10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

A cross-linguistics study on how emotion is perceived in sports commentaries: comparing prosodic cues from Japanese and French

Carolyn Fontagnol[1], *Catherine Mathon*[2]

[1]FoReLLIS (EA 3816), University of Poitiers, France
[2]CLLE-ERSSàB (UMR 5263), University Bordeaux Montaigne, France
carolyn.fontagnol@univ-poitiers.fr, catherine.mathon@u-bordeaux-montaigne.fr

Abstract

The purpose of this paper is to provide a better understanding of the characteristics of sports emotion and how it can be related to the five main emotion categories: anger, joy, fear, sadness and disgust. This paper relates three listening tests conducted on French subjects who were asked to judge a set of stimuli taken from a rugby match with both French or Japanese commentary. The selection of the stimuli was based on the main rugby moves (i.e. the ones that score points). The subjects were given three evaluation scales for each stimulus to be evaluated: arousal, valence & intensity. The results show that arousal is the criterion which most frequently characterizes sports emotion. Furthermore, the listening tests demonstrate that the subject can recognize and evaluate sports emotion even when having access only to acoustic features (with semantic information neutralized).

Index Terms: sports emotion, prosody, listening test, cross-cultural study, sports commentary

1. Introduction

Emotional prosody research is a large and important field at the interface of different disciplines such as psychology, neurology, natural speech processing and linguistics. Thanks to extensive literature on this subject, the academic community has gained a good insight into emotional prosody processing, and numerous studies describe prosodic patterns for emotions (anger, sadness, joy, fear and disgust) [1], including mixed emotions or social affects (doubt, authority or surprise for example, as in [2]).

As perception experiments have been frequently used and described, various methodological protocols have been established.

In the research field of phonogenre, sports commentary has been an interesting research object, presenting linguistic specificities such as a profusion of non-verbal phrases [3] [4] and particular prosodic patterns [5]. Some of these studies acknowledge the presence of sports emotion as an element of suspense or tension relative to dramatic speech [5], while others consider sports emotion as a genre constraint linked to the media aspects of such discourse [6]. However none of these studies have tested how sports emotion is perceived by listeners.

Thus, this paper presents a perceptive experimentation on sports emotion. The aims of the study are:

- To acknowledge the existence of sports emotion through detection and evaluation by listeners;
- To characterize sports emotion using traditional semantic scales usually related to emotions;
- To evaluate the impact of prosodic information on the recognition of sports emotion.

2. Corpus

In order to conduct our experiments, we used stimuli taken from a large corpus of TV commentaries of 2007 Rugby World Cup matches [7]. The corpus presents some interesting features:

- it is a multimodal corpus including video and audio data that requires a multimodal analysis to explain the linguistic specificities of a discourse under strict constraints;
- it is a comparable corpus, which is not so frequent in the case of an oral corpus, with commentaries available in French, English and Japanese. For the same footage, the commentary is available either in French and English or in French and Japanese.

The 2007 Rugby World Cup was organized in France and was a major mediatic event that put rugby into the limelight as a national and not merely regional sport.

2.1. Working corpus

We selected a specific match - Japan against Fidji - for which commentaries were available in French and Japanese. This match was chosen for two reasons:

- Commentary was provided by Japanese journalists, who are more likely to take sides for their team and express subjectivity;
- It was a very important match for the Japanese team, and the first time they attracted the public's attention. The team played well, the score was very close and there was a lot of suspense.

2.2. Selection of stimuli

Stimuli were selected according to the final score online report, which indicates the key game plays, the player involved and the timing.

Table 1: *Main rugby game plays during the Japan-Fidji match (2007).*

Japan	**31 - 35**	**Fidji**
Try: 2		Try: 4
Try conversion: 2		Try conversion: 3
Penalty kick: 4		Penalty kick: 3

10.21437/SpeechProsody.2020-10

As said previously, this match produced a lot of suspense because the score was close throughout, including during extra time: Some stimuli were taken from strong attacking team moves which took place during extra time.

21 stimuli were chosen from the video using game plays as time frames for each stimulus. The final repartition of the stimuli in relation to game actions is quite homogeneous as shown in Table 2.

Table 2: *Repartition of the stimuli (%) regarding to game plays.*

Try	29%
Penalty kick	28%
Try conversion	24%
Attacking moves	19%

Three sets of stimuli were extracted from the video commentary in Japanese and French.

- 1 set of stimuli in Japanese;
- 1 set of stimuli in French;
- 1 set of French stimuli which was put through a low-pass filter in order to mask any lexical content.

3. Experimental protocol

3.1. Tasks and subjects

Each set of stimuli was presented to two different groups of 19 and 22 subjects respectively:

- Group FF judged French filtered stimuli ;
- Group J judged Japanese stimuli.

A control group of 6 French subjects was presented with the French stimuli without any filter, so they could have access to lexical content as well as prosodic cues.

Subjects had to evaluate each stimulus according to three five-point scales:

- An arousal scale, from no arousal (1) to full arousal (5) ;
- A valence scale, from very negative (-2) to very positive (+2) with a neutral point (0) ;
- An intensity scale, from very weak (1) to very strong (5).

Subjects are French students between 19 and 25 years old, studying for a Bachelor's degree in Linguistics or a Master's degree in French as a Foreign Language. None of them speaks or understands Japanese.

Each hearing test was conducted online using the LMS (Learning Management System) Moodle platform. Tests were formatted as Moodle xml quizzes containing 21 audio stimuli embedded in 21 questions, each divided into three sub-questions (one for each scale: arousal, valence, intensity).

3.2. Test conditions

Each test lasted approximately ten minutes. Subjects were asked to answer the questions as spontaneously as possible, without too much hesitation. They were also required to complete the test in one sitting, in a calm environment, without any risk of being interrupted. The time taken to complete the tests was then checked to ensure that no subject took too long to answer the questions, which would have indicated they had not done it in optimal conditions (discontinuity, perturbations, etc.).

Answers were collected in a csv file from which data can be filtered and organized in order to obtain descriptive statistics such as standard deviation (see 4.1.) or to observe answers repartition.

4. Results

4.1. Judgments consistency

The first results we were looking for concerned the relevance of using arousal, valence and intensity scales to characterize sports emotion. In order to find out if the answers given by various subjects showed some element of consistency for a given stimulus, we calculated the standard deviation concerning arousal, valence and intensity, for each set of answers and for each stimulus. The results show that there is consistency, the standard deviation never exceeding 1. This means that for the most difficult stimuli to judge, subjects were hesitating between two close points on the scale. We also noticed that group J obtained the best score in terms of consistency, with the lowest standard deviation rates, especially for Japanese tries (standard deviation = 0). The valence scale seemed the most difficult to judge (with less consistent answers) for group J as well as for group FF.

4.2. Sports emotion and game plays

In this section, we use arousal, valence and intensity scales to compare the emotional detection linked to each game plays (try, try conversion, penalty and attacking moves) .

First, the arousal scale (figure 1) shows the highest level of arousal is detected in commentaries over tries and attacking moves. FF group and J group show the same tendencies, but Japanese stimuli covering triess are detected with a higher level of arousal (A5 at 55%). Penalty kicks are also detected as high-level stimuli regarding arousal for group J (A4-A5 at 50%) whereas it only reaches medium level for group FF.

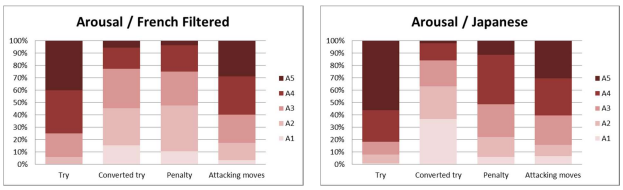

Figure 1: *Levels of arousal detected for commentary over four different game plays, based on French filtered stimuli (left) and Japanese stimuli (right).*

Figure 2 below shows how subjects from group FF and group J evaluate the valence of the commentaries over each type of game plays. We observe that a try is the game play that is perceived as the most positive by both groups (around 70% of the judgements). Attacking moves are very close behind (around 65% of the judgements) but only for group FF.

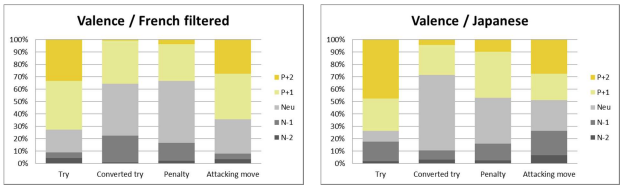

Figure 2: *Levels of valence from very negative (N-2) to very positive (P+2) with a neutral state (Neu) detected for commentary over four different game plays, based on French filtered stimuli (left) and Japanese stimuli (right).*

We can also note that a try conversion is perceived as a neutral game play mostly by group J. This observation could be related to the low level of arousal (lowest scores) perceived for this game play (cf. Fig. 1).

The intensity scale shows that a try and an attacking move create the most emotion. It's interesting to note that intensity and arousal scores seem correlated : the repartition of high intensity levels (I4 & I5) and of high arousal levels (A4 & A5) are quite similar for a try (from 70% to 80%) and an attacking move (from 50% to 60%), whether the signal is filtered (group FF) or not (group J).

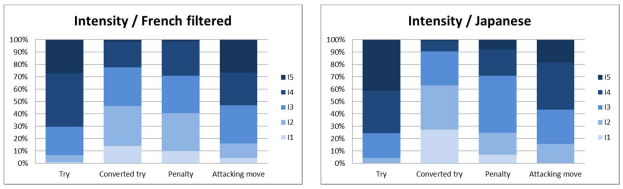

Figure 3: *Levels of intensity from very weak (I1) to very strong (I5) detected for commentary over four different game plays, based on French filtered stimuli (left) and Japanese stimuli (right).*

4.3. Sports emotion and team allegiance

One of the reasons that made us choose the Japan vs Fidji match was the fact that we had a corpus of commentaries in French and in Japanese. Thus, we expected to be able to evaluate the impact of team allegiance (from the Japanese journalists) on the expression of emotions during the match depending on which team scored. We chose to study biased emotion in the Japanese commentaries of tries, which is the game action that enables a team to score the most points (5 points). It is also the game play which registers the highest levels of perceived arousal and intensity (see section 4.2). During the Japan vs Fidji match, 6 tries were scored, 4 by the Fijians and 2 by the Japanese.

Figure 4 below shows the levels of arousal (in shades of red), valence (from grey to yellow) and intensity (in shades of blue) detected by the subjects listening to six Japanese stimuli (group J) corresponding to the commentary of the 6 tries scored. It is very clear from figure 4, which shows tries in chronological order from left to right, that the highest levels of arousal and intensity are detected and the valence is always positive to very positive when the Japanese score a try.,. However, for Fijian tries, even though arousal and intensity are still in the higher levels, the repartition is more balanced between levels 4 and 5, while the valence is more neutral or even negative.

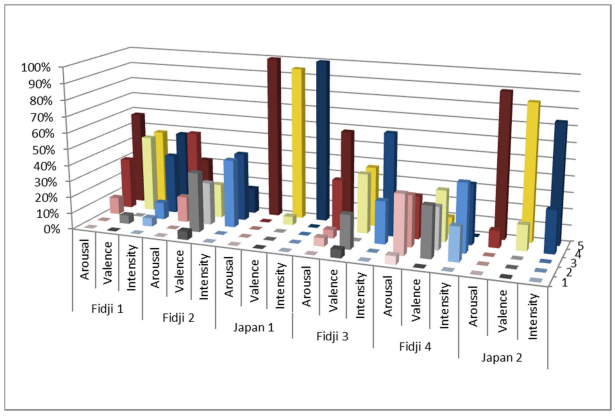

Figure 4: *Levels of arousal, valence and intensity detected in Japanese stimuli corresponding to the commentary over 6 tries, presented in chronological order.*

A Mann-Whitney test was run to confirm the presence of a significant difference between the answers given for the 4 Fijian tries and those given for the 2 Japanese tries. For arousal as well as valence and intensity, the p-value is well above 0.05 and even 0.1. However, these poor significance scores can result from the facts that the number of answers was too small and that the difference between the number of tries scored by each team (2 vs 4) was too big.

4.4. Recognition of sports emotion: the impact of prosody

Many studies have shown that prosody alone can enable us to recognize various kinds of emotions very accurately, from the big five (anger, joy, sadness, fear and disgust) to more complex cultural centered ones. One aim of this study was to confirm that subjects supplied only with prosodic cues could still recognize and characterize sports emotion. For this purpose, we used two methods to hide the semantic content of the stimuli we tested:

- French stimuli were altered with a low-pass filter;
- Japanese stimuli were evaluated by French subjects who did not speak or understand this language.

Thus, we compared the results obtained through both methods with the results we obtained from the control group (French subjects who had access to French stimuli without any alteration).

Results obtained from the control group show few differences between stimuli with lexical content and filtered stimuli. Indeed, the evaluation of arousal shows similar tendencies: high levels of arousal are detected for tries and attacking moves while converted tries and penalty conversions are perceived with a lower level of arousal (A2-A3).

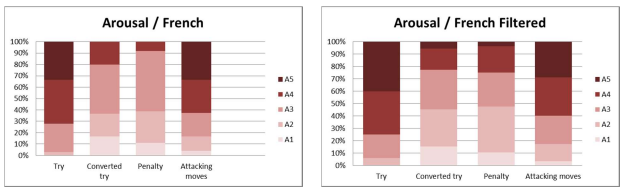

Figure 5: *Levels of arousal detected for four different game plays commented on, based on French stimuli (left) and French filtered stimuli (right).*

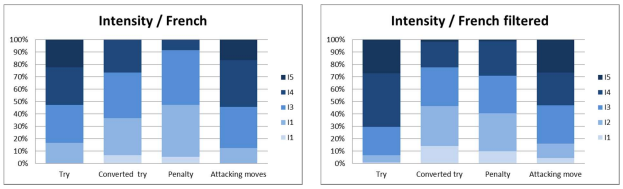

Figure 6: *Levels of intensity detected for commentary over four different game plays, based on French stimuli (left) and French filtered stimuli (right).*

The intensity results present some differences. Indeed, intensity seems to be perceived as much stronger in the French filtered stimuli, especially for tries, and this was unexpected considering that low-pass filtering tends to diminish signal intensity. We can form the hypothesis that in this case intensity is a mirror of the arousal perceived in the same conditions and that, without sufficient input, subjects confuse arousal and intensity.

5. Discussion

This paper presents the first results of an exploratory perceptive study on sports emotion recognition using a comparable corpus of sports commentaries in French and Japanese. The results presented here are therefore leads for further studies with statistically-significant sized groups.

According to these first results, sports emotion can be characterized as presenting a high-level of arousal and intensity, similar to anger or joy. Evaluating the valence of sports emotion proved more difficult, especially when the subjects had no access to lexical content. They did however show tendencies to evaluate sports emotion as positive, closer to joy than anger. The results concerning the high-level of arousal are coherent with previous characterizations such as in [8]. The difficulties to discriminate a positive emotion from a negative one the subjects met on the valence scale were expected because they had been reported in other studies as indicated in [9].

It is interesting to note that sports emotion, like other more traditional emotions, can be detected using prosodic cues only and evaluated with the same traditional scales (arousal, valence and intensity). Two different methods were used to conceal the lexical content (low-pass filter, and foreign language). Given that prosodic expressions of emotions are universal regardless of languages [10], asking the subjects to evaluate emotion from a foreign language they do not understand seems to be an efficient way to avoid the signal alteration induced by low-pass filtering.

Using a corpus consisting of commentaries in two different languages for the same visual object enabled us to make some interesting comparisons. The most relevant ones stemmed from comparing expressions of emotions enhanced by the importance of the game plays in terms of scoring and strategy, particularly as half of them were colored with team allegiance. Sports commentary belongs to the journalistic genre of discourse, which is characterized by objectivity. Even then, it seems like team bias plays a part in sports commentary, as a reinforcement of arousal and intensity. It could be that sports commentary is overall a media genre of discourse, which implies constraints on the discourse such as making the commentary dramatic enough to keep the viewers interested throughout the match.

Regarding the details of sports emotion related to prominent game plays, it seems that not all scoring game plays induce the same level of emotional expressions. This nuances the first hypotheses made in [11]. Indeed, if tries commentary, as anticipated, convey high levels of arousal and intensity, try conversions or penalty kicks are more neutral. It is a counterintuitive observation, especially in this particular match, since penalty kicks were the key game plays that nearly won the match for the Japanese. On the contrary, attacking moves, even if failing to score, are still an object which generates the expression of sports emotion. So, scoring is not the only issue that triggers emotional expression. Pragmatic circumstances could also shape discourse: as shown in [12], correlations can be made between game rhythm and prosodic patterns. Unlike tries, penalty conversions, , are slow game plays in the first phase of preparation. In this case, the result comes as soon as the ball has been kicked. As the commentary follows the game play, it is mainly "color-commentary" [13] in the first phase of the conversion and very short "play-by-play" [13] in the second phase, which carries the only element of excitement. This could explain why penalty conversions, even if they increase the score, are not assessed with the same emotional charge.

6. Conclusions

It would prove interesting to further inquire into and confirm the trends shown in this exploratory study with larger groups, to get statistically significant results.

Our study has shown that sports emotion was mainly perceived as a highly aroused and intense emotion, such as anger or joy. Indeed, the valence, arousal and intensity scores given to actions tend to prove that either no emotion is perceived (neutral valence) or strong emotion is perceived.
This will guide acoustic analysis in further studies to see if the prosodic patterns for sports emotion are the same as those for anger and joy.

Prosodic information seems sufficient to allow the recognition of sports emotion, or at least to highlight the fact that something exciting and intense is happening. As for the valence of the emotion (positive or negative), it seems more difficult to judge with prosodic material only. It seems plausible that lexical content and an access to a wider context are needed to make a more accurate evaluation of the valence. A linguistic analysis will be conducted in order to assess the part played by semantics in the recognition of emotion more accurately.

7. Acknowledgements

The authors would like to thank the students who have taken part in this experimentation as willing subjects, with no other reward than the satisfaction of having participated in a research project (and chocolate!).

8. References

[1] Scherer, K.R., Vocal Communication of Emotion: A Review of Research Paradigms. *Speech Communication* 40, 227-256, 2003.

[2] Albert Rilliard, Takaaki Shochi, Jean-Claude Martin, Donna Erickson, Véronique Aubergé. Multimodal Indices To Japanese And French Prosodically Expressed Social Affects. *Language and Speech*, SAGE Publications (UK and US), 52 (2&3), pp.223-243, 2009.

[3] Deulofeu, J., Les commentaires sportifs constituent-ils un "genre", au sens linguistique du terme ? In Bilger, M. (ed.),

Corpus: Méthodologie et applications linguistiques, Paris : Champion, 2000.

[4] Augendre, S.,Kupsc, A., Boyé, G.,Mathon, C., Live TV sports commentaries: specific syntactic structures and general constraints, in (Legallois, D. Charnois, T., Larjavaara, M., eds.) *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, De Gruyter Mouton, Trends in Linguistics, p. 194, 2018.

[5] Audrit, S., Pršir, T., Auchlin, A. & Goldman, J.-P. (2012), Sport in the media: a contrasted study of three sport live media reports with semiautomatic Tools, *Proceedings of Speech Prosody 2012*.

[6] Mathon, C., Boyé, G., Augendre, S., Kupść, A., Contraintes sur le discours et genre de discours contraint : le commentaire sportif télévisé en direct, *Discours,* 17, 2015.

[7] Lortal, G., Mathon, C., Motion and Emotion or How to Align Emotional Cues with Game Actions, in *Proceedings of EMOT Worshop*, LREC 2008, Marrakech, 2008.

[8] Trouvain, J., Barry, W.J., The prosody of excitement in horse race commentaries, in *Proceedings of ITRW on Speech and Emotion 2000*, 86-91, 2000.

[9] Bąk H., The State of Emotional Prosody Research—A Meta-Analysis. In: *Emotional Prosody Processing for Non-Native English Speakers.* The Bilingual Mind and Brain Book Series, vol 3. Springer, Cham, 2016.

[10] Ekman, P., Sixteen enjoyable emotions. *Emotion Researcher*, 18, 6–7, 2003.

[11] Augendre, S., Mathon, C., Boyé, G., Kupsc, A., Influence des contraintes extra-linguistiques sur le discours : le cas du commentaire sportif télévisé, in *Proceedings of CMLF 2014*, Berlin, 19-23 juillet 2014.

[12] Mathon, C., Boyé, G., Kupsc, A. (2016), Commentaire sportif en direct : Etude des correspondances entre le rythme du jeu et le rythme de parole, in F. Neveu, G. Bergounioux, M.-H. Côté, J.-M. Fournier, L. Hriba et S. Prévost (Eds.), *Proceedings of CMLF 2016*, Tours, 4-8 juillet 2016.

[13] Hartmann, C., *Pre-fabricated Speech Formulas as Long-term Memory Solutions to Working Memory Overload,* PHD Dissertation, University of Zurich, 2011.