ELECTROPHYSIOLOGICAL CORRELATES OF PROSODIC BOUNDARIES AT DIFFERENT LEVELS IN BRAZILIAN PORTUGUESE

Oyedeji Musiliyu [1], Annie C. Gilbert [2], Shari R. Baum [3] and Miguel Oliveira Jr. [4]

[1,4] Federal University of Alagoas, Maceió, Brazil
[2,3] McGill University, Montreal, Canada
bodeses@yahoo.fr, gilbert.annie.c@gmail.com, shari.baum@mcgill.ca, miguel@fale.ufal.br

ABSTRACT

In tasks involving spoken language comprehension, prosodic cues serve as a guide to correct linguistic processing. In a statement, for instance, a set of prosodic cues is used to segment and organize speech into intonational phrases (IP) and phonological utterances (U). Though these two highest prosodic constituents are delimited by the same set of cues, their respective boundaries coincide with different syntactic boundaries. The present study investigated phonetic differences between boundaries associated with these constituents and aimed to provide evidence of detection of these differences by way of an ERP experiment examining Closure Positive Shifts (CPS). The results revealed prosodic cue differences between IP and U boundaries, as reflected in acoustic parameters. CPSs were elicited in response to processing of these boundaries that were modulated as a function of the prosodic cue differences between them. This result is further evidence that, in speech comprehension, listeners are sensitive to prosodic cues.

Keywords: Speech Comprehension; Prosodic Cues; Prosodic Boundary; CPS.

1. INTRODUCTION

Written language generally has a clear structure, due to the use of typographic conventions such as punctuation marks, like commas and periods. Speech, however, involves a fleeting series of connected sounds, with no obvious boundary markers. There are specific mechanisms that are used to signal the structure of spoken discourse, nonetheless. Several studies have shown that prosody is often used to organize speech into a series of hierarchically arranged, coherent macro units [19], [8], [9]. Attempts to provide phonological evidence to support prosody as a guide in structuring speech led to the development of Prosodic Phonology theory [21], [15], [13] which proposes that the flow of speech is organized into a finite set of phonological units, or “prosodic constituents”, composed of hierarchically arranged components from lowest to highest, as follows: syllable (σ), foot (Σ), phonological word (ω), clitic group (C), phonological phrase (ϕ), intonational phrase (IP) and phonological utterance (U).

With respect to the existence of a systematic difference between IP and U, [16] marked them as distinct on the basis of differential application of certain phonological rules. Importantly, U consists of at least one IP, and usually extends through the length of the string dominated by the highest node in the syntactic tree, thus being referred to as X^n. However, U is not simply the phonological counterpart of X^n since it can combine two or more sentences into a unit of a greater level.

Studies on Brazilian Portuguese (henceforth BP) propose that the IP boundary (henceforth *IPB*) and U boundary (henceforth *UB*) are marked mainly by three major prosodic cues: final lengthening [22], pitch variation [26] and, although not required, a pause [5]. Since the role of prosody as a device in structuring spoken discourse is well described with respect to production, this study aims to investigate its impact in perception from a neurocognitive perspective.

With the advent of techniques such as Event-Related Potentials (henceforth ERPs), [24] first found the Closure Positive Shift (CPS) ERP response for the processing of *IPBs*. The majority of CPS studies supports that acoustic prosodic cues are primarily responsible for the generation of the CPS [4], while linguistic cues modulate its features [11]. In order to investigate further the modulation of the CPS, we measured and compared it at *IPB* and *UB*.

2. METHOD

2.1. Stimuli

2.1.1 Design of contexts

The experimental items consist of statements that contained an *IPB* and *UB*. Three versions of each item were created. The first version (Type A) was the basis for the other two versions (Type B and Type C) as follows in Table 1, where **(#)** stands for *IPB*, **(#*)** for “*no IPB” - NIPB* -, **(%)** for *UB* and (%*) for “*no UB*”- *NUB* -. The analyses only focus on “target words” shown in Table 1 in bold italic font. These words were always a trisyllabic verb complement, presenting with a paroxytone stress pattern (indicated in uppercase in the penultimate syllable). A total of 120 experimental items were selected after a norming study (with 30 students, native speakers of BP) to ensure the naturalness and acceptability of the items. In addition, 120 fillers were created.

Table 1: Example of experimental item.

Type A
Assim que Paula viu sua ***aMIga #*** *ela fechou a* ***jaNEla %*** *Foi abrir a porta.* (As soon as Paula saw her friend, she closed the window. She then opened the door)
Type B
Assim que Paula viu sua ***aMIga #**** *de infância* ***#*** *ela fechou a janela* ***%*** *Foi abrir a porta.* (As soon as Paula saw her childhood friend, she closed the window. She then opened the door)
Type C
Assim que Paula viu sua amiga # ela fechou a ***jaNEla %**** *da sala* ***%*** *Foi abrir a porta.* (As soon as Paula saw her friend, she closed the living room window. She then opened the door)

2.1.2 Stimuli recording

A male professional announcer, native speaker of BP, recorded the stimuli using an omnidirectional microphone in adequate acoustic conditions with a normal speech rate (M = 5.589 syllable per second (syll/s), SD= 0.4), consistent with previous average speech rates (ranging from 3.2 to 5.5 Syll/s) for a native fluent speaker of BP [17], [14]. The recording was digitized at 44100 Hz with a bit depth of 16 bits per sample.

2.1.3 Acoustic analyses

We ran a Praat [1] script called "Analyse_tier" [10] to measure and compare prosodic cues associated with the stressed syllables, marking *IPB*, *NIPB*, *UB* and *NUB*, which we categorized as follows, respectively (where "*Str*" stands for stressed syllable): (i) *Str_IPB*; (ii) *Str_NIPB*; (iii) *Str_UB*; (iv) *Str_NUB*. We recorded syllable lengthening, pitch variation, mean values of fundamental frequency (F0) and intensity. The data for each cue were subjected to two-sample *t*-tests. An overview of the results is presented in Table 2, where (*) stands for p≤ 0.05; (**), for p≤ 0.01; (***), for p≤ 0.001. This illustration is related with the fact that ERP responses are time-locked to the onset of the stressed syllables of "target words", as detailed in the following "EEG experimental paradigm" subsection of this study.

2.2 EEG experimental paradigm

2.2.1 Participants

Thirty volunteer students (15 males; mean age: 24.3 years; SD= 3), native speakers of BP, from the Federal University of Alagoas (henceforth, UFAL), participated in the ERP experiment. They were right-handed with no hearing disorder or previous history of neurological or psychiatric disorders based on self-declaration.

2.2.2 EEG recording procedure

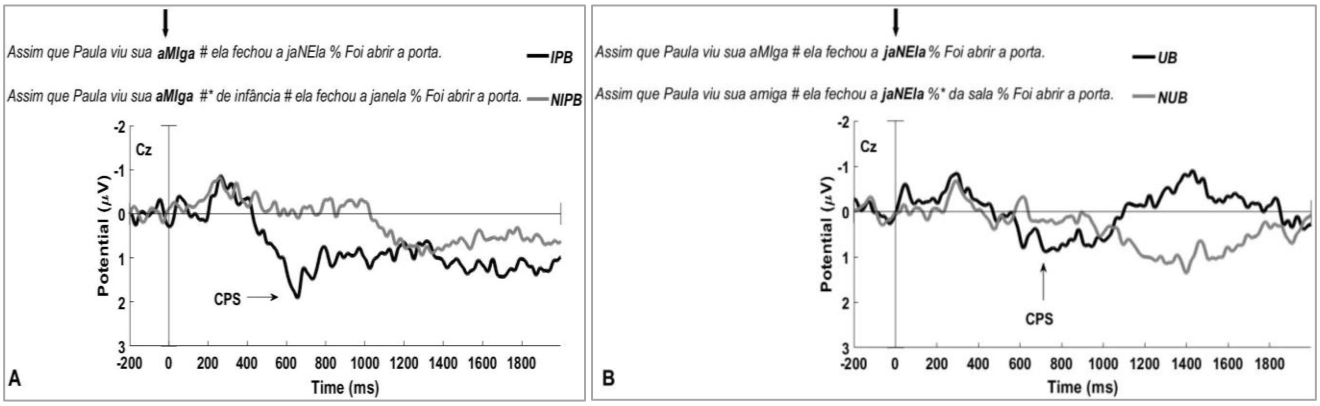
The experiment was conducted at UFAL. We used a Geodesic EEG System 400 (with a *HydroCel Sensor Net* of 256 channels), relying on the enhanced 10-20 system. Experimental stimuli along with filler items were presented in a pseudorandom order via headphones (*Sennheiser* hd280 pro) using *E-Prime*, in blocks of 40 items of no more than 5 minutes, with rest pauses between blocks and the constraint that the same type of stimulus was not presented twice in a row. The sound loudness was controlled and set at 22 % of the maximum volume of the computer sound card for all the participants. For each participant, 5 % of the stimuli were randomly followed by a written prompt (a word). In the task, the participants were asked to indicate via a key press if the prompted word was present or absent from the immediately preceding stimulus. This task was given to ensure that the participants were paying attention while listening to the sentences. Participants had to look at a fixation point to avoid eye-movements and blinks until the offset of a stimulus.

Electrodes were adjusted till their impedances were kept below 5kΩ. EEG data were recorded along with trigger codes, with a high-pass filter at 0.1 Hz and sampling rate set at 1000 Hz. The online recording reference for all sensors was Cz.

Table 2: Statistical analysis results of comparison of acoustic analysis data associated with *IPB, NIPB, UB* and *NUB.*

	Lengthening (z-score)	Pitch variation (Hz)	Mean F0 (Hz)	Mean intensity (dB)
Comparisons of stressed syllables	t (df) Mean (SE)	t (df) Mean (SE)	t (df) Mean (SE)	t (df) Mean (SE)
Str_IPB vs. Str_NIPB	10.4 (238)*** 1.1 (0.05) vs. 0.2 (0.05)			3.5 (238)*** 75.9 (0.2) vs. 74.6 (0.2)
Str_UB vs. Str_NUB	6.9 (238)*** 0.7 (0.05) vs. 0.2 (0.03)	4.3 (236)*** 10.3 (0.7) vs. 15.9 (1)	16.6 (236)*** 84.7 (0.7) vs. 116 (1.7)	5.9 (237)*** 72.9 (0.2) vs. 75.1 (0.2).
Str_IPB vs. Str_UB	2.5 (238)** 1.1 (0.05) vs. 0.9 (0.05)		12.1 (236)*** 100.4 (0.9) vs. 84.7 (0.7)	8.1 (239)*** 75.9 (0.2) vs. 72.9 (0.2)

Figure 1: Grand averaged ERPs at Cz electrode for *IPB vs. NIPB* and *UB vs. NUB*.



2.2.3 Measuring and analysing the ERPs

In offline analysis, the EEG data were filtered using a digital low-pass filter (30Hz) and re-referenced to an average reference. Ocular artefact correction was performed using independent components analysis (ICA) as implemented in EEGLAB (‘eeg_runica’ function) [6]. Independent components with known features of eye blinks (based on activity power spectrum, scalp topography, and activity over trials) were identified visually for each participant. The contributions of these components were then removed from the epoched EEG data. Artefacts were detected and removed automatically by using a moving window peak to peak procedure, with a 200 ms moving window, a 100 ms window step, and a 100 µV voltage threshold.

Artefact-free EEG segments were divided into sections (epochs) of -200 ms to 2000 ms, relative to the onset of the stressed syllables of the “target words”. ERPs were time-locked to the onset of the stressed syllables of the “target words”. This approach was adopted following [3] that considered the last stressed syllable as the “onset of the prosodic boundary”, and as the theoretically most appropriate time-locking point for the CPS analysis. Epochs were averaged to produce an ERP for each prosodic boundary condition. Individual ERP waves were averaged to get grand averaged ERPs for each condition.

Table 3: Significant effects of ANOVAs for mean amplitudes across time-window of [400–1200 ms] for the CPS component of ERPs.

	Midline electrodes			Lateral ROIs		
Tws (ms)	Effect	*F*(df)	Mean (SE) (µV) B vs. NB	Effect	*F*(df)	Mean (SE) (µV) B vs. NB
400 – 600	Cond	4.2 (1,348)*	0.2 (0.08) vs. -0.1 (0.07)			
	Cond × Boun	5.5 (1,348) **				
	IP	10.5 (1,178)**	0.3 (0.1) vs. -0.14 (0.11)			
600 – 800	Cond	23.2 (1,348)***	0.6 (0.10) vs. 0.03 (0.08)	Cond	10.2 (1,1896)**	0.05 (0.0) vs. -0.14 (0.04)
	Cond × Elec	3.9 (2,348)*		Cond × Hem	11.9 (1,1896)***	
	Cz	19.4 (1,118)***	1.05 (0.1) vs. 0.01 (0.1)	Right Hem	19.9 (1,958)***	0.24 (0.0) vs. -0.17 (0.06)
	Pz	12.6 (1, 118)**	0.5 (0.14) vs. -0.1 (0.1)	Cond × Boun	6.6 (1,1896)**	
				IP	9.9 (1,958)***	0.14 (0.07) vs. -0.14 (0.05)
800–1000	Cond	9.7 (1,348)***	0.4 (0.1) vs. 0.007 (0.08)	Cond	31.6 (1,1896)***	0.16 (0.05) vs. -0.22 (0.04)
	Cond × Elec	4.4 (2,348)**		Cond × Hem	14.6 (1,1896)***	
	Cz	11.2 (1,118)**	0.7 (0.18) vs. -0.02 (0.15)	Right Hem	40.5 (1,958)***	0.37 (0.07) vs. -0.27 (0.06)
	Pz	9 (1,118)**	0.44 (0.15) vs. -0,1 (0.14)			
1000–1200				Cond	32.3 (1,1896)***	0.24 (0.05) vs. -0.19 (0.05)
				Cond × Hem	6.7 (1,1896)**	
				Right Hem	31.4 (1,958)***	0.38 (0.08) vs. -0.24 (0.07)
				Cond × Loca	9.18 (2,1896) ***	
				Anterior Loca	25.8 (1,718)***	0.34 (0.1) vs. -0.4 (0.1)
				Central Loca	10.4 (1,238)**	0.4 (0.1) vs. -0.13 (0.1)

Since the pairs of conditions – *IPB vs. NIPB* and *UB vs. NUB* – contained the same lexical and prosodic information up to the onset of the stressed syllables of the "target words", the comparisons of grand averaged ERPs in our time-windows [0–2000 ms] would reflect the processing of prosodic cues marking the *IPB* and *UB*.

3. RESULTS

Grand Averaged ERPs at the Cz electrode for *IPB vs. NIPB* and *UB vs. NUB*, are presented Figure 1A and Figure 1B. The boldened arrows above the words at the top of the figure indicate the time-locking point to measure the ERPs and the stressed syllables of "target words", which are in uppercase. From the lowest positive point after the initial negativity peak, we observed a broad positive deflection, as marked by the arrows in Figure 1A and Figure 1B, for *IPB* (from ~500 to 1200 ms) and *UB* conditions (from ~600 to 1100 ms). We assumed this is the CPS.

For both the *IPB* and *UB* conditions, the CPS effects were observed at midline, and largely over right hemisphere, central and anterior locations. The later positivity in the *NUB* condition is the CPS effect associated with the subsequent utterance boundary (indicated with % in Figure 1B). Statistical analysis substantiated the findings, as presented in Table 3 (where Tws = time-windows; B = boundary condition; NB = "no boundary" condition; Cond = condition; Hem = hemisphere; Elec = electrode; Boun= boundary; Loca = location; × = interaction. (*) stands for $p \leq 0.05$; (**), for $p \leq 0.01$; (***), for $p \leq 0.001$).

4. DISCUSSION AND CONCLUSION

The findings showed differences in the strength of the prosodic cues between *IPB* and *UB* stimuli, consistent with the literature [26], [7], [22], [5], indicating that from the perspective of production, prosodic boundaries at different levels of the prosodic hierarchy are characterized differently. Behavioral findings revealed CPS effects in response to processing of *IPB* and *UB* when contrasted against the absence of these boundaries. Prior to the CPS, the "initial negative peak" we found is considered the pre-CPS negativity also observed in previous studies [18], [12], [2], [20]. This "early negativity" may be a consequence of the processing of early prosodic cues marking prosodic boundaries.

The scalp distribution of the CPS reported for both the *IPB* and the *UB* at midline, and predominantly over right hemisphere, central and anterior locations, indicates that there is a similarity in processing of *IPB* and *UB*. In addition, we observed early onset latency, longer duration and relatively higher amplitude for *IPB* stimuli. A plausible account for the onset latency and amplitude differences might be that the stronger cues in the *IPB* condition (as reflected in the acoustic analyses) led to faster processing of that boundary *IPB*. As for the duration, the difference may be attributable to the pre-final syllable lengthening difference for *IPB* as compared to *UB*. Taking the whole discussion into consideration, we may conclude that the amplitude, onset latency and duration of the CPS effects reflected the extent to which acoustic prosodic cues were perceived, in keeping with previous findings [25], [23].

This study presents evidence for the fact that, in speech comprehension, listeners are sensitive to prosodic cues of different constituent levels, namely, the intonational phrase (IP) and the utterance (U) in BP. This was observed through the examination of the prosodic differences between the boundaries signaling the two constituents, and the analysis of neurophysiological responses to processing of cues signaling them by way of an experiment inspecting the Closure Positive Shift (CPS). The study contributes to the research fields of speech processing and spoken language comprehension by providing evidence for the difference in processing utterance-final and phrase-final prosodic boundaries (as reflected in the different parameters of CPS components observed at the two distinct boundaries). The comparison of the two CPS responses is a relevant contribution to the field, providing insight into the neurocognitive processes at hierarchically and prosodically different phrasal constituents.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2018. *Praat: doing phonetics by computer [Computer program]*. Version 6.0.37, retrieved 14 March 2018 from http//www.praat.org/

[2] Bögels, S., Schriefers, H., Vonk, W., Chwilla, D.J., & Kerkhofs, R. 2010. The interplay between prosody and syntax in sentence processing: the case of subject- and object-control verbs. *J. Cogn. Neurosci*. 22, 1036–1053.

[3] Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J. 2011. Prosodic breaks in sentence processing investigated by event-related potentials. *Linguist. Lang. Compass*, 5, 424–440.

[4] Brouwer, H., Fitz, H., Hoeks, J. 2012. Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, V. 1446, 29 March, 127-143.

[5] Chacon, L., Fraga, M. 2014. Pausas na interpretação teatral: delimitação de constituintes prosódicos. *Filol. e Linguística Port*. São Paulo, V.16, n. 1, 121-146.

[6] Delorme, A., Makeig, S. 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*, v. 134, 9–21.

[7] Fernandes, F. R. 2007. *Ordem, focalização e preenchimento em português: sintaxe e prosódia*. (Ph.D. Dissertation). Universidade Estadual de Campinas.

[8] Geluykens, R., Swerts, M. 1994. Prosodic cues to discourse boundaries in experimental dialogues. *Speech Commun*. 15, 69–77.

[9] Gilbert, A. C., Boucher, V. J., & Jemel, B. 2015. The perceptual chunking of speech: a demonstration using ERPs. *Brain Research, 1603*, 101-113.

[10] Hirst, D. 2012. *Analyse_tier.praat*. Aix-en-Provence: PRAAT script.

[11] Kerkhofs, R., Vonk, W., Schriefers, H., Chwilla, D. 2007. Discourse, syntax, and prosody: the brain reveals an immediate interaction. J. *Cogn. Neurosci*. 19, 1421–1434.

[12] Kerkhofs, R., Vonk, W., Schriefers, H., Chwilla, D. J. 2008. Sentence processing in the visual and auditory modality: Do comma and prosodic break have parallel functions? *Brain Res.* 1224, 102–118.

[13] Ladd, D. 1986. Intonational phrasing: the case for recursive prosodic structure. *Phonology* 3, 311.

[14] Martins, V., Andrade, F. 2008. Perfil evolutivo da fluência da fala de falantes do Português brasileiro. *Pró-Fono*, v. 20, n. 1, 7–12.

[15] Nespor, M., Vogel, I. 1983. *Prosodic Structure Above the Word*. Springer, Berlin, Heidelberg, 123–140.

[16] Nespor, M., Vogel, I. 1986. *Prosodic Phonology*. Linguistic Inquiry. Foris, Dordrecht.

[17] Oliveira, M. 2000. *Prosodic features in spontaneous narratives.* (Ph.D. Dissertation). Simon Frazer University, Vancouver, Canada.

[18] Pannekamp, A., Toepel, U., Alter, K., Hahne, A., Friederici, A. D. 2005. Prosody-driven sentence processing: an event-related brain potential study. *J. Cogn. Neurosci.*, 17(3), 407-421.

[19] Passonneau, R., & Litman, D. 1993. Intention-based Segmentation: Human Reliability and Correlation with Linguistic Cues, in: Proceeding of the 31st Annual Meeting of the Association for Computational Linguistics (ACL-93). Columbus, Ohio, 8.

[20] Pauker, E., Itzhak, I., Baum, S. R., Steinhauer, K. 2011. Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *J. Cogn. Neurosci.*, 23(10), 2731-2751.

[21] Selkirk, E. 1984. *Phonology and Syntax: The relation between sound and structure*. Cambridge, MA: The MIT Press.

[22] Serra, C., Frota, S. 2009. Fraseamento prosódico no Português do Brasil: pistas para a percepção. *Encontro Nac. Da Assoc. Port. Linguística, xxv*, Universidade de Lisboa, Lisboa, Portugal.

[23] Steinhauer, K. 2003. Electrophysiological correlates of prosody and punctuation. *Brain Lang*. 86(1), 142–164.

[24] Steinhauer, K., Alter, K., Friederici, A. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature euroscience*. 2(2), 191–196.

[25] Steinhauer, K., Friederici, A. D. 2001. Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *J. Psycholinguist. Res*. 30(3), 267–295.

[26] Tenani, L. 2002. *Domínios prosódicos no português do Brasil: Implicações para a prosódia e para a aplicação de processos fonológicos*. (Ph.D. Dissertation). Universidade Estadual de Campinas.