Sport in the media: a contrasted study of three sport live media reports with semi-automatic tools

Stephanie Audrit[1], Tea Pršir[2,3], Antoine Auchlin[3], Jean-Philippe Goldman[3]

[1] CENTAL, UCLouvain, Belgium, [2] VALIBEL, UCLouvain, Belgium
[3] Department of Linguistics, University of Geneva, Switzerland

Stephanie.Audrit@uclouvain.be - Tea.Prsir,Antoine.Auchlin,Jean-Philippe.Goldman@unige.ch

Abstract

This communication presents a phonostylistic description of sport live media reports, contrasting three different sports, basketball, football ('soccer') and rugby. It is based on 12-minute long recordings for each sport. The corpus is aligned at a phone level with *EasyAlign*, and manually annotated for the type of event in the game. It is analyzed with *ProsoReport* and *ProsoDyn* tools, a set of Praat scripts that aims at detecting in a dynamic way macro-prosodic "ambiance changes" in speech, based on robust relativized prosodic measures (F0, duration, intensity). Dynamic macro-prosodic contrasts between sports show their specificity, as well as what their family resemblance rests on, as contrasted to another speech genre (reading aloud). At the same time, this communication evaluates the semi-automatic *ProsoDyn* toolkit's robustness on noisy data such as sport live report.

Index Terms: sport live report; phono-genre prosody; phonostylistic dynamic description; semi-automatic prosodic description tools

1. Introduction

Sport live media report (SLMR) has received recent attention in prosodic studies from different perspectives. [1] evaluate this verbal activity as for its "phono-genre" consistency, as contrasted to three other speech genres (viz. radio news, political discourse, and religious discourse), through an identification test. Besides this observation on shared common knowledge on kinds of verbal activities, SLMR has been devoted descriptive attention, for various sports (horse race [2]; rugby [3]; soccer [4]; quidditch [5]) and media conditions: radio ([4]; [5]), tv [3]; from an interactional [4]; [5]; [6]; or phenomenological perspective [3].

At large, all studies agree on the existence of a consistent phono-stylistic generic picture; they also agree on a first level internal distinction between "off-line commentary" and a gross "on-line", or "time-critical" [4] speech activity. Further subdivision on smaller sequences and their legitimation is issued by [4], for soccer radio report, through phonostylistic "building up suspense" and "presenting a climax" distinction. To what extent is this subdivision sport, or media, dependent, or cross-sports and media is not clearly stated.

This communication reports on prosodic and phonostylistic properties of three different sports live TV reports corpora. Each sport constitutes a "sub-genre" whose description is conducted with *ProsoReport* [7] that provides a global prosodic description, and with *ProsoDyn* [8][9][10] that aims at detecting, representing, and describing macro-prosodic ambiance changes, on one or many parameters. More precisely we look for prosodic correlations with game's action phases and moments.

2. Material

The corpus of this study is composed on one hand by three different sports commentaries in French, extracted from video match SLMRs retrieved from the Internet. We extracted out audio files of around 12 minutes each.

On the other hand it contains 3 recordings of oral reading, from "C-prom" corpus [11], as external phonostylistic reference. Table 1 gathers each recording's features:

Table 1. *Description of 4 genres, duration and num. of speakers*

Speech genre	Content	Time	Num. of speakers (% of time)
Basket	Liège-Ostend, 2011	11'32''	1 (100%)
Football	France-Brazil, 1998	13'36''	2 (52%+47%)
Rugby	France-New Zealand, 2007 World Cup	10'07''	3 (71%+26%+3%)
Oral reading	C-prom multigenre speech corpus	6'41''	3 (28%+35%+37%)

3. Methodology

Each recording has been orthographically transcribed, tagged for sport events, phonetically aligned, and then analyzed with several tools developed by *ProsoReport* and *ProsoDyn* as described below. Alignment has been manually verified in order to avoid possible temporal errors.

3.1. Manual annotation of sequences: typology of annotation

Observation in [4] is based on a sub-jacent typology that may be summarized as Figure 1 below:

Figure 1: *Sport media report phonostylistic taxonomy*
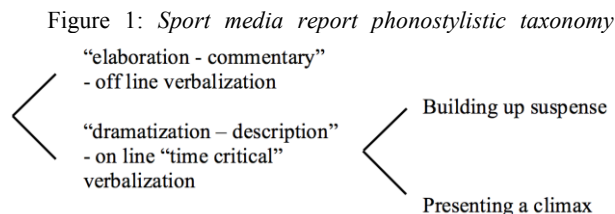
The first level distinction meets most researchers' agreement. As for "dramatization" internal distinctions, the two phonostyles "building up suspense" and "presenting a climax" is both an intuitive hypothesis and a result of semi-formal data observation that should be sustained by more formal parameters measurement and contrast.

However, the present paper's strategy is to differentiate phonostyles, or ambiances, on the basis of semi-automatic procedures. Such differentiation should be an outcome of global analysis. For this purpose, the authors manually tagged events of the three SLMRs independently of prosody. The proposed tagging set is constituted of labels, which aim at describing in a very factual way the events described by the sports reporter. The tags correspond to what is said by the sports commentator, not to what occurs actually within the game. This choice's reason is that we want to study the sports commentator's prosody, and our starting point is sound, not video. The tags (Table 2) were initially developed for basketball, then adapted in order to be general enough to cover the sports studied here, i.e. basketball, football and rugby.

Table 2. *Description of event tags*

Event type	Description
context	Factual items, more or less strongly related to what happens on the ground (e.g. players' or coaches' biographical events)
opinion	Reporter's assessments. Opposite to context, include a part of subjectivity
shot	Attempt to mark a goal, successful or not
possession	A given team or player possesses the ball
score	Expression of the score or of a score trend
offence	Any kind of offence
confrontation	A team attacks or defends
abort	Something goes wrong and breaks the progress of the game
time out	Nothing remarkable happens

3.2. Automatic analysis

The usual process to describe macro-prosodic parameters phonetically aligned [7] is based on successive analyses performed by scripts developed within Praat [12]:
- *EasyAlign* segments the speech recording into phonetic segments, syllables, and words; optionally add morpho-syntactic information to each syllable [13];
- *ProsoGram* stylizes F0 curve, and provides a simplified representation corresponding to perception in syllable nuclei [14];
- *ProsoProm* automatically detects prominent syllables [15]
- *ProsoReport* computes many parameters, including speech and articulation duration, speech rate, syllables mean duration, mean and range of pitch register, proportion of prominent over non-prominent syllables (=*density*).

This set of automatic tools provides a global, detailed, but static prosodic description.

3.3. Graphical representation of prosodic parameters

ProsoDyn [8][9][10] is a new tool that performs a macro prosodic report on a reduced set of parameters (F0, syllable duration, speech rate, and prominence density variation) – but in a dynamic way. An initial two-pass pitch detection and a manual checking prevent from pitch errors. Prosodyn also reports mean and standard deviation for those parameters. It gives two alternative and complementary strategies for the graphical representation of variation:
- a smoothed curve of each parameter based on a "sliding window" technique (SW) with adjustable length and steps (default is 15-syllable analysis window, and 1-syllable-step);
- local mean values for independently pre-defined annotated speech macro-units (MU) that may be genre-dependent or sample-specific (e.g. speaker turns, or match events).

Each strategy corresponds to a specific use. SW strategy is heuristics, as it lets appear parameters variations as more or less coupled to each other. Pre-defined MU strategy is descriptive, and allows for quantification for macro-units and their contrasts. As for the present study, the chosen independent macro-units are based on the game's events. However we report on both strategies' results.

4. Results

Sport report, whether audio or audiovisual, is modeled by the speaker "trying to keep up with the action" ([5]: 239), that is taking an active part to the action through verbal activity arousal. This is the common point to [4]'s both "dramatization" phonostyles.

Our external categorization of events however aims at exploring some subtler differences on prosodies according to event-types. We hypothesize that event-type categorization is somehow prosodically reflected. The non-normality of our data (given by a Shapiro-Wilk test) prevents us from using a statistical t-test. For that reason, we focus on graphical analysis of the data.

4.1. Statistics about manual annotations of events

Table 3 describes the repartition of events by three sport genres and their mean duration (in sec.). The "reading" genre (not in table) has 121 "events" (interpausal speech segments, i.e. portions of speech between pauses of at least 200ms). Their mean duration is 2.55 seconds, with a standard-deviation of 1.31 seconds.

Table 3. *Num. of events by sports (and mean duration in sec.)*

	Basket	Football	Rugby
abort	3 (2.6)	6 (3.6)	0 (0)
confrontation	4 (2.8)	7 (8.2)	13 (8.2)
context	18 (6.2)	34 (6.9)	20 (7.8)
offence	6 (2.8)	4 (5.4)	7 (4.9)
opinion	28 (4.0)	44 (4.8)	14 (8.6)
possession	18 (1.7)	46 (5.3)	9 (11.7)
score	18 (2.7)	5 (2.5)	3 (18.5)
shot	20 (3.7)	9 (3.8)	2 (11.7)
time out	1 (0.7)	0 (0)	0 (0)
Total	116	155	68

4.2. Comparison by events

We suppose that each event (MU tag) shows specific prosodic characteristics, such as (i) articulation ratio, (ii) syllabic duration, (iii) pitch mean, (iv) pitch range and (v) prominent syllables density . As can be seen in figure 2a, "reading" genre has significantly higher speech rate, whereas "possession" event shows a significantly lower speech rate. In figure 2b "reading" and "possession" are again significantly different from the others, but this time grouped together and not opposed, with lower pitch range values. Possession event is constructed as a slow enumeration of players' names according to ball passes rhythm. It is a representative part of SMLR, corresponding to [4]'s "building up suspense".

Fig. 2a and 2b: *Rate and pitch range boxplots by events (contrasted with "lec", reading aloud)*
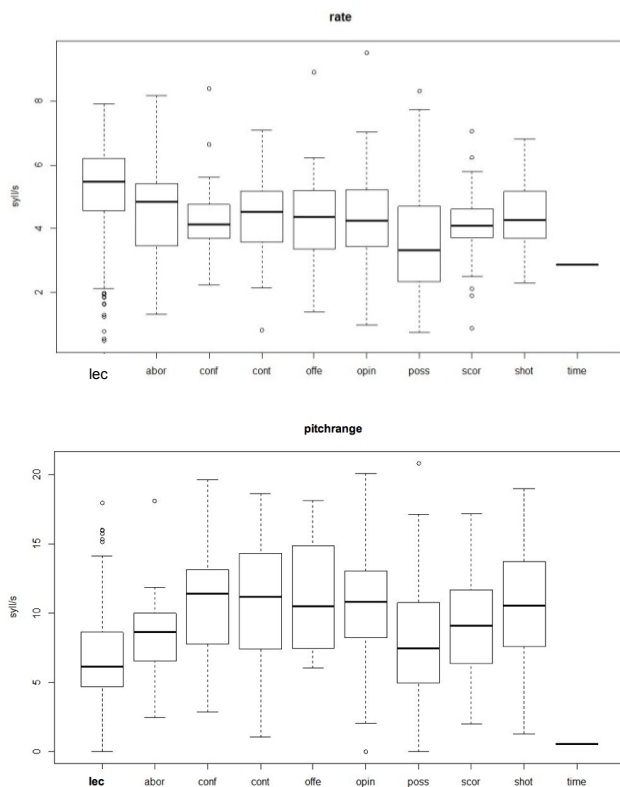

4.3. Comparison by speech genre

ProsoReport is used to compare the global prosodic features of three sport genres between them and with reading speech.

The results of *ProsoReport* show that the break ratio (percentage of pause time) is more important for sport sub-genres (33% for basketball, 33.9% for football, 29.7% for rugby) than for reading aloud (24.5%). Mean F0 for sport sub-genres (92ST) is for 12ST higher than reading aloud one (80.5ST); F0 range is larger for sports (16ST) than for reading (9.4ST), and shows a gradual shift from basketball (15.1ST) to rugby (16.5ST) and football (17.3ST). Articulation rate (syll/s), is higher for reading (6.1) compared to sports (5.2).

These results reflect homogeneity of global prosodic features of three sport commentaries. They reflect the fact that the dynamic of speech of sport commentaries follows the events on the field. The more detailed analysis of comparison by sports and comparison by events will show their inner specificities.

Fig. 3a-c: *Rate, pitch range and articulation ratio boxplots for three sports and reading*
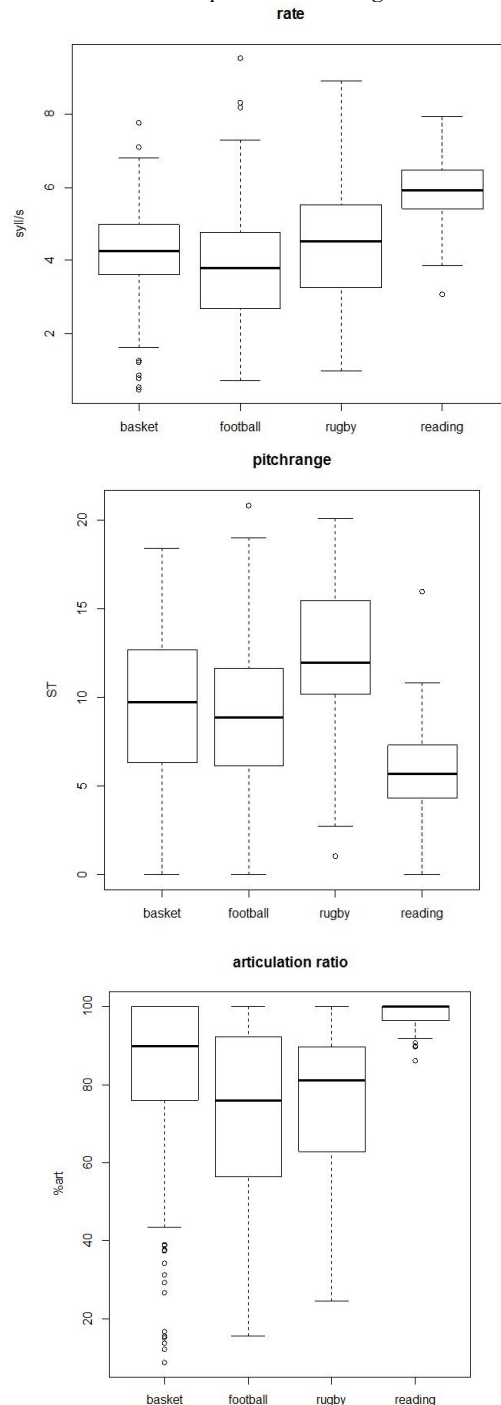

The reading differs significantly from the three sports by its strongly higher speech rate. Rugby's rate is slightly higher than football one. The articulation ratio of reading is particularly high, while that of football is quite low. As for pitch range, the contrast between the 4 styles is stronger: each style differs from each other. Reading has the lowest pitch range, rugby the highest. Sports commentaries and reading strongly differentiate through multiple prosodic parameters.

4.4. Illustration with *ProsoDyn*

Graphical representation of global F0 variation by *ProsoDyn* illustrates how this prosodic parameter is exploited differently in football and in basketball. For football (Fig.4), there is an extended excursion of F0 that follows major sport events (e.g. goal or confrontation). For basketball (Fig.5), F0 is more stable and changes more often according to a bigger and faster number of actions.

The vertical lines in Fig. 4 (football) indicate the distribution of speech between two commentators (1,2). Speaker (2) has a significantly higher pitch than speaker (1) especially during the first 110sec.

Figure 4 & 5: *Global F0 (in ST) variation for football and basket with the same time span (300seconds) and pitch scale*
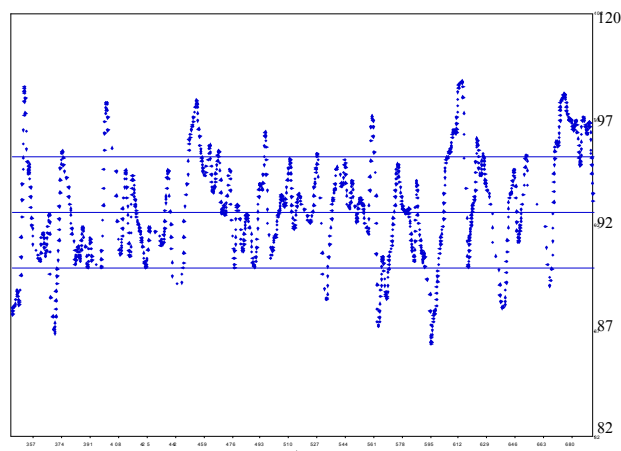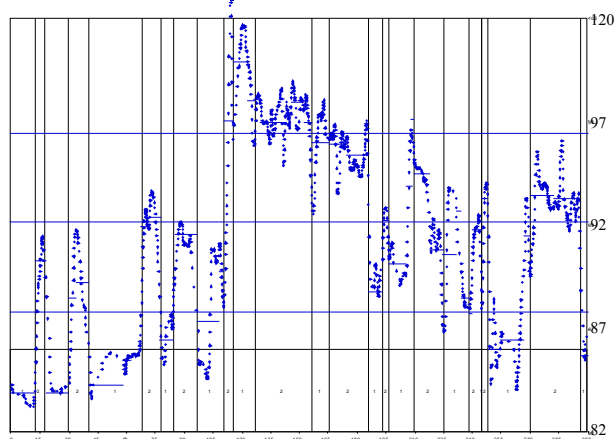
5. Conclusion

Cross analysis of three SMLR sub-genres - basketball, football and rugby - showed similarity and differences at various levels. Even though the differences are subtle, we can regroup rugby and football at one side and basketball at the other. The first two sub-genres are marked by punctual prosodic variation (F0 and rate), whereas basketball is more homogeneous.

We attempt to establish sport events categories shared by three genres. The categories such as "confrontation" or "shot" are clearly distinguished from "context" or "opinion", regarding their lexical and prosodic content.

As for global prosodic analyses, *ProsoReport* provided evidence that the three sport genres analyzed are easily discriminated from the other speech genre (oral reading). Further studies will include more speakers per sport in order to avoid the idiosyncratic bias and will also control and measure the alleged phonostylistic differences between TV and radio. Finally, these recordings initiated new developments for the *ProsoDyn* tool, especially to cope with noisy environment and multispeaker discourses.

7. References

[1] Obin, N., Dellwo, V., Lacheret, A. and Rodet, X., "Expectations for Discourse Genre Identification: a Prosodic Study", Interspeech 2010, Makuhari, Japan, 2010.

[2] Trouvain, J. and Barry W.J., "The prosody of excitement in horse race commentaries", Proceedings ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research, Belfast, Northern Ireland, 86-91, 2000.

[3] Mathon, C. and Boulakia, G., "Le commentaire sportif en direct : une combinatoire de différentes fonctions de la prosodie", in Yoo, H-Y and Delais-Roussarie, E. [eds.], Actes d'IDP 2009, Paris, ISSN 2114-7612, 287-301, 2009.

[4] Kern, F., "Speaking dramatically. The prosody of life radio commentary of football matches", in Barth-Weingarten, D., Reber E., and Selting M. [eds.] Prosody in interaction, John Benjamins, 217-237, 2010.

[5] Wagner, J., "Commentating fictive and real sports", in Barth-Weingarten, D., Reber E. and M. Selting [eds.] Prosody in interaction, John Benjamins, 239-241, 2010.

[6] Müller, T., Football, language and linguistics, Tübingen: Narr, 2007.

[7] Goldman J.-P., Auchlin, A., Simon A.C. and Avanzi, M., "ProsoReport: an automatic tool for prosodic description. Application to a radio style", Proceedings of Speech Prosody 2008, Campinas, Brazil, 2008.

[8] Simon, A.C., Goldman, J.-P. and Auchlin A., "Analyse des variations macro-prosodiques dans un corpus de parole expressive : détection de changements d'ambiance proso-discursive", Symposium Prosody-Discourse Interface, 2011.

[9] Goldman J.-P., Auchlin A., and Simon A.C., "Discrimination de styles de parole par analyse prosodique semi-automatique", in Yoo, H-Y and Delais-Roussarie, E. [eds.], Actes d'IDP 2009, Paris, ISSN 2114-7612, 207-221, 2009.

[10] Goldman J.-P., "ProsoDyn: a graphical representation of macroprosody for phonostylistic ambiance change detection", Proceedings of Speech Prosody, Shangai, China, 2012.

[11] Avanzi, M., Simon, A.C., Goldman, J.-P. and Auchlin, A., "C-PROM. An annotated corpus for French prominence studies", Proceedings of Prosodic Prominence: Perceptual and Automatic Identification, Speech Prosody 2010 Workshop, Chicago, Illinois sites.google.com/site/corpusprom

[12] Boersma, P. and Weenink, D., Praat: doing phonetics by computer (version 5.3), 2011, http://www.praat.org.

[13] Goldman, J.-P., "EasyAlign: a friendly automatic phonetic alignment tool under Praat", Proceedings of Interspeech Conference 2011, Florence, Italy, 2011.

[14] Mertens, P., "Un outil pour la transcription de la prosodie dans les corpus oraux", Traitement Automatique des langues 45 (2) : 109-130, 2004.

[15] Avanzi, M., Goldman, J.-P., Lacheret-Dujour, A., Simon, A.C. and Auchlin, A., "Méthodologie et algorithmes pour la détection automatique des syllabes proéminentes dans les corpus de français parlé", Cahiers of French Language Studies 13/2, 2007.