How is your laugh today?

Maurizio Mancini
University of Genoa
viale Causa 13
16145, Genoa, Italy
maurizio.mancini@unige.it

Gualtiero Volpe
University of Genoa
viale Causa 13
16145, Genoa, Italy
gualtiero.volpe@unige.it

Giovanna Varni
University of Genoa
viale Causa 13
16145, Genoa, Italy
giovanna.varni@unige.it

Antonio Camurri
University of Genoa
viale Causa 13
16145, Genoa, Italy
antonio.camurri@unige.it

Radoslaw Niewiadomski
University of Genoa
viale Causa 13
16145, Genoa, Italy
radoslaw.niewiadomski@dibris.unige.it

CHI 2014, Apr 26 - May 01 2014, Toronto, ON, Canada

DOI http://dx.doi.org/10.1145/2559206.2581205

Abstract

Despite its relevance for human-human communication, laughter has been quite under-investigated and under-exploited in human-machine interaction. Nevertheless, endowing machines with the capability of analyzing laughter (i.e., to detect when the user is laughing, to measure intensity of laughter, to distinguish between different laughter styles and types) in ecological contexts is a very challenging task. An approach to laughter recognition consisting of the real-time analysis of a single communication modality, i.e., body, is presented in this paper and positive results of an evaluation study are discussed.

Author Keywords

laughter detection; body movement analysis;

ACM Classification Keywords

H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous

Introduction

Leonard is in a particularly stressing period: he recently moved to another country, he has some strict deadlines to meet, he feels frustrated and nervous about his situation. So his new colleagues and friends invite him to spend some time watching funny movies and playing amazing

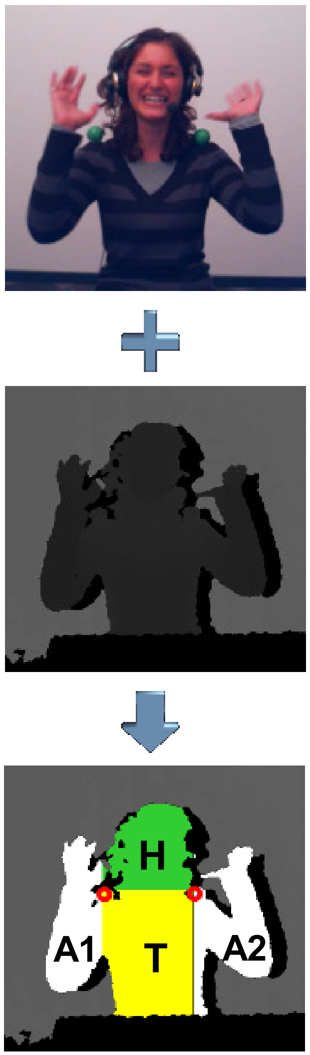

Figure 1: Input RGB and depth images are pre-processed to isolate user's head, trunk, arms and to track shoulder's markers.

board games. During these activities they laugh a lot and Leonard feels better, recovering his well-being. Indeed, the positive effects of laughter have already been observed and measured, e.g., in [12][3]. Also, social contexts could facilitate eliciting laughter [14]. He then decides to probe whether laughter could be automatically elicited and detected by machines. Beside monitoring user's physical and psychological state, such "laughter-sensitive" machines could be used to elicit and measure laughter by involving the user in funny activities.

This is a very challenging task which is starting to be addressed by researchers, see for example the EU Project ILHAIRE (`www.ilhaire.eu`) on laughter detection and synthesis. It is noteworthy to take into account that: laughter is highly multimodal; in social context such multimodality could affect laughter detection (e.g., distinguishing and analyzing users' voices in multi-party interaction is an open challenge; facial activity can not be tracked in ecological contexts); it is not clear if it is possible to distinguish different types of laughter, both at general (e.g., ironic, fearful) and individual (e.g., introvert, extrovert) level, from expressive or morphological multimodal features. The presented study consists of the real-time automated analysis of laughter intensity from a single modality, that is, body movement. It differentiates from previous work on laughter recognition that focuses on other/multiple modalities [7][15]. Recently, it was shown that it is possible to distinguish laughter only from body movement [9]. Intensity has been chosen as it can be evaluated for any laughter type. This is a work-in-progress: further expressive characteristics of laughter, such as up/down-regulation, will be addressed with the same approach as intensity, allowing researchers to overcome the issues about laughter listed above and, in a long-term view, to build "laughter-sensitive" machines.

The approach presented in the remainder of this paper is based on computer vision and machine learning. First, the positions of body parts significant in laughter are extracted from input video streams. Then, low level body features describing laughter are computed. Finally, these features constitute the input vector for two neural networks: the former to distinguish between laughter and non-laughter; the latter to measure laughter intensity on a four-step scale.

Body Laughter Features

Body and its movements are important indicators of laughter which have been widely neglected in the past. Ruch and Ekman [17] observed that laughter is often accompanied by one or more (i.e., occurring at the same time) of the following body behaviors: "rhythmic patterns", "rock violently sideways, or more often back and forth", "nervous tremor ... over the body", "twitch or tremble convulsively". Becker-Asano and colleagues [2] observed that laughing users "moved their heads backward to the left and lifted their arms resembling an open-hand gesture". Markaki and colleagues [11] analyzed laughter in professional (virtual) 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".

Preprocessing

Feature extraction starts from: (i) RGB video captured by a webcam 640x480 @ 30 fps (upper panel of Figure 1); (ii) BW depth map video (each pixel is a 16 bit value indicating the distance from camera) captured by Kinect 640x480 @ 30 fps (middle panel of Figure 1); (iii) two green polystyrene markers on user's shoulders. The data is captured and processed in real-time using EyesWeb XMI [16][10]. Shoulder's markers are automatically extracted

Figure 2: From top to bottom: examples of trunk, head, shoulder features.

by thresholding the RGB video components. Similarly, the user's silhouette is automatically thresholded from the depth map. The green markers' position helps one to separate head from trunk and arms in the user's silhouette. Such video processing is necessary because Kinect SDKs (e.g., OpenNI, Microsoft) fail to detect changes of shoulder's position during shoulder trembling, as the authors tested. The final result of the process, that is, the areas labeled **H**, **T**, **A1**, **A2**, is shown in the lower side of Figure 1. These areas have been considered in previous studies on laughter body movement [9].

Head features
Analysis of head movement starts from the head's silhouette, that is, the region labeled **H**. The Center of Gravity (CoG) of the region is detected and its 2D coordinates are extracted; CoG horizontal and vertical speed are computed. The maximum values of such speeds over a 2 seconds long (with 1 second overlap) time window are body laughter features *F1* and *F2*.

Trunk features
The focus is on: (i) trunk *leaning*, i.e., a slow, wide and repetitive front/back or side-to-side movement of trunk; (ii) trunk *throwing*, i.e., a quick, abrupt and non-repetitive front/back or side-to-side movement of trunk. Analysis of trunk movement starts from a comparison between head's and trunk's silhouette distance from the camera. These distances come from the depth image segmentation. More specifically, the difference D between the averaged distances of areas **H** and **T** is computed. Then, the standard deviation of D over a 2 seconds long (with 1 second overlap) time window is used as a first hint of trunk leaning/throwing. If such a kind of movement is detected (i.e., standard deviation of D is above a threshold) then the following trunk features are extracted:

- *F3* is the periodicity of D in the time window described above; it is high if a prominent frequency is detected in the amplitude vs. frequency spectrum of D, it is low otherwise;
- *F4* is the maximum amplitude of D in the time window;
- *F5* is the impulsiveness of D, that is, the ratio between the prominent peak amplitude of D in the time window and the duration of movement, as described in [5];

Shoulder features
Shoulder *trembling* is a quick and repetitive movement often displayed by laughing people. Three features, based on shoulders' vertical coordinates $y1$ and $y2$, are extracted on a 2 seconds long (with 1 second overlap) time window:

- *F6* is the maximum value of Kinetic Energy, that is, the squared sum of both shoulders' vertical speed (i.e., the time derivative of $y1$ and $y2$)
- *F7* is the shoulders' correlation, that is, correlation between $y1$ and $y2$;
- *F8, F9* is the left/right shoulder's periodicity; it is high when a prominent frequency is detected in the amplitude vs. frequency spectrum of $y1$ and $y2$, it is low otherwise;

The above features are enabled if trunk leaning is not detected, i.e., standard deviation of D is under a prefixed threshold. The threshold is computed in some preliminary sessions in which people perform trunk movements with variable speed. It seems reasonable to neglect time intervals in which shoulder trembling movement is induced by trunk leaning back and forth.

Dataset and Annotation

Five participants were asked to participate in two different tasks: an individual one, that is, watching video clips alone; a social one, that is, playing a game called *yes/no*. The rules of the game are the following: the experimenter can ask the participant any questions and she is obliged to answer them without using any words "yes" and "no". Choice of such tasks was inspired by a recent work [13], showning that they could successfully elicit laughter. Each participant performed each task in an experimental room equipped with a pc having internet connection, LCD screen, a webcam (640x480 @ 30 fps) and Kinect (640x480 @ 30 fps). The participant, sitting alone in front of the pc, wore a head mounted microphone, headphones and two green markers on her shoulders. At the beginning, the participant was invited (i) to play the yes/no game via Skype with one of the experimenters. Then, the participant was asked (ii) to choose and watch from internet a funny clip she liked (e.g., tv shows, clips from movies), lasting about 4-6 minutes, and (iii) to watch a funny clip the experimenters previously selected. Finally, the participant had (iv) to play for a second time the yes/no game.

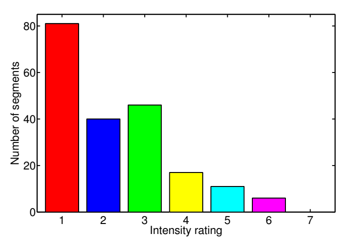

Figure 3: Histogram of the intensity ratings

All material was segmented (by considering multiple modalities such us facial expression and body movement) into 201 laughter and 164 non-laughter segments. Two experts on body movement analysis separately rated the intensity of each laughter segment by using a 7-points Likert scale from 1 to 7. The inter-rater agreement between raters was computed. The resulting weighted Cohen's kappa indicated substantial agreement, k=0.78 [8] in laughter intensity ratings. Due to this substantial agreement, the provided ratings could be used as labels for gold standard in the performance evaluation of the classifier. When the ratings of a laughter segment differed between the two raters, the highest one was chosen as gold standard. Figure 3 depicts the histogram of the intensity ratings. It shows a strong imbalance: only very few segments were rated 5 or 6, none segment was rated a 7. If a classifier is built on such kind of data, the most frequent ratings (e.g., 1 and 2) would tend to prevail in the classification results. Consequently: (i) less frequent ratings were re-grouped following the schema showed in Table 1; (ii) a random sampling method was applied to the most frequent ratings.

Ratings	
Original	Re-mapped
1	1
2	2
3,4	3
5,6,7	4

Table 1: Re-mapping of laughter intensity ratings.

Real-time automated detection

Two Kohonen's self-organising maps (SOMs) were trained with laughter and non-laughter instances extracted from the dataset. The first map is aimed at recognising laughter from non-laughter, the second one at providing a classification of laughter intensity. The training instances (125 for the laughter vs. non-laughter SOM and 425 for the intensity SOM) consist of features F1-F9 computed on randomly picked segments of the dataset. Finally, each instance was standardised to have zero average and unitary standard deviation. Both SOMs consist of 8-by-8 rectangularly oriented units with codebook vectors randomly initialised with numbers in $[0, 1]$. Training sets were presented 900 times to each map. Learning rate and size of the neighborhood exponentially decrease, respectively, from 0.1 and 0.3 to 0 during the training.

Performance evaluation
To evaluate how the SOMs classification matches the gold standard, that is, to provide accuracy of the maps, *Adjusted Rand Index (ARI)* [6] and *Adjusted Mutual Information (AMI)* [18] were computed on 100 bootstrapped instance sets. ARI and AMI are two commonly used measures in the literature for this purpose. These measures, already adjusted for chance, range from 0 to 1, where 0 means chance class label assignment and 1 means that all the predicted class labels agree with the gold standard. Table 2 shows the averaged value of ARI and AMI and their standard deviation for each of the two SOMs, the one for laughter vs. non-laughter classification and the one for laughter intensity classification, respectively. The overall results are promising as both ARI and AMI are strongly above chance level with a small standard deviation.

	ARI		AMI	
SOM	avg	std	avg	std
laughter/non-laughter	0.52	0.10	0.43	0.10
laughter intensity	0.44	0.05	0.49	0.04

Table 2: SOM Performance evaluation results

Conclusion and Future Works

A technique for automated laughter intensity detection from body movements is discussed. Taking into account the preliminary nature of this work, results are promising: objective performance evaluation showed the validity of the proposed approach. Use of dynamic analysis allowing to consider the temporal evolution of laughter and its intensity over time is planned, e.g., by using HMMs. Nevertheless, results should be confirmed on a larger data-set containing, for example, recordings of people acting in different contexts when they laugh. Further, a limited number of laughter body movement features is here considered. Person-specific laughter styles and different laughter types have to be explored in collaboration with psychologists. In conclusion, it is worth noticing that the ability of detecting laughter and its characteristics would have wider applications. For instance, several research shows that people, when interacting with other humans [1], but also with virtual agents [4], adapt their behavior with the interlocutor by copying movements and/or movement features (e.g., speed/amplitude of gestures).

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 270780.

References

[1] Bailenson, J. N., and Yee, N. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science 16*, 10 (2005), 814–819.
[2] Becker-Asano, C., Kanda, T., Ishi, C., and Ishiguro, H. How about laughter? perceived naturalness of two laughing humanoid robots. In *3rd International*

Conference on Affective Computing and Intelligent Interaction, IEEE (2009), 1–6.

[3] Bennett, M. P., Zeller, J. M., Rosenberg, L., and McCann, J. The effect of mirthful laughter on stress and natural killer cell activity. *Nursing Faculty Publications* (2003).

[4] Bousmalis, K., Mehu, M., and Pantic, M. Spotting agreement and disagreement: A survey of nonverbal audiovisual cues and tools. In *3rd International Conference on Affective Computing and Intelligent Interaction*, vol. 2 (Los Alamitos, September 2009).

[5] Castellano, G., Camurri, A., Mazzarino, B., and Volpe, G. A mathematical model to analyse the dynamics of gesture expressivity. In *Proceedings of AISB* (2007).

[6] Hubert, L., and Arabie, P. Comparing partitions. *Journal of Classifications 2*, 1 (1985), 193–218.

[7] Ito, A., Wang, X., Suzuki, M., and Makino, S. Smile and laughter recognition using speech processing and face recognition from conversation video. In *Cyberworlds, 2005. International Conference on*, IEEE (2005).

[8] Landis, J., and Koch, G. The measurement of observer agreement for categorical data. *Biometrics 33*, 1 (1977), 159–174.

[9] Mancini, M., Hofmann, J., Platt, T., Volpe, G., Varni, G., Glowinski, D., Ruch, W., and Camurri, A. Towards automated full body detection of laughter driven by human expert annotation. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on* (Sept 2013), 757–762.

[10] Mancini, M., Varni, G., Glowinski, D., and Volpe, G. Computing and evaluating the body laughter index. In *Human Behavior Understanding*. Springer Berlin Heidelberg, 2012, 90–98.

[11] Markaki, V., Merlino, S., Mondada, L., and Oloff, F. Laughter in professional meetings: the organization of an emergent ethnic joke. *Journal of Pragmatics 42*, 6 (2010), 1526–1542.

[12] Mora-Ripoll, R. The therapeutic value of laughter in medicine. *Alternative therapies in health and medicine 16*, 6 (2010).

[13] Niewiadomski, R., Mancini, M., Baur, T., Varni, G., Griffin, H., and Aung, M. S. Mmli: Multimodal multiperson corpus of laughter in interaction. In *Human Behavior Understanding*. Springer International Publishing, 2013, 184–195.

[14] Owren, M. J., and Bachorowski, J.-A. Reconsidering the evolution of nonlinguistic communication: The case of laughter. *Journal of Nonverbal Behavior 27* (2003), 183–200.

[15] Petridis, S., and Pantic, M. Audiovisual discrimination between laughter and speech. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, IEEE (2008), 5117–5120.

[16] Piana, S., Mancini, M., Camurri, A., Varni, G., and Volpe, G. Automated analysis of non-verbal expressive gesture. In *Human Aspects in Ambient Intelligence*. Atlantis Press, 2013, 41–54.

[17] Ruch, W., and Ekman, P. The expressive pattern of laughter. In *Emotion, qualia and consciousness*, A. Kaszniak, Ed. World Scientific Publishers, Tokyo, 2001, 426–443.

[18] Vinh, N., Epps, J., and Bailey, J. Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *The Journal of Machine Learning Research 11* (2010), 2837–54.