Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Reading aloud: eye movements and prosody

Isabel Falé[1,3], *Armanda Costa* [1,2], *Paula Luegi* [1,2]

[1] Centro de Linguística da Universidade de Lisboa, Portugal
[2] Faculdade de Letras, Universidade de Lisboa, Portugal
[3] Universidade Aberta, Lisboa, Portugal

ifale@campus.ul.pt, armandacosta@letras.ulisboa.pt, paulaluegi@letras.ulisboa.pt

Abstract

This study aims to connect data from ocular movements and reading aloud speech to syntactic and discursive properties of texts, in order to understand integrative cognitive processes during reading for understanding and to identify prosodic and eye movements' indicators of reading fluency. Assuming that in reading aloud there is a close interaction between syntax structure and speech prosody, we collected eye movements and reading speech data from 17 native EP speakers. Eye movements and reading speech produced simultaneously were analyzed and our results show that eyes and voice are both responsive to text complexity and to syntactic and discursive critical *loci*, as key points of information integration.

Index Terms: reading comprehension, eye movements, prosody, text complexity, syntactic and discursive structure.

1. Introduction

In this research, we study reading comprehension processes through a reading aloud task to capture the integration of syntactic and discursive information, combining both the prosodic analysis of "reading speech" and eye movements analysis (see [1] and [2] for examples of work with a similar methodology). By "reading speech" we mean a planned speech instance, in which prosody is strongly constrained by punctuation and layout, resulting in a temporal organization most predictable than in spontaneous speech.

We thereby intend to identify linguistic information integration processes undergoing in reading for understanding, in particular related to the oral reading mode, and the effect of linguistic and discursive properties of the read texts. Finding related measures of prosodic features in reading aloud and of eye movements will be useful for reading fluency assessment and teaching.

By choosing a methodology that crosses speech analysis with eye movements, we pursue our goal of deepening our knowledge about the possible interconnection of different behavioral indicators and see how these may signal the same cognitive processes. This approach gives us information on mental, linguistic and non-linguistic mental representations, which are activated on the processes required to recognize and decoding the print signal. The reader and reading processes have been thoroughly studied and a lot is known about what is involved (see, among others, [3] and [4]), but little is known about interconnected processes in reading.

2. Experimental study

The purpose of our study is to identify behavioral indicators of different nature and relate them with cognitive operations underlying perception, storage and integration of information in reading aloud, in the simultaneous execution of two tasks: reading and speech production. We assume that ocular movements and voice, besides the well-known eye-voice span effect, must be interconnected and have correlated manifestations.

Two hypotheses were proposed. The first undertakes that an increase in visual reading time (provided by fixations and saccades times') and in "reading speech" time (production time, including pauses) in critical regions are indicators of textual complexity effect. To verify this hypothesis, two passages differentiated by subject and vocabulary complexity were used as experimental materials. In accordance with the hypothesis of textual complexity, longer values for reading times are expected in the more complex text, namely at given discursive or syntactic *loci*.

The second hypothesis assumes that larger chunks of information are processed at strategic points, marked both by syntactic structure and by punctuation marks. We consider two conditions - syntactic boundary (SB), at the right edge of a syntactic constituent – noun phrase, verbal phrase or prepositional phrase, and discourse boundary (DB) marked by a period. Taking measures from speech (stressed vowel length, F0) and from eye movements (first fixation duration, first pass and total time word fixation) we expect *wrap-up* effects and stronger prosodic indicators at discourse boundary than at syntactic boundary *loci*. As baseline condition, we consider the same measures at a head of a Noun Phrase (SN), which is taken as a point of structure building and not of information integration.

2.1. Participants

The study included 17 European Portuguese native speakers, students, all female and proficient readers, attested by prior questionnaire on reading habits.

2.2. Materials

In this experiment we used two texts, which have been used in several studies on reading research in Portuguese [5, 6]. Both texts are similar in length, syntactic and information structure and layout; they are crucially different at vocabulary and theme levels: the easiest addresses a current and familiar theme to most readers - a description of a Lisbon neighborhood -, *Campo de*

 doi: 10.21437/SpeechProsody.2016-169

Ourique (CO); and the hardiest a technical subject - thermo-acoustic proofing -, *Isolamento Termo-acústico* (ITA). We assume though that thematic and lexical dimensions contribute to a greater level of complexity of ITA text (cf. standards for linguistic complexity, Standard 10, Range and level of text complexity, in [7]). Also assume that this complexity will affect the processing at micro and macro linguistic levels [8].

Word length and word frequency

Considering the perceptual reading span, the concept of processing unit and the reading modality, the two texts were compared in different layers. Both texts are identical in number of words: 205 CO and 203 ITA. However, they differ in terms of number of characters *per* word with the presence of longer words in the more complex text (monosyllabic words: CO - 82/ ITA - 79; disyllabic words: CO – 57/ ITA -36; four or more syllables words: CO - 24 / ITA - 45). In phonological terms, namely in syllable patterns and its complexity [9], both texts' are identical. Phonological measures were obtained with *FrePOP* [10].

Analyzing vocabulary frequency of the words in the experimental texts, we conclude that this is connected to the topic specificity of each text. Comparing trisyllabic words that occur in identical number in both texts (30), we confirm that ITA words are less frequent in the language lexicon. Word frequency was evaluated with the *Multifunctional Computational Lexicon of Contemporary Portuguese* [11].

Informational structure

Texts are typical descriptive passages consisting of a title that represents the theme, and six paragraphs, each one introducing a new subtopic.

Syntactic structure

Sentence syntactic structures are similarly distributed in both texts. The texts include typical syntactic properties of Portuguese, a *pro drop* language, such as: declarative sentences, some of them with a null subject; WH- interrogative sentences with post-verbal subject; clitic pronouns in enclitic and proclitic position; co-referential chains including the rotation of overt and covert pronouns as anaphoric expressions.

Prosodic matrix

Although is not isomorphic, the prosodic phrasing in European Portuguese tends to project the syntactic structure. Taking this into account, we make a proposal of a prosodic matrix *per* text that allow to compare a given reading with the expected reading, We intend to contribute for scales of reading fluency that operate in a more intuitive way,

2.3. Procedure

Experiment was carried out in Psycholinguistics Laboratory (CLUL) at the Faculty of Arts, University of Lisbon. Eye movements were recorded with the SMI IVIEW X ™ HI-SPEED system, at a 1250Hz speed and each participant calibration error never exceeded 0.5° of visual angle in average of horizontal and vertical deviation (x and y coordinates). Sound was recorded with a Logitech® Webcam Pro 9000 camera.

Stimuli were presented with the Experiment Center software (from SMI), divided into two blocks of text, formatted with New Courier font in size 22, with two paragraphs spacing between rows, in a 17-inch screen. In addition to the initial calibration, the equipment was recalibrated for each part of the experiment (training, first text reading, second text reading) and whenever necessary. Before the task, participants were informed about the procedure, namely that they should read aloud the text, at a normal speed, and that by the end of each text they would have to answer a short comprehension questionnaire. The presentation of the texts was alternated among the participants.

3. Analysis

Data analysis of "reading speech" was performed with the Praat [12] software and the prosodic annotation ToBI_PE ([13] and [14]) was used. Ocular behavior analysis was performed with BeGaze software for data extraction, and subsequently the data was analyzed in R version 3.0.2 [15], using Rstudio, version 0.98.1062.

According to other studies, we assume that the first fixation in a word is reactive to its orthographic, phonological and possibly morphological properties, allowing specific processes to visual word recognition (VWR). Mean first fixation duration (FF) is our first dependent variable. All the time spend in a word, before leaving it, could tap processes besides VWR, such as those involved in less frequent or long words processing. Therefore our second measure is first pass (FP), which includes first fixation and other fixations before moving the eyes to right or left regions. Finally, our third measure is total time of word fixation (TTF), which includes FP and any fixation time spent in the word, no matter if it comes from left or right regions. TTF must reflect word integration in a semantic-discursive mental representation, and can be taken as a window for *wrap-up* effects (see, amongst others, [16]). Within texts, we do not expect an effect of position in FF, regarding its value as an indicator of VWR, but there should be differences between texts given vocabulary specificity. Position effects on FP and TTF are expected within and between texts.

Concerning speech analysis, we choose as dependent variables mainly acoustic parameters usually associated with the presence of prosodic boundaries in Portuguese, such as stressed vowel length (SVL), indicating the proximity of a high level prosodic boundary ([13] and [17]) and fundamental frequency of the vowel (F0), indicating the position of the word in the sentence.

As already noted, we considered two independent variables: Text and Position. The variable Text has two levels: less complex text (CO) and more complex text (ITA). The variable Position unfolds on three levels: Syntactic Nucleus (SN), Syntactic Boundary (SB) and Discursive Boundary (DB).

Acoustic and eye tracking measures were extracted from three critical regions, each one corresponding to a word located in a strategic syntactic position (head or right edge), or a discourse unit (before a period). The selection of these regions is justified by the assumption that in each of these points occur different integrative processes. We expect that Discursive Boundary is a location for possible *wrap-up* processes, resulting from integration and storage of information contained in the sentence ending ([16], [18] and [19]). At Syntactic Boundary we expect to find effects of structure integration. The Syntactic Nucleus position was selected as a baseline condition, since it corresponds to a syntactic head, that, when perceived, has not yet been projected, so it should not reflect integrative processes.

The fundamental frequency analysis points to the existence of interaction between Text and Position variables, through the fall of these values from SN to SB and from SB to DB, from CO to ITA text. This interaction indicates a close relationship between both Text and Position variables.

3.1. Results

3.1.1. Syntactic and Discursive position effects on eye movements

To assess the effects of syntactic and discursive position on eye movements, we conducted analyzes of variance (ANOVA), considering the following variables: Text at two levels (CO and ITA), and Position at three levels (SN, SB and DB). Following standard procedures, fixations shorter than 80 ms were excluded (see, for example, [20]). Figure 1 presents the values for the FF, FP and TTF on each position and each text. The summary of the analysis of variance results are presented in Table 1.

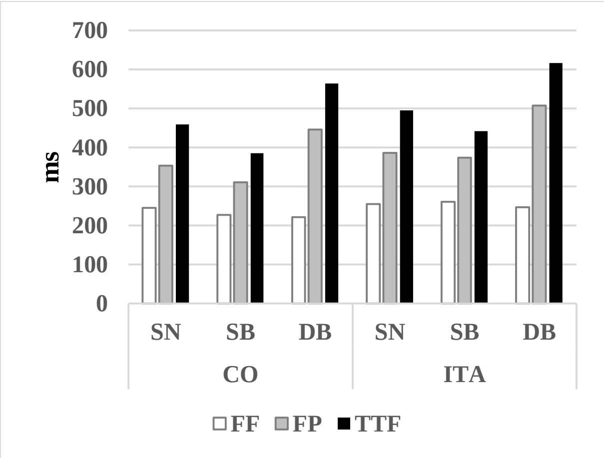

Figure 1: *Mean values (ms) for visual reading variables (FF, FP and TTF) by Text (CO, ITA) and by Position (SN, SB, DB).*

There is a Text effect in FF ($F(1, 16) = 6.931$, $p = 0.02$), in FP ($F(1, 16) = 17.539$, $p < 0.01$), and in TTF ($F(1,16) = 18.050$, $p < 0.01$) (Figure 1). Results show a strong effect of Text condition, with significant differences between the two texts in all dependent variables measured, with lower times in CO than in ITA. These results confirm the complexity gradient between the two texts, and support the hypothesis of speakers' sensitivity to linguistic and discursive complexity.

		FF		FP		TTF	
	df	*F*	*p*	*F*	*p*	*F*	*p*
Text	16	6.931	**0.018**	17.539	**0.001**	18.050	**0.001**
Position	32	1.289	0.290	14.035	**0.000**	20.040	**0.000**
Text*Position	32	0.625	0.542	0.423	0.659	0.118	0.890

Table 1. *ANOVA results, in the 3 measured variables (FF, FP and TTF), in the different conditions of Text and Position and in the interaction between Text and Position. Values of* p*<0.05 are signaled with bold.*

Position effect was found in FP ($F(2, 32) = 14.035$, $p < 0.01$), with differences between SN and DB ($p<0.01$), and between SB and DB ($p<0.01$), with longer reading times for DB in both cases. A position effect was also found in TTF ($F(2,32) = 20.040$; $p < 0.01$), with differences between all conditions: SN higher than SB ($p = 0.03$) and lower than DB ($p < 0.01$); SB lower than DB ($p < 0.01$). Regarding the Position effect, there are no differences between the three regions analyzed in respect to FF.

During the first pass, there were differences between Syntactic Nucleus and Discursive Boundary and between Syntactic Boundary and Discursive Boundary, but there were no differences between Syntactic Nucleus and Syntactic Boundary. Concerning the TTF we found differences between all positions, being the longest value for Discursive Boundary and the shortest for Syntactic Boundary. The results also reveal that the *wrap-up* processes are mainly registered in Discursive Boundary, as expected, and are identifiable mainly in total time measures [19].

3.1.2. Syntactic and Discursive position effects on "reading speech"

To assess the effects of syntactic and discursive position in reading speech, we conducted analyzes of variance, considering the two variables used for eye movements: Text and Position. Figures 2 and 3 present the mean values for the Stressed Vowel Length (in ms) and Fundamental Frequency (in Hz), on each position and each text. The summary of the analysis of variance is presented in Table 2.

There is a Text effect in SVL ($F(1, 16) = 39.397$; $p < 0.01$) with lower values in CO than in ITA (Figure 2). There is no effect of Position, although we observed an effect of the interaction between Text and Position ($F(2, 32) = 12.329$; $p < 0.01$).

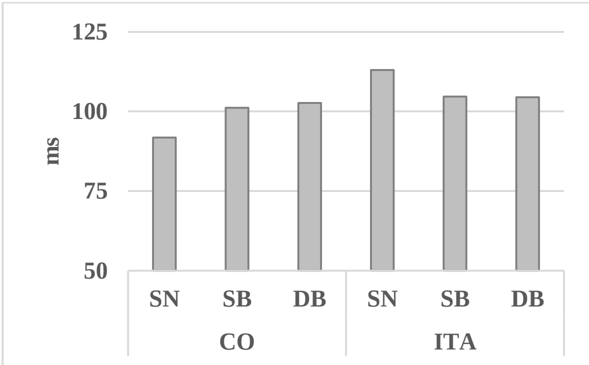

Figure 2: *Mean values for "reading speech" variable SVL for both texts, by Position (SN, SB, DB).*

The interaction between Text and Position in Stressed Vowel Length could be explained by SVL values at the Syntactic Nucleus, which present the shortest values in CO and the longer values in ITA. Syntactic Nucleus values vary clearly between texts, while the remaining values are similar.

Concerning F0, there are effects of Position ($F(2, 32) = 205.68$, $p < 0.01$) (Figure 3). There are differences between all the conditions, being SN higher than SB ($p < 0.01$), and than DB ($p < 0.01$), and SB higher than DB ($p < 0.01$).There is also an interaction between Text and Position ($F(2, 32) = 12.329$, $p < 0.01$), with higher values in SN and SB for CO than for ITA and lower values in DB in CO than in ITA. The F0 values, when comparing the two texts, are higher in CO both in SN and SB, but lower in DB, the discursive information integration region.

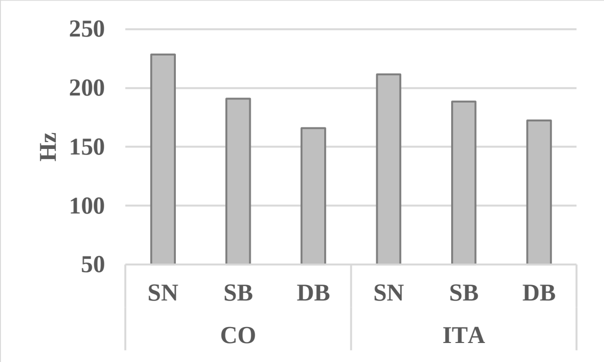

Figure 3: *Mean values for "reading speech" variable F0 for both texts, by Position (SN, SB, DB).*

		SVL		F0	
	df	*F*	*P*	*F*	*P*
Text	16	39.397	**0.000**	3.567	0.077
Position	32	0.219	0.804	205.680	**0.000**
Text*Position	32	12.329	**0.000**	10.572	**0.000**

Table 2. *ANOVA results, in the 2 measured variables (SVL and F0), in the different conditions of Text and Position and in the interaction between Text and Position. Values of p<0.05 are signaled with bold.*

3.1.3. Eye movements and "reading speech" variables correlation

Returning to the hypothesis of the existence of a possible relationship between behavioral indicators of "reading speech" and patterns of eye movements during the processing of the same linguistic stimuli, we proceeded to a correlation analysis (Pearson correlation coefficient) among the variables related to eye movements and reading speech production (except TTF) by Position and Text. From tests carried out it was found that all correlations were positive.

In the easiest text (CO), the value of the First Fixation correlates with both F0 and Stressed Vowel Length. FF with F0, has a correlation in the overall analysis ($r = 0.33$; $p = 0.02$; 11%), i.e., irrespective of the position where the measurement was taken. We also found a correlation at Syntactic Boundary ($r = 0.60$; $p = 0.01$; 37%). FF with SVL just has a correlation with Syntactic Nucleus position ($r = 0.58$; $p = 0.02$; 33%). FP correlates with F0 in the Syntactic Boundary ($r = 0.57$; $p = 0.02$; 33%) and, in the general analysis, with SVL ($r = 0.31$; $p = 0.03$, 9%). In the more complex text (ITA), there is only a correlation between FP and Syntactic Boundary at SVL ($r = 0.69$; $p < 0.01$; 48%).

4. Discussion and Conclusions

This study explores the possibility that visual and prosodic variables may relate to cognitive processes triggered by processing textual properties. To test the hypothesis concerning the effect of Text complexity and syntactic and discursive Position on vocal and visual behavior, we proceed to an analysis of variance where we examined results in eye movements - First Fixation, First Pass and Total Time of word Fixation -, and in reading speech - Stressed Vowel Length and Fundamental Frequency. Finally, we run an analysis of correlation between eye movements and speech variables to discuss the hypothesis of a possible relation between speech and eyes.

If we accept that First Fixation and First Pass are responsive to processes concerning the visual word recognition and lexical access, and that Total Fixation Time can unveil *wrap-up* effects of previous information integration, we should expect Text effect on all visual reading variables. The results confirmed the expectations: there are always lower values for CO than for ITA that is, reading times are lower in the most accessible text, in terms of topic and vocabulary familiarity. We interpret these results as a consequence of two properties of textual complexity impacting on visual perception and information integration. First, higher values in First Fixation and in First Pass may indicate the specificity of the technical vocabulary in ITA, which is due to the existence of more low frequency words. Second, complexity is also explained by the impact of the degree of familiarity with the text subject: less familiar theme (ITA) must induce more costs for integration, due to the issue novelty, and for working memory, given that lexical specificity delays structure building. The difference between the mean TTF, regardless position, at CO and at ITA (470ms vs. 518ms, respectively) confirms this effect: less familiar theme means that there is more information still available at Discursive Boundary; CO information is being integrated and resolved before reaching the main boundary marked by the lowest score in TTF.

Considering that texts were read at a normal speed rate and that they are different at vocabulary frequency and specificity, we would expect to find a Text effect on SVL. This was confirmed by the lengthening of stressed vowels within words of the most complex test. One way to explain this result relies on the possibility that reading a text with a less frequent vocabulary delays visual word recognition, affecting speech planning time to their production, lengthening the stressed vowel.

The similarity between texts in structural levels – syntactic and discursive – allows us to expect an effect of Position but not a Text effect. Hence, we did not find any effect of Text, however we found an effect of Position in each text. Being F0 an indicator of prosodic/syntactic boundary - a locus of closure of structure building and information integration - we registered significant variation comparing the three positions inside each text. These expected results confirm F0 as an indicator of prosodic phrasing in EP, responsive to syntactic and discursive cues, and not to lexical frequency.

Taking up the question of the relationship between speech and eyes, we have some indicators that the lengthening of the stressed vowel can be associated with the first fixation in the visual word, both sensible to lexical properties, while F0 is better associated with total time of word fixation, both revealing integration of information previously processed and *wrap-up* processes.

5. Acknowledgements

This work was done with Fundação para a Ciência e a Tecnologia financial support through PEst-OE/LIN/UI0214/2013 budget.

6. References

[1] A. Inhoff, M. Solomon, R. Radach, and B. Seymour, "Temporal dynamics of the eye-voice span and eye movement control during oral reading," *Journal of Cognitive Psychology*, vol. 23, no. 5, pp. 543-558, 2011.

[2] J. Ashby, J. Yang, K. Evans, and K. Rayner, "Eye movements and the perceptual span in silent and oral reading," *Attention, Perception, & Psychophysics*, vol. 74, no. 4, pp. 634-640, 2012.

[3] C. Perfetti, "Comprehending written language: the blueprint of the reader," In C. Brown and P. Hagoort (Eds.), *The neurocognition of language*. Oxford University Press, pp. 167-208, 1999.

[4] F. Vellutino, "Literacy: Reading (Early Stages)," *Encyclopedia of Cognitive Science*, 2006.

[5] A. Costa,"Leitura: conhecimento linguístico e compreensão." In M. R. Delgado Martins et al. *Para a Didáctica do Português, seis estudos de Linguística*. Lisboa: Colibri, pp. 75-117, 1992.

[6] A. Costa, I. Falé, I. Hub Faria, and T. Condelipes, "Assessing oral reading fluency and comprehension," Poster presented at the *10th International Symposium in Psycholinguistics*, Donostia – San Sebastian, 13-16 April, 2011.

[7] *Common Core State Standards for English Arts & Literacy in History/Social Studies, Science, and Technical Subject. Common Core State Standards Initiative*. Preparing America's Students for College & Career. http://www.corestandards.org/ELA-Literacy.

[8] A. Costa, "Processamento sintáctico e compreensão na leitura – análise de estratégias prosódicas na Leitura Oral," *Actas do VIII Encontro da Associação Portuguesa de Linguística*. Lisboa, pp. 112-125, 1992.

[9] P. Stenneken, M. Conrad, and A. Jacobs, "Processing of syllables in production and recognition tasks," *Journal of Psycholinguistic Research*, vol. 36, no. 1, pp. 65-78, 2007.

[10] S. Frota, M. Vigário, F. Martins and M. Cruz, *FrePOP* (Frequency Patterns of Phonological Objects in Portuguese), Version 1.0, Laboratório de Fonética (CLUL), Faculdade de Letras da Universidade de Lisboa, 2010.

[11] *Multifunctional Computational Lexicon of Contemporary Portuguese*, available at http://www.clul.ul.pt/en/resources/88-project-multifunctional-computational-lexicon-of-contemporary-portuguese-.

[12] P. Boersma and D. Weenink, "Praat: doing phonetics by computer" [Computer program]. Version 5.3.82, retrieved 27 July 2014 from http://www.praat.org/.

[13] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg, "ToBI: A standard for labeling English prosody," *Proceedings of the International Conference on Spoken Language Processing*, pp. 867-870, 1992.

[14] M.C. Viana and S. Frota (Coords.), "Towards a P_ToBI", available at: http://ww3.fl.ul.pt//laboratoriofonetica/sonsemelodias/P-ToBI/P-ToBI.htm, 2007.

[15] R Development Core Team, *R: A language and environment for statistical computing*, Vienna: R Foundation for Statistical Computing, 2011. Available at http://www.r-project.org/.

[16] K. Rayner, B. Juhasz, and A. Pollatsek, "Eye movements during reading," In M. Snowling and C. Hume (Eds.) *The Science of reading:a handbook*. Oxford: Blackwell Publishing Ltd, 79-97, 2005.

[17] C. Gussenhoven and T. Rietveld, "Intonation contours, prosodic structure, and preboundary lengthening," *Journal of Phonetics*, vol. 20, pp. 283-303, 1992.

[18] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychological Bulletin*, 124, pp. 372–422, 1998.

[19] M. Hirotani, L. Frazier, and K. Rayner, "Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements," *Journal of Memory and Language*, 54, pp. 425-443, 2006.

[20] A. Staub, S.White, D. Drieghe, E. Hollway and K. Kayner, "Distributional Effects of Word Frequency on Eye Fixations", *Journal of Experimental Psychology: Human Perception Performance*, 36, 5, pp. 1280-1293, 2012.