



The prosody of the TV news speaking style in Brazilian Portuguese

Luciana Castro, Ben Serridge, João Moraes, Myrian Freitas
Federal University of Rio de Janeiro, Brazil

Abstract

This study characterizes the prosodic structure of the TV news speaking style in Brazil and compares it to the speech of interview subjects on a television talk show. Fifteen distinct metrics, designed to characterize both temporal and melodic characteristics of speech, were evaluated on the two speaking styles. The results of the analysis show that the TV news speaking style is characterized by a lack of filled pauses, a fast speaking rate, few and short silent pauses, and a high percentage of dynamic tones. However, the only metric for which the difference in speaking styles was shown to be statistically significant was the intra-vocalic mean melodic movement.

Key words: prosody, TV news, speaking styles, professional speech.

Introduction

TV news is a professional activity that relies heavily on oral communication. Previous studies suggest the existence of specific melodic lines that characterize this professional speaking style (Fónagy, 1976; Léon, 1993). This study attempts to quantify the prosodic variations of the speech of Brazilian TV news anchors.

Method

The data used in this study consist of ten minutes of labelled speech, one minute each from five TV news anchors and five interview subjects. All speech was recorded directly from Brazilian television. The interview speech consists of responses given during a television talk show, and was included in this study to allow for comparison of professional read speech and non-professional spontaneous speech. The speakers, all native speakers of Brazilian Portuguese with no apparent communication disorders, are all male between the ages of 46 and 68. Details regarding the recordings, speech data captured and transcription of the speech are described in Castro et al. (2010a).

A perception test using the same corpus demonstrated that listeners are able to distinguish between these two speaking styles with 90% accuracy, even when semantic and lexical information is removed from the signal by a low-pass filter (Castro et al., 2010b). This suggests that there are sufficient

differences in the temporal characteristics and in the fundamental frequency contour to allow listeners to distinguish these speaking styles. If so, it should be possible to define metrics that capture these differences.

As a starting point for the metrics described below, the data were analyzed based on the segmentation of the corpus into phonetic sequences and pauses. Praat was used to calculate the fundamental frequency for individual frames of speech, spaced at intervals of 10 msec (Boersma & Weenik, 2008).

The metrics used to quantify the temporal characteristics of speech were: number of silent and filled pauses per minute; average pause duration; total pause time; average phonetic sequence duration; number of syllables per phonetic sequence; and speaking rate (Freitas, 1992). To characterize the fundamental frequency contour of the TV news speaking style, the metrics used in this study were: the average and standard deviation of the fundamental frequency; the interquartile range, which is the range measured after excluding the highest and lowest 5% of F0 values; the percentage of static, rising, or falling tones (calculated only for stressed vowels); and the intra-vocalic mean melodic movement, expressed in semitones per second, which measures the sum of the absolute value of changes in fundamental frequency from one frame to the next, considering only the values within vowels (Goldman et al., 2008).

Finally, an attempt was also made to quantify the occurrence of pitch reset, which was calculated, for each pause, as the difference between the average F0 of the first stressed vowel after the pause and the average F0 of the last stressed vowel before the pause (Heldner et al, 2004). Based on this definition, two metrics were calculated: the percentage of pauses that exhibit a pitch reset of more than two semitones, and the pitch reset averaged over all pauses

Results

A striking characteristic of the TV news speaking style is the complete absence of filled pauses. In the five minutes of TV news speech, there was not a single filled pause, compared to 19 in the interview speech. Regarding silent pauses, in this corpus there were 19.8 silent pauses per minute of speech, with an average duration of 327 ms, for a total pause time of 11% of the recorded speech. The average duration of the phonetic sequences in the TV news speech was 2306 milliseconds, with 15.6 syllables per phonetic sequence and a speaking rate of 6.0 syllables per second. As a point of comparison, Uglova & Shevchenko (2006) observed a slower rate of speech (5.1 sil / sec) in English TV news broadcast.

To put these results in perspective, the same temporal metrics were evaluated on the speech of interview subjects, and a Student's t-test was then performed on each metric. Table 1 shows the results of the comparison.

Surprisingly, the temporal characteristics of professional TV news speech do not differ significantly from those of spontaneous interview speech. As expected, the TV news speaking style exhibits slightly fewer silent pauses per minute, and these pauses are slightly shorter, but the speaking rate is approximately the same and the differences observed in pause time are not statistically significant.

Table 1: Results and p-value of the Student’s t-test of the temporal metrics.

Metric	TV news	Interview	p-value
number of silent pauses per minute	19.8	21.70	0.49
average pause duration (ms)	327	387	0.16
total pause time (%)	11	14	0.09
average phonetic sequence dur (ms)	2306	2216	0.56
number of syl per phonetic sequence	15.6	15.8	0.76
speaking rate (syllables / second)	6.0	6.1	0.80

The results of the metrics related to the fundamental frequency contour are presented in Table 2. As before, a Student’s t-test was applied to compare the two speaking styles. Note that the only measure that differs in a statistically significant way between the two speaking styles was the intravocalic mean melodic movement. This may indicate that the professional speakers make more use of short-range pitch movements, although other related metrics, such as the percentage of rising or falling tones, do not reinforce this conclusion with statistical significance.

Table 2: Results and p-value of the Student’s t-test of the F0 metrics.

Metric	TV news	Interview	p-value
mean fundamental frequency (Hz)	139	155	0.35
standard deviation of F0 (ST)	4.3	4.0	0.41
interquartile range (ST)	14.1	12.2	0.10
mean melodic movement (ST/seg)	36.3	30.7	0.04
percentage of pauses with pitch reset > 2 ST	55	42	0.07
average pitch reset (ST)	2.6	3.1	0.72
percentage of static tones (%)	46	54	0.24
percentage of rising tones (%)	26	25	0.79
percentage of falling tones (%)	28	21	0.34

Conclusion

In summary, the results of the analysis presented here show that the TV news speaking style is most distinctly characterized by a lack of filled

pauses. It is relatively fast speech, with few and short silent pauses, and a high percentage of dynamic tones. When compared with the speech of talk show interview subjects, the metrics defined in this study are not capable of differentiating the two speaking styles. Of the 15 prosodic metrics evaluated in this study, the only one that exhibited a statistically significant difference was the intra-vocalic mean melodic movement. However, in a perception test based on the same corpus, listeners were able to distinguish the two speaking styles from low-pass filtered speech with an accuracy of 90%, suggesting that there must be prosodic differences between the two speaking styles that are not captured in the metrics proposed here. Perhaps the relevant prosodic markers are intermittent and therefore diluted in the global metrics commonly proposed in the literature? Further research is required to develop more sophisticated metrics that capture the perceptually relevant aspects of prosody that listeners depend upon to identify the characteristics of professional speaking styles.

References

- Boersma, P., Weenink, D. PRAAT: doing phonetics by computer (5.0.21). Available at: <<http://www.praat.org/>>. Accessed: April, 23rd.
- Castro, L., Serridge, B., Moraes, J., Freitas, M. 2010a. Characterizing variation in fundamental frequency contours of professional speaking styles. In: *Proceeding of Speech Prosody*, Chicago.
- Castro, L., Serridge, B., Moraes, J., Freitas, M. 2010b. Listeners' Ability to identify professional speaking styles based on prosodic cues. In: *Proceeding of Speech Prosody*, Chicago.
- Fónagy, J. 1976. Prosodie professionnelle et changements prosodiques. *Le Français Moderne*, XLIV(3): 193-228.
- Freitas, M.J. 1992. Contributo para o estudo de padrões de estruturação temporal da fala no português europeu. In: Pereira, I. (org.) *Estudos em prosódia*. Lisbon: Colibri, 74-103.
- Goldman, J.-Ph., Auchlin, A., Avanzi, M., Simon, A.C. 2008. ProsoReport: an automatic tool for prosodic description. Application to a radio style. In: *Proceedings of Speech Prosody*, Campinas, 701-704.
- Heldner, M., Edlund, J., Björkenstam, T. 2004. Automatically extracted F0 features as acoustic correlates of prosodic boundaries. *Proceedings of FONETIK*, Stockholm, 52-55 .
- Léon, P. 1993. Variation situationnelle et voix professionnelles. In: *Précis de phonostylistique*. Paris: Nathan, 157-184.
- Shevchenko, T., Uglova, N. 2006. Timing in news and weather forecasts: implications for perception. *Proceedings of Speech Prosody*, Dresden, 5-8.