diacrítica
REVISTA DO CENTRO DE ESTUDOS HUMANÍSTICOS

Vol. 33, n.º 2, 2019, p. 4-18. DOI: doi.org/10.21814/diacritica.401

PHONOLOGIC REPRESENTATION AND SPEECH PERCEPTION: THE ROLE OF PAUSE

REPRESENTAÇÃO FONOLÓGICA E PERCEPÇÃO DE FALA: O PAPEL DA PAUSA

Geovana Soncin*
geovanasoncin@gmail.com

Luciani Tenani**
luciani.tenani@unesp.br

Larissa Berti***
berti.larissa@gmail.com

We investigate the perception of pauses at intonational phrase (IP) boundaries in Brazilian Portuguese (BP) and address the discussion about the relation between speech production and perception. Twenty adult subjects who were native speakers of BP, with no language disorders and no hearing complaints, took part in an experimental pause identification test. We used auditory stimuli in which the IP-boundary was marked by a combination of pause and pitch variation or pitch variation only. The results are as follows: (i) when stimuli consisted of a combination of pause with pitch variation at the IP-boundary, the pauses were significantly identified; (ii) when the stimuli did not have pitch variation at IP boundaries, the pauses were not significantly identified; and (iii) when only pitch variation occurred, without pause production, the pauses were identified at the tested boundary. These results support the argument that speech perception does not entirely depend on recovering an acoustic pattern and provide evidence of the importance of phonologic representation for the perception and organization of the perceived auditory stimulus. Based on the results, we argue that pause perception results from a perceptual illusion marked by the combination of different types of linguistic information at a phonetic-acoustic and representational level.

Keywords: Pause. Prosodic boundary. Intonation. Auditory perception. Brazilian Portuguese.

Investigamos a percepção de pausas em fronteiras de frase entonacional (IP) no Português Brasileiro (PB) e abordamos a relação entre produção e percepção da fala. Vinte sujeitos adultos, falantes nativos do PB, sem distúrbios de linguagem e sem queixas auditivas, participaram de um teste experimental de identificação de pausa. Usamos estímulos auditivos nos quais a fronteira IP foi marcada por uma combinação de pausa e variação de altura ou apenas variação de altura. Os

* Phonetics Laboratory, Department of Linguistic and Literary Studies, São Paulo State University (UNESP), São José do Rio Preto, São Paulo, Brazil. ORCID: https://orcid.org/0000-0003-4903-1919
** Phonetics Laboratory, Department of Linguistic and Literary Studies, São Paulo State University (UNESP), São José do Rio Preto, São Paulo, Brazil. ORCID: https://orcid.org/0000-0002-8487-0825
*** Speech-Language and Audiology Department, São Paulo State University (UNESP), Marília, São Paulo, Brazil. ORCID: http://orcid.org/0000-0002-4144-2804

resultados são os seguintes: (i) quando os estímulos consistiam em uma combinação de pausa com variação de altura no limite IP, as pausas foram identificadas significativamente (ii) quando os estímulos não tinham variação de altura nos limites IP, as pausas não foram identificadas significativamente, (iii) quando somente ocorreu variação de altura, sem produção de silêncio, foram identificadas pausas na fronteira testada. Esses resultados corroboram o argumento de que a percepção de fala não depende inteiramente da recuperação de um padrão acústico e evidencia a importância da representação fonológica para a percepção e organização do estímulo auditivo percebido. Com base nos resultados, argumentamos que a percepção de pausa resultaria da ilusão perceptiva marcada pela combinação de diferentes tipos de informações linguísticas, em nível fonético-acústico e representacional.

Palavras-chave: Pausa. Fronteira prosódica. Entoação. Percepção auditiva. Português brasileiro.

•

1. Introduction

The complexity involved in speech perception has been addressed by various research works dealing with different linguistic phenomena to answer a common driving question: what types of analysis procedures do listeners perform as they process the speech signal (Borden, Harris & Lawrence 1994)? Based on the results obtained, these works discuss different theoretical views on the unit of perception.

Traditionally, two theoretical approaches can be identified in speech perception studies: one assumes that speech perception is correlated with the recovery of an acoustic pattern, whereas the other conceives of speech perception as a co-relation between perception and production.

The first theoretical approach is represented mainly by the studies of Fant (1967), Morton and Broadbent (1967), and Stevens and Blumstein (1978). From this perspective, perception is conceived of as a sensory process related to the identification of an acoustic pattern from the physical signal presented to the human ear. Therefore, the physical signal plays a key role in the speech perception process, since it provides everything required for the listener to make distinctive evaluations and retrieve from memory an acoustic pattern for the perception of a given event. It follows that the auditory procedure supports the assumptions of this theoretical current and that the unit of perception is sensory-auditory.

The second approach has among its representatives Liberman and Mattingly (1985), Fowler (1986, 1996), Goldstein and Fowler (2003), Galantucci, Fowler and Turvey (2006). Apart from a few differences between these works, they share the concept that speech perception does not essentially consist of the recovery of an acoustic pattern through sensory-auditory identification; rather, recovering aspects of speech production is of key importance. To sustain this theoretical point of view, counterarguments for auditory theories are provided, mainly acoustic signal variability, coarticulation and the non-univocal relation between the acoustic signal and the phonetic segment.

First, the scholars consider the acoustic signal to vary to such an extent that the acoustic pattern cannot be identified exclusively through the relation between hearing and memory, since a given sound could be represented by several acoustic patterns, which should in turn comprise the wide arsenal supposed to represent our memory. Second, regarding coarticulation, considering that a superposition of elements occurs in the articulation of speech, they argue that one must take into account that the articulatory information of each segment can expand over a period of time and thus the ear should find the means to manage the identification of the articulated segments. However, studies show that coarticulation also allows us to observe non-univocity between the acoustic signal and the phonetic segment, given that, according to Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967), speech is not an acoustic alphabet whose elements are identified on a term-by-term basis by hearing but is in fact an effective code in which one particular piece of information in the acoustic signal can be identified as different phonetic segments or vice versa.

By pointing out these issues, the motor theories of speech perception indicate the insufficiency of the physical signal in regard to understanding the perception process and, as a result, they develop a theoretical framework uniting knowledge about speech production with perception. The manner in which production relates to perception varies according to the versions of the theory. For instance, Fowler (1986, 1996) and Goldstein and Fowler (2003) present the motor theory with some distinctive aspects compared to the classical version of Liberman and Mattingly (1985). They propose the realist theory of direct perception, in which perception and production share the same unit, consisting of the relation between physical and mental aspects; in other words, the unit of perception is equivalent to the unit of production and comprises phonetic and phonologic aspects at once, which had previously not been considered.

In any case, the concept of parity between perception and production proposed by Liberman and Mattingly (1985) and more recently theorized by Liberman and Whalen (2000) drives the underlying concept of the different versions of the motor theory, since the unit of perception is perceptual/motor, that is, perception and production are intrinsically related in the process of speech perception. Therefore, in the theoretical framework, motor gestures serve as references for perceptual interpretation. As a result, the motor theory settles previously unresolved cases for the auditory theory by showing that, despite the acoustic signal variability, the production of motor gestures would not change. Thus, in the cases in which the acoustic patterns are different but the articulatory gestures that have originated them in speech are the same, or vice versa, perception follows the decisions posed by articulation (Galantucci *et al.* 2006).

In the domain studied in this paper, that is, the investigation domain of perception of prosodic phenomena, several studies have suggested that the perception of the IP-boundary (Nespor & Vogel 1986, 2007; Selkirk 1984) in different languages, whose relevance is essential for language processing, is identified by listeners even when some acoustic information that defines that boundary is missing.[1]

[1] In different languages, the IP boundary is consensually performed by three acoustic cues: preboundary lengthening, pitch change and pause (Beckman & Pierrehumbert 1986; Hirst & di Cristo 1998; Ladd 1996; Pierrehumbert 1980).

According to Steinhauer, Alter and Friederici (1999), in German, even though the IP-boundary is typically set by a combination of three acoustic cues (preboundary lengthening, pitch change and pause), listeners have flexibility regarding the type of acoustic cue that helps them perceive this prosodic boundary, since it could be recognized even in the absence of a pause – the most robust cue to perceive this boundary (Peters 2005).[2] Concerning the pause, Martin (1970) analysed the judgement of pauses in spontaneous speech for English and observed that beyond the identified pauses that matched the instant of silence in the acoustic signal, some pauses were identified when no instant of silence existed and some instants of silence in turn were not identified as pauses.

Combined, the research performed by Steinhauer and others (1999), Peters (2005) and Martin (1970), regarding prosody-related phenomena, adds evidence against the arguments of the auditory theories, given that, on the one hand, the IP-boundary is identified even when no cues considered as important for marking this boundary are produced in the acoustic signal; on the other hand, pauses can be identified even when they are missing from the acoustic signal.

Within the context of theoretical discussions about speech perception, accounting for the results on the identification of intonational phrases in English and in German, this paper presents an experimental study about the relevance of pauses in the perception of this prosodic boundary in Brazilian Portuguese (BP).

In this language, the intonational phrase is characterized by a potential pause and pitch variations, which, from the perspective of intonation, function as nuclear accents—primarily as H+L* or L+H* (Frota & Vigário 2000; Serra 2009; Tenani 2002).[3] Conversely, preboundary lengthening did not have statistical relevance for the studies carried out, as opposed to the case of European Portuguese (Frota 2000, 2003; Moraes 2007; Serra 2009). Very few studies deal with perception of the intonational phrase in BP. However, Serra (2009) presents a result showing that the IP-boundary is perceived much more significantly when a pause exists.

To perform this experimental study, the following hypotheses have been taken into account:

> Hypothesis 1: since the identification of the IP-boundary is not restricted exclusively to the identification of existing pauses, we expect that the pitch variation that marks the IP-boundary in the acoustic signal in BP—especially the boundary tone L%—could be identified as a pause;
>
> Hypothesis 2: both pitch variations and the simultaneous presence of pause and pitch variations allow the identification of the IP-boundary; however, the reaction time spent identifying the boundary marked only by the presence of pitch variation would be greater in comparison to identifying the boundary through the presence of both acoustic cues.

[2] For German, Peters (2005) identified a perceptual hierarchy of acoustic cues to identify the intonational boundary, according to which the pause would be the most robust cue, followed by pitch changes and, finally, by preboundary lengthening.
[3] Nuclear accent, in the framework of Intonational Phonology (Ladd 1996), is defined as the most prominent pitch accent in a series of adjacent pitch accents.

This study, thus, aims to verify to what extent pitch variations at the IP-boundary could be perceived as pauses. Consequently, we are interested in checking whether the perception of a pause at an IP-boundary could be considered the effect of a perceptual illusion of other phonetic-acoustic and grammatical pieces of linguistic information.

As input for the experiments, we used sentences combined in pairs, marked by the same segmental chain, but different in terms of syntactic-semantic interpretation, whose identification is defined by prosodic phrasing.

Among its contributions, this paper introduces new data to descriptive studies related to prosodic boundaries and intonational phenomena in BP, a variety of Portuguese spoken by over 206 million people (IBGE 2016) yet scarcely studied, and presents evidence that allows discussions within the theoretical framework of speech perception about what would be a pause from a perceptual point of view.

2. Methods

The participants were twenty adult subjects, seven male and thirteen female, who were native speakers of BP. Ten of them were between the ages of 18 and 24, and the other ten were between 30 and 55 years old. All subjects participated in the experiment as volunteers and signed a document declaring their consent to participate in the study. The subjects were selected on the basis of convenience according to the following inclusion criteria: undergraduate linguistics students and high school teachers in the area of humanities, due to the ease of recruiting participants. We adopted the following exclusion criteria: detection of some language disorder by the first author and a reference to hearing or understanding complaints.

The experiment consisted of a perceptual test of pause identification in the auditory stimulus, characterized by a forced-choice task.

2.1. Experimental procedure

In the task, each participant answered whether he/she heard a pause within the sentences presented in the auditory stimulus. The answer was indicated by pressing the keys 1 or 2. The computer screen indicated that key 1 referred to the answer “with pause”, whereas key 2 referred to the answer “no pause”.

This test was performed using the software PERCEVAL (Version 5.0.30; André, Ghio, Cavé & Teston 2009). Installed on the computer, whose screen was placed before the participant during the execution of the experiment, the software simultaneously presented the auditory and visual stimulus and prompted each participant to answer the task. In addition to automatically controlling the presentation of the stimuli, the software measured the subjects’ reaction time to the requested task. During the experimental procedure, the participants used high-precision headphones connected to the computer that they used. The test was applied by including a practice sentence to familiarize the participants with the task. At the practice stage, the answers were not computed.

2.2. Stimuli

Three types of stimuli were prepared through sentence recording and editing, performed at the Phonetics Laboratory at São Paulo State University (LabFon/UNESP). The complete set of auditory stimuli was formed by the eight sentences listed below that were divided into four pairs, as presented in Table 1. The sentences of each pair are characterized by the same segment sequence. However, they offer two syntactic-semantic interpretation possibilities whose identification is defined by prosodic phrasing. Here ae the sentences of stimuli:

(1) Não, mereç-o saber
NEG deserve[PRS.1SG] know-INF
'No, I deserve to know.'

(2) Não mereç-o saber
NEG deserve[PRS.1SG] know-INF
'I do not deserve to know.'

(3) Aceit-o, obrigad-o
Accept[PRS.1SG] thank[PRS.1SG]
'I accept, thanks.'

(4) Aceit-o obrig-ad-o
Accept-[PRS.1SG] force-PTCP[M.SG]
'I'm forced to accept.'

(5) Isso apenas, ele resolve.
DEM[M.SG] only, he solve-PRS[3SG]
'Only this he can solve.'

(6) Isso, apenas ele resolve.
DEM[N.SG], only he solve-PRS[3SG]
'This [problem] only he can solve.'

(7) Vamos perder, nada foi resolv-id-o.
go-AUX[PRS.1PL] lose-INF nothing go-AUX[PRF.3SG] solve-PTCP[M.SG]
'We will lose [it], nothing was solved.'

(8) Vamos perder nada, foi resolv-id-o
go-AUX[PRS.1PL] lose-INF nothing go-AUX[PRF.3SG] solve-PTCP[M.SG]
'We will lose nothing, it was solved.'

Table 1. Pairs of contrastive sentences and prosodic phrasing.

Pairs	Sentences	Prosodic Phrasing
A	(1) Não, mereço saber.	[[não**]IP** [mereço saber]IP]U
	(2) Não mereço saber.	[[não mereço saber]IP]U
B	(3) Aceito, obrigado.	[[aceito**]IP** [obrigado]IP]U
	(4) Aceito obrigado.	[[aceito obrigado]IP]U
C	(5) Isso apenas, ele resolve.	[[isso apenas**]IP** [ele resolve]IP]U
	(6) Isso, apenas ele resolve.	[[isso**]IP** [apenas ele resolve]IP]U
D	(7) Vamos perder, nada foi resolvido.	[[vamos perder]**IP** [nada foi resolvido.]IP]U
	(8) Vamos perder nada, foi resolvido.	[[vamos perder nada**]IP** [foi resolvido.]IP]U

Source: Elaborated by the authors.

A male 35-year-old actor, a native speaker of BP, recorded these eight sentences. By editing the recorded audio, the three stimuli were obtained by manipulating two particular pieces of information: pitch variation and pause. This manipulation occurred at the IP-boundary. As shown in Table 1, among the eight sentences, six sentences (1, 3, 5, 6, 7, 8) display a non-final IP-boundary, and two (2 and 4) do not display this boundary.[4]

The description of the three types of stimuli, named A, B and C, is presented in Table 2.

Table 2. Types of auditory stimuli.

Type of stimulus	Description
Type A	sentence uttered with the sequence of tonal events H+L* L% and no pause production at the tested boundary of *I*.
Type B	sentence uttered with the sequence of tonal events H+L* L% and with pause production at the tested boundary of *I*.
Type C	sentence without the tested boundary of *I* and therefore without the sequence of tonal events H+L* L% and no pause inside.

Source: Elaborated by the authors.

The six sentences structurally characterized by the tested IP-boundary generated type A and type B stimuli, adding up to twelve stimuli, six of each type (6 x 2 = 12). Conversely, the two sentences without an IP-boundary generated only type C stimuli, totalling two stimuli of this type (2 x 1 = 2). Thus, the total sum of obtained stimuli was fourteen (12 + 2 = 14). The fourteen stimuli were presented by the software with three repetitions (14 x 3 = 42) to measure the consistency of responses.

Table 3 displays the sentences and stimuli obtained through them along with the acoustic description of the controlled variables. The sentences 1, 3, 5, 6, 7 and 8 generated type A and B stimuli, whereas the sentences 2 and 4 generated type C stimuli.

[4] We consider a "non-final IP boundary" to be an intonational phrase whose ending does not match the end of the phonological utterance, the highest constituent of the prosodic hierarchy proposed by Nespor and Vogel (1986, 2007). In this study, the final IP boundary which does not coincide with the final phonological utterance boundary is the prosodic context under study.

Table 3. Acoustic description of the auditory stimuli.

Sentence	Sentence phrasing	Type of stimulus	Pause length at the I-boundary (ms)	F0 variation at the I-boundary (Hz)
1	[[não]I [mereço saber]I]U	A	no pause	35.3
		B	860	27.3
2	[[não mereço saber]I]U	C	no pause	–
3	[[aceito]I [obrigado]I]U	A	no pause	48.9
		B	860	26.0
4	[[aceito obrigado]I]U	C	no pause	–
5	[[isso apenas]I [ele resolve]I]U	A	no pause	33.2
		B	860	56.5
6	[[isso]I [apenas ele resolve]I]U	A	no pause	56.5
		B	860	31.5
7	[[vamos perder]I [nada está resolvido.]I]U	A	no pause	72.3
		B	860	24.2
8	[[vamos perder nada]I [está resolvido.]IU	A	no pause	33.0
		B	860	33.2

Source: Elaborated by the authors.

The stimuli characterized by a pause combined with F0 variation were manipulated to obtain an instant of silence lasting 860 milliseconds. Therefore, a pause was produced by the actor according to the instructions given to him during the recording, and afterwards, the pause was manipulated through the software *Sound Forge 8.0* to achieve the duration mentioned above. A total of 860 milliseconds was identified by Chacon and Fraga (2014) as the average time for pause duration at the IP-boundary in BP. Regarding F0, this phonetic factor was not manipulated. The value presented in Table 3 was obtained through the Praat software (Version 4.2.28; Boersma & Weenick 2005) by considering the relative difference between the values of F0 peak and F0 valley at the word that received the nuclear accent of the intonational phrase in the sentence produced by the actor.

2.3. Data analysis

The results were obtained through an analysis of the following variables: (i) percentage of pause identification per auditory stimulus and (ii) average reaction time spent by the participants on each type of stimulus.

The data were treated based on descriptive and inferential statistical analyses. To compare the percentage of pause identification according to the type of stimulus, the Kruskal-Wallis nonparametric test was used, whereas for the comparison of the reaction time in pause identification according to the type of stimulus, the repeated measures ANOVA and the Bonferroni post hoc test were used. A 0.05 α value was established.

3. Results

The results related to the percentage of identified and non-identified pauses according to the three types of auditory stimuli are presented in Table 4.

Table 4. Pause identification by type of auditory stimulus.

Type of stimulus	Identified pause		Non-identified pause		Total
	Frequency	Percentage	Frequency	Percentage	
Stimulus A	194	53%	166	47%	360
Stimulus B	350	97%	10	3%	360
Stimulus C	4	3.3%	116	96.70%	120

Source: Elaborated by the authors.

The participants identified pauses mainly in stimulus B, marked by a combination of the tonal event sequence H+L* L% with pause at the IP-boundary, and did not identify pauses in stimulus C, marked by the absence of IP-boundary and pause. In the case of stimulus A, in which the sequence of tonal events H+L* L% was not combined with the pause at the IP-boundary, the participants tended to identify the pause, but at a lower rate.

The results received inferential statistical treatment by the software *Statistica* (version 7.0). Table 5 presents the results obtained. The Kruskal-Wallis nonparametric test was used to verify the significance level of the pause identification (dependent variable) according to the auditory stimuli (independent variable).

Table 5. Comparison between pause identification and auditory stimulus.

Type of stimuli	Pause identification (absolute value and %)	Sum of ranks (Kruskal-Wallis test)	Nonparametric Kruskal-Wallis test
A	194/360 (53%)	15480.00	
B	350/360 (97%)	22060.00	H (2.280) = 142.09, $p < 0.00$
C	4/120 (3.3%)	1800.00	

Source: Elaborated by the authors

The results of the nonparametric Kruskal-Wallis test show that a significant difference exists in the identification of presence/absence of pauses according to stimuli (H (2.280) = 142.09, $p < 0.00$). In Table 5, the sum of ranks shows a gradient in the identification of pauses between stimuli A, B and C. The highest sum of ranks occurred for stimulus B, followed by stimulus A and, finally, stimulus C, at the lowest rate. From this result, it follows that (i) the best result in detecting the presence of the boundary occurred when sentences were heard in the auditory stimulus with the sequence of tonal events H+L* L% and the presence of pauses at the I-boundary (stimulus B); (ii) the presence of the boundary was detected when sentences were heard in the auditory stimulus without a pause and only with the sequence of tonal events H+L* L% at the I-boundary (stimulus A); (iii) the performance of boundary detection was poor when sentences were heard in the auditory stimulus without pause and without the sequence of tonal events H+L* L% inside, that is, without the I-boundary (stimulus C).

In a multiple comparison, that is, when comparing all groups to each other, a significant difference between them was detected ($p < 0.00$ in all stimuli). Thus, the first hypothesis assumed for the execution of the experiment has been fully confirmed: the identification of I-boundaries in BP is not restricted to exclusive identification of the presence of pauses. Therefore, it was expected that the pitch variation that marks the I-boundary in the acoustic signal in this variety of Portuguese – especially by the tonal events H+L* L% – could be identified as a pause, though at a lower percentage than if the pitch variations were combined with a pause: a fact confirmed through the results obtained by the experiment.

Regarding the analysis criterion referring to reaction time, the repeated measures ANOVA had a significant effect only for the type of stimulus (Z (2.275) = 31.07, $p < 0.00$). No significant effect has been observed for repetitions, categorization of the presence/absence of pauses or the corresponding interactions. On the other hand, the Bonferroni post hoc analysis indicated that the reaction time spent differed significantly between all stimuli (all $p < 0.05$), according to the following descending order: Stimulus A > Stimulus B > Stimulus C. In Graph 1, the average reaction time (in milliseconds) is shown for the three types of stimuli, indicating a significant difference between them.

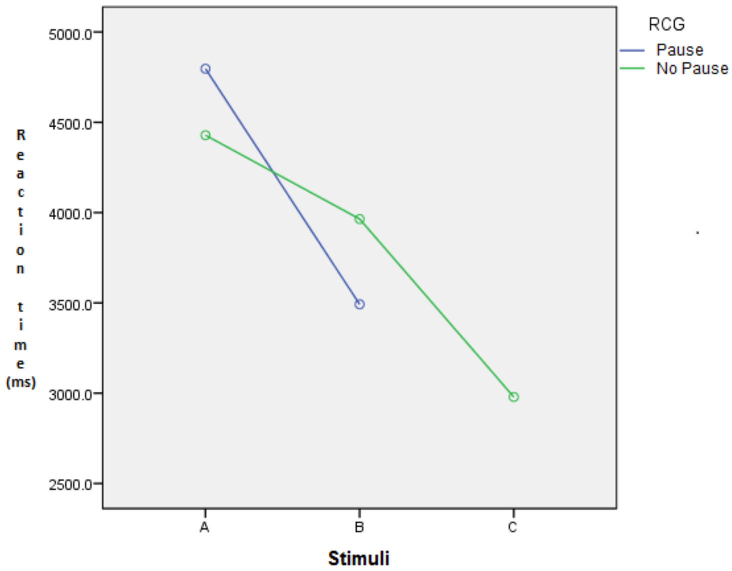

Figure 1. Average reaction time in milliseconds by type of stimulus.

The data concerning reaction time, shown in Figure 1, fully shows the second assumed hypothesis, given that, as expected, both pitch variations and the simultaneous presence of pause and pitch variations led to identification of the IP-boundary; however, the reaction time spent identifying the boundary marked only by the presence of pitch variation (stimulus A) was greater in comparison to the time spent identifying the boundary marked by the presence of both acoustic cues (stimulus B).

4. Discussion

The results of the auditory perception test indicate that in the theoretical field of speech perception studies perception does not entirely depend on the recovery of an acoustic pattern, since pauses were identified even in cases in which the auditory stimulus did not include any instants of silence (type A stimuli). As such, the results obtained on the perception of prosodic boundaries with data from BP add to the evidence provided by Martin (1970), Steinhauer and others (1999) and Peters (2005) with data from English and German. These results also give room to question the theoretical-analytical assumptions of the auditory theories of speech perception (Fant 1967; Morton & Broadbent 1967; Stevens & Blumstein 1978), which centre around the physical signal, which is in turn captured by the listener through his/her auditory sensorial capability to retrieve an acoustic pattern. It follows that from this perspective, symbolic aspects, such as the grammatical organization of a language, are disregarded, since the physical signal holds the required information for perception.

Refuting this disregard for symbolic aspects, our results point out the role of the prosodic organization in BP, considering that the identification of an intonational phrase, precisely its boundary, is a relevant piece of information for identifying a pause. In other words, the results highlight the relevance of phonological representation, since they suggest that the identification of a pause is conditioned on the presence of the tested phonological context: the I-boundary. This tendency is observed in the high percentage of pauses identified in stimuli A and B that display this boundary and is reinforced, in contrast, by the low pause identification level in stimulus C, which does not have the abovementioned boundary.

The average reaction time by type of auditory stimulus, presented in Figure 1, confirms the argument that perception is not limited to retrieving an acoustic pattern but is instead a more complex process, as argued by motor theories, since the average reaction time was significantly greater in answers given to the type A stimulus that did not have any pause production; nevertheless, its boundaries were mostly interpreted as occurring with pause. Consequently, the higher average reaction time for stimulus A indicates the complexity involved in the task of identifying a pause, given that it allows us to construe that other grammatical criteria, such as the phonological representation of an intonational phrase, help identify a pause even when it is not acoustically present, requiring a significantly greater time to do so.

Hence, the empirical results on pause identification and the average reaction time by acoustic stimulus allow us to interpret that the intonational phrase, a prosodic constituent, works in combination with the sensorial-auditory capability of a speaker/listener in the perception of pauses. As such, we argue that the test performed in the prosodic domain reinforces the view proposed by Fowler (1986, 1996) and Goldstein and Fowler (2003) for speech perception, according to which perception is composed of the relation between phonetic and phonological aspects, thus joining physical aspects of sound and symbolic aspects of sound in a given language. In this relation, perception intersects with production/representation.

From a phonetic-acoustic perspective (*i.e.*, considering production), the intonation contour plays a key role in the characterization of an intonational phrase. Specifically, in BP, the tonal sequence H+L* L% allows us to categorize the abovementioned prosodic constituent (Frota & Vigário 2002; Fernandes 2007; Serra 2009; Tenani 2002). Pitch changes, interpreted as discrete tone units, consist of a phonetic cue marking the intonational phrase. In particular, when the pause is combined with this tonal sequence, as controlled in stimulus B, the boundary is more easily perceived (Serra 2009).

However, the results of the test performed in this research have allowed us to advance further in qualitative terms regarding the understanding of how the perception of the I-boundary is interpreted in Portuguese. Beyond the information that producing a pause helps perceive the boundary, the results for stimulus A show that the tonal sequence H+L* L% could be identified as a pause at an I-boundary, even if this pause was not performed. The obtained percentage and the significance level detected in the statistical analysis ($p < 0,00$) indicate that the relation between the type of stimulus and response is not random. Thus, specifically in terms of type A stimuli, the results indicate that the pause would be a perceptual effect of the tonal sequence H+L* L%, which characterizes the I-boundary. The descending contour, phonetically marked by the fall of F0, allows us to identify grammatical information in BP coded by the end of an I-boundary, which defines, in general, just as in the analysed sentences, the end of an utterance, which also corresponds to the end of a syntactic unit and unit of meaning.

In the case of the sentences that represented the auditory stimuli, considering their organization in pairs, the presence of an I-boundary or a change in the position of a boundary within the same segmental chain leads to a syntactic-semantic difference that marks contrastive meanings. For example, let us consider the pair A. The pair A [[não]I [mereço saber]IP]U (‘No, I deserve to know’) *versus* [[não mereço saber]]IP]U (‘I do not deserve to know’) illustrates how the existence of the IP-boundary in the first sentence allows us to identify a meaning opposed to that of the second sentence, which does not have that boundary. Identifying one meaning or the other is possible thanks to the prosodic boundary, given that its presence in the first sentence blocks the syntactic relation existing between the expression of the negation *não* and the verbal predicate *mereço*, ‘I deserve’, which can be identified when there is no such boundary, as in the case of the second sentence. Considering that the IP-boundary in the first sentence is phonetically marked by pitch changes (by H+L* L% tonal events), it follows that this acoustic variable, combined with the prosodic context, defines the relation of syntactic non-dependency between *não* and its verbal predicate, thereby enabling its identification, which is in turn translated into semantic information.

Therefore, beyond the argumentation that perception does not exclusively correspond to the recovery of an acoustic signal and that a symbolic component is involved in perception – in this research, the prosodic structure – the results of the test have the potential to challenge the concept of pause.

From the perspective of perception, evidence is provided that the pause is not a synonym for an instant of silence, since the identification of a pause depends, to some extent, on the symbolic information resulting from the phonological representation of the intonational phrase. This information is interpreted by speakers/listeners as a pause, even when there is no instant of silence, because in the prosodic organization of BP, the pause can delimit a construction defined in the syntactic-semantic domain; however, it is primarily characterized by intonation, particularly by the falling intonation contour, in the case of the analysed data. In other words, we argue that the identification of a pause results from the perceptual illusion marked by the combination of different types of linguistic information at the phonetic-acoustic and grammatical levels.

5. Conclusion

This paper has shown by means of the results of an experimental test that pause perception in BP is not restricted to exclusive identification of instants of silence present in the acoustic signal, since pitch variations at IP boundaries also led to pause perception. Regarding this result, we highlight the fact that the identification of pauses has also been conditioned to grammatical/symbolic information related to the phonological representation, which in turn intersects with syntactic-semantic information.

Although similar tendencies have already been pointed out for different languages by various means of investigation (for English: Edwards & Beckman 1990; Gussenhoven & Rietveld 1992; Martin 1970; among others; for French: Duez 1985; Simon & Christodoulides 2016; among others; for German: Steinhauer and others 1999; Peters 2005; Männel, Schipke & Friederici 2013; Männel & Friederici 2016; among others), the results presented in this research are unprecedented for BP. Based on the achieved results, we argue that speech perception is not solely limited to the apprehension of acoustic events, since perceiving speech implies considering it from a phonological representation, meaning that speech perception is a process that comprises both physical and symbolic aspects of the sound produced.

Funding: This research was funded through grants awarded to GS and LT by the São Paulo Research Foundation – FAPESP (Process Number: 2014/24778-3). The translation to the English version of the paper was funded by the Graduate Programme “Estudos Linguísticos” of the São Paulo State University (UNESP), supported by PROEX-CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), Brazil.
A transversal prospective study was approved by the Research Ethics Committee at São Paulo State University (UNESP, Brazil) under number 57016116.3.0000.5466.

Acknowledgements: We thank Mr. Márcio Santana da Silva and Mr. Rubens Leão for their technical assistance in recording the stimuli and in programming the scripts for stimuli presentation.

References

André, C., Ghio, A., Cavé, C. & Teston, B. (2009). PERCEVAL: PERCeption EVALuation Auditive & Visuelle (Version 5.0.30) [Computer software]. Aix-en-Provence: Laboratoire Parole et Langage.

Beckman, M. & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–310. https://doi.org/10.1017/S095267570000066X

Boersma, P. & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.2.28) [Computer software]. Amsterdam, NL: University of Amsterdam.

Borden, G. J., Harris, K. S. & Lawrence, J. R. (1994). *Speech science primer: Physiology, acoustics and perception of speech.* Baltimore: Williams & Wilkins.

Brazilian Institute of Geography and Statistics [IBGE] (2016). *Anuário estatístico do Brasil*. Rio de Janeiro: IBGE.

Chacon, L. & Fraga, M. (2014). Pausas na interpretação teatral: Delimitação de constituintes prosódicos. *Filologia e Linguística Portuguesa*, *16* (1), 121–146. https://doi.org/10.11606/issn.2176-9419.v16i1p121-146

Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and Speech*, *28* (4), 377–389. https://doi.org/10.1177/002383098502800403

Edwards, J. & Beckman, M. E. (1988). Articulatory timing and the prosodic interpretation of syllable duration. *Phonetica*, *45*, 156–174. https://doi.org/10.1159/000261824

Fant, G. (1967). Auditory patterns of speech. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 111–125). Cambridge: M.I.T. Press.

Fernandes, F. R. (2007). *Ordem, focalização e preenchimento em português: Sintaxe e prosódia* (Doctoral thesis, University of Campinas, Campinas, Brazil).

Fowler, C. (1986). An event approach to the study of speech perception. *Journal of Phonetics*, *14* (1), 3–28. https://doi.org/10.1016/S0095-4470(19)30607-2

Fowler, C. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, *99* (3), 1730–1741. https://doi.org/10.1121/1.415237 ·

Frota, S. (2000). *Prosody and focus in European Portuguese*. NewYork: Garland Publishing.

Frota, S. (2003). The phonological status of initial peaks in European Portuguese. *Catalan Journal of Linguistics*, *2*, 133–152. https://doi.org/10.5565/rev/catjl.47

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). Braga: APL.

Galantucci, B., Fowler, C. & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13* (3), 361–377. https://doi.org/10.3758/BF03193857

Goldstein, L. & Fowler, C. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production*: *Differences and similarities* (pp. 159–207). Berlin: Mouton de Gruyter.

Gussenhoven, C. & Rietveld, A. C. M. (1992). Intonation contours, prosodic and preboundary lengthening. *Journal of Phonetics*, *20* (3), 283–303. https://doi.org/10.1016/S0095-4470(19)30636-9

Hirst, D. & Di Cristo, A. (1998). *Intonation systems: A survey for twenty languages*. Cambridge: Cambridge University Press.

Ladd, R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. S. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74* (6), 431–461. https://doi.org/10.1037/h0020279

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36. https://doi.org/10.1016/0010-0277(85)90021-6

Liberman, A. M. & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4* (5), 187–186. https://doi.org/10.1016/S1364-6613(00)01471-6

Männel, C. & Friederici, A. D. (2016). Neural correlates of prosodic boundary perception in German preschoolers: If pause is present, pitch can go. *Brain Research*, *1632*, 27–33. https://doi.org/10.1016/j.brainres.2015.12.009

Männel, C., Schipke, C. S. & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience*, *5*, 86–94. https://doi.org/10.1016/j.dcn.2013.01.003

Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of verbal learning and verbal behavior*, *9* (1), 75-78. https://doi.org/10.1016/S0022-5371(70)80010-X

Moraes, J. (2007). *Nuclear and pre-nuclear contours in Brazilian Portuguese intonation*. Workshop presented at Phonetics and Phonology in Iberia, University of Minho, Braga, Portugal.

Morton, J. & Broadbent D. E. (1967). Passive versus active recognition models or is homunculus really necessary? In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 103–110). Cambridge: M.I.T. Press.

Nespor, M. & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.

Nespor, M. & Vogel, I. (2007). *Prosodic phonology: with a new foreword*. Berlin: Walter de Gruyter.

Peters, B. (2005). Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache [Further investigations to prosodic borders in German spontaneous language.]. In K. J. Kohler, F. Kleber & B. Peters (Eds.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel [Work reports of the Institute for phonetics and digital speech processing of Kiel University]*: *Prosodic structures in German spontaneous speech* (Vol. 35, pp. 203–345). Kiel: IPDS.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Cambridge: M.I.T. Press.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: M.I.T. Press.

Serra, C. R. (2009). *Realização e percepção de fronteiras prosódicas no português do Brasil: Fala espontânea e leitura* (Unpublished doctoral thesis, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil).

Simon, A. C. & Christodoulides, G. (2016). Perception of prosodic boundaries by naïve listeners in French. *Proceedings of Speech Prosody, USA*, *8*, 1158–1162. http.//doi.org/10.21437/SpeechProsody.2016-238

Steinhauer, K., Alter, K. & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2* (2), 191–196. https.//doi.org/10.1038/5757

Stevens, K. N. & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64* (5), 1358–1368. https://doi.org/10.1121/1.382102

Tenani, L. E. (2002). *Domínios prosódicos no português do Brasil*: *Implicações para a prosódia e para a aplicação de processos fonológicos* (Doctoral thesis, University of Campinas, Campinas, Brazil).

[received on June 1, 2019 and accepted for publication on November 27, 2019]