Computing and evaluating the Body Laughter Index

Maurizio Mancini, Giovanna Varni, Donald Glowinski, and Gualtiero Volpe

InfoMus - University of Genoa
maurizio.mancini@dist.unige.it
giovanna@infomus.org
donald.glowinski@dist.unige.it
gualtiero@infomus.org

Abstract. The EU-ICT FET Project ILHAIRE is aimed at endowing machines with automated detection, analysis, and synthesis of laughter. This paper describes the Body Laughter Index (BLI) for automated detection of laughter starting from the analysis of body movement captured by a video source. The BLI algorithm is described, and the index is computed on a corpus of videos. The assessment of the algorithm by means of subject's rating is also presented. Results show that BLI can successfully distinguish between different videos of laughter, even if improvements are needed with respect to perception of subjects, multimodal fusion, cultural aspects, and generalization to a broad range of social contexts.

1 Introduction

Traditional Human Computer interfaces are frequently perceived as "cold, incompetent, and socially inept". According to Zeng and colleagues, this results from the fact that they ignore the user's affective state and consequently miss a key component of human-human communication [1]. This is why, in the last years, progress was made toward the creation of emotional Human-Computer interfaces, see for example Affective Computing [2] and Kansei Information processing [3].

Laughter is a relevant component in human-human nonverbal communication and it is a powerful trigger for facilitating social interaction. Indeed, Grammer [4] suggests that it conveys signals of social interest and reduces the sense of threat in a group [5]. Further, laughter seems to improve learning of new activities from other people [6] and facilitates sociability and cooperation [7].

For the above reasons, the newly started EU-ICT FET Project ILHAIRE (`http://www.ilhaire.eu`) aims to investigate how machines can decode laughter (i.e., to know when the user is laughing, to measure intensity of laughter, to distinguish between different types of laughter) and also how Embodied Conversational Agents can communicate laughter.

In our work, we mainly focus on the detection and on the analysis of the movement descriptors (e.g., speed, direction, periodicity, and so on) that are deemed to characterize laughter. Very few researchers investigated the role that body plays in human laughter, even if all of them agree that body configuration and dynamics contribute to the communication of different types of laughter.

Ruch and Ekman [8] observed that laughter is often accompanied by one or more (i.e., occurring at the same time) of the following body behaviors: "rhythmic patters (five pulses per second)", "initial forced exhalation", "rock violently sideways, or more often back and forth", "nervous tremor ... over the body", "twitch or tremble convulsively". Becker-Asano and colleagues [9] observed that laughing users "moved their heads backward to the left and lifted their arms resembling an open-hand gesture". De Graaf [10] observed that laughing consists of a deep inspiration followed by a rapid convulsive expiration whereas de Melo et al. [11] implemented a virtual agent that "convulses the chest with each chuckle". Finally, Markaki and colleagues [12] analyzed laughter in professional meetings: the user laughs "accompanying the joke's escalation in an embodied manner, moving her torso and laughing with her mouth wide open" and "even throwing her head back".

A pioneering system including automatic detection of laughter is the *Affective Multimodal Mirror* [13][14]. This system "tries to induce positive emotions in users by showing a distorted ("funny") representation of their face" [13]. The system senses and elicit laughter, based on a vocal and a facial affect-sensing module, whose outputs are integrated by a fusion module.

The above studies suggest that it should be possible to develop systems for automatic detection of laughter and even differentiate between different types of laughter (e.g., hilarious vs. social [12]). In this paper, we present a preliminary work in this direction in the framework of the ILHAIRE Project: we conceived and implemented the Body Laughter Index (BLI), an index that, by combining together movement descriptors, allows to automatically determine whether a user is laughing or not. We also describe a pilot evaluation study we conducted on the BLI.

2 Computation of the Body Laughter Index

Figure 1 depicts the software architecture for computing BLI. Next subsections provide details on the major software modules. All of them have been implemented in the EyesWeb XMI platform (`http://www.eyesweb.org`) and in the EyesWeb Expressive Gesture Processing Library [15]. EyesWeb is a software platform that allows developers to implement software modules for automatic analysis of user's expressive movement in an intuitive, visual way.

Based on the above literature, since laughter implies deep breathing (e.g., [10]) and possible rhythmic patterns, the initial set of descriptors taken into account for developing BLI includes shoulders correlation and energy of body movement, integrated with a measure of periodicity of movement.

2.1 Tracking of head and shoulders

Starting from an input video source (e.g., recorded video or camera), we detect and track the 2D position of user's head and shoulders. Head and arms

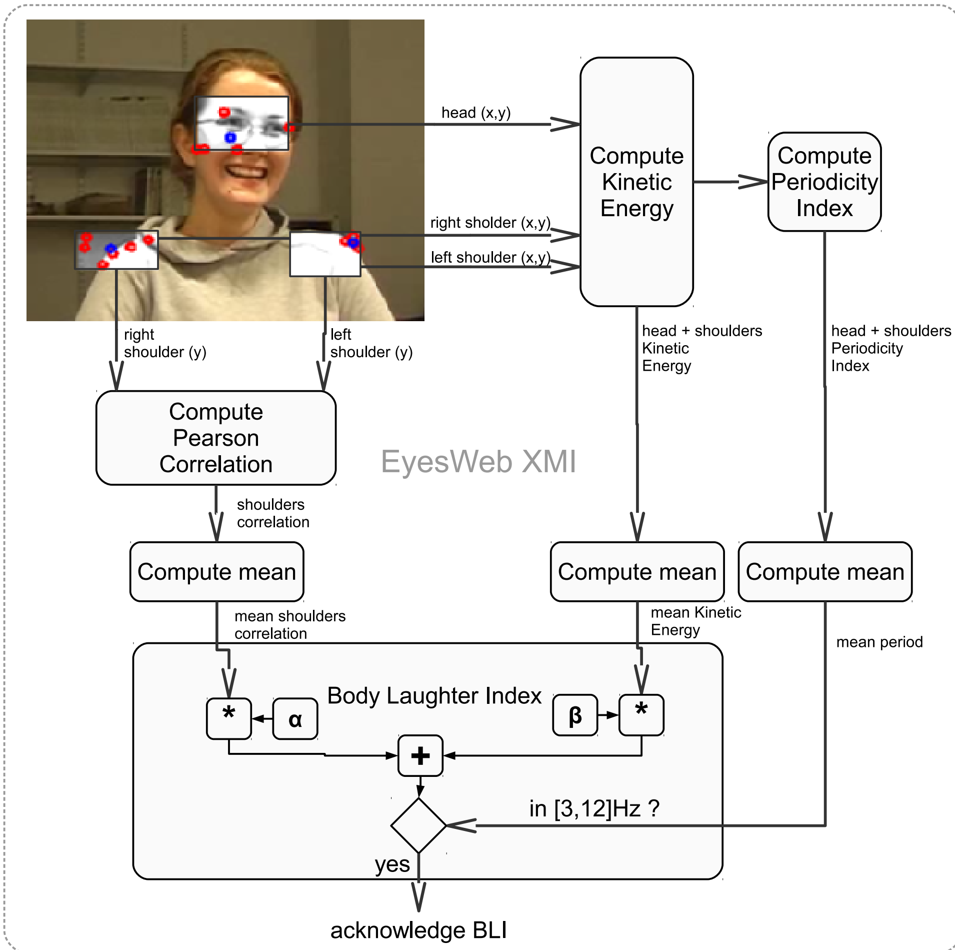

Fig. 1. The software architecture for computing BLI. Firstly, tracking of head and shoulders is carried out: the cloud of red points determines the Regions Of Interest (ROIs) head and shoulders are located in. The blue dots are the geometrical barycenters of each cloud. The boxes are the major software modules extracting and processing movements descriptors.

movements are useful hints to detect one's affect [16]. We manually identify the Regions Of Interest (ROIs) user's head and shoulders are located in (see the light areas in Figure 1). Standard computer vision tracking techniques are applied to each ROI, resulting in a cloud of points for each of them (see the red dots in Figure 1). Then we compute the geometrical barycenter of the cloud (see the blue dots in Figure 1) and we extract its x and y coordinates.

2.2 Low-level descriptors

We extract two low-level descriptors of the head and shoulders movement: *kinetic energy* and *correlation of shoulders movement.*

- *Kinetic energy* (E) is computed from the speed of the head (v_h), of each shoulder's barycenter (v_{ls} and v_{rs}), and their percentage masses (m_h, m_{ls}, and m_{rs}). These are derived from anthropometric tables as referred by [17]. In particular, kinetic energy is computed as:

$$E = \frac{1}{2}\sum_{i=1}^{3} m_i v_i^2 = \frac{1}{2}(m_h v_h^2 + m_{ls} v_{ls}^2 + m_{rs} v_{rs}^2) \tag{1}$$

- *Correlation of shoulders movement* (ρ_s) is computed as the Pearson correlation coefficient between the vertical position of the user's left shoulder and the vertical position of the user's right shoulder. Vertical positions are approximated by the y-coordinate of each shoulder's barycenter extracted as mentioned above.

2.3 Periodicity Index

Kinetic energy is serialized in a sliding window time-series having a fixed length. *Periodicity Index* is then computed on such time-series. The Periodicity Index (PI) is a real-time implementation of the Periodicity Transforms described in [18]. The input data is decomposed into a sum of its periodic components by projecting data onto periodic subspaces. Periodicity Transforms also provide a measure of the relative contribution of each periodic signal to the original one. Among many algorithms for computing Peridiocity Transforms, we choose *mbest*. It determines the m periodic components that, subtracted from the original signal, minimize residual energy. With respect to the other algorithms, *mbest* also provides a better accuracy and does not need the definition of a threshold. Figure 2 shows an example of computation of the Periodicity Index in EyesWeb for the following input signal:

$$I(t) = sin(t) + N(t) \tag{2}$$

where $N(t)$ is a uniform random function generating values in $[0, 0.6]$ to simulate random noise. Such a range for $N(t)$ is chosen so that the noise is strong enough for simulation, but not so strong to destroy the original signal. The Periodicity Index value for such an input function is 20 frames, as Figure 2 shows.

2.4 Body Laughter Index

As mentioned above, the *Body Laughter Index* (BLI) stems from the combination of the averages of the low-level descriptors, integrated with the Periodicity Index. Such averages are computed over a fixed range of frames. However such a range could be automatically determined by applying a motion segmentation

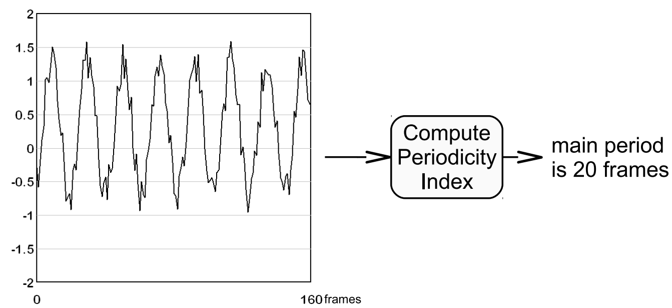

Fig. 2. An example of Periodicity Index computation: the input time-series (on the left) has a periodicity of 20 frames.

algorithm on the video source. A weighted sum of the mean correlation of shoulders movement and of the mean kinetic energy is carried out as follows:

$$BLI = \alpha\overline{\rho_s} + \beta\overline{E} \tag{3}$$

As reported in [8], rhythmical patterns produced during laughter usually have frequencies around $5Hz$. In order to take into account such rhythmical patterns, the Periodicity Index is used. In particular, the computed BLI value is acknowledged only if the mean Periodicity Index belongs to the arbitrary range $[\frac{fps}{8}, \frac{fps}{2}]$, where fps is the input video frame rate (number of frames per second).

2.5 Example

We ran our algorithm for BLI computation on 8 short input videos at 25 fps taken from: (1) a previously recorded video corpus named "The Belfast Induced Natural Emotion Database", collected by the Queen's University of Belfast [19]; (2) the YouTube website (videos generated with the Skype Laughter Chain application, `www.skypelaughterchain.com`). The videos show users laughing while watching funny images on TV. They smile and laugh, tilting their head and producing rhythmic body movements.

Table 1 summarizes the results: the first column reports the video *id*; the second and third columns show the average values of the low-level descriptors (mean kinetic energy and mean Pearson correlation of shoulders movements); the fourth column shows the computed BLI value; the last column reports the mean Periodicity Index.

In this example, parameters for BLI were set to $\alpha = 0.7$ and $\beta = 0.3$, respectively. These are arbitrary values, argued from the literature reported in Section 1. An in-depth study for optimal values of these parameters will be needed in future work.

Table 1. An example of computation of BLI. Kinetic energy $\overline{E}$ ranges in $[0, +\infty)$, correlation of shoulders movements $\overline{\rho_s}$ ranges in $[-1, 1]$, BLI ranges in $[0, +\infty)$, and Periodicity Index $\overline{PI}$ ranges in $[0, w]$, where w is the time window length in frames.

Video id	$\overline{E}$	$\overline{\rho_s}$	BLI	$\overline{PI}$
1	40.7472	0.312	12.44256	16.2778
2	172.6268	0.358	52.03864	16.5362
3	117.4252	0.3508	35.47312	19.6532
4	14.458	0.6982	4.82614	7.874
5	0.5064	0.3092	0.36836	10.7522
6	0.1112	0.0664	0.07984	6.8234
7	250.8674	-0.2226	75.1044	18.9312
8	2.1034	0.5064	0.9855	10.293

3 Evaluation of the Body Laughter Index

BLI was also tested on 8 participants that watched and rated the 8 videos stimuli of Section 2.5. The stimuli were randomized using a balanced latin square. Participants were asked to rate on the following two 3-point Likert items: **Q1** "Did you perceive energetic body movements, involving shoulders rhythmically moving together?" and **Q2** "How fast was the rhythmic movement you perceived?". These items were aimed at an initial assessment of BLI and of its components. Both items were rated from 0 (*not at all*) to 2 (*very much/fast*).

3.1 Video samples

We first checked whether the ratings between videos are significantly different from one another. In other terms, we aimed at checking whether the 8 video samples, submitted to the participants, offered a sufficiently variable level of laughs for the pilot. We ran a Friedman test to observe possible differences between the participants' ratings for items Q1 and Q2 (see Figure 3).

Results show a significant effect for item Q1, $\chi^2(7, n = 8) = 16.492$, $w = 1.4$, $p < .05$, but no effect for item Q2, $\chi^2(7, n = 8) = 12.388$, $w = 1.24$, $p > .05$ ($p = .089$).

Post-Hoc tests were conducted to put in evidence possible differences between videos with respect to their Q1 ratings. The Bonferroni correction was applied to the levels of statistical significance (p-values) to control the inflation of type 1 error probability due to multiple comparisons. A significant difference was found between the ratings of video 2 and video 3 ($p = .032$).

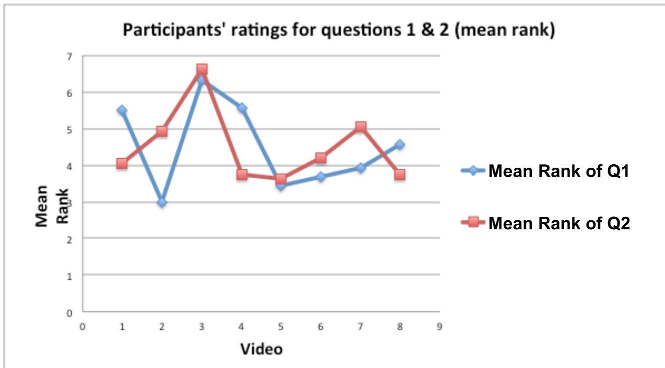

Fig. 3. Participants' ratings for Q1 and Q2.

Table 2. Computation of bivariate Kendall's tau-b correlation between movement descriptors and participants' ratings

		Participants' rating	
		Q1	Q2
Descriptors	BLI	.07	-.09
	PI	-.14	-.25

3.2 Correlation of movement descriptors with participants' ratings

We were interested in evaluating Periodicity Index and Body Laughter Index with respect to the participants' ratings. We conducted a set of bivariate Kendall's tau-b correlations, whose results are shown in Table II. Findings show the highest negative relationship between the Periodicity Index (PI) and Q2: $\tau = -.25$. Smaller relationships were also found between PI and Q1, $\tau = .14$, and between Body Laughter Index (BLI) and Q1, $\tau = .07$, and Q2, $\tau = -.09$.

4 Conclusion

In this paper we presented the implementation and evaluation of the Body Laughter Index, a body descriptor of laughter. Evaluation results show that some improvements are needed to reach successful automatic detection of laughter. The outcomes of BLI computation, reported in Table 1, indicate that BLI, combined with the Periodicity Index, allows us to successfully distinguish between different videos of laughter. However the evaluation of these videos by human participants, reported in Table 2, reveals that BLI and PI only partially match human perception. A possible reason is that laughter is a complex construct depending upon many features, as demonstrated by several studies.

In the future, in the framework of the EU-ICT FET Project ILHAIRE, we aim to carry out multimodal (audio, face, and body) fusion of descriptors: if audio signals analysis, facial expression detection and BLI computation agree with a high statistical significance, then we could claim that the user is laughing. We also aim to automatically differentiate between, for example, hilarious and social laughter. Moreover, cultural aspects need to be considered as modulators of the interpretation of human movement. The resulting multimodal fusion will be assessed with a new set of experiments and the concerning evaluation.

An important issue to be taken into account is the context (activity which is performed, personality of the user, social environment) for laughter detection. Whereas BLI was computed with reference to a specific context (watching funny images on TV) and was evaluated in laboratory conditions, more research is needed to assess to what extent it is able to generalize to other, more general contexts.

From the implementation point of view, we aim to detect user movement with a higher resolution and more reliable systems, enabling to distinguish between different body parts (head, shoulders, and so on), such as Qualisys Mocap (http://www.qualisys.com) and Microsoft Kinect (http://www.xbox.com). An initial real-time implementation of BLI from live video input, using color tracking techniques, was developed and tested at the eNTERFACE'12 Summer Workshop on Multimodal Interfaces (Metz, France, July 2012).

Acknowledgments

We would like to acknowledge I. Sneddon and G. McKeown from Queen's University of Belfast for allowing us to use videos from the Belfast Induced Natural Emotion Database. This work was partially supported by the FP7 EU-ICT FET Project ILHAIRE (Incorporating Laughter into Human Avatar Interactions: Research and Experiments, `http://www.ilhaire.eu`).

References

1. Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on **31**(1) (january 2009) 39 –58
2. Picard, R.: Affective Computing. MIT Press (1997)
3. Hashimoto, S.: Kansei as the third target of information processing and related topics in japan. In: Proceedings of the International Workshop on KANSEI: The technology of emotion. (1997) 101–104
4. Grammer, K.: Strangers meet: Laughter and nonverbal signs of interest in opposite-sex encounters. Journal of Nonverbal Behavior **14**(4) (1990) 209–236
5. Owren, M.J., Bachorowski, J.A.: Reconsidering the evolution of nonlinguistic communication: The case of laughter. Journal of Nonverbal Behavior **27** (2003) 183–200
6. Fredrickson, B., et al.: The broaden-and-build theory of positive emotions. Philosophical Transactions - Royal Society of London Series B (2004) 1367–1378
7. Dunbar, R.: Mind the gap: Or why humans are not just great apes. In: Proceedings of the British Academy. Volume 154. (2008)

8. Ruch, W., Ekman, P.: The expressive pattern of laughter. Emotion, qualia, and consciousness (2001) 426–443
9. Becker-Asano, C., Kanda, T., Ishi, C., Ishiguro, H.: How about laughter? perceived naturalness of two laughing humanoid robots. In: Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on, IEEE (2009) 1–6
10. De Graaf, V.: Human Anatomy. 6th edn. McGraw-Hill, New York, NY, USA (2002)
11. de Melo, C., Kenny, P., Gratch, J.: Real-time expression of affect through respiration. Computer Animation and Virtual Worlds **21**(3-4) (2010) 225–234
12. Markaki, V., Merlino, S., Mondada, L., Oloff, F.: Laughter in professional meetings: the organization of an emergent ethnic joke. Journal of Pragmatics **42**(6) (2010) 1526–1542
13. Shahid, S., Krahmer, E., Swerts, M., Melder, W., Neerincx, M.: You make me happy: Using an adaptive affective interface to investigate the effect of social presence on positive emotion induction. In: Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on. (sept. 2009) 1 –6
14. Melder, W.A., Truong, K.P., Uyl, M.D., Van Leeuwen, D.A., Neerincx, M.A., Loos, L.R., Plum, B.S.: Affective multimodal mirror: sensing and eliciting laughter. In: Proceedings of the international workshop on Human-centered multimedia. HCM '07, New York, NY, USA, ACM (2007) 31–40
15. Camurri, A., Mazzarino, B., Volpe, G.: Analysis of expressive gestures: The eyesweb expressive gesture processing library. In: Gesture-Based Communication in Human-Computer Interaction. Volume LNAI 2915. Springer Verlag (2004) 460–467
16. Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., Scherer, K.: Toward a minimal representation of affective gestures. Affective Computing, IEEE Transactions on **2**(2) (april-june 2011) 106 –118
17. Winter, D.: Biomechanics and motor control of human movement. John Wiley & Sons, Inc., Toronto (1990)
18. Sethares, W., Staley, T.: Periodicity transforms. Signal Processing, IEEE Transactions on **47**(11) (1999) 2953–2964
19. Sneddon, I., McRorie, M., McKeown, G., Hanratty, J.: The belfast induced natural emotion database. Affective Computing, IEEE Transactions on **3**(1) (jan.-march 2012) 32 –41