo de foco.

Palavras-chave: *focalização; prosódia audiovisual; Português do Brasil.*

1. Introduction

Focalization has been widely approached in linguistic studies as a prosodic phenomenon, traditionally associated with intonation in different linguistic perspectives. Bolinger (1954) describes focus as the most informative part of a sentence, drawing attention to its association to a prosodic prominence. Halliday (1967) presented the information focus as new information, textually and situationally, which is marked out by stressed prominence, indicating where the new element ends. In a formal approach, Chomsky (1971) refers to focus as the expression that carries the nuclear stress of a sentence, and Jackendoff (1972) affirms that new information is expressed within a sentence as a product of accent and intonation. This relation between focus and prosodic features has also been developed in Lambrecht (1994) and Krifka (2008), according to their own framework.

However, as emphasized in Krahmer and Swerts (2009), pitch accents are not the only relevant cues in focus perception, since specific visual cues, such as eyebrow and head movements, may contribute to prominence perception. In fact, it has been advocated, in many works, that there is a close relationship between prosody and gestures, which are then traditionally called co-speech gestures, since they have a temporal alignment with speech (Kendon, 1980; McNeill, 1992). Thus, for instance, the peak prominence of a gesture (apex) is usually aligned with pitch accents (Alexanderson et al., 2013; De Ruiter, 1998; Esteve-Gilbert et al., 2017; Loehr, 2012; Pouw & Dixon, 2019). Pouw et al. (2021) state that gesture physical impulses influence the speech system, reinforcing previous studies that claim that speech and gestures are programmed together (Esteve-Gibert et al., 2017; Esteve-Gibert & Prieto, 2013; Shattuck-Hufnagel & Ren, 2018; Wagner et al., 2014).

Prieto et al. (2011) investigate the relevance of different gestures (eyebrows and head movements) to contrastive focus perception in Catalan, concluding that the visual components not only accompany the acoustic one but are decisive in identifying the semantic value of contrast. Borràs-Comes and Prieto (2011) show that contrastive focus and echo questions can be visually distinguished in Catalan, while acoustic cues play a secondary role. In contrast with Catalan participants, who are highly sensitive to facial cues to identify incredulity yes-no

questions, Dutch participants rely more on intonational cues, thus showing that the weight of auditory and facial cues is relative and language dependent (Crespo-Sendra et al. 2013). In French, according to Dohen and Lœvenbruck (2009), the integration of audio and visual channels improves the contrastive focus identification, and reaction times decrease significantly when both modalities, audio and visual, are simultaneously displayed. For the Portuguese language, Cruz et al. (2015) report that eyebrow-raising movements mark narrow focus statements for two European varieties, the standard variety, and the Azores insular variety.

Although firmly established for some languages, the multimodal perception of focus is still understudied in Brazilian Portuguese (BP). This paper aims at providing a multimodal analysis of both production and perception of different focus types, increasing the awareness and the understanding of this audiovisual phenomenon in BP. Thus, our main goal is evaluating the relevance of the presentation modalities – audio (A), visual (V), and audiovisual (AV) – in the production and identification of five focus types: (a) Informational Focus, (b) Contrastive Focus, (c) Attenuated focus, (d) Interrogative focus, and (e) Surprise focus.

Focus Types

The focus types are defined as follows:

(a) Informational Focus is referred to by Gussenhoven (2006) as Presentational Focus. It is defined as a simple answer to a previous question, with a prominence on the new information in a discursive context. For instance, in context (1), the question of speaker A elicits speaker B's answer, that contains the new information (It's John who wakes up early every day). Thus, the subject John is the sentence's focus.

- (1) A: Who wakes up early every day?
B: **John**_(informational focus) wakes up early every day.

(b) Contrastive Focus expresses a correction with regard to a previous information considered as erroneous. Dik (1980) defines it as Counter-assertive focus, used when one substitutes information, while Gussenhoven (2006) highlights that it implies a straight rejection of an alternative, as in context (2) in which speaker B corrects speaker A's assertion about the identity of the subject.

- (2) A: Mary wakes up early every day.
B: **John**_(contrastive focus) wakes up early every day.

(c) Attenuated Focus corresponds to the proposal of an alternative solution concerning previous information, both being potentially true, as defined in Moraes (2006) when referring to non-exclusive focus in Brazilian Portuguese. Also, Elordieta and Irurtzun (2009) use the same terminology, investigating this type of focus in the Basque language. In context (3), an attenuated correction would imply, after speaker A's assertion about Mary, that, as far as speaker B knows, it is true for John, but that B has no information concerning Mary's habits. This proposition, produced as a weaker contrast to the previous information, does not exclude possible alternatives limiting the information to what is known to the speaker.

- (3) A: Mary wakes up early every day.
B: **John**_(attenuated focus) wakes up early every day.

(d) Interrogative Focus requests the confirmation of a new information under the form of a question specifying a previous statement (Carnaval et al., 2019; Moraes et al., 2015), as shown in context (4) which illustrates that, after speaker A unspecific statement, speaker B searches to fill in an information gap, by asking if it is John that has this habit.

- (4) A: He wakes up early every day.
B: **John**_(interrogative focus) wakes up early every day?

(e) Surprise Focus corresponds to the "Surprise Question" in Truckenbrodt et al. (2009). It occurs when speaker B casts doubt on

the information previously given by speaker A, as shown in context (5), concerning the constituent, *John*.

- (5) A: John wakes up early every day.
B: **John**_(surprise focus) wakes up early every day?

Our main hypothesis is that the visual channel, when associated with the audio, improves the identification of these semantic values, since, as mentioned above, many previous studies show that speech and gestures are simultaneously set. Focus extension and position in the sentence are analyzed to test if those structural factors influence focus production and perception, as well as its interaction with multimodality.

2. Method

Ethical statement

The protocol of this study (recording and perceptual evaluation) was submitted to UFRJ ethical committee and validated under process 98728718.6.0000.5286. All participants (speakers and perceivers) signed an informed consent form and were informed they can withdraw at any time from the study and require their data to be deleted.

Corpus

Four L1 speakers (2 female/2 male) of Brazilian Portuguese produced the sentence “*O professor de Literatura vai aplicar a prova final*” (The literature professor will give the final test) with the five mentioned types of narrow focus: informational focus, contrastive focus, attenuated focus, interrogative focus, and surprise focus. The speakers’ age groups ranged from 28 to 30 years old, except for one elderly speaker, who was 66 years old⁷. All four speakers presented previous experience with multimodal corpus recording. However, only one of them knew about the corpus issue. This feature allowed us to

7. As age was not a relevant factor in our analysis, this difference did not influence our methodology.

verify if there would be relevant production differences between them, which were not observed.

To make the corpus recording a more spontaneous task, speakers were told that they would have to pronounce the sentence “The literature professor will give the final test” in response to different previously presented contexts. Thereby, a given communicative situation elicited their production, to which they would have to answer with the mentioned sentence, varying their intonation and visual expression, according to the context itself. Speakers produced three repetitions for each given context. Nonetheless, only one production was chosen for analysis and to be included in the perceptual test data set, considering all utterances would result in an exhaustive task. Thus, to compose our final *corpus*, the second production was always chosen, maintaining the same methodological criteria.

According to the different contexts, the speakers were elicited to specifically target the following elements as the sentence focus: the noun phrase “*O professor*”, the prepositional phrase “*de literatura*”, the head of the verbal phrase “*vai aplicar*”, the noun phrase “*a prova*” and the adjectival phrase “*final*”. Then, in order to test the extension, we also asked to consider the syntactic constituents subject (“*O professor de literatura*”) and object (“*a prova final*”) as the sentence focus. Also, the sentence was produced with a broad focus both in declarative and interrogative sentence types, considered as default stimuli for further analysis. However, as our aim with this perception test is centered on the identification of focus type, for limiting the number of the elements to be evaluated, we only consider here focus over the following elements: the noun phrase “*O professor*” (at the beginning of the sentence), the verbal phrase “*vai aplicar*” (at the middle of the sentence), and the adjectival phrase “*final*” (at the end of the sentence) as simple focus, and the subject “*O professor de literatura*” as complex focus, all potential narrow focus units, while the broad focus correspond to the entire sentence. It allowed us to evaluate focus production and perception in different sentence positions and different focus extensions, two relevant factors in our analysis. This reduced set of four utterances allowed us to analyze a potential role of the focalized structure’s position and size on the identification ratio of the semantic type, considering the three presentation modalities (A, V, AV). Altogether, 80 utterances

(4 focalized elements x 5 focus types x 4 speakers) were selected for analysis, and each one presented in three modalities, totalizing 240 stimuli for the perception task.

Acoustic and visual analysis

The following parameters were measured for the four speakers' production of the sentence in each of the five focus types, at the four positions (and two extensions) presented during the test: (i) for the audio modality, fundamental frequency (F0), and syllabic duration; (ii) for facial movements, Action Units (AU) 01 and 04 (respectively "inner brow raiser" and "brow lowerer") (Ekman et al., 2002), the distance between the two brows (as a proxy of frowning – expressed in z-score), and head movement along the three axes (i.e., head nod, roll and yaw). Acoustic measurements were made using the Praat program (Boersma & Weenink, 2020) with the default pitch detection algorithm for F0. For the syllabic duration measurement, phonemes were manually segmented. Visual gestures were estimated by virtue of the OpenFace program (Baltrušaitis et al., 2018), and its capacity to estimate action units from video recordings (Baltrušaitis et al., 2015); the inter-brow distance and head movements were estimated from the three-dimensional output of the program, that estimates position of landmarks on the face, corrected for head rotations. Rotations and inter-brow distances were expressed in z-score of the output estimated for each speaker in order to correct for individual characteristics.

The F0, the inter-brow distance and head movements were corrected for potential inter-speaker differences by expressing them as z-score of the distribution observed for each speaker; action units, as already expressed on a common 0 to 4 scale, were left unchanged. The same criterion was applied to syllabic duration because only one sentence is used as a basis for all the performances, making comparison with the broad focus version straightforward. All the measurements (but syllabic duration) were averaged at the level of the syllable, using the median value of all observations along the syllable (for visual parameters), or along the vowel (for F0), for each syllable. The variation of these mean values was then observed given the following factors:

- The presence or absence of focus on each element of the sentence: the elements considered are each the heads of the syntactic phrases of the sentence (stressed syllables are indicated in bold font): “*O professor / de literatura / vai aplicar / final.*”, with four possibilities of receiving a focus (respectively on “*o professor*”, “*o professor de literatura*”, “*vai aplicar*”, and “*final*”).

- Four syllable positions in the words corresponding to the heads of the syntactic phrases: the stressed syllable, the pre-stressed (syllable immediately preceding the stressed one), the post-stressed, or “other” for syllables prior the pre-stressed one. For instance, in the phrase “*de literatura*”, we considered the stressed syllable [tu], the pre-stressed syllable (immediately previous to the stressed one) [fa], the post-stressed syllable [te] and as “other” the non-immediately previous pre-stressed syllables, as [te], [li] and also the preposition [dʒi].

A given syllable thus pertains or not to a focalized element, and has a specific position in this element (one of the four types of syllables).

Experimental Design and Participants

To evaluate the capacity of perceivers in interpreting these semantic variation of narrow focus categories, and the influence of the various factors (presentation modality, position, and extension), as previously exposed, the following perceptual experiment was set up. We considered the five focus types and the different modalities in which they were presented, that is, audio-only (A), visual-only (V), and audiovisual (AV) presentations. The main goal was to evaluate if the semantic-pragmatic values of the mentioned focus types could be acoustically and visually identified, and if there is a potential enhancement linked to audiovisual presentations.

As the *corpus* is based on a relatively large number of stimuli, a Latin square design (Cochran & Cox, 1992) was set to distribute the stimuli across different perceiver groups, spread across speakers: participants had to evaluate each focus type, in each position and extension, presented in three modalities (A, V, AV) – but a single combination is presented to a group with the performance of one speaker only. The different groups were presented to the same

sentence, with a specific focus type and focused element, although the performances were from different speakers. Four participants' groups were set, to which were presented twenty stimuli in the three mentioned modalities, resulting in sixty stimuli evaluated by each group. The four speakers were evenly spread across the groups.

The experiment was run with the *Qualtrics* platform (Qualtrics, Provo, UT), where the participants were presented with the following task: after hearing or/and seeing a stimulus, they had to identify which of the five categories best corresponded to the performance: a simple answer, an explicit (strong) correction, an attenuated (weak) correction; a simple question or a surprise question. Those labels correspond, respectively, to the semantic values of Informational Focus, Contrastive Focus, Attenuated Focus, Interrogative Focus, and Surprise Focus. Definition and examples of the five possible interpretations were presented to the participants prior the test begins.

The stimuli presentation occurred in three blocks, according to the presentation modality. Half the participants were presented with the (A, V, AV) modality order, the other half with the (V, A, AV) order to balance potential presentation order effect. Participants were informed which modality they would evaluate, although the presentation of the 20 stimuli was randomized inside each block.

This forced-choice test lasts between 20 and 30 minutes to be completed. Each stimulus could be replayed as many times as considered necessary.⁸

Statistical Analysis of perceptual results

Perceptual results were analyzed with a multinomial regression model (Gries, 2013; Venables & Ripley, 2002) using the “nnet” library of R software (R Core Team, 2021), considering three factors having a potential effect on the dependent variable (proportion of answer in each category of focus type) - the focus type (5 levels); the place of the focused element (4 levels); the modality (3 levels) - as well as the double and triple interactions between them. A simplification process

8. See Appendix 1 for an example of the experiment interface.

(following Crawley, 2013, and comparing models using likelihood ratio (LR) tests) was applied. Simplification steps are detailed in section 3, together with the statistical analysis).

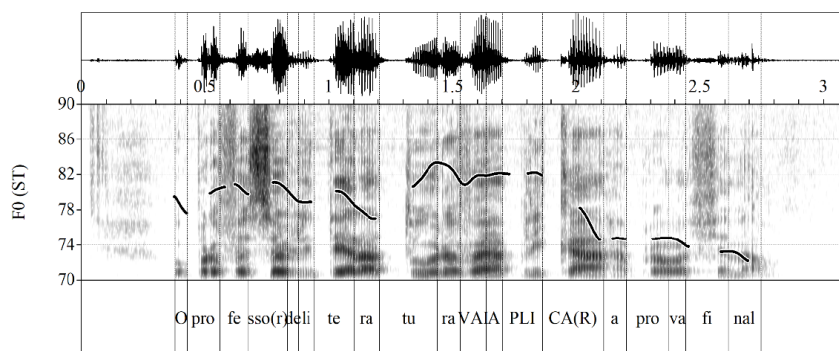
3. Results

Focus Production

In this subsection, we provide, for each focus type, a general description of its acoustic and visual patterns. Thereby, it details the stimuli evaluated in the perceptual test, which results are presented in the next section.

Figure 1 illustrates the informational focus prosodic pattern, which is mainly characterized by a plateau followed by a falling pitch movement on the stressed syllable of the focused element ([*'ka*] in the head of the verbal phrase *vai aplicar* (will give) of the proposed example), followed by a low deaccented post-focal melody until the end of the sentence.

Figure 1 – The utterance *O professor de literatura vai aplicar a prova final* (The literature professor will give the final test), with an Informational Focus on *vai aplicar* (will give), produced by the male speaker M2



Regarding the Informational Focus visual realization, no recurrent facial movements were observed on the recordings across speakers.

The consistent identification level in the visual modality for IF reveals that neutrality may be enough to match visual settings with this focus type (thus its use as a default answer). Figure 2 presents the visual frames for IF of the four speakers whose audiovisual materials were evaluated in the perceptual experiment. The snapshots were taken over the focalized stressed syllable ([^hka] in the phrase *vai aplicar*).

Figure 2 – Visual production of the four speakers of Informational Focus (IF) on the phrase *vai aplicar*. Snapshots were taken during the production of the focused stressed syllable [^hka]



Figures 3 and 4 present the melodic pattern for Contrastive Focus and Attenuated Focus, respectively. Both patterns are characterized by a melodic rise on the pre-stressed syllable (here on [pli] in “*aplicar*”) and a fall on the stressed one (here on [^hka] in “*aplicar*”). However, in Contrastive Focus, this pattern is followed by a deaccenting of the post-focal material. We could also highlight a difference between both strategies regarding the falling movement along the stressed syllable: while for Contrastive Focus a steep fall is observed, for Attenuated Focus a shallower movement is extended until the end of the sentence. Furthermore, the duration increase is clear for both types, however, it is larger for Contrastive Focus.

Figure 3 – The utterance *O professor de literatura vai aplicar a prova final* (The literature professor will give the final test), with a Contrastive Focus on *vai aplicar* (will give), produced by the male speaker M2

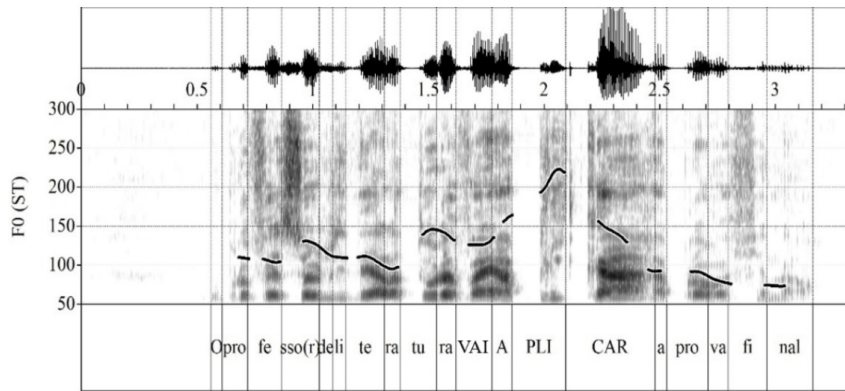
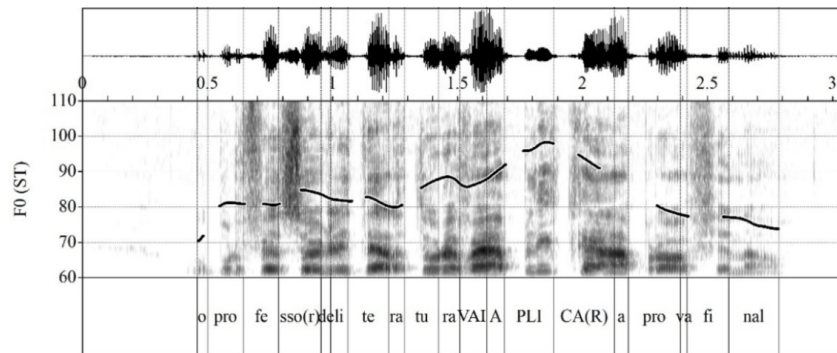


Figure 4 – The utterance *O professor de literatura vai aplicar a prova final* (The literature professor will give the final test), with an Attenuated Focus on *vai aplicar* (will give), produced by the male speaker M2



Concerning their visual performances, both types present more distinctive cues than Informational Focus, with eyebrow and head movements. Figure 5 illustrates the visual productions for Contrastive Focus, with speakers producing a rising eyebrow movement alongside a head nod, both movements synchronized with the melodic pattern, ascending on the pre-stressed syllable (here the [pli] of “*aplicar*”) and falling on the stressed syllable (the ['ka] of “*aplicar*”). Figure 6 shows the Attenuated Focus visual performances, with a side tilt of the head and an asymmetric brows movement along the sentence.

Figure 5 – Visual production of the four speakers for Contrastive Focus (CF) on the phrase *vai aplicar*. Snapshots were taken over the production of the focus pre-stressed (left side) and stressed (right side) syllables, [pli] and ['ka], respectively, to show the dynamic eyebrow and nod movements

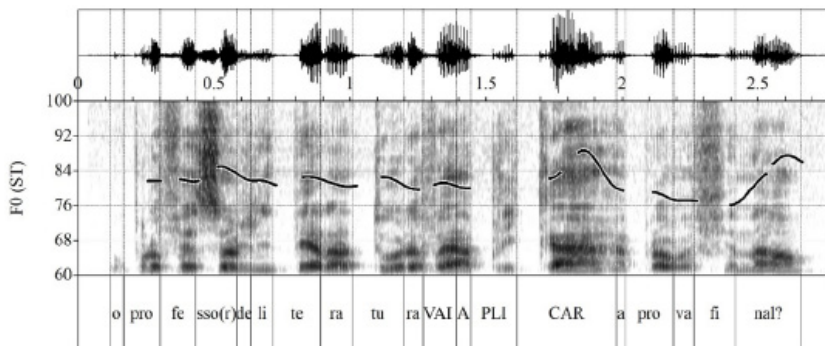


Figure 6 – Visual production of the four speakers for Attenuated Focus (ATTF) on the phrase *vai aplicar*. Snapshots were taken over the production of the focused stressed syllable ['ka]



The Interrogative Focus melodic pattern is characterized by a melodic rise on the stressed syllable of the focused element, in addition to the nuclear interrogative rise at the end of the sentence, as can be observed in Figure 7 (note that the final rise does not play a role in the localization of the focus, but may be relevant for the identification of its semantics).

Figure 7 – The utterance *O professor de literatura vai aplicar a prova final?* (Will the literature professor give the final test?), with an Interrogative Focus on *vai aplicar* (will give), produced by the male speaker M2



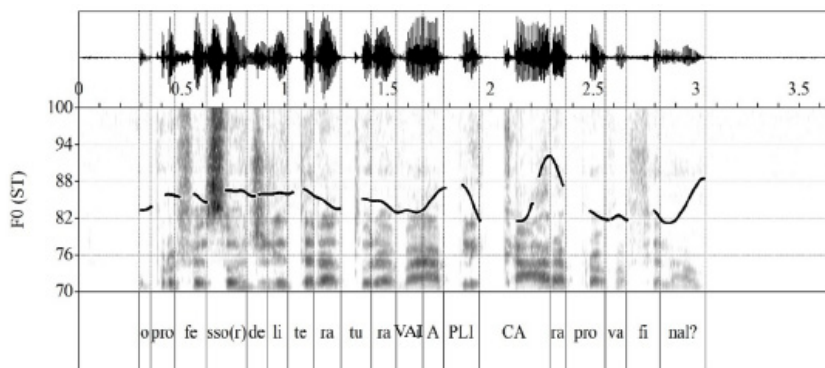
Interrogative Focus visual performances did not bring type-specific cues for its identification, being mostly associated with the Surprise and Informational Focus labels (when presented alone), and vary depending on each speaker's production, as illustrated in Figure 8.

Figure 8 – Visual production of the four speakers for Interrogative Focus (INTF) on the phrase *vai aplicar*. Snapshots were taken over the production of the focused stressed syllable [ˈka]



Surprise Focus (SF) presents distinctive acoustic patterns and visual performances. Its melodic contour may be described as a double melodic peak over the focalized element (see Figure 9): the first peak occurs on the pre-stressed syllable (here the [pli] of “aplicar”) while the second one reaches its maximum at the end of the stressed syllable (here the [ˈka] of “aplicar”). The focalized element also presents an expressive increase of syllabic duration, as can be observed on Figure 9.

Figure 9 – The utterance *O professor de literatura vai aplicar a prova final?* (Will the literature professor give the final test?), with a Surprise Focus on *vai aplicar* (will give), produced by the male speaker M1



Surprise Focus visual productions are also distinctive, presenting a frown movement which marks the beginning of the focalized element, gradually losing intensity until the end of the sentence. Figure 10 illustrates the four speakers' visual performance for Surprise Focus, with a brow frowning over the focalized stressed syllable.

Figure 10 – Visual production of the four speakers for Surprise Focus (SF) on the phrase *vai aplicar*. Snapshots were taken over the production of the focused stressed syllable ['ka]



This qualitative analysis of the melodic and visual patterns is validated by a quantitative analysis from acoustic and visual measurements, which were described in the Method section. It seeks underlining the behavior of the most expressive parameters in the production of these five focus types. The following figures (figs. 11, 12, and 13) present the mean and confidence interval of the values observed for the different categories of syllables (stress, pre-stress, post-stress, other), for each focus type (IF, CF, ATF, INTF, SF), as well as for the broad focuses with assertive and interrogative sentence types (BFa, BFi).

Figure 11 presents the results for F_0 and syllabic duration. It is possible to observe raised F_0 values over the focalized elements, with a peak on the pre-stressed syllable for the assertive focus types (IF, CF, ATF), on the post-stressed syllables for INTF, and a double peak (on the pre- and post-stressed syllables) for SF. Differences of F_0 increase are also observed from IF up to ATF, which is performed with the highest mean values. Duration changes in focalized elements mostly affect the stressed syllables, with two levels of lengthening: IF, ATF, and INTF present stressed syllables with a mean duration of 200-300ms, while CF and SF present stressed syllables that last more than 300ms; focalized pre-stressed syllables are almost as long as non-focalized stressed syllables.

Figure 11 – Mean and confidence interval of the F_0 (top) and syllabic duration (bottom) values observed for each focus type (individual panels), for each syllable type (stress, pre-stressed, post-stressed or other, indicated by shapes) in focalized or non-focalized elements (indicated by colors)

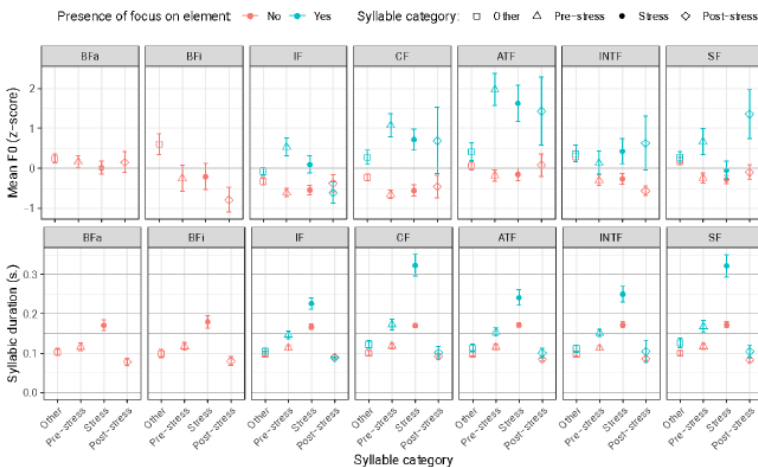


Figure 12 presents the intensity of action units 01 and 04, and the distance between the brow (a measure taken to reflect frowns). AU01, inner brow raising, is typically observed along the focalized elements of contrastive focus. AU04, brow lowering, is enacted on surprise focus, but lasts until the end of the sentence, not being timely restricted to the focalized element. Frown – reduced distance between the brows – is typical of SF, with values slightly smaller along the focalized elements than outside this time-window, but generally recalling results of AU04; an extreme distance between brows is observed for CF focalized elements, which may result in raised brows (AU01).

Figure 12 – Mean and confidence interval of the AU01 (top) and AU04 (bottom) values observed for each focus type (individual panels), for each syllable type (stressed, pre-stressed, post-stressed or other, indicated by shapes) in focalized or non-focalized elements (indicated by colors)

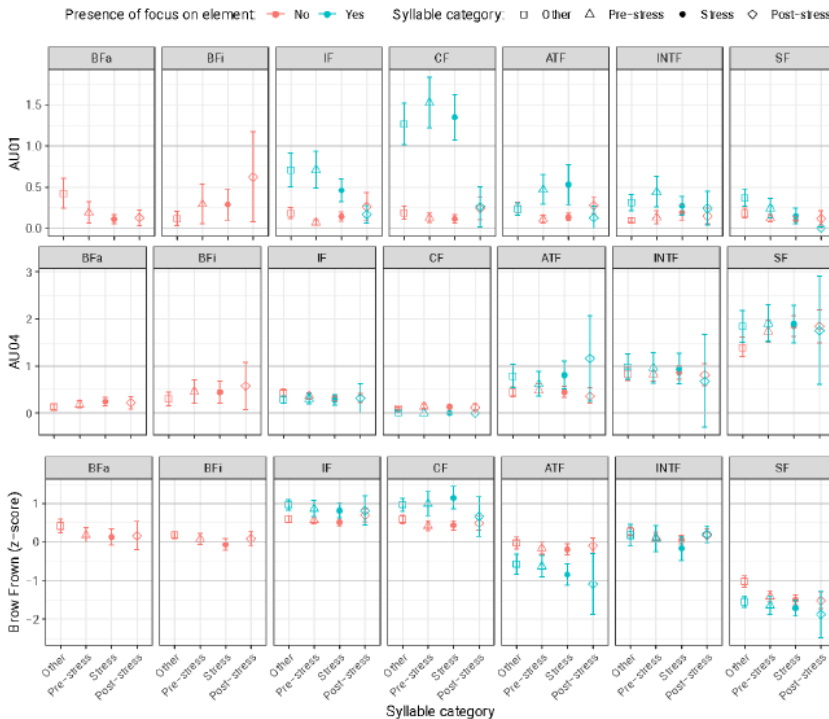
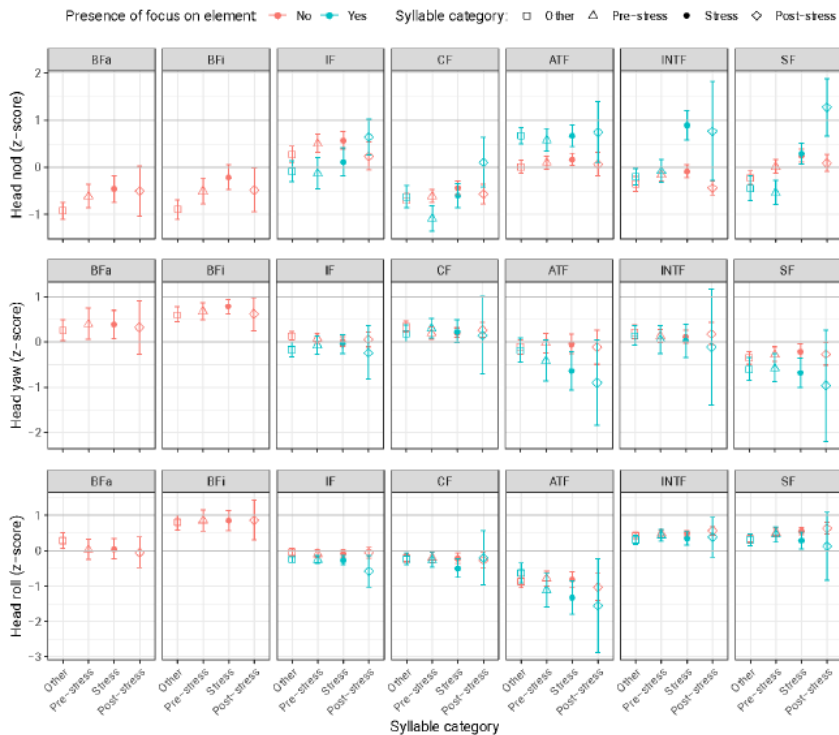


Figure 13 presents head rotation movements. It shows an up and down (nod) movement of head during the CF, while there is a tendency of speaker to turn their head on the left side (yaw movement) and roll it on the right (roll movement) for ATF; these movements are not specific to the focalized element.

Figure 13 – Mean and confidence interval of head rotations (from top to bottom: nod, yaw, roll) values observed for each focus type (individual panels), for each syllable type (stressed, pre-stressed, post-stressed or other, indicated by shapes) in focalized or non-focalized elements (indicated by colors). NOTE: head nod movement is negative for head up, positive for head down; yaw: is positive for head turn on the right side of the speaker, negative on left; roll: tilt of the head on the left is positive, on the right is negative.



A summarizing table follows, with the main acoustic and visual features for each focus type:

Focus types in Brazilian Portuguese

PARAMETERS	ACOUSTIC		VISUAL	
	F0	DURATION	EYEBROWS	HEAD
IF	Melodic falling movement on the focus stressed syllable	Expressive duration increase on the focus stressed syllable	X	X
CF	Rise-fall pattern over the focus stressed and prestressed syllables, respectively, followed by a <i>deaccenting</i> in the post focal position	Very expressive duration increase over the focus stressed syllable	Eyebrows raise, synchronized to the F0 movement.	Head nod, synchronized to the F0 movement
ATF	Rise-fall pattern over the focus stressed and prestressed syllables, respectively, with a gradual falling movement until the end of the sentence	Expressive duration increase over the focus stressed syllable	Eyebrows assymetry, marking the beginning of the focused element and gradually losing intensity	Head yaw and roll
INTF	Melodic rise over the focus stressed syllable	Expressive duration increase over the focus stressed syllable	No type-specific cues	X
SF	Double melodic rise over the focused element.	Very expressive duration increase over the focus stressed syllable	Brows frown, marking the beginning of the focused element	X

Focus Perception

As mentioned in section 2, a simplification process of the multinomial regression model was applied. Simplification steps led to a minimal adequate model based on the three main factors, plus the double interaction between focus type and presentation modality (the factor modality not having a significant main effect, but being kept in the model as part of a significant interaction). The triple interaction, as well as the double interactions between type & place, and between place &

modality do not improve significantly the model (the results of the LR tests were respectively: $LR\chi^2(96) = 69.8$, $p = 0.98$; $LR\chi^2(48) = 59.6$, $p = 0.12$; $LR\chi^2(24) = 32.7$, $p = 0.11$, and were thus deleted so to get this minimal adequate model, the ANOVA table of which is presented below (with type III sums of squares).

Table 1 – Analysis of deviance table for the effect of the minimal adequate model, presenting for each main factor and the interaction the respective likelihood ratio's χ^2 , the associated degrees of freedom (*df*), and probability (*p*)

	LR χ^2	<i>df</i>	<i>p</i>
Focus type	1621.8	16	< 0.05
Presentation modality	8.2	8	= 0.41
Focus place	94.9	12	< 0.05
Type * Modality	193.9	32	< 0.05

As we can observe in Table 1, the statistical analysis of the perception test answers highlighted significant effects of three factors: (i) the distribution of the participants' answers within the five focus categories; (ii) the position of the focalized constituent in the sentence; and (iii) the interaction between the type of focus and the presentation modality. The effect of these factors on the distribution of answers is displayed in Figures 14 and 15.

Figure 14 – Probability estimated by the model for each answer category (color / shapes), for each level of the factor “Focus place” (x-axis: “O professor”, “O professor de literatura”, “aplicar”, “final”)

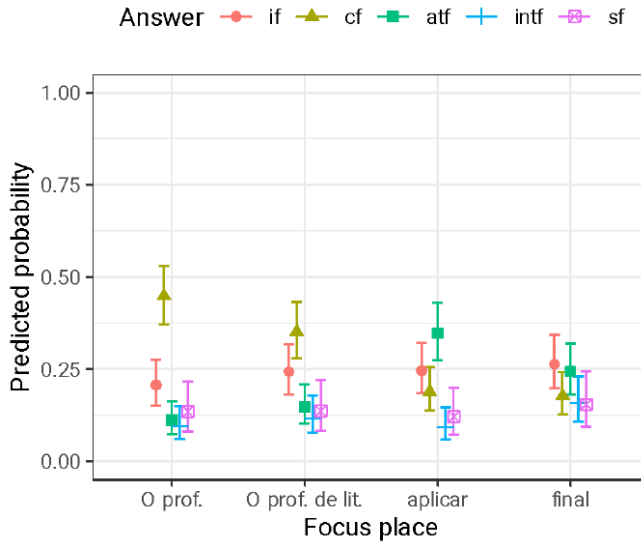


Figure 14 presents the effect of focus position on the probability distribution of answers: the effect is linked to the probability of contrastive and attenuated focus answers, which show a reverse evolution along the sentence. Participants do not change much the way they interpret Informational Focus, Interrogative Focus, or Surprise Focus according to the focus position in the sentence. The probability of Contrastive Focus answers is higher at sentence initial position, and show a marked decrease when focus is placed on “vai aplicar” or “final”. Inversely, the probability of Attenuated Focus answers is lower at sentence initial positions, and higher on “vai aplicar”, and to a lesser extent on “final”. The Contrastive Focus (CF) received a 0.6 overall identification ratio, with most confusions going to the Attenuated Focus (ATF) type, while ATF received the lowest ratio of correct guesses (about 0.4), with a considerable number of confusions with the two other assertive focus (IF and CF).

Figure 15 presents the effect of the interaction between the focus type and the presentation modality. The focus type has an important influence on the most probable type of answer selected by the participants, who were generally more likely to answer with the expected type,

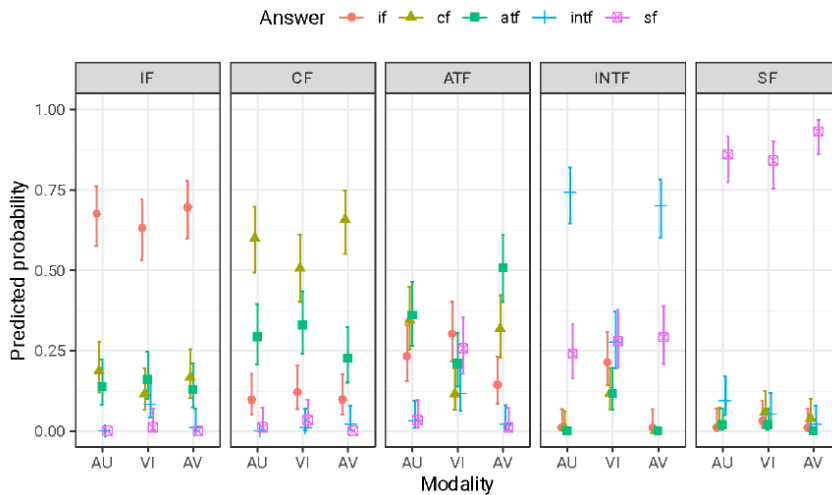
with differences linked to modality and type. Informational Focus (IF) and Surprise Focus (SF) show the clearest distinction among answer categories, since the identification levels are high for the expected type (respectively above 0.6 and 0.8) and consistent for the three modalities.

Informational Focus (IF) presents a clear identification ratio (around 0.7) in the three modalities. Noticeable confusions with IF as an answer were found among the other assertive focus types (i.e., CF and ATF), but also in visual-only presentation of INTF. The informational focus category may have been used as a default answer, in cases where prosodic meaning was more difficult to infer. This result could be explained by the Informational Focus's broad meaning, which may fit most stimuli with low informational content (e.g., visual INTF), as well as by its relatively unmarked prosodic and visual patterns which present the most neutral performances when compared to other types.

Both audio and visual cues individually allowed the identification of Contrastive Focus; identification ratio is however improved in the audiovisual modality, revealing the integration between both cues for this focus type. Regarding Attenuated Focus, results show that visual only condition does not allow the identification of the focus category (leading to a considerable number of confusion), while they are adequately interpreted when occurring simultaneously to audio cues, in the audiovisual modality, with identification levels higher than the audio-only condition.

We suppose that there is an ambiguity between Contrastive and Attenuated Focus meanings due to a generic interpretation of both types as a "contrast", since both focus categories convey a contrasting meaning with regard to the previous information. However, in the audiovisual modality, this ambiguity is reduced and the perceivers are better identifying an explicit (strong) contrast meaning for Contrastive Focus and an attenuated (weak) contrast meaning for Attenuated Focus.

Figure 15 – Estimated probability for each answer category (color / shapes), for the interaction between “Focus type” (individual panels) and modality (x-axis)



This semantic/pragmatic similarity between Contrastive and Attenuated Focus could be related to the similar melodic patterns found in these focus types, as previously described for Figures 3 and 4.

For Interrogative Focus (INTF), the effect of modality presentations was also observed, with audio cues only allowing its identification – visual cues do not contribute to the identification when presented alone, or in the audio-visual condition (no increase of the identification ratio in the AV condition, compared to the audio-only one, as shown in Figure 15). This pleads for the perceptual relevance of acoustic cues of INTF, but also shows that the visual cues are not detrimental: they are just not specific, and may thus have an impact in degraded auditory conditions (Miranda et al., 2021). Interrogative Focus visual performances also vary depending on each speaker’s production, as illustrated in Figure 8. The Interrogation Focus (INTF) received a global ratio of correct identification above 0.6, with most confusions being with SF, with which it shares its sentence type.

5. Discussion and Final Remarks

The analysis of acoustic and visual parameters on focus production in Brazilian Portuguese, along with the results of focus type perceptual identification, leads to some interesting findings. Multimodality plays a relevant role in focus production and perception, as observed during the data analysis. However, different parameters contribute to convey distinct semantic and pragmatic values. This general result reaffirms that speech and gesture are also aligned to provide different semantic and pragmatic meanings (Bergmann et al., 2014; Kelly et al., 2010; Özyürek et al., 2007).

The first semantic type described is Informational Focus (IF), which consists of a simple answer to a question previously asked. It was probably used, in the perceptual experiment, as a default answer for stimuli whose meaning was unclear to the perceivers. This correlation between Informational Focus and a neutral semantic meaning is linked to its acoustic and visual realizations, characterized by less marked performances. Its melodic pattern presents a falling movement over the focus stressed syllable, also performed with a duration lengthening. Nonetheless, the fundamental frequency reaches less prominent levels, and the duration increase is smaller, compared to the other focus types. For visual parameters, there were no systematic facial movements detected.

Contrastive Focus (CF) and Attenuated Focus (ATF) established a complementary relationship in the perceptual scope. Those types present a close semantic value since both convey a contrast notion. Contrastive Focus provides a stronger correction meaning to the utterance, considering previous information as false while Attenuated Focus (ATF) presents a weaker, more complex, semantics since it implies that both previous and new information are potentially true in the discourse. The perceptual experiment showed a correlation between the contrastive focus meaning and the focus position at the beginning of the sentence as well as between the attenuated focus meaning and the middle and final positions of focus elements in the sentence. Confusions between CF and ATF often occur in the visual modality. However, when acoustic cues are simultaneous to the visual ones (AV condition), perceivers improve their identification of both types,

evidencing a collaboration between modalities. Visual cues for ATF, not identified in the visual modality (even at the ATF preferred positions), are used for identification along the acoustic ones in the AV condition. As for CF, its identification is improved in the audiovisual condition; the improvement is particularly strong at the middle or end of the sentences, positions tending to a higher ATF in the A and V conditions.

Considering the acoustic production of CF and ATF, their semantic similarities are reflected in their similar melodic realizations, both presenting a rise-fall pattern over the stressed syllable, with a different implementation, though, on the post-focal part. The rising movement over the focus pre-stressed syllable is followed by a steep falling movement on the contrastive focus stressed syllable, presenting a deaccenting in the post-focus position, while for attenuated focus the falling movement is shallower, reaching its lower level at the end of the sentence. We hypothesize that this distinct implementation of the falling movement may explain the complementary distribution between CF and ATF in the sentence position: when the focused element is at the beginning of the sentence, the CF identification reaches the highest levels, since the post-focal material shows a longer, clearer, deaccenting movement produced until the end of the sentence; when the focus is produced in the middle or final positions, the post-focal extension is comparatively reduced or absent, diminishing or impeding a clear observation of deaccenting. Apart from the melodic realization, there is also a distinction in the duration parameter, with lengthening mainly in the focus stressed syllable. This duration lengthening reaches a higher level in the contrastive focus than in the attenuated focus. Regarding the visual parameters, contrastive and attenuated focus present both eyebrow raise and head movements, although with different configurations. Contrastive Focus is characterized by raised eyebrow, as already observed by Cruz et al. (2015) for the European Portuguese, and a head nod synchronized to the melodic pattern, reaching its peak at the pre-stressed syllable, followed by a decrease over the stressed one. For the Attenuated Focus, the main movement detected was a head-side tilt (yaw and roll movement), marking the beginning of the focused element and gradually losing intensity until the end of the sentence. As presented in the Introduction section, many studies show this alignment between prosodic aspects, as pitch accents and also duration increase in prominent syllables, with the gestures

apexes (Alexanderson et al., 2013; De Ruiter, 1998; Esteve-Gilbert et al., 2017; Loehr, 2012; Pouw & Dickson, 2019; Pouw et al., 2021). This seems to be reinforced by our research.

For the interrogative focus types (Interrogative Focus and Surprise Focus), we highlighted different melodic patterns: a simple rise on the focus stressed syllable for Interrogative Focus, and a double melodic rise, over the pre-stressed and the stressed syllables for Surprise Focus. Yet, SF presents an important duration increase, mainly on the stressed syllable. For their visual performance, the Surprise Focus presents a distinctive frown, while Interrogative Focus does not show systematic visual characteristics. Perceptual results showed that Surprise Focus is the most prototypical type, presenting high identification levels in all three modality conditions, while Interrogative Focus show a higher identification ratio in the acoustic condition, opposed to a non-functional visual cue in both V and AV conditions.

Thus, the melodic implementation specificities and the duration differences, as well as distinct visual performances are aspects to be investigated further, typically to be systematically tested with a resynthesis approach, which will allow establishing the relevance of each parameter for the perception of each focus type, as conducted in Kraemer and Swerts (2004) and in Prieto et al. (2015). For the present paper, its main goal was to provide an acoustic and visual description of the five focus types in Brazilian Portuguese, both from a production and perception perspectives.

Conflict of interests

The authors declare they have no conflict of interest.

Credit Author Statement

We, Manuella Carnaval, João Antônio de Moraes and Albert Rilliard, hereby declare that we do not have any potential conflict of interest in this study. We have all participated in the study conceptualization, methodology study design, writing and editing. Specifically, Manuella Carnaval was responsible for data collection and validation, João Antônio de Moraes supervised the project and Albert Rilliard conducted statistical data analysis.

References

- Alexanderson, S., House, D., & Beskow, J. (2013). Aspects of cooccurring syllables and head nods in spontaneous dialogue. *Proceedings of 12th International Conference on Auditory-Visual Speech Processing (AVSP2013)*. https://www.isca-speech.org/archive_v0/avsp13/papers/av13_169.pdf.
- Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015). Cross-dataset learning and person-specific normalisation for automatic action unit detection. *Proceedings of 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG 2015)*. <https://doi.org/10.1109/FG.2015.7284869>.
- Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. *Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 59-66). <https://doi.org/10.1109/FG.2018.00019>.
- Bergmann, K., Kahl, S., & Kopp, S. (2014). How is information distributed across speech and gesture? A cognitive modeling approach. *Cognitive Processing*, 15(1: Special Issue: Proceedings of KogWis 2014), S84-S87. <https://pub.uni-bielefeld.de/record/2700040>.
- Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (Version 6.1.16) [Computer software]. <http://www.praat.org/>.
- Bolinger, D. L. (1954). English prosodic stress and Spanish sentence order. *Hispania*, 37(2), 152-156. <https://doi.org/10.2307/335628>.
- Borràs-Comes, J., & Prieto, P. (2011). 'Seeing tunes.' The role of visual gestures in tune interpretation. *Laboratory Phonology*, 2(2), 355-380. <https://doi.org/10.1515/labphon.2011.013>.
- Carnaval, M., Moraes, J. A., & Rilliard, A. (2019). Marcação de foco estreito e o acento secundário em interrogativas totais no português do Brasil. *Working Papers em Linguística*, 19(2), 136-167. <https://doi.org/10.5007/1984-8420.2018v19n2p136>.
- Chomsky, N. (1971). Deep structure, surface structure and semantics interpretation. In D. D. Steinberg, & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 183-216). Cambridge University Press.
- Cochran, W. G., & Cox, G. M. (1992). *Experimental designs*. 2nd ed., Wiley classics library ed. Wiley.
- Crawley, M. J. (2013). *The R book*. 2nd ed. Wiley.
- Crespo-Sendra, V. C., Kaland, C., Swerts, M., & Prieto, P. (2013). Perceiving incredulity: The role of intonation and facial gestures. *Journal of Pragmatics*, 47(1). <https://doi.org/10.1016/j.pragma.2012.08.008>.

- Cruz, M., Swerts, M., & Frota, S. (2015). Variation in tone and gesture within language. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. <http://repositorio.ul.pt/bitstream/10451/25020/1/ICPHS0452.pdf>.
- De Ruiter, J. P. (1998). *Gesture and speech production* [Doctoral dissertation]. Katholieke Universiteit.
- Dik, S. (1980). On the typology of focus phenomena. In T. Hoekstra, H. van der Hulst, & M. Moortgat (Eds.), *Perspectives on Functional Grammar* (pp. 41–74). De Gruyter. <https://doi.org/10.1515/9783112329603-005>.
- Dohen, M., & Lævenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, 52(2–3), 177–206. <https://doi.org/10.1177/0023830909103166>.
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system: The manual*. Research Nexus.
- Elordieta, G., & Irurtzun, A. (2009). The prosody and interpretation of non-exhaustive narrow focus in Basque. *Anuario Del Seminario de Filología Vasca Julio de Urquijo*, 43(1–2), 205–230. <https://ojs.ehu.eus/index.php/ASJU/article/view/1692>.
- Esteve-Gibert, N., Borrás-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America* 141, 4727–4739. <https://doi.org/10.1121/1.4986649>.
- Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J Speech Lang Hear Res*, 56(3), 850–864. [https://doi.org/10.1044/1092-4388\(2012/12-0049\)](https://doi.org/10.1044/1092-4388(2012/12-0049)).
- Fernandes, F. R. (2007) *Ordem, focalização e preenchimento em português: Sintaxe e prosódia* [Doctoral dissertation]. State University of Campinas. <https://doi.org/10.47749/T/UNICAMP.2007.398459>.
- Frota, S., & Vigário, M. (2000). Aspectos de prosódia comparada: Ritmo e entoação no PE e no PB. In R. V. Castro, & P. Barbosa (Eds.), *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística* (pp. 533–555). Associação Portuguesa de Linguística (APL). <https://apl.pt/wp-content/uploads/2017/12/1999-35.pdf>.
- Gries, S. T. (2013). *Statistics for linguistics with R: A practical introduction*. 2nd revised ed. De Gruyter Mouton.
- Gussenhoven, C. (2006). Types of focus in English. In C. Lee, M. K. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 83–100). Springer.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, 3(2), 199–244. <https://doi.org/10.1017/S0022226700016613>.

- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. 6th print. The MIT Press Classics.
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21(2), 260-267. <https://doi.org/10.1177/0956797609357327>.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Kay (Ed.), *The Role of Nonverbal Communication* (pp. 207-227). De Gruyter Mouton.
- Krahmer, E., & Swerts, M. G. J. (2004). More about brows: A cross-linguistic analysis-by-synthesis study. In Z. Ruttkey, & C. Pelachaud (Eds.), *From brows to trust: Evaluating Embodied Conversational Agents, Human-Computer Interaction Series*, (pp. 191-216). Kluwer Academic Publishers.
- Krahmer, E., & Swerts, M. G. J. (2009). Audiovisual prosody: Introduction to the special issue. *Language and Speech*, 52(2-3), 129-133. <https://doi.org/10.1177/0023830909103164>.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3-4), 243-276. <https://doi.org/10.1556/ALing.55.2008.3-4.2>.
- Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511620607>.
- Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Journal Laboratory Phonology*, 3, 71-89. <https://doi.org/10.1515/lp-2012-0006>.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/H/bo3641188.html>.
- Moraes, J. A. (2006). Variações em torno de tema e rema. *Cadernos do IX Congresso Nacional de Lingüística e Filologia, Universidade do Estado do Rio de Janeiro*, 279-289.
- Moraes, J. A., Carnaval, M., & Coelho, A. B. B. (2015). A manifestação prosódica do foco em interrogativas totais no Português do Brasil e sua percepção. *ReVEL*, 10(spe), 170-194. <http://revel.inf.br/files/25628f323ed484f9952532a1604fbb93.pdf>.
- 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), 3-23. <https://doi.org/10.1177/0023830919898886>.

- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Foris Publications.
- Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, 19(4), 605-616. <https://doi.org/10.1162/jocn.2007.19.4.605>.
- Pouw, W., & J. A. Dixon. (2019). Quantifying gesture–speech synchrony. *Proceedings of the 6th Meeting of Gesture and Speech in Interaction* (pp. 68-74). Germany.
- Pouw, W., De Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., & Dixon, J. A. (2021). Gesture– speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, 1491(1), 89-105. <https://doi.org/10.1111/nyas.14532>.
- Prieto, P., Pugliesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. (2011). Crossmodal prosodic and gestural contribution to the perception of contrastive focus. *Interspeech 2011*, 977-980. <https://doi.org/10.21437/Interspeech.2011-397>.
- Prieto, P., Pugliesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *Journal of Phonetics*, 49, 41-54. <https://doi.org/10.1016/j.wocn.2014.10.005>.
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, 3, 371-405. <https://doi.org/10.1017/S0952675700000695>.
- Shattuck-Hufnagel, S., Ren, A., & Tauscher, E. (2010). Are torso movements during speech timed with intonational phrases? *Proceedings of the Speech Prosody 2010* (paper 974). https://www.isca-speech.org/archive/speechprosody_2010/shattuckhufnagel10_speechprosody.html.
- Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01514>.
- Tenani, L. (2002). *Domínios prosódicos do português do Brasil: Implicações para a prosódia e para a aplicação de processos fonológicos* [Doctoral dissertation]. State University of Campinas. <https://doi.org/10.47749/T/UNICAMP.2002.253138>.
- Truckenbrodt, H., Sandalo, F., & Abaurre, B. (2009). Elements of Brazilian Portuguese intonation. *Journal of Portuguese Linguistics*, 8(1), 75-114. <https://doi.org/10.5334/jpl.122>.

- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. 4th ed. Springer.
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209-232. <https://doi.org/10.1016/j.specom.2013.09.008>.

Recebido em: 26/10/2021

Aprovado em: 02/03/2022

