This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Automated Laughter Detection From Full-Body Movements

Radoslaw Niewiadomski, Maurizio Mancini, Giovanna Varni, Gualtiero Volpe, and Antonio Camurri

***Abstract*—In this paper, we investigate the detection of laughter from the user's nonverbal full-body movement in social and ecological contexts. Eight hundred and one laughter and nonlaughter segments of full-body movement were examined from a corpus of motion capture data of subjects participating in social activities that stimulated laughter. A set of 13 full-body movement features was identified, and corresponding automated extraction algorithms were developed. These features were extracted from the laughter and nonlaughter segments, and the resulting dataset was provided as input to supervised machine learning techniques. Both discriminative (radial basis function-support vector machines, k-nearest neighbor, and random forest) and probabilistic (naive Bayes and logistic regression) classifiers were trained and evaluated. A comparison of automated classification with the ratings of human observers for the same laughter and nonlaughter segments showed that the performance of our approach for automated laughter detection is comparable with that of humans. The highest F-score (0.74) was obtained by the random forest classifier, whereas the F-score obtained by human observers was 0.70. Based on the analysis techniques introduced in the paper, a vision-based system prototype for automated laughter detection was designed and evaluated. Support vector machines (SVMs) and Kohonen's self-organizing maps were used for training, and the highest F-score was obtained with SVM (0.73).**

***Index Terms*—Automated analysis of full-body movement, body expressivity, detection, laughter, motion capture, multimodal interaction.**

I. Introduction

Laughter is a powerful signal capable of triggering and facilitating social interaction. Grammer [1] suggests that it may convey social interest and reduce the sense of threat in a group [2]. Further, laughter seems to improve learning of new activities from other people [3], creativity [4], and it facilitates sociability and cooperation [5]. Healthy positive effects of laughter have been observed with people living with stress or depression [6]. The EU-ICT FET Project ILHAIRE[1] aims to study how machines could interact with users through laughter: for example, to know when the user is laughing [7], [8], to measure intensity of laughter [9], and to distinguish between different types of laughter [10] by means of laughter enabled virtual agents [11], [12].

In this paper, we propose models and techniques for the automated detection of laughter from the user's full-body movement in social and ecological contexts. Whereas research has focused on speech and facial expression as major channels for detecting laughter (e.g., [13], [14]), capturing them reliably in a social and ecological context is a complex task. Consider an example involving a small group of friends standing and conversing where robust capture of facial expressions is challenging and/or costly. This situation requires multiple cameras capturing the face of each user with enough detail to perform analysis. Due to the user's movement, the cameras also need to either track and follow the movements or continuously zoom into the location containing the user's face. In relation to speech, the well-known *cocktail party effect* [15] describes how people are capable of focusing attention on a single conversation by filtering out other conversations and noise. Audio source separation techniques are still an open research area, and their output is unlikely to be reliable enough for laughter analysis. In contrast, low-cost motion tracking and analysis systems can track and analyze the full-body movement of each user. For example, by analyzing depth images, Microsoft Kinect can reliably retrieve the silhouette of each user and her body skeleton, including the 3-D displacement of each body joint, at a frame rate of 30 frames/s.

In this study, we analyze laughter by focusing on full-body expressive movement captured with a motion capture system. We do not distinguish among different laughter types nor determine laughter intensity. Our study demonstrates that, when data from other modalities are not available or are noisy, the body is a robust cue for automated laughter detection. We also present and evaluate a practical application of the results of our study that uses a real-time system prototype based on low-cost consumer hardware devices. The prototype is developed with the freely available EyesWeb XMI[2] research platform [16], [17] and applies real-time algorithms for automated laughter detection starting from data captured by RGB-D sensors (Kinect and Kinect2).

In Section II, we describe the state of the art of laughter analysis. Our study on laughter detection from motion capture data is described in Section III. Section IV presents a real-time system prototype for automated laughter detection. We conclude the paper in Section V.

Manuscript received November 25, 2014; revised August 10, 2015; accepted September 9, 2015. This work was supported by the EU 7th Framework Program under Grant 270780 ILHAIRE. This paper was recommended by Associate Editor Z. Yu.

R. Niewiadomski, M. Mancini, G. Volpe, and A. Camurri are with the Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi, Università degli Studi di Genova, 16146 Genova, Italy (e-mail: radoslaw.niewiadomski@dibris.unige.it; maurizio.macini@unige.it; gualtiero.volpe@unige.it; antonio.camurri@unige.it).

G. Varni is with the Institute for Intelligent Systems and Robotics, University Pierre and Marie Curie, 75005 Paris, France (e-mail: varni@isir.upmc.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2015.2480843

[1]http://www.ilhaire.eu

[2]http://eyesweb.infomus.org

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

II. State of the Art

Laughter can be expressed with acoustic, facial, and full-body cues. Most research on laughter expressive patterns focuses on audio and facial expressions. Nevertheless, results of our preliminary experiment [18] show that people are able to recognize laughter from body movements only. In the experiment, ten animations displaying full-body motion capture data corresponding to laughter and nonlaughter episodes were shown to participants. The results showed a recognition rate over 79%. McKeown *et al.* [10] investigated the human capability to distinguish between four different laughter types in a perceptive evaluation. People were able to correctly classify four laughter types with an accuracy rate of 28% (chance level was 25%). In order to check whether it is possible to distinguish between different laughter types only from body movements, Griffin *et al.* [8] conducted a perceptual study with the use of avatars animated with MoCap data. Thirty-two participants categorized 126 stimuli using five labels: hilarious, social, awkward, fake, or nonlaughter. The agreement rates between the participants varied from 32% (fake laughter) to 58% (hilarious laughter).

Body movements of laughter were described by Ruch and Ekman [19]. According to them, most of the body movements in laughter are related to respiration activity. These may include "the backward tilt of the head," "raise and straighten of the trunk," and "shaking of the shoulders and vibrations of the trunk" [19]. Other body movements (OBMs) are also observed in laughter, which are not related to respiration such as "rocking violently sideways" or "hands throwing" [19]. A more formal description of body movements in laughter was proposed by Ruch *et al.* [20]. They developed an annotation scheme that specifies, for each part of the body (head, trunk, arms, legs), the shape of movement as well as its dynamic and expressive qualities. For example, descriptors such as "shaking," "throwing," or "rocking" characterize velocity of movement or its tendency to be repetitive. Among the movements observed in laughter are: head nodding up and down, or shaking back and forth; shoulders contracting forward or trembling; trunk rocking, throwing backward and forward or straightening backward; arm throwing; and knees bending.

Existing laughter detection algorithms mainly focus on audio (e.g., [21], [22]), physiological (e.g., [23]), or combined facial and audio laughter detection (e.g., [13], [14]). Such work supports classifying laughter segments offline, but also provides automatic online segmentation and detection. Importantly, most do not include body movements data.

Aiming at detecting laughter from audio, Truong and Leeuwen [21] compared the performance of different acoustic features (i.e., perceptual linear prediction features, pitch and energy, pitch and voicing, and modulation spectrum features) and different classifiers. Gaussian mixture models trained with perceptual linear prediction features performed the best with equal error rate (EER) ranging from 7.1% to 20.0%. Knox and Mirghafori [24] applied neural networks to automatically segment and detect acoustic laughter from conversation. They used mel frequency cepstral coefficients (MFCC) and the fundamental frequency as features and the obtained EER was 7.9%. Salamin *et al.* [22] proposed an automatic detection of acoustic laughter in spoken conversations captured with mobile phones. They segmented audio recordings into four classes: laughter, filler, speech, and silence. Hidden Markov models (HMMs) combined with statistical language models were used, and reported F-scores for laughter varied between 49% and 64%.

With respect to multimodal detection and fusion, Escalera *et al.* [14] applied stacked sequential learning for audio-visual laughter detection. Audio features were extracted from the spectrum and complemented with accumulated power, spectral entropy, and fundamental frequency. Facial cues included the amount of mouth movement (between consecutive frames) and the laughter detection obtained from a classifier trained on principal components extracted from a labeled dataset of mouth images. Results showed an accuracy between 77% and 81%, depending on the type of data (multimodal or audio only). Petridis *et al.* [13] proposed an algorithm based on the fusion of audio and facial modalities. Using 20 points (facial features), six MFCCs, and zero crossing rate (audio features), they trained a neural network for a two-class (laughter versus speech) discrimination problem, and they showed the advantage of a multimodal approach over video-only detection (with accuracy of 83.3% for video-only and of 90.1% for multimodal analysis). Scherer *et al.* [25] compared the efficacy of various classifiers in audio–visual offline and online laughter detection in natural multiparty conversations. SVM was the most efficient in the offline classification task, while HMM received the highest F-scores in online detection (72%). Tatsumi *et al.* [23] argued that people may hide their amusement (and laughter), and that physiological cues may be indicators of such inhibited laughter. They detected inhibited laughter using facial electromyogram, skin conductance, and electrocardiogram data. Cosentino *et al.* [26] detected laughter expressions using the data from inertial measurement units and EMG sensors placed directly on participant torso.

Body movements of laughter were rarely considered in laughter detection algorithms. Mancini *et al.* [7] proposed the body laughter index (BLI). Their algorithm, based on a small number of nonverbal expressive body features extracted with computer vision methods, tracks the position of the shoulders in real time and computes an index, which tends to 1 when laughter is more likely to occur. The BLI is a linear combination of the kinetic energy of shoulders, of the Pearson's correlation between the vertical positions of the shoulders, and of the periodicity of movement. Griffin *et al.* [8] proposed to detect different types of laughter from motion capture data of body movements. They used 126 segments from the UCL body laughter dataset of natural and posed laughter in both standing and sitting postures. The segments were divided into five classes according to a perceptual study with stick-figure animations of motion capture data. They extracted 50 features: 1) low-level features corresponding to distances and angles between the joints, 2) high-level features, e.g., kinetic energy of certain joints, spectral power of shoulder movements, or smoothness of shoulders trajectory. Features took into consideration both upper and lower body parts. In the last step, they applied a variety of classifiers. Results show efficacy in laughter detection above chance level for three classes: hilarious laughter (F-score: 60%), social laughter

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

(F-score: 58%), and nonlaughter (F-score: 76%), using random forest (RF).

The described laughter detection algorithms mainly focus on acoustic and facial cues. In ecological multiparty interaction, the audio extraction of a single person's laughter and noninvasive face tracking is still challenging. It is easier to track the users' body movements during the interaction. Further, de Gelder *et al.* [27] suggest that bodily cues are particularly suitable for communication over larger distances, whereas facial expressions are more suitable for a fine-grained analysis of affective expressions. This suggests that full-body movement can play an important role in social communication.

Our study on automated full-body laughter detection aims:

1) to detect laughter from full-body movement only;
2) to detect laughter occurring in natural spontaneous contexts;
3) to distinguish laughter from other bodily expressive activities that may occur in the same contexts.

Similar work was carried out by Griffin and colleagues [8]. Both our and their work focus on laughter full-body movements in natural and spontaneous contexts. While the main focus of our study is on discriminating laughter from nonlaughter expressions, Griffin *et al.* [8] propose an automatic recognition system for discriminating between the body expressions that are perceived as different laughter types and the ones that are perceived as nonlaughter. Second, we use a top-down approach for feature selection. Our set of high-level features is based on the body annotation schema of laughter presented in [20]. Our movement features capture the dynamics of movement, e.g., its periodicity or suddenness. Such features are representative of biological motion and, consequently, have a meaningful interpretation. To define ground truth, segments labeling was performed by taking into account the available synchronized data, i.e., motion capture, video, and audio. Next, we compare the results of automated classification and humans' classification of laughter stick-figure animations against the ground truth. A larger set of laughter segments was used in our study (801), compared with Griffin *et al.* (i.e., 126). Additionally, we present a real-time system prototype based on the results of our study.

III. Automated Laughter Detection: Data Set and Experiments

This section describes a study in which we recorded people while performing activities involving laughter and nonlaughter movements (see Section III-A). Then, we segmented data corresponding to such movements to generate a set of laughter (see Section III-B1) and nonlaughter (see Section III-B2) segments. The data of these two sets were used to define feature vectors (see Section III-C) which were, next, provided as input to supervised machine learning algorithms (see Section III-D). We compared machine with human laughter classification ability on the same dataset (see Section III-E).

A. Multimodal and Multiperson Corpus of Laughter

We used the multimodal and multiperson corpus of laughter in interaction (MMLI) corpus, recorded in collaboration with ILHAIRE partners from Telecom ParisTech, University of Augsburg, and University College of London [28]. This corpus consists of full-body data collected with high precision motion capture technology. The corpus is also characterized by high variability of laughter expressions (variability of contexts, many participants). It contains natural behaviors in multiparty interactions, mostly spontaneous laughter displays. The creation of the experimental protocol was inspired by the previous works carried out within the ILHAIRE Project. McKeown *et al.* [29] proposed guidelines for laughter induction and recording. They stressed the importance of creating a social setting that is conducive to laughter generation by avoiding the formality of the laboratory environment, recruiting participants having strong affiliation, or using social games as the laughter elicitation instrument.

To capture laughter in different contexts, we invited groups of friends to perform six enjoyable tasks (T1–T6). In addition to classical laughter inducing tasks, such as watching comedies, participants were asked to play social games, i.e., games regulated by one simple general rule in which players are left free to improvise. According to [29], a lack of detailed rules could encourage easy-going spontaneous behavior.

1) Tasks: Participants were asked to perform the following tasks: T1) watching comedies together, T2) watching comedies separately, T3) "Yes/no" game, T4) "Barbichette" game, T5) "Pictionary" game, and T6) tongue twisters.

T1 and T2 are classic laughter-inducing tasks, i.e., watching comedies selected by experimenters and participants. Compared with other laughter corpora (e.g., [30]), participants were not alone; they could talk freely (e.g., comment videos) and hear each other. In T2, a curtain impeded one participant to see the other ones during task execution, still allowing her to hear them. Tasks T3 and T4 consisted of two social games that were carried out in turns with participants switching between different roles and competing against each other. In T3, one of the participants had to quickly respond to questions from the other participants without saying sentences containing either "yes" or "no." The role of the other two participants was to ask questions and distract her, in an attempt to provoke the use of the "forbidden" words. T4 is a French game for children whose aim is to avoid laughing. Two participants faced each other, made eye contact, and held the other person's chin. Participants were allowed to talk, move, and perform facial expressions, always maintaining physical and eye contact. The person who laughed first lost the game. In T5, one participant drew words printed on a piece of paper extracted from an envelope. Her task was to convey the word to the other participant by drawing on a large board. T6 consisted of participants pronouncing tongue twisters in different languages.

2) Technical Setup: During corpus collection, we captured full-body movement of up to three human participants at the same time. For this purpose, we recorded:

a) the motion data of two participants using the Xsens MVN Biomech system 4. The system consists of 17 inertial sensors placed on velcro straps. Data were recorded at 120 frames/s; each frame consisting of the 3-D position of 22 body joints;
b) audio samples captured with wearable microphones (Mono, 16 kHz) placed close to the participants' mouth;

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

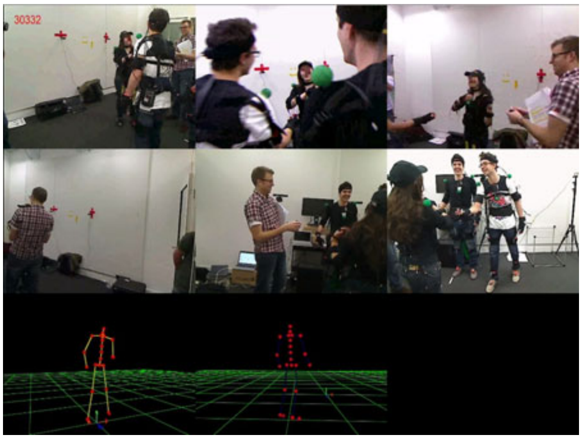


Fig. 1. Synchronized data view.

c) four video streams captured with Logitech Webcam Pro 9000 (640×480, 30 frames/s) recorded the room from different viewpoints in order to get the frontal view of the participants;

d) two high-frame rate video streams captured with Philips PC Webcam SPZ5000 (640×480, 60 frames/s) placed over tripods recorded close-ups of the participants' face.

3) Protocol: We recruited groups of friends. Participants were selected from university (Master and Ph.D.) students. Data collection consisted of recording all interactions. We also recorded participants during pauses between tasks. The whole corpus consists of six sessions with 16 participants: four triads and two dyads, age 20–35; three females; eight French, two Polish, two Vietnamese, one German, one Austrian, one Chinese, and one Tunisian. Participants were allowed to speak the language they used to communicate with each other most of the time.

B. Segmentation

We analyzed and segmented data from ten participants (eight men, two women) involved in four tasks (T1, T3, T4, and T5). We skipped the data recorded during two tasks: T2 (watching comedies separately), because some groups did not perform it, and T6, because during tongue twisters people laughed while speaking; therefore, it was particularly difficult to precisely segment and annotate this task.

For each participant and each task, the synchronized streams of motion capture data (visualized through a graphical representation of a skeleton), six RGB videos, and the corresponding audio recordings were used for performing segmentation (see Fig. 1). We implemented software tools for streams synchronization and segmentation by developing modules for EyesWeb XMI. These tools are available for research purposes on the EyesWeb XMI forum.[3] Segments were annotated depending on whether they contained laughter body movements (LBMs) or other kinds of body movements occurring during spontaneous interaction.

[3]http://forum.eyesweb.infomus.org

1) Laughter Body Movements: This set consists of 316 segments in which participants perform full-body movements during laughter. Observers watched and segmented the data corpus, performing a two-phases process.

a) Laughter segmentation: An observer watched and listened to all recorded and synchronized data from the MMLI corpus, isolating laughter segments, where laughter could be observed or heard from at least one modality (i.e., face, audio, or full-body movement). Isolating laughter segments by taking into account the synchronized modalities was indispensable to establish ground truth. The result of the process was a set of 404 laughter segments that could contain full-body-only or audio-only laughter cues.

b) Laughter annotation: Two raters watched the 404 laughter segments resulting from the segmentation. They observed a graphical interface showing the output of six cameras as well as the graphical representation of a skeleton; see Fig. 1. The two raters did not hear any audio. They focused on the body movement cues of laughter [19], [20] (see also Section II):

i) *F1—head side movement*: head movements on the frontal plane, the plane dividing the body into front and back halves;

ii) *F2—head front/back movement*: head movements on the sagittal plane, the plane dividing the body into left and right halves;

iii) *F3—weight shift*: a change in body posture during which the user switches the leg on which body weight is mainly applied;

iv) *F4—knee bending*: leg movement during which one or two legs are bent at the knee;

v) *F5, F10, and F13—abdomen, arm, and shoulder shaking*: according to [12] and [19], a laughter episode can exhibit several repetitive body pulses that are caused by forced exhalation; these pulses can induce a repetitive fast contraction/vibration of user's abdomen, arm, and/or shoulder; we define such a movement type as a *shaking* movement;

vi) *F6 and F12—trunk and arm straightening*: trunk/arm is extended, that is, a rotation is performed at the pelvis/elbow level, increasing the angle between respectively, trunk/upper arm and legs/lower arm;

vii) *F7 and F11—trunk and arm rocking*: according to [12] and [19], during laughter, contraction/vibration induced by forced exhalation can be accompanied by OBMs such as sideways trunk and arm rocking, which are, however, slow and repetitive; we define such a movement type as a *rocking* movement;

viii) *F8 and F9—trunk and arm throwing*: quick movement of trunk/arm, in any direction, that is, a quick modification of head/hand position in space.

The result of the annotation process is the LBM set, consisting of 316 laughter segments exhibiting visible full-body movements (movements in which one or more of the cues F1–F13 were observed). In the excluded 88 segments, none of the cues F1–F13 was observed by any rater. The interrater agreement between the two raters, measured with Cohen κ, was 0.633, which is considered a "good" result [31]. In case of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

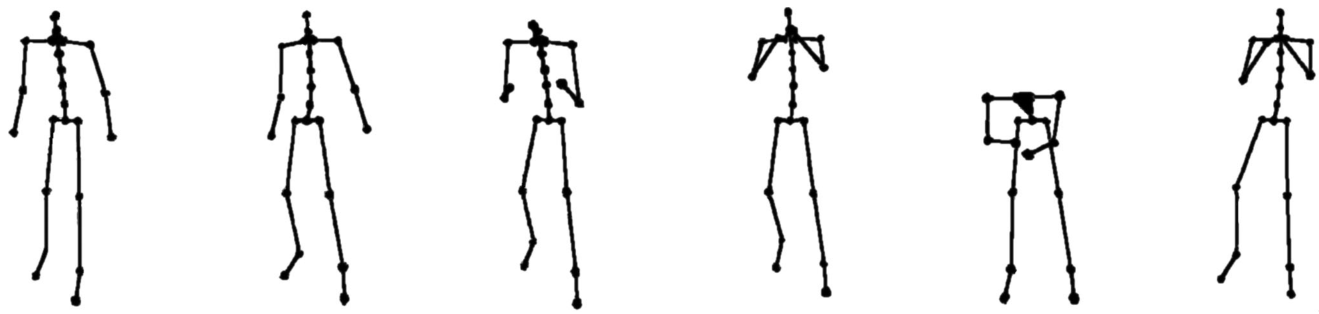


Fig. 2. Some frames of a laughter episode. Trunk throwing (F8) and knee bending (F4) can be observed.

TABLE I
DESCRIPTIVE STATISTICS

Type	No episodes/no participants	Total duration (s)	Min duration (s)	Max duration (s)	Avg (s)	Std (s)
Laughter Body Movements	316/10	27 min 3 s	1.4 s	46.4 s	5.13 s	4.28 s
Other Body Movements	485/10	46 min 18 s	1.4 s	23.3 s	5.72 s	2.59 s

disagreement between raters (e.g., only one rater observed LBMs) such a segment was also included into the LBM set. In total, 254 segments were evaluated by both raters as displaying full-body movement cues of laughter; 62 segments on which the two raters did not agree were also added to the set. Statistical information on the LBM set is presented in Table I .

2) Other Body Movements: The same observer performed another segmentation by isolating segments exhibiting full-body movements that did not occur during laughter such as folding/unfolding arm gestures, walking, or face rubbing. All available modalities (audio, video, MoCap) were observed and listened to during the segmentation process. The result of the segmentation process is the OBM set, consisting of 485 segments of full-body movements occurring *without* laughter. The statistical information on the OBM set is presented in Table I. All 801 segments containing MoCap data can be downloaded from http://www.infomus.org/ILHAIRE/mmli.

C. Feature Vector

Starting from the LBM and OBM sets, we built a feature vector to be provided as input to classification models described in Section III-E. The feature vector contains the 13 full-body movement features presented in Section III-B1. The algorithms for extracting these features are based on a common set of *primitive functions*. We first provide a description of such primitives; then, each feature is computed as a combination of primitives. Algorithms are implemented in MATLAB. Each feature is extracted on the entire length of each LBM or OBM segment: For each of the 801 segments, we obtained a 13-value feature vector.

1) Primitive Functions

a) *Distance*

$$D = \text{Distance}(J1, J2). \tag{1}$$

Given two body joint labeled as $J1$ and $J2$, it returns a 1-D vector D in which the ith value is the distance between the two joints at frame i.

b) *Speed*

$$S = \text{Speed}(J1). \tag{2}$$

Given one body joint labeled as $J1$, it returns a 1-D vector S in which the ith value is the joint's speed at frame i. Speed is computed with the MATLAB `diff` function, and then, it is filtered to remove spikes and noise. We apply a low-pass Savitzky–Golay filter [32] to the speed of the participant's joints. We do not apply any filter to positional data. In particular, we run the following MATLAB function on the participant's joints speed data: `sgolayfilt(speed_data,3,41)`. The parameters N=`3`,M=`41` define a filter with a cutoff frequency of about 1 Hz.

c) *Normalize*

$$V_N = \text{Normalize}(V). \tag{3}$$

The provided 1-D vector V is normalized in $[0, 1]$ by 1) subtracting the minimum element from all elements contained in the vector, and 2) dividing all elements of the vector by the maximum element of the vector:

d) *Threshold_Check*

$$A = \text{Threshold\_Check}(v, t, f). \tag{4}$$

The value of v is compared with the threshold t. However, a tolerance factor $f \in [0, 1]$ is taken into account. If v is lower than t, then $A = 0$; if v is higher than t but lower than $t + (t * f)$, then $A = (v - t)/(t * f)$; $A = 1$ otherwise.

e) *Range_Check*

$$A = \text{Range\_Check}(v, r_1, r_2, f). \tag{5}$$

This function compares the input value v with the range $[r_1, r_2]$, taking into account a tolerance factor

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

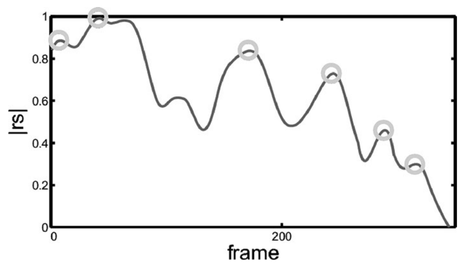


Fig. 3. Detected peaks of the participant's right shoulder position $|\mathbf{rs}|$ are the local maxima exhibiting slope values higher than a given threshold. In the graph, peaks are highlighted by a gray circle, that is, the algorithm does not detect all local maxima as peaks. The approximate peaks frequency is then computed as the ratio between the segment length and the number of peaks. In the example, the segment length is approximately 320 frames, that is, 2.6 s at 120 frames/s. The approximate peaks frequency is $6.0/2.6 = 2.30$ Hz.

$f \in [0, 0.5]$. If v is lower than r_1 or higher than r_2, then $A = 0$; if v is higher than r_1 but lower than $r_1 + (r_2 - r_1) * f$, then $A = (v - r1)/((r_2 - r_1) * f)$; if v is lower than r_2 but higher than $r_2 - (r_2 - r_1) * f$, then $A = -(r2 - v)/((r_2 - r_1) * f)$; $A = 1$ otherwise.

f) *Frequency_Range_Check*

$$C = \text{Frequency\_Check}(V, f_1, f_2). \tag{6}$$

The goal of this function is to compare the frequency of variation of the 1-D vector V provided as input with the range of frequencies $[f_1, f_2]$. Estimation of the frequency of variation of the input 1-D vector is performed as follows.

We apply to the input vector V a function to find peaks.[4] We apply a least squares curve fitting to find all local maxima in the input data in which the fitted curve exhibits a slope higher than a given threshold. We fixed this threshold to find peaks corresponding to frequencies higher than f_L.

If 0 or 1 peaks are found, then we set $C = 0$ and the algorithm terminates. If two or more peaks are detected, then we compute their approximate frequency F (in Hz) of repetition as the ratio between the number of peaks and the length of the segment. For example, if three peaks are detected in a segment lasting 4 s, we estimate a peaks frequency of $F = 3/4 = 0.75$ Hz.

We finally compare the computed frequency F with $[f_1, f_2]$ by applying the Range_Check primitive. If F is outside the range, then we set $C = 0$. Otherwise, the value of C will tend to reach the value of 1 as long as the value of F tends to reach the center of the interval $[f_1, f_2]$.

Fig. 3 illustrates the computation of participant's right shoulder frequency of movement.

2) Movement Features Extraction: The MATLAB implementation of the 13 full-body movement features is illustrated in Fig. 4. On the left side of the figure, skeleton joints labels are reported, except for joint $(0, 0, 0)$, which refers to the world's center. All features algorithms are based on the primitive functions, which are reported in the middle of the figure. The functions var and $cumsum$ correspond to, respectively, the MATLAB variance and integral (cumulative sum) functions. On the right, the computed movement features names are reported.

[4]For further details, see http://terpconnect.umd.edu/∼ toh/spectrum/PeakFindingandMeasurement.htm

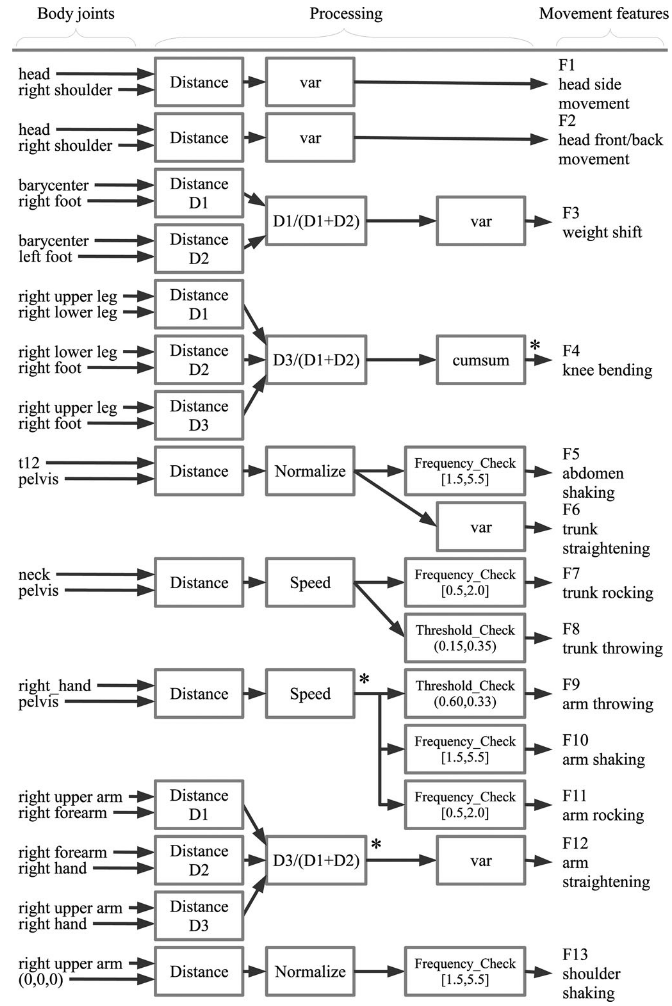


Fig. 4. Movement features extraction algorithms: On the left, body joints are selected; then, their positional data are provided as input to the algorithms in processing portion; the computed features names are reported on the right.

In Fig. 4:

a) algorithms marked with an * are computed two times, both on the joints reported on the left and on the same joints belonging to the opposite side of body; then, the computed quantities are summed before continuing with the algorithm. For example, for feature *F4*, the cumulative sum (the block marked with an *) is computed two times, on joints *right upper leg*, *right lower leg*, and *right foot*, and on joints *left upper leg*, *left lower leg*, and *left foot*. Then, the resulting cumulative sums are summed to compute *knee bending*.

b) *Threshold_Check* is performed two times: on the neck-pelvis distance speed to compute *trunk throwing* and on

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

the sum between right hand-pelvis distance speed and left hand-pelvis distance speed to compute *arm throwing*. The two thresholds, 0.15 and 0.60, respectively, were determined empirically by measuring the two speed values on movements that, according to annotation, exhibited the *trunk throwing* and *arm throwing* movement features.

c) *Frequency_Range_Check* is performed four times to check whether some distances vary with a frequency in a given range. In particular, we focused on two ranges: $[0.5, 2.0]$ Hz and $[1.5, 5.5]$ Hz. The first one corresponds to frequencies typical of *rocking* movements, and the second one corresponds to *shaking* movements. According to [12] and [19], the frequency of trunk and limbs rocking during laughter varies in the first range, while the frequency of abdomen and shoulders shaking varies in the second one.

We checked the pairwise correlations between features F1–F13 on the whole dataset. The mean absolute correlation is 0.075, and the standard deviation is 0.14 (only 27 out of 78 pairs had significant correlations). The highest correlations were observed for the pairs: F10, F11 ($r = -0.731, p < 0.001$), F1, F2 ($r = 0.6024, p < 0.001$) and F5, F13 ($r = 0.4257, p < 0.001$). The pair F10, F11 corresponds to arm shaking and rocking. The high negative correlation is not surprising, as these two features are measuring two different types of repetitive movements (slow and quick) defined with two different ranges of frequencies. The pairs F1, F2 corresponds to the head movements on the different axes. As the head movements cannot be performed exclusively on one plane in real-life settings, a higher correlation can be expected also in this case. Finally, in the case of the pair F5, F13, both features measure repetitive movements having the same ranges of frequencies. The higher correlation could be also expected in this case, as these movements are related to the respiration pattern [19].

D. Automated Classification

The performances of five supervised machine learning algorithms were tested to classify LBM versus OBM segments: radial basis function-support vector machine (rbf-SVM), k-nearest neighbor (k-NN), random forest (RF), naive Bayes (NB), and logistic regression (LR). We chose these algorithms in order to evaluate how both discriminative (SVM, k-NN, RF) and probabilistic algorithms (NB, LR) work on our dataset. The averaged performance of each classifier was assessed via a multiple-run *k*-fold (nested) stratified cross validation. In our study, we adopted five run and ten folds. The inner loop of the cross validation aimed at performing model selection. The parameters of rbf-SVM and k-NN were estimated via a grid search approach with a fivefold stratified cross validation. A fivefold cross validation was used to tune the number of trees composing the RF, while the number of attributes for each tree in the forest was chosen equal to the square root of the number of features. For the NB classifier, the likelihood of the features is assumed to be Gaussian.

Table II reports average confusion matrices for k-NN and SVM algorithms. Table III shows the performance of each classifier in terms of Precision, Recall and F-score. Tables IV and V show the same metrics for LBM and OBM classes, respectively. All the classifiers were able to discriminate LBM from OBM well above chance level (50%). To determine whether one of the learning algorithms outperforms the other ones on our dataset, we carried out a five-run tenfold cross validation in the *use all data* version as described in [33]. The *use all data* approach with calibrated degrees of freedom is a successful method to compensate for the difference between the desired Type I error and the true Type I error. It was chosen because it is the conceptually simplest test for comparing supervised classification algorithms. Further, it outperforms on power and replicability other common tests such as, for example, $5X2$ cross-validation, resampling, and *k*-fold cross validation [33].

TABLE II
AVERAGE VALUES OF CONFUSION MATRICES FOR k-NN AND SVM ALGORITHMS

	k-NN		SVM	
	Laughter	Nonlaughter	Laughter	Nonlaughter
Laughter	189.6	126.4	213.2	102.8
Nonlaughter	72.6	412.4	126.2	358.8

TABLE III
WEIGHTED AVERAGE PRECISION, RECALL, AND F-SCORE FOR ALL SEGMENTS (CLASSES LBM+OBM)

	Avg. Precision	Recall	F-score
SVM	0.72	0.71	0.71
k-NN	0.75	0.75	0.74
RF	0.73	0.73	0.72
LR	0.72	0.71	0.71
NB	0.69	0.65	0.59

TABLE IV
PRECISION, RECALL, F-SCORE FOR LAUGHTER BODY MOVEMENT SEGMENTS (CLASS LBM ONLY)

	Precision	Recall	F-score
SVM	0.62	0.67	0.64
k-NN	0.72	0.60	0.65
RF	0.66	0.67	0.66
LR	0.64	0.64	0.64
NB	0.72	0.21	0.32

TABLE V
PRECISION, RECALL, AND F-SCORE FOR OTHER BODY MOVEMENT SEGMENTS (CLASS OBM ONLY)

	Precision	Recall	F-score
SVM	0.78	0.73	0.75
k-NN	0.76	0.85	0.80
RF	0.78	0.76	0.77
LR	0.77	0.76	0.76
NB	0.65	0.94	0.77

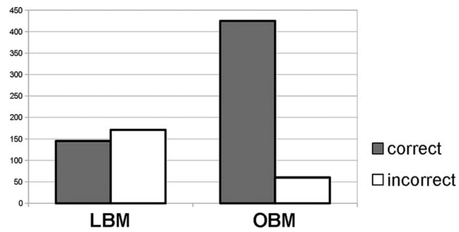


Fig. 5. Results of human classification of 801 segments.

F-score values were computed for each algorithm, and the differences among these values were then computed for each pair of algorithms. Such resulting differences are used as independent samples for Z-tests. Bonferroni adjustment of α was used where necessary to compensate for multiple comparisons when Z statistics are calculated. We chose to compare between the algorithms belonging to the same class, that is discriminative or probabilistic and, then, in case of significant differences, to compare the two winning algorithms.

The Z-tests indicated no difference among all the discriminative nor among all the probabilistic classifiers. The Z-test between one discriminative (SVM) and one probabilistic (NB) classifier showed a significant difference ($Z = 5.514$, $p < 0.00001$). We conclude that the discriminative classifier outperforms the probabilistic classifier on our dataset.

E. Machine Versus Human Classification

To evaluate our approach, we measured the human ability to recognize laughter from body movements. We asked to label the segments of our dataset in an evaluation study that was carried out through an online questionnaire consisting of videos and questions. Participants had to watch stick-figure animations of a skeleton (i.e., with *no* audio and *no* facial expressions; see Fig. 2) and answer to the question: "Do you think that the person represented in the video is laughing?"

A web page displayed one full-body skeleton animation of motion capture data corresponding to one segment among the segments of both LBM and OBM (i.e., the whole machine learning dataset). Participants could watch each animation as many times as they wanted and they had to decide whether the displayed skeleton was laughing or not. Each participant could evaluate any number of animations. Evaluation was performed by keeping the participants unaware of the cause, of the mechanisms, and of the context of laughter [34]. Animations were displayed in a random order: Each new animation was chosen among the animations that received the smaller number of evaluations. This way, we obtained a balanced number of evaluations for all segments.

In total, 801 stick-figure animations were used in this study. We collected 2403 answers from anonymous participants. Each animation was labeled three times. Next, for each segment, the simple majority of the votes was considered to assign it to a class. Fig. 5 shows the final results. Most of the OBM segments were classified correctly (i.e., 425 out of 485). About half of the LBM segments were incorrectly labeled as nonlaughter segments (i.e., 171 out of 316). Our participants tended to often use the "nonlaughter" label. The accuracy of the human classification is 0.71, the global F-score is 0.695. The results are presented in Table VI.

TABLE VI
AVERAGED AND SINGLE-CLASS PRECISION AND F-SCORE OF MACHINE AND HUMAN CLASSIFICATION

	Precision (F-score)		
Measure Class	Weighted Avg.	Laughter	Nonlaughter
SVM	0.72(0.71)	0.62(0.64)	0.78(0.75)
k-NN	0.75(0.74)	0.72(0.65)	0.76(0.80)
RF	0.73(0.72)	0.66(0.66)	0.78(0.77)
LR	0.72(0.71)	0.64(0.64)	0.77(0.76)
NB	0.69(0.59)	0.72(0.32)	0.65(0.77)
Human	0.71(0.70)	0.70(0.55)	0.71(0.78)

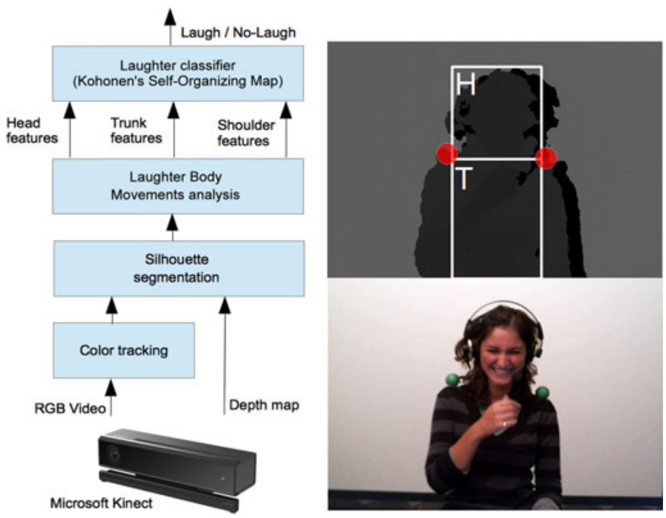
There is a difference between the selection done by raters skilled in nonverbal body movements (see Section III-B1) and the results of this study. However, that these two tasks are different: the elements of the LBM set were chosen using precise criteria that were explicitly explained to the raters (i.e., cues F1–F13), whereas in the perceptive study, we asked participants to express their overall feeling about animations. In order to check whether this difference depends on a specific subject, we carried out additional analyzes on the LBM segments only. The percentage of correctly annotated laughter segments ranges from 5% to 61%. The participants did not recognize most of the laughs of subjects S9 (5% correctly recognized animations), S6 (30%), and S10 (33%).

F. Discussion

The results of our study show that it is possible to build a machine that can recognize laughter from full-body movements. Both humans and machines exhibited similar performance in such a task: They are both well above chance level (50%). Interestingly, when comparing the Recalls and F-scores of automatic classification and human observers (see Table VI), the number of true positives and true negatives in automatic classification is more balanced than for the human observers (e.g., recall for SVM is 0.67 (laughter) versus 0.73 (nonlaughter), while for human classification is 0.46 versus 0.88). Whereas humans were not particularly good in detecting laughter segments, some classification algorithms (e.g., SVM) were able to classify on average more laughter segments correctly (but less nonlaughter segments).

An limitation of our study is that the number of segments per participant in our dataset was not balanced. During the recordings, important differences in number and intensity of laughs between participants were observed (see [28]). The personal laughing styles of the participants, who more frequently appear in the dataset may have influenced the models the machine learning algorithms generated. An advantage of our approach is that the features we compute strictly follow the latest theoretical works on the expressive pattern of laughter.

It would be interesting to compare the classification accuracy when using different techniques and modalities, e.g., audio, video. However, direct comparison of our results with other



(a) (b)

Fig. 6. Our system prototype for automated laughter detection from full-body movement: (a) the prototype system architecture, (b) a user is sitting in front of the system: the user's body silhouette is extracted and segmented in two parts: head (upper box marked by H) and trunk (lower box marked by T).

laughter detection algorithms is not possible, because: 1) such algorithms were trained and tested on different datasets (and full-body movement data are not available), and 2) it is difficult to record at the same time different modalities (e.g., spontaneous facial expressions and body movements) with the existing technology (see Section I). We made our training set publicly available to facilitate future research in this area. Laughter detection from acoustic cues is around 70–80%, whereas multimodal (facial and audio) detection can even reach the accuracy of 90%. Even if the results of our classifiers are lower than results of other classifiers trained on acoustic or multimodal data, our classifiers can be used when data from other modalities is not available or can be noisy.

Comparing with the results of Griffin *et al.* [8], we obtain comparable F-scores on our dataset. They obtained the best results using RF (F-score: 0.60 for laughter class, F-score: 0.76 for nonlaughter class), and SVR (F-score: 0.63 for laughter, 0.61 for nonlaughter), while our best F-scores were: 0.66 (laughter), 0.77 (nonlaughter) using RF, and 0.65 (laughter), 0.80 (nonlaughter) using k-NN. Their results were obtained on a dataset including both sitting and standing participants, and the results on standing participants only were lower than ours. In our study, we only use the standing data (thus, potentially more difficult case, as Griffin *et al.*, showed in [8]).

IV. System Prototype

We applied the results of our study to design and implement a system prototype using low-cost consumer hardware and lightweight algorithms to detect laughter from body movements. The architecture of our system prototype is depicted in Fig. 6. We exploit a Kinect sensor,[5] a laptop, two polystyrene markers (to simplify tracking of shoulder movement), and the freely available EyesWeb XMI platform.

A. Setup

In Fig. 6(b), the user sits on a stool in front of a computer screen with a Kinect device on top of it, wearing lightweight green polystyrene markers on her shoulders. The user's position puts some constraints on her degree of movement (the user has to remain seated and look at the screen), introducing some limitations on the features we can extract in the prototype. For example, legs are not visible, arms never move because of the table, the user's head, and trunk are always facing the camera. However, head and trunk movements are measurable, as well as the shoulders due to the green markers.

Tracked markers are highlighted in red on the user's silhouette in Fig. 6. The user's silhouette, automatically extracted by Kinect, is segmented in two regions based on the position of the markers: head and trunk (H and T areas, respectively, in Fig. 6). The Kinect SDK also provides as output the distance of the user's silhouette from the sensor: We consider head and trunk distance in a separate way; we define D as the difference between head and trunk distances (i.e., it approximates trunk leaning).

B. Feature Vector

With respect to the 13 features F1–F13 described in Section III-B1, our real-time system uses nine features K1–K9 computed in real-time with EyesWeb XMI. The first two (K1 and K2) are the same as before (F1 and F2): they measure the head's horizontal and vertical displacement of the head's 2-D barycenter. Three features (K3, K4, and K5) measure torso movements: 1) periodicity of trunk (K3) approximates abdomen shaking (F5) and trunk rocking (F7) by checking whether distance D (head versus trunk distance) varies in a periodic way; 2) maximum amplitude of distance D (K4) measures trunk straightening (F6); and 3) trunk impulsiveness (K5), computed as the ratio between peaks height and duration of D, corresponds to trunk throwing (F8). Considering the limitations and constraints on the user's degree of movement, we implemented an analysis of the user's shoulders to overcome the missing information about the user's legs and arms. Left and right shoulder periodicity (K8 and K9), computed by checking whether shoulder vertical position varies in a periodic way, correspond to shoulder shaking (F13). Two new features were introduced, inspired by Mancini *et al.*, [7]: shoulder energy (K7) and correlation (K8). These two features benefit from the prototype setup: With the user sitting in front of a camera, it is easier to compute them. Features regarding legs (F3 and F4) and arms (F9 to F12) cannot be computed with this prototype setup.

C. Automated Classification and Discussion

A dataset consisting of 367 laughter and nonlaughter segments from five participants was created. Participants were asked to perform two different tasks from those presented in Section III-A: an individual one, that is, watching video clips

[5] http://www.xbox.com

alone; and a social one, that is, playing the "Yes/no" game via Skype. At the beginning, the participant was invited to play the "Yes/no" game via Skype with one of the experimenters. Then, the participant was asked to choose and watch from internet a comedy clip she liked (e.g., tv shows, clips from movies), lasting about 4–6 min, and then a comedy clip that the experimenters previously selected. Finally, the participant had to play for a second time the "Yes/no" game.

Two classifiers were trained and run on the dataset: SVM and Kohonen's self-organizing map (SOM). The first one is described previously, the second one exhibits two main differences: 1) it executes quickly; and 2) the configuration of the map can be updated in real time: that is, it can adapt to the movement features values that characterize a user. We did not yet exploit the latter capability in our prototype, but previous work showed that this approach can be used to create *reflexive* interfaces [35]. SVM had a performance (F-score) of 0.73 (Precision 0.75, Recall 0.73). For the SOM, the F-score was 0.68 (Precision 0.65, Recall 0.73) and Accuracy 0.69.

The classification results are comparable with the results obtained in our study presented in Section III. However, in this setup, we use less precise data (Kinect and video instead of MoCap), and the setup has some constraints: participants are sitting, and their movements are limited. In such a setup, the laughter detection from full-body movement might be easier, as Griffin *et al.* showed in [8]. Thus, while the first aspect could influence negatively the detection, the second might counterbalance the lower performance of the input sensors.

V. Conclusion

In this paper, we have presented techniques to detect laughter solely from body movements. For this purpose, we developed laughter detection algorithms based on 13 full-body movement features extracted from motion captured data and grounded in an LBM annotation schema [20]. The algorithms were applied to a dataset of 801 manually segmented laughter and nonlaughter episodes with a total duration of 73 min. These episodes consisted of spontaneous full-body behaviors collected during social multiperson activities (e.g., social games). In this context, the use of other modalities to detect laughter is challenging since different participants' utterances (i.e., speech and laughter) overlap each other continuously and participants are very mobile, making face tracking difficult. The dataset is available for research purposes. The obtained classification results improve the current state of the art: discriminative classifiers (SVM, RF, k-NN) outperformed probabilistic classifiers (NB, LR) and slightly higher classification results were obtained in comparison to the results of previous work. Moreover, in our work, on laughter detection, we compare automated detection with the human ability to recognize laughter from body movements on the same dataset. We found that the overall performance of our algorithms was similar to the performance of the human observers but automatic classification algorithms obtained better scores for laughter detection (although they were worse for nonlaughter detection). Thus, machines can surpass humans in laughter detection from full-body movement in situations involving sensory deprivation (e.g., when no audio modality is available). A prototype system for automating the detection of laughter using low-cost motion tracking was introduced and evaluated.

To create laughter-sensitive interfaces, several open research questions remain unanswered. The automatic real-time laughter segmentation of continuous body movement is still an open challenge. Fusion algorithms must take into account the entire palette of human interaction modalities: Initial work in this direction proposed by Petridis *et al.* [13] does not yet consider body movement. Classification of different laughter types also has to be addressed. Initial work on this topic was carried out by Griffin *et al.* [36], who tried to distinguish between "hilarious" and "social" laughter. Future research should also address the detection of different communicative intentions of laughter, to communicate irony for example, from body movement. The analysis of full-body movement can be particularly useful for detecting behavior regulation, that is, when one tries to inhibit laughter.

Acknowledgment

The authors would like to thank T. Baur (University of Augsburg), H. Griffin, and M. S. H. Aung (University College of London) who supported the recording of the MMLI corpus.

References

[1] K. Grammer, "Strangers meet: Laughter and nonverbal signs of interest in opposite-sex encounters," *J. Nonverbal Behavior*, vol. 14, no. 4, pp. 209–236, 1990.

[2] M. J. Owren and J.-A. Bachorowski, "Reconsidering the evolution of nonlinguistic communication: The case of laughter," *J. Nonverbal Behavior*, vol. 27, pp. 183–200, 2003.

[3] B. Fredrickson, "The broaden-and-build theory of positive emotions," *Philosoph. Trans. - Roy. Soc. London Series B*, pp. 1367–1378, 2004.

[4] L. W. Hughes and J. B. Avey, "Transforming with levity: Humor, leadership, and follower attitudes," *Leadership Org. Develop. J.*, vol. 30, no. 6, pp. 540–562, 2009.

[5] R. Dunbar, "Mind the gap: Or why humans are not just great apes," in *Proc. Brit. Academy*, vol. 154, pp. 403–423, 2008. DOI:10.5871/bacad/9780197264355.003.0015

[6] R. Mora-Ripoll, "The therapeutic value of laughter in medicine," *Alternative Therapies Health Med.*, vol. 16, no. 6, pp. 56–64, 2010.

[7] M. Mancini, G. Varni, D. Glowinski, and G. Volpe, "Computing and evaluating the body laughter index," in *Human Behavior Understanding* (ser. Lecture Notes in Computer Science), vol. 7559, A. Salah, J. Ruiz-del Solar, C. Meriçli, and P.-Y. Oudeyer, Eds. Berlin, Germany: Springer, 2012, pp. 90–98.

[8] H. Griffin, M. Aung, B. Romera-Paredes, C. McLoughlin, G. McKeown, W. Curran, and N. Berthouze, "Perception and automatic recognition of laughter from whole-body motion: Continuous and categorical perspectives," *IEEE Trans. Affective Comput.*, vol. 6, no. 2, pp. 165–178, Apr.–Jun. 2015.

[9] M. Mancini, G. Varni, R. Niewiadomski, G. Volpe, and A. Camurri, "How is your laugh today?" in *Proc. Extended Abstr. Human Factors Comput. Syst.*, 2014, pp. 1855–1860.

[10] G. McKeown, W. Curran, D. Kane, R. Mccahon, H. Griffin, C. McLoughlin, and N. Bianchi-Berthouze, "Human perception of laughter from context-free whole body motion dynamic stimuli," in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interaction*, Sep. 2013, pp. 306–311.

[11] Y. Ding, K. Prepin, J. Huang, C. Pelachaud, and T. Artières, "Laughter animation synthesis," in *Proc. Int. Conf. Auton. Agents Multi-Agent Syst.*, 2014, pp. 773–780.

[12] R. Niewiadomski, M. Mancini, Y. Ding, C. Pelachaud, and G. Volpe, "Rhythmic body movements of laughter," in *Proc. 16th Int. Conf. Multimodal Interaction*, 2014, pp. 299–306.

[13] S. Petridis, B. Martinez, and M. Pantic, "The MAHNOB laughter database," *Image Vis. Comput.*, vol. 31, no. 2, pp. 186–202, 2013.

[14] S. Escalera, E. Puertas, P. Radeva, and O. Pujol, "Multi-modal laughter recognition in video conversations," in *Proc. Comput. Vis. Pattern Recog. Workshop*, Jun. 2009, pp. 110–115.
[15] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica United Acustica*, vol. 86, no. 1, pp. 117–128, 2000.
[16] A. Camurri, P. Coletta, G. Varni, and S. Ghisio, "Developing multimodal interactive systems with EyesWeb XMI," in *Proc. Conf. New Interfaces Musical Expression*, 2007, pp. 302–305.
[17] S. Piana, M. Mancini, A. Camurri, G. Varni, and G. Volpe, "Automated analysis of non-verbal expressive gesture," in *Proc. Human Aspects Ambient Intell.*, 2013, pp. 41–54.
[18] M. Mancini, J. Hofmann, T. Platt, G. Volpe, G. Varni, D. Glowinski, W. Ruch, and A. Camurri, "Towards automated full body detection of laughter driven by human expert annotation," in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interaction*, Sep. 2013, pp. 757–762.
[19] W. Ruch and P. Ekman, "The expressive pattern of laughter," in *Emotion, Qualia and Consciousness*, A. Kaszniak, Ed. Tokyo, Japan: World Scientific, 2001, pp. 426–443.
[20] W. F. Ruch, T. Platt, J. Hofmann, R. Niewiadomski, J. Urbain, M. Mancini, and S. Dupont, "Gelotophobia and the challenges of implementing laughter into virtual agents interactions," *Frontiers Human Neuroscience*, vol. 8, p. 928, 2014.
[21] K. P. Truong and D. A. van Leeuwen, "Automatic discrimination between laughter and speech," *Speech Commun.*, vol. 49, no. 2, pp. 144–158, 2007.
[22] H. Salamin, A. Polychroniou, and A. Vinciarelli, "Automatic detection of laughter and fillers in spontaneous mobile phone conversations," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 4282–4287.
[23] S. Tatsumi, Y. Mohammad, Y. Ohmoto, and T. Nishida, "Detection of hidden laughter for human-agent interaction," *Procedia Comput. Sci.*, vol. 35, pp. 1053–1062, 2014.
[24] M. T. Knox and N. Mirghafori, "Automatic laughter detection using neural networks." in *Proc. INTERSPEECH*, 2007, pp. 2973–2976.
[25] S. Scherer, M. Glodek, F. Schwenker, N. Campbell, and G. Palm, "Spotting laughter in natural multiparty conversations: A comparison of automatic online and offline approaches using audiovisual data," *ACM Trans. Interact. Intell. Syst.*, vol. 2, no. 1, pp. 4:1–4:31, Mar. 2012.
[26] S. Cosentino, T. Kishi, M. Zecca, S. Sessa, L. Bartolomeo, K. Hashimoto, T. Nozawa, and A. Takanishi, "Human-humanoid robot social interaction: Laughter," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2013, pp. 1396–1401.
[27] B. de Gelder, J. V. den Stock, H. K. Meeren, C. B. Sinke, M. E. Kret, and M. Tamietto, "Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions," *Neurosci. Biobehavioral Rev.*, vol. 34, no. 4, pp. 513–527, 2010.
[28] R. Niewiadomski, M. Mancini, T. Baur, G. Varni, H. Griffin, and M. S. Aung, "MMLI: Multimodal multiperson corpus of laughter in interaction," in *Human Behavior Understanding* (ser. Lecture Notes in Computer Science), vol. 8212, A. Salah, H. Hung, O. Aran, and H. Gunes, Eds. Berlin, Germany: Springer, 2013, pp. 184–195.
[29] G. McKeown, W. Curran, C. McLoughlin, H. Griffin, and N. Bianchi-Berthouze, "Laughter induction techniques suitable for generating motion capture data of laughter associated body movements," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recog.*, Apr. 2013, pp. 1–5.
[30] J. Urbain, R. Niewiadomski, E. Bevacqua, T. Dutoit, A. Moinet, C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner, "AVLaughterCycle: Enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audiovisual laughter animation," *J. Multimodal User Interfaces*, vol. 4, no. 1, pp. 47–58, 2010.
[31] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
[32] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.
[33] R. R. Bouckaert, "Choosing between two learning algorithms based on calibrated tests," in *Proc. 20th Int. Conf. Mach. Learning*, 2003, pp. 51–58.
[34] H. Bergson, *Le Rire: Essai sur la Signification du Comique*. Paris, France: F. Alcan, 1901.
[35] G. Varni, G. Volpe, R. Sagoleo, M. Mancini, and G. Lepri, "Interactive reflexive and embodied exploration of sound qualities with besound," in *Proc. 12th Int. Conf. Interaction Design Children*, 2013, pp. 531–534.
[36] H. Griffin, M. Aung, B. Romera-Paredes, C. McLoughlin, G. McKeown, W. Curran, and N. Bianchi-Berthouze, "Laughter type recognition from whole body motion," in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interaction*, 2013, pp. 349–355.

Radoslaw Niewiadomski received the Ph.D. degree in computer science from the University of Perugia, Perugia, Italy, in 2007.

Since June 2013, he has been involved in research on expressive gesture analysis with the DIBRIS, University of Genoa, Genoa, Italy. His research interests are in the area of affective computing, and include embodied conversational agents, detection and synthesis of emotions, interactive multimodal systems, and evaluation of affective interaction.

Maurizio Mancini received the Ph.D. degree in computer science from the University of Paris 8, Saint-Denis, France, in 2008.

Since 2001, he has carried out his research activity in the framework of several EU projects in FP5-7 and H2020. In 2008, he joined the DIBRIS, University of Genoa, Genoa, Italy. His research interests include the definition and implementation of models and algorithms for automated expressive movement analysis and synthesis in the field of HCI.

Giovanna Varni received the M.Sc. degree in biomedical engineering and the Ph.D. degree in electronic, computer, and telecommunications engineering from the DIBRIS, University of Genoa, Genoa, Italy, in 2005 and 2009, respectively.

She is currently a Postdoctoral Research Assistant with the Institute for Intelligent Systems and Robotics, University Pierre and Marie Curie, Paris, France. Her research interests include the area of HCI, especially on social signals analysis and expressive gesture.

Gualtiero Volpe received the M.Sc. degree in computer engineering and the Ph.D. degree in electronic and computer engineering from the DIBRIS, University of Genoa, Genoa, Italy in 1999 and 2003, respectively.

Since 2014, he has been an Associate Professor with the DIBRIS, University of Genoa, Genoa, Italy. His research interests include intelligent and affective human–machine interaction, social signal processing, sound and music computing, modeling and real-time analysis and synthesis of expressive content, and multimodal interactive systems.

Antonio Camurri received the Ph.D. degree in computer engineering from the University of Genoa, Genoa, Italy.

He is currently a Full Professor with DIBRIS, University of Genoa, Genoa, Italy. He is the Founder and Scientific Director of InfoMus Lab and of Casa Paganini (www.infomus.org). He is a Coordinator and Local Project Manager of several EU projects (FP5-FP7, H2020, Culture 2007, Cost Actions). His research interests include multimodal intelligent interfaces and interactive systems; sound and music computing; computational models of expressive gesture, emotion, and social signals; and multimodal systems for theater, music, dance, museums, and health.