forum acusticum 2026
8-12 September · Graz, Austria

Exploring Intonemes in Brazilian Portuguese

Hansjörg Mixdorff[1], Pablo Arantes[2], Luciana Lucente[3]

[1]*Berlin University of Applied Sciences, Germany*
[2]*Federal University of São Carlos, Brazil*
[3]*Federal University of Minas Gerais, Brazil*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Isačenko and Schädlich describe German F0 contours as communicatively motivated tone switches aligned with accented syllables. Stock and Zacharias link these switches to phonologically distinct intonational elements, or intonemes, with three main classes distinguished by sentence modality: information, contact and non-terminal intoneme. Using the Fujisaki model, tone switch timing is captured relative to the syllabic grid mostly by the alignment and amplitudes of the accent component. In the current study we further explore the tone switch approach by applying it to Brazilian Portuguese. Based on acoustic results we developed synthetic stimuli which covered a continuum of utterances between the three categories. Analysis of stimuli by prosody experts of BP showed that the corner stimuli indeed reflect the three intended categories well. In an ensuing perceptual study, we examined whether native speakers of BP were also able to decode the three different basic intonemes. While Information and Contact intoneme were robustly identified, the Non-terminal intoneme was more elusive in naïve perception, despite consistent expert recognition. This asymmetry reflects functional differences between sentence mode and discourse-level intonational cues and the constraints by the identification task that brought these two dimensions together in one experiment.

Keywords: *Fujisaki model, Brazilian Portuguese, intonemes, prosody perception*

1. Introduction

Mixdorff [1] combined a phonological system with a suitable F0 parameterization approach. He followed the works of Isačenko and Schädlich [2] and Stock and Zacharias [3] in which a given F0 contour in German is described as a sequence of communicatively motivated tone switches, major transitions of the F0 contour aligned with accented syllables. Tone switches can be thought of as the phonetic realization of phonologically distinct intonational elements, the so-called "intonemes". In the original formulation by Stock and Zacharias, depending on their communicative function, three classes of intonemes are distinguished, namely the N↑intoneme ("non-terminal intoneme", signaling incompleteness and continuation, rising tone switch), I↓ intoneme ("information intoneme" at declarative-final accents, falling tone switch, conveying information), and the C↑ intoneme ("contact intoneme") associated, for instance, with question-final accents, rising tone switch, establishing contact). Hence intonemes in the original sense mainly distinguish sentence modality, although there exists a variant of the I↓ intoneme, I↓(E) which denotes emphatic accentuation and occurs in contrastive, narrowly focused environments. Intonemes for reading style speech are predictable by applying a set of phonological rules to a string of text as to word accentability and accent group formation. Other F0 transitions – termed 'pitch interrupters' by Isačenko and Schädlich - will occur at phrase boundaries or in unstressed syllables where they do not have the same prominence-lending effect as tone switches).

Based on this concept, Mixdorff & Jokisch [4] developed a model of German prosody, anchoring prosodic features such as F0, duration, and intensity to the syllable as a basic unit of speech rhythm. In order to quantify the interval and timing of the tone switches with respect to the syllabic grid, the framework adopts the Fujisaki model for parameterizing F0 contours [6]. Brazilian Portuguese has been extensively studied with respect to intonation and sentence modality, and characteristic contours associated with declaratives, questions, and other pragmatic functions have been described in

Corresponding author: hmixdorff@bht-berlin.de.

considerable detail. The present study does not aim to establish the existence of such distinctions in Brazilian Portuguese. Rather, it investigates whether these communicative functions can be interpreted within the phonological-parametric framework of intonemes originally proposed for German. More specifically, we ask whether a small set of functionally defined intoneme categories can account for the observed Brazilian Portuguese patterns and whether these categories remain perceptually recoverable after controlled resynthesis. In this sense, the study constitutes an analysis-by-synthesis evaluation of intoneme categories in Brazilian Portuguese.

2. Methods

2.1 Speech material

Our experiences with the realization of intonemes in German showed that they mainly differentiate sentence modality. Hence, as a starting point for defining the underlying intonemes in BP we examined their realizations in the pre-existing BP corpus of attitudinal prosody [7]. The obvious advantage is that we always look at the same target phrase "uma ba'nana" and will be able to closely examine how the prominent syllables interact with the F0 contour as represented by the underlying accent commands. As a working hypothesis we assume that neutral yes/no questions and neutral statements are characterized by opposites of intonemes, that is, the Contact Intoneme and the Information Intoneme. As we stated earlier, in German the Contact Intoneme features a rise of F0 to a high level, later in the underlying accented syllable.

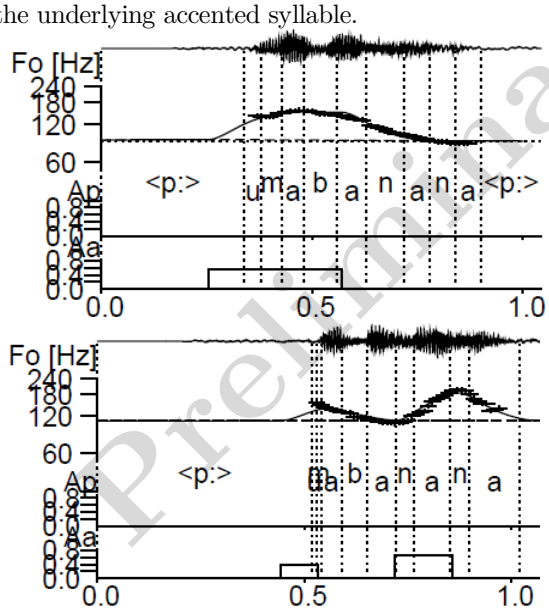

Figure 1: Examples of analysis for statement (top) and question (bottom).

In contrast, Brazilian Portuguese features a rise-fall pattern that is closely aligned with the underlying syllable. Acoustically it resembles the emphatic Information Intoneme in German. Statements in BP are mostly associated with a fall into the accented syllable. We find these two patterns well represented in the 12 speakers of BP that we examined. Figure 1 shows examples produced by male speaker S47. As can be seen, the onset of the accent command in the statement case (information intoneme) is far away from the accented syllable, the accent command even ends before the syllable onset of the (first, accented) syllable "na" to realize the fall into its low nucleus. This is something we also witnessed with other speakers. In the question case, the accent command is tightly aligned with the accented syllable. In the question case in Figure 1, we also see a smaller prominence on "uma" associated with a second accent command at utterance onset.

Based on the considerations from the attitude data, we recorded productions from two male and two female native speakers of BP. In line with the BP attitude data, we asked the speakers to produce the target phrase "uma banana" as an assertion, a question and as part of a list, yielding tokens of the non-terminal intoneme.

The participants read each of the following sentences ten times:

João olhou a fruta e afirmou: uma banana.

(engl. João looked at the fruit and affirmed: a banana.)

João olhou a fruta e perguntou: uma banana?

(engl. João looked at the fruit and asked: a banana?)

João listou frutas: banana, laranja e maca.

(engl. John listed fruits: banana, orange, and apple.)

Hence, in the continuation/non-terminal condition, banana occurred list-medially, ensuring a post-tonic rise typical of non-terminal-phrasing. Each token was recorded and the target phrase excised from the context, segmented on the phone level using WebMAUS Basic [7], and then F0 contours were extracted using Praat [9] with the standard method at a spacing of 10 ms. F0 contours were checked and then decomposed using the Fujisaki model [10]. For our purpose, we concentrated on the final accent command associated with the accented syllable 'na', that is, the second syllable of "ba**na**na". We calculated means and standard deviations of the distance between the syllable onset and the accent command onset, as well as the syllable offset and the accent command offset, but also accent command amplitude *Aa*. The three configurations are displayed in Table 7. As can be seen, the Information intoneme implies an accent command offset on average 22 ms after the onset of 'na'. The Contact intoneme has its accent command onsets and offsets close to the syllable onset and offset, respectively, whereas the Non-terminal intoneme is characterized by an accent command onset at 121 ms into the syllable 'na', and an offset much later towards the end of the word, at 146 ms after the accented syllable. Mean *Aa* at 0.471 is largest for the contact intoneme, but it does not differ greatly from the Information and Non-terminal intonemes, respectively

(0.427 and 0.435), and corresponds roughly to an interval of 6 semitones.

2.2 Stimulus Creation

Based on the values from the natural utterances we created three corner stimuli by superimposing the accent command onto a standard phrase baseline that remained unchanged for the continuum of intermediate stimuli that we subsequently produced. Figure 9 shows a schematic of the F0 contours associated with the three corner stimuli. Instead of manipulating the F0 contour of one of the natural utterances, we rather employed a monotonous synthetic substrate at F0=122.7 Hz, generated with MBROLA diphone synthesis [11] male Brazilian Portuguese voice **br1**, using segmental durations from the natural utterances that were left unchanged, and manipulated F0 within the *FujiParaEditor* tool [12]. The stimulus material was produced at 16 kHz/16bit.

Table 1: Alignment of accent command with accented syllable 'na' for the three types of intonemes, as well as accent command amplitude, command duration as well as syllable duration. Alignment mentioned in the text is marked in red.

Intoneme		t1relon [ms]	t2relon [ms]	t1reloff [ms]	t2reloff [ms]	*Aa*	accent duration [ms]	syllable duration [ms]
Information	mean	-220	22	-394	-152	0.427	242	174
	SD	88	20	103	33	0.182	81	29
Contact	mean	69	241	-138	34	0.471	172	207
	SD	63	41	56	36	0.169	34	25
Non-terminal	mean	121	,384	-117	146	0.435	263	238
	SD	137	186	115	162	0.211	70	36

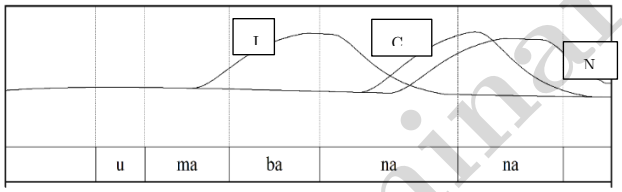

Figure 2: Schematic of F0 contours for corner stimuli for **I**nformation, **C**ontact and **N**on-terminal intonemes and their alignment with underlying syllables, drawn with Praat.

This yielded 9 linearly spaced intermediate stimuli between information and contact intoneme (see Figure 3), as well as 4 intermediate stimuli between contact intoneme and non-terminal intoneme.

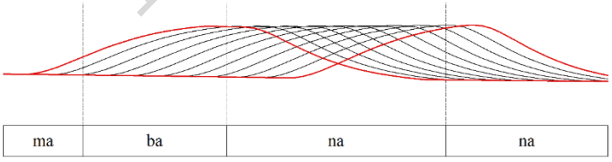

Figure 3: Schematic of F0 contours for stimulus continuum between **I**nformation and **C**ontact intonemes (in red) and their alignment with underlying syllables, drawn with Praat. There are nine intermediate stimuli, numbers I_C_01-09 in Table 2.

The different numbers of intermediate stimuli are due to the temporal difference between the former two being much larger than for the latter ones, see Figure 2. Furthermore, we added stimuli where we only modified the offset of the accent command, but not the onset, to examine whether the offset was perceptually more salient than the onset (see Figure 3).

This resulted in 22 different stimuli. Since the non-terminal intoneme occurred relatively seldom, we added the series between contact and non-terminal intoneme a second time, yielding 27 stimuli which were repeated once, resulting in a total of 54 trials.

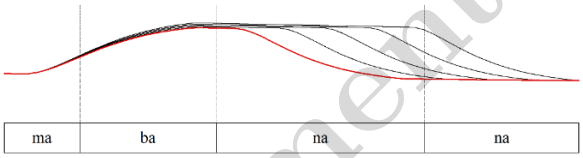

Figure 4: Schematic of F0 contours for stimuli starting from Information Intoneme (in red), where only T2 is modified, drawn with Praat, stimuli I_C_14-17 in **Table 2**.

2.3 Experiment Design

A *Psytoolkit* [13][14] experiment was programmed to run the study. Participants listened to each stimulus once and had to decide whether it sounded like a statement, a question or an unfinished utterance. The latter was described as sounding like a speaker being interrupted or reciting the first items in a list. The motivation of the study was given as exploring whether speech synthesis conveyed the three different categories convincingly. Only after giving their consent to participate explicitly by clicking a checkbox, subjects gained access to the experiment. We collected bio-information such as age, gender and native language. Each trial consisted of stimulus playback and subsequent choice of one of the three categories, by pressing the respective number button. A progress bar was displayed on each trial.

3. Stimulus Evaluation by Experts

Five experts in Brazilian Portuguese prosody independently evaluated the synthesized corner and intermediate stimuli. All experts consistently identified three distinct intoneme categories corresponding to Information, Contact, and Non-terminal contours. In particular, the Non-terminal stimuli were repeatedly described using terms such as *continuação, não-final*, or *frase interrompida* (continuation, non-final, interrupted phrase), indicating a shared interpretation of non-finality, despite presentation as an isolated word.

Beyond categorical labeling, expert judgments converged on a set of perceptual dimensions relevant for intoneme identification. Category assignment was primarily guided by contour shape and tonal alignment

rather than absolute pitch height. In particular, continuation judgments were associated with rising or sustained post-tonic movement, whereas interrogative interpretations involved rising movement aligned to the tonic syllable. Variations in pitch height or accent command amplitude alone did not alter intoneme identity but instead modulated paralinguistic interpretations such as emphasis or evidentiality.

Several experts explicitly noted that the Non-terminal contour appeared less natural when presented in isolation, attributing this impression to the absence of phrase-level continuation beyond the target word. Importantly, such remarks concerned stimulus naturalness rather than category identity, which remained stable across experts. This dissociation suggests that continuation in Brazilian Portuguese relies on prosodic cues distributed across a broader phrase domain, rendering isolated realizations perceptually attenuated while preserving expert recognition.

Taken together, the expert evaluations establish the phonological validity of the three modeled intonemes and clarify the perceptual dimensions underlying their identification. These judgments provide a reference point for interpreting the naïve perception results reported in the following section.

4. Results of Perception Test

4.1 Naïve listener judgment

Having established the intoneme categories and their defining perceptual dimensions based on expert evaluation, we now turn to a perception experiment with naïve listeners using the same stimulus set. In contrast to the expert judgments, which were self-paced and allowed repeated playback and analytic inspection of individual stimuli, naïve listeners performed a randomized one-shot three-choice task, making a single categorical decision per presentation without replay. The experiment thus probes how robust the modeled contours are for untrained listeners under strict trial-by-trial decision constraints.

We collected 810 responses from 15 participants, all native speakers of BP (7 males, 8 females, aged 19-73, median of 41); 14 time-outs were excluded. Table 2 reports response proportions for all stimuli, with continua between Information-Contact (I_C) and Contact-Non-Terminal (C_N), as well as stimuli in which only the accent command offset T2 was modified (see Figure 3).

Corner stimuli for Information and Contact intoneme yielded clear majorities for assertion and question, respectively. Along the I→C continuum, judgments shifted monotonically from assertion to question, with a clear crossover toward the Contact pattern.

In contrast, responses along the C→N continuum showed only a modest increase in non-terminal judgments, which never exceeded 33%, even for the Non-terminal corner stimulus, where question judgments remained dominant.

If we look at the right-hand side of the table, it becomes obvious that extending T2 towards the value for the Contact intoneme does not significantly increase the votes for "question". That is, the falling flank of the accent command still signals an assertion. Hence, the Contact intoneme actually requires the narrow rise-fall pattern coupled with the accented syllable to be perceptible. In the case between Contact and Non-terminal intoneme, delaying the fall, does not change the perception in favor of "non-terminal".

Binary logistic regressions confirmed these patterns: Along the I→C continuum, the alignment predictor yielded a significant slope and a categorical boundary, whereas along the C→N continuum no crossover occurred within the tested range, with Contact judgments dominating throughout.

Reaction times were analyzed as an attention check. After excluding time-outs, RTs were consistent with engaged listening. Responses were faster for stimuli strongly cuing a dominant category ($r = -.22$, $p < .01$), but RTs did not vary systematically along the continua, indicating that response speed reflects category certainty rather than acoustic position.

Table 2: Results from the perception test, given as ratios for choices assertion, question and non-terminal for all stimuli involved. Right-hand side stimuli share T1 with the corner stimulus (information/contact) and T2 with the stimulus on the same line at the left.

stimulus	assert-ion	quest-ion	non-terminal	stimulus	assert-ion	quest-ion	non-terminal
Informat-ion	0.80	0.00	0.20				
I_C_01	0.92	0.04	0.04				
I_C_02	0.84	0.04	0.12	I_C_14	0.88	0.00	0.12
I_C_03	0.80	0.00	0.20				
I_C_04	0.83	0.00	0.17	I_C_15	0.68	0.08	0.24
I_C_05	0.68	0.08	0.24				
I_C_06	0.54	0.27	0.19	I_C_16	0.77	0.12	0.12
I_C_07	0.23	0.54	0.23				
I_C_08	0.08	0.76	0.16	I_C_17	0.77	0.12	0.12
I_C_09	0.12	0.76	0.12				
contact	0.12	0.85	0.04				
C_N_10	0.12	0.75	0.14	C_N_18	0.08	0.76	0.16
C_N_11	0.19	0.65	0.15				
C_N_ 12	0.14	0.57	0.29	C_N_19	0.20	0.56	0.24
C_N_13	0.14	0.59	0.27				
non-terminal	0.15	0.52	0.33				

5. Discussion and Conclusions

In line with our acoustic results for BP, experts consistently recognized all three intoneme types. The continuation pattern could be differentiated based on tonal alignment cues, although it was sometimes described as *não-terminal, porém pouco natural* (non-terminal, but not very natural) reflecting reduced naturalness rather than category ambiguity.

In contrast, the perceptual study showed that **Information** and **Contact** intonemes were reliably identified along the manipulated continuum, whereas the non-terminal category proved to be perceptually unstable. Even the non-terminal corner stimulus elicited only a minority of non-terminal responses, while such responses occurred at low rates across much of the stimulus set, indicating a diffuse rather than categorical perceptual profile.

We described the effect of incompleteness as "somebody being interrupted or enumerating the first items in a list". Although we repeated the corner stimulus and those adjacent to it once, they were still perceived as appearing quite seldom in the experiment. As we already stated, evidence shows that the experts recover the intended three-way contrast even when naïve listeners do not.

We could attribute this instability to the fact that the continuation contour was realized list-medially in natural speech. In the resynthesized stimuli, the crucial post-tonic rise into the following word was often truncated, resulting in a subtle word-final downtrend that weakens the perceptual cue.

Looking at Figure 2 we see that in the prototypical stimulus for the **Non-terminal** intoneme the underlying accent command ends before the end of the word, leading to a subtle F0 downtrend which, of course, is even stronger for the preceding stimuli, towards the **contact** intoneme. In natural utterances, word-final amplitude typically tapers off, reducing the perceptual salience of the underlying F0 trajectory. In the MBROLA synthesis, however, by design all diphones are given the same rms energy, making the contour audible right until the end of the word.

An important methodological consideration is that the three-choice identification task brings together two functional dimensions. Information vs Contact primarily encodes sentence modality (statement vs question), whereas the Non-terminal intoneme encodes discourse continuation and incompleteness. A forced three-way choice therefore requires listeners to map stimuli onto possibly overlapping categories, which naturally favors robust speech-act interpretations and yields diffuse use of the continuation label. This task-induced conflict provides a principled account of why Information and Contact were identified reliably while non-terminal responses remained sparse and distributed across the stimulus set.

Taken together, our findings support the hypothesis that BP possesses a **distinct Non-terminal intoneme**, realized by a post-tonic high or plateau that signals syntactic or pragmatic continuation. Its weak perceptual salience in isolated tokens underscores its dependence on **phrase-level and contextual cues**. For naïve listeners, who relied mainly on the local shape (rise–fall vs. fall), the stimuli were ambiguous between question and assertion; for experts, the underlying tonal grammar remained accessible. It is also possible that knowledge of a dialect or foreign language with rising intonation in questions utterance-finally influenced the participants' judgments. The present findings do not imply that intonemes are realized phonetically in the same way across languages. On the contrary, the Brazilian Portuguese Contact intoneme differs substantially from its German counterpart in terms of contour shape and alignment. What appears to be preserved is not a specific F0 pattern but the communicative function associated with the category. The results therefore support the view that intonemes may be understood as functionally defined prosodic categories whose phonetic implementation is language-specific. From this perspective, the present study provides evidence that the phonological-parametric framework developed for German can be extended to Brazilian Portuguese despite substantial differences in surface realization. Future work will rebuild the non-terminal stimuli from the full list productions to preserve post-tonic continuation cues, and test whether contextual presentation restores naïve identification of the Non-terminal intoneme.

6. Acknowledgments

We thank João de Moraes, Plinio Barbosa, and Sandra Madureira for their expert commentaries, and all participants in the perceptual study. The perception experiments were conducted under the umbrella of Luciana Lucente's project "Investigação sobre as funções executiva e moduladora do oscilador glotal", Faculdade de Letras da UFMG, Brazil.

References

[1] H. Mixdorff, Intonation Patterns of German - Model-based Quantitative Analysis and Synthesis of F0-Contours. PdD thesis submitted to TU Dresden, 1998

[2] A. Isačenko, and H.J. Schädlich. Untersuchungen über die deutsche Satzintonation, Akademie-Verlag, Berlin, 1964.

[3] E. Stock, and C. Zacharias. 1982. Deutsche Satzintonation, VEB Verlag Enzyklopädie, Leipzig, 1982.

[4] H. Mixdorff, and O. Jokisch, "Building An Integrated Prosodic Model of German," Proceedings of Eurospeech 2001, vol. 2, pp. 947-950, Aalborg, Denmark, 2001.

[5] H. Fujisaki, and K. Hirose, "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," J. Acoust. Soc. Japan (E), Vol.5, No.4, pp.233-242, 1984.

[6] A. Rilliard, J. de Moraes, D. Erickson, and T. Shochi. "Prosodic analysis of Brazilian Portuguese attitudes," Speech Prosody 2012, 2012.

[7] T. Kisler, U.D. Reichel, and F. Schiel, "Multilingual processing of speech via web services," In: Computer Speech & Language, Volume 45, pp. 326–347, 2017

[8] P. Boersma, "Praat, a system for doing phonetics by computer". *Glot International* **5:9/10**, 341-345, 2001.C. Author, "The title of the conference paper," in *Proc. of the International Symposium on Topic,* (Chania, Greece), pp. 000–111, 2000.

[9] H. Mixdorff, "A novel approach to the fully automatic extraction of Fujisaki model parameters," Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP2000), vol. 3, pp. 1281.1284, Istanbul.

[10] T. Dutoit, V. Pagel, N. Pierret, F. Bataille and O. van der Vrecken. 1996, "The MBROLA project: towards a set of high-quality speech synthesizers free of use for non-commercial purposes," *Proceedings of Fourth International Conference on Spoken Language Processing.* ICSLP '96, Philadelphia, PA, USA, pp. 1393-1396 vol.3, doi: 10.1109/ICSLP.1996.607874, 1996.

[11] FujiParaEditor: http://public.bht-berlin.de/~hmixdorff/thesis/fujisaki.html.

[12] G. Stoet, "PsyToolkit - A software package for programming psychological experiments using Linux," Behavior Research Methods, 42(4), 1096-1104, 2001.

[13] G. Stoet, "PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments," Teaching of Psychology, 44(1), 24-31, 2017.