Expressive *vs* neutral prosody in reading aloud: from descriptive binary to continuous features

Julien Magnier[1], *Maya Gratier*[2], *Anne Lacheret*[3]

[1, 2] Laboratoire Ethologie, Cognition, Développement, Université Paris Ouest Nanterre, France
[3] Laboratoire Modyco, Université Paris Ouest Nanterre, France

jmupon@gmail.com, gratier@gmail.com, anne@lacheret.com

Abstract

In this paper, we propose to compare expressive and neutral oral renditions of a children's tale in French by examining the segmentations performed by twelve high level French readers. We used a software dedicated to this kind of analysis (Analor) which takes into account different parameters (pause, pitch gesture, pitch jump) and their relative strength to determine pertinent prosodic units (phrases). The extraction of these phrases and their features enables us to observe the influence of both the type of oralisation (expressive or neutral) and punctuation signs on the organization of speech flow. Results show there are more prosodic phrases in the expressive readings (specially at comma locations), that their boundaries are more clearly demarcated, and that they have more varied contours than in neutral readings.

Index Terms: expressive prosody, neutral prosody, narrative, oral segmentation, phrasing

1. Introduction

Vocal expressivity, which we define as the capacity to express feelings, intentions and attitudes, is an important issue for research on oral communication and lies at the interface of computational linguistics, phonetics and psycholinguistics [1]. A number of prosodic resources can be used to impart expressivity to written text. [2] Some of these are universal features of all languages (variations in melody, rhythm and intensity, etc…) while others are language-specific (the nature of the variations, the use of melodic rather than temporal cues or vice versa). We focus in this paper on the prosodic process of grouping and phrasing in French performed by readers in a narrative task during a controlled study of oralized reading (Section 2). The aim of this study was to measure the variations in phrasing and prosodic grouping, and the nature of their correlation with punctuation signs[1], which distinguish two modes of oralisation, a neutral reading and an expressive reading.

Several studies confirm the influence of punctuation in the marking of breaks in oralized text [3], [4] and the variable durations of these breaks according to the type of punctuation sign [5] or the level of closeness of the sequence [6], [7]. However, other studies highlight the role of vocal expressiveness in the placement of these breaks, produced sometimes in places not oriented by punctuation marks or by discourse structure, but rather by the emotional qualities of the text [8], [9], [10]. Reading aloud is an activity which requires an emotional and physical involvement on the part of the reader [11], and we may assume that the structuring of breaks and intonation groups varies according to the degree of this involvement.

It is thus relevant to question the role of vocal expressiveness in the processes of segmentation in reading. Would the prosodic structuring of a text be less marked by readers who've been asked to read a story in a neutral manner ?

To answer this question, we tested 3 hypotheses: i) the number and location of phrase boundaries should be greater in the expressive reading condition; ii) the relative strength of phrase boundaries should be greater in the expressive condition; and iii) expressive reading should involve more pitch variation, i.e. specific contour types.

2. Corpus

2.1. Participants and task

Participants were 12 adults (6 men and 6 women ranging from 20 to 55 years in age) who considered themselves to be 'good readers'. They were asked to read aloud a children's tale (*the laughter of the frog* – 426 words, 23 full stops, 42 commas) in two styles: an expressive style, varying the tone of voice to convey emotion, and a neutral one, i.e. without expressing emotion.

(…) Dans sa folie des grandeurs, la grenouille avait asséché toute la planète. Les êtres vivants mourraient de soif et commençaient à suffoquer. (…)

Example 1.*Text extract from the tale*

The text was presented on a white sheet of paper in a single block, without paragraph breaks. Each reader was given the time he/she needed to familiarize him/herself with the tale and once he/she was ready, performed each of the 2 tasks (reading styles) as often as necessary until he/she was satisfied. The readings were recorded by means of a digital audio device (Roland BR600) with a sampling frequency of 44.1 kHz. The final corpus included 24 recordings (12 neutral readings and 12 expressive readings) with a total duration of 62 minutes.

2.2. Prosodic Segmentation

Each sample was segmented semi-automatically into prosodic phrases with the help of a software program called Analor[2] based on a method that has previously been tested [12]. The algorithm used by Analor is based on a global and multiparametric approach. The segmentation procedure is as follows: only melodic variations in time and breaks are used for segmentation into phrases, regardless of any segmental and syntactic data. Each break of at least 0.1 sec. is assigned a

[1] These markers co-exist with others (such as intonation contour variation, prototypicality of contours, melodic and tempo variation or vocal quality). They are considered easily observable indices for a model of vocal expressivity that precedes the phonetic level of analysis.

[2] http://www.lattice.cnrs.fr/analor

10.21437/SpeechProsody.2014-16

temporary marking and becomes a potential candidate for a phrase boundary. A break is a necessary but not a sufficient marker to locate a potential phrase boundary. In other words, the approach is global because the localization of a boundary can be envisaged only with respect to the combination of several parameters. Two other criteria are also used: (i) the detection of an ample melodic gesture; the trait [±ample] is fixed according to the melodic interval, measured in semitones, between the last extreme F0 value (before the boundary break) and the average F0 over the whole segment preceding the break; and (ii) the detection of a melodic jump which corresponds to the melodic interval which separates the points of F0 before and after the break (melodic resetting). An interesting characteristic of the algorithm is that the decision to place a boundary and its relative strength do not depend on the thresholds of each parameter taken independently but on the interaction between a set of parameters (Table 1, Figure 1).

PAUSE (seconds)	GESTURE (semi tones)	JUMP (semi tones)	Strength of parameter
x < 0,25	x < 2	x < 2	- 1
0,25 < x < 0,6	2 < x < 5	2 < x < 3,5	0
0,6 < x < 1	5 < x < 8	3,5 < x < 7	1
1 < x	8 < x	7 < x	2

Table 2.*Grid of the parameter values used to segment speech into phrases. The strength of the phrase boundary depends on the sum of the strength of each parameter. Each potential phrase that has a general index ≥ 0 must be validated by the experimenter.*

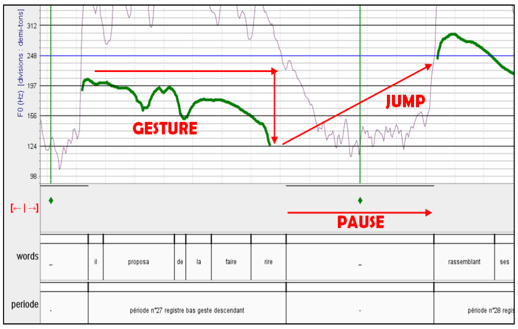

Figure 1: Screenshot of the software Analor and schematization of the criteria used for phrase segmentation

3. Data analysis

3.1. Analysis criteria

The comparisons between expressive and neutral readings were conducted, first, on the basis of the number of phrases produced by the subjects in the readings. We also compared the readings with regard to phrase boundaries taking into account text-based punctuation signs, i.e. 23 full-stops (excluding the final stop) and 42 commas. The tale contains 5 other punctuation marks, but these were excluded from analysis. The relative strength of phrase boundaries was also compared across all of the renditions, taking into account the punctuation signs. Finally, we quantified various types of general contours of phrases, according to the contour description provided by the Analor algorithm (flat, rise, fall). The contour of a phrase is considered flat if the amplitude of the gesture does not reach a minimal threshold (2 semitones (Table 1)) and it is considered as rising or falling if the absolute value of amplitude exceeds the given threshold. Given the size of our sample, the non parametric Mann Whitney test was used to compare the averages.

3.2. Number and location of phrases and phrase boundaries

Figure 2 shows that, on average, readers produced many more phrases in the expressive reading (m=55.16, sd = 14.10) compared with the neutral reading condition (m=33.16, sd = 10.65). This difference is significant (*U*=11, *p*< 0.001). When we take into account the location of phrases and their correspondence with punctuation marks, it appears that full-stops almost systematically correlate with prosodic phrase boundaries in both conditions (m(exp) = 21.91, sd = 0.28, et m(neu) = 20.16, sd = 1.40). However, we find a significant difference between the expressive and neutral renditions at the comma level (*U* = 14, *p*< 0.001). The boundaries afforded by this punctuation mark are much less respected in neutral reading (m=8, sd = 6.68) than in expressive reading (m=21.66, sd = 7.61). Lastly, participants marked on average a slightly higher number of boundaries at unpunctuated locations in expressive reading (m=7.25, sd = 7.67) that in neutral reading (m=2.16, sd = 3.35) but this difference is not significant (*p* = 0.07).

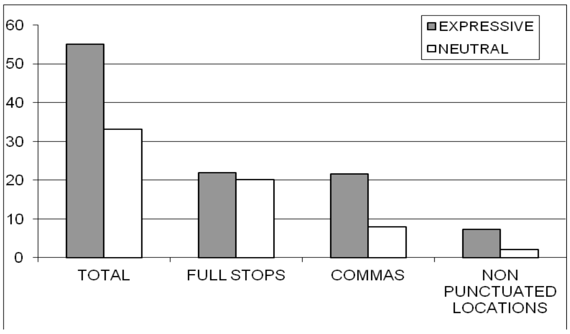

Figure 2: Mean number of phrases produced according to the versions and with respect to their locations

3.3. Strength of phrase boundaries

Based on the multiparametric index of boundary strength provided by Analor, it appears that the marking of boundaries is significantly stronger in the expressive readings (Figure 3) for all phrases (m(exp) = 2.40, sd = 0.46 and m(neu) = 1.41,sd = 0.54, *U* = 15, *p*< 0.001), with regard to full-stops (m(exp) = 2.83, sd = 0.63 and m(neu) = 1.65, sd = 0.66, *U* = 16, p < 0.01), and with regard to commas (m(exp) = 2.15, sd = 0.59,and m(neu) = 0.67, sd = 0.54, *U* = 4, *p*< 0.001). In both versions, boundaries are more strongly marked at full-stop locations than at comma locations (*U*(exp) = 31, *p*< 0.05 and *U*(neu) = 19, *p*< 0.01).

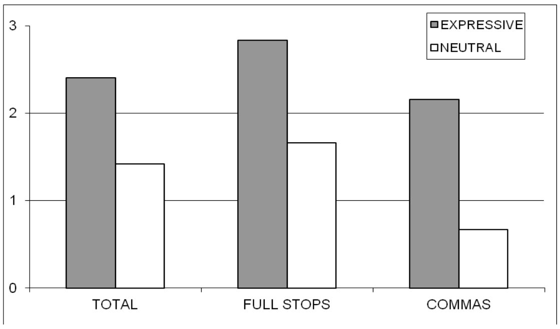

Figure 3: Mean index of cut strength according to the versions and with respect to their locations

3.4. Phrase contours

As shown in figure 4, the mean proportions of contour types are different in the two conditions. A higher proportion of flat contours is produced in neutral reading (m(exp) = 0.19, sd = 0.11, and m(neu) = 0.48,sd = 0.28,$U = 26.5$, $p< 0.01$), whereas a higher proportion of rising contours is produced in expressive reading (m(exp) = 0.17, sd = 0.12,and m(neu) = 0.04, sd = 0.05, $U = 29.5$, $p> 0.05$). Furthermore, a higher proportion of falling contours is produced in expressive reading, but this difference is only a trend (m(exp) = 0.63, sd = 0.12, and m(neu) = 0.46, sd = 0.27, $p = 0.15$).

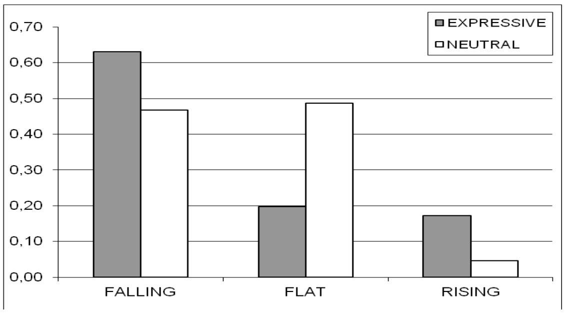

Figure 4: Mean proportions of contour types in the two versions

4. Conclusion and discussion

The aim of this study was to compare the prosodic grouping processes involved in expressive and neutral reading.

Our findings reveal a set of general factors that determine phrase segmentation in both conditions. Phrase boundaries, for instance, essentially coincide with punctuation marks and the full-stop is almost always a phrase boundary indicator. With regard to the strength of boundary markers, full-stops more frequently entail a segmentation than commas. This set of results was highly predictable as it illustrates the facilitative role of punctuation in segmenting a text into semantic units during oralisation. In addition, more than 45% of the phrases produced by the readers have a falling contour in both conditions. The high proportion of this contour type can be explained by the natural phenomenon of declination associated with the assertive modality of most of the clauses in the text.

However, our findings reveal a number of interesting differences between the expressive and neutral renditions of the text. It appears, first, that half of the commas give rise to prosodic boundary markers in the expressive condition as opposed to a little less than a quarter in the neutral reading condition. Readers thus use punctuation as a resource for conveying meaning and emotion in the oral rendering of written text. Furthermore, boundaries are more clearly marked (in terms of strength), whether they coincide with full-stops or commas, in the expressive reading condition.

These two observations point to a general principle associated with neutral reading, that of a minimal segmentation into prosodic units. The more frequent and marked segmentation in expressive reading, which we find in other genres such as political speeches, oratory or sermons, suggests that rhythm is used in oral discourse to support intersubjective engagement and emotional bonding. In effect, the neutral style can be considered an artefact of the experimental situation, and to be ecologically invalid, especially in a narrative genre where the plot itself pushes the reader to tell the story with emotion[1]. For this reason, the major implication of our findings is that expressive segmentation should be treated not in binary terms but rather as a continuous process of combining variable prosodic features.

Finally, the low frequency of flat contours in expressive reading compared to neutral reading shows that the prosodic structure of a text is made richer and more contrasted in the expressive rendition. These features may then facilitate the on line processing of prosody by the receiver and, it can be hypothesized, its comprehension and memorization. Further research should be undertaken to test this idea.

[1] Because lexical and syntactic properties of text carry intrinsic expressivity independently of prosody [13].

5. References

[1] Suciu, I., Kanellos, I. and Moudenc, T., "Expressivité et synthèse vocale. Isotopies expressives, cohérence discursive et structures prosodiques," *Nouveaux Cahiers de Linguistique Française*, vol. 28, pp. 199-206, 2007.

[2] Patel, R. & Mc Nab, C., "Displaying prosodic text to enhance expressive oral reading," *Speech Communication*, vol. 53, pp. 431-441, 2011.

[3] Rossard, B. & Cosnier, J., « Etude des pauses dans la lecture orale, » *Psychologie Française*, vol.26, pp. 54-67, 1981.

[4] O'Connell, D. & Kowal, S., "Use of punctuation for pausing : Oral readings by German radio homilists," *Psychological Research*, vol. 48, pp. 93-98, 1986.

[5] Martin, P., "Ponctuation et structure prosodique," *Langue Française*, vol. 172, pp. 99-114, 2011.

[6] Zvonik., E. & Cummins, F., "Pause duration and variability in read texts," in *Proc. 2002 International Conference on Spoken Language Processing*, 2002.

[7] Smith, C. , "Topic transitions and durational prosody in reading aloud : production and modelling," *Speech Communication*, vol.42, pp. 247-270, 2004.

[8] Campione, E. & Véronis, J., « Etude des relations entre pauses et ponctuations pour la synthèse de la parole à partir de texte », in *Proc. 2002 Congrès TALN*, 2002.

[9] Wang, X., Li, A., Yuan, C., "A preliminary study on silent pauses in Mandarin Expressive Speech", in *Proc. 4th Conference on Speech Prosody*, 2008.

[10] Viola, I. & Madureira, S., "The roles of pause in speech expression", in *Proc. 4th Conference on Speech Prosody*, 2008

[11] Sterponi, L., "Reading as involvement with text : Insights from a study of high functioning children with autism", *Rivista di Psicolinguistica Applicata*, vol. 3, pp. 87-114, 2007.

[12] Lacheret, A. & Victorri, B., "La période intonative comme unité d'analyse pour l'étude du français parlé : modélisation prosodique et enjeux linguistiques," *Verbum*, vol.24, pp. 55-73, 2002.

[13] Lacheret, A. & Legallois, D., "Expressivité vocale et grammaire : comment le symbolique construit le prosodique ? " in Gaudemar, M. (eds.), *Les plis de la voix*, pp. 45-54, 2013.