Multimodal Analysis of laughter for an Interactive System

Jérôme Urbain[1], Radoslaw Niewiadomski[2], Maurizio Mancini[3], Harry Griffin[4], Hüseyin Çakmak[1], Laurent Ach[5], Gualtiero Volpe[3]

[1] Université de Mons, Place du Parc 20, 7000 Mons, Belgium
jerome.urbain@umons.ac.be
[2] LTCI UMR 5141 - Telecom ParisTech, Rue Dareau, 37-39, 75014 Paris, France
[3] Università degli Studi di Genova, Viale Francesco Causa, 13, 16145 Genova, Italy
[4] UCL Interaction Centre, University College London,
Gower Street, London, WC1E 6BT, United Kingdom
[5] LA CANTOCHE PRODUCTION, rue d'Hauteville, 68, 75010 Paris, France

Abstract. In this paper, we focus on the development of new methods to detect and analyze laughter, in order to enhance human-computer interactions. First, the general architecture of such a laughter-enabled application is presented. Then, we propose the use of two new modalities, namely body movements and respiration, to enrich the audiovisual laughter detection and classification phase. These additional signals are acquired using easily constructed affordable sensors. Features to characterize laughter from body movements are proposed, as well as a method to detect laughter from a measure of thoracic circumference.

Key words: laughter, multimodal, analysis

1 Introduction

Laughter is an important signal in human communication. It can convey emotional messages, but is also a common back-channeling signal, indicating, for example, that we are still actively following the conversation. In dyadic conversations, each participant laughs, on average, every 2 minutes [1]. Recent works have also discovered the positive impact of a laughing virtual agent on users experiencing human-machine interactions [2].

Our long-term objective is to integrate laughter into human-machine interactions, in a natural way. This requires building an interactive system able to efficiently detect human laughter, analyze it and synthesize an appropriate response. The general system architecture of our application is displayed in Figure 1. We distinguish 3 types of components: input components, decision components and output components.

The input components are responsible for multimodal data acquisition and real-time laughter analysis. In our previous experiments [2], only the audio modality was used for laughter detection. This resulted in two types of detection errors: a) false alarms in presence of noise; b) missed detections when the

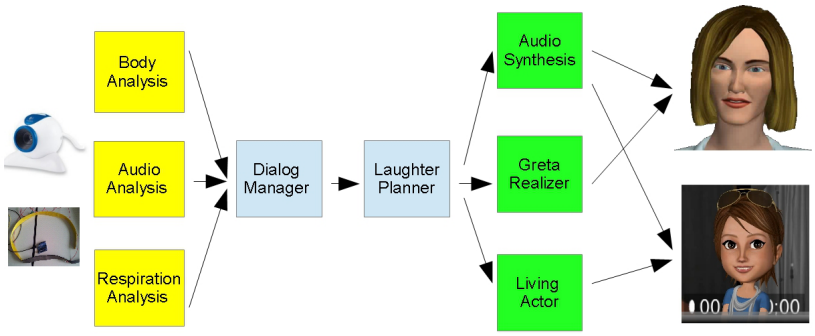

Fig. 1. Overall architecture composed of input components (in yellow), decision components (in blue) and output components (in green).

laugh is (almost) silent. This is why in this work we are introducing new modalities to make the laughter detection more robust. The input components now include laughter detection from body movements and respiration in addition to audio detection and intensity estimation. The data on user behavior (see Table 1) are captured with two devices: a simple webcam and the respiration sensor developed at University College London (see Section 4).

Table 1. Recorded signals.

Recording device	*Captured signal*	*Description*
Webcam	Video	RGB, 25 fps
	Audio	16 kHz, 16 bit, mono
Respiration Sensor	Respiration	120Hz, 8 bit

The laughter-enabled decision making modules decide, given the information from the input components, when and how to laugh so as to generate a natural interaction with human users. At the moment, two decision components are used to decide the agent audiovisual response. The first one (Dialog Manager) receives the information from the input components (*i.e.*, laughter likelihoods and intensity) as well as contextual information and it generates the instruction to laugh (or not) with high-level information on the laugh to produce (*i.e.*, its duration and intensity). The second component, Laughter Planner, controls the details of the expressive pattern of the laughter response by choosing, from the lexicon of pre-synthesized laughter samples, the most appropriate audiovisual episode, *i.e.* the episode that best matches the requirements specified by the Dialog Manager module.

Finally, the output components are responsible for the audiovisual laughter synthesis that generates avatar laugher when the decision components instruct them to do so. For this purpose two different virtual characters are used: Greta

Realizer [3] and Living Actor by Cantoche[1]. At the moment the acoustic and visual modalities of laughter are synthesized separately using the original audiovisual signals from the AVLaughterCycle (AVLC) corpus of human laughter [4]. All synthesized episodes are stored in the agent lexicon, and can then be displayed in real-time. In more details, audio is synthesized with the use of the HMM-based Speech Synthesis System (HTS). HMMs have been trained on the AVLC database and its phonetic annotations [5]. The facial animation in the Greta Realizer was created with two different approaches [6]. First, a procedural approach was used: the AVLC videos were manually annotated with FACS [7], then the animations were resynthesized with the Greta system, able to control the intensity and duration of each action unit. The second approach - a data-driven synthesis - was realized by applying a freely available face tracker to detect facial landmarks on the AVLC videos and then by mapping these landmarks displacements to the facial animation parameters of the virtual character.

The facial animation of Living Actor virtual characters is similar to speech synthesis, where information about phonemes or visemes is sent by the Text to Speech engine along with the audio signal. For laughter, the visemes are composed of lip deformation but also cheek and eye movements. Pseudo-phoneme information is sent using a chosen nomenclature of sounds depending on the synthesis functions. Figure 2 displays examples of laughter poses.


```
<?xml version="1.0" encoding="iso-8859-1" ?>
<LivingActor>
<item time_offset="0.223" morph_name="Laugh_Hum" intensity="1"/>
<item time_offset="0.521" morph_name="Laugh_a" intensity="0.5"/>
<item time_offset="2.190" morph_name="Laugh_a" intensity="0.3"/>
<item time_offset="3" morph_name="Laugh_a" intensity="0.1"/>
</LivingActor>
```

Fig. 2. Using laughter visemes for facial animation

A demo of the system can be viewed at `https://www.youtube.com/watch?v=fElP2_c8vJU`. Further details on the components (acoustic laughter detection, decision and audiovisual synthesis), the communication middleware as well as experimental results can be found in [2].

The rest of this paper focuses on the new input components of our system, with the objective of improving laughter detection robustness through multimodal decisions. In Section 2 we present related work for laughter detection. Section 3 discusses laughter detection from body cues while Section 4 shows

[1] http://www.cantoche.com

how we can use respiration which is a very important element of laughter expressive pattern. Finally, Section 5 presents the conclusions and future works.

2 Related work

In the last decade, several systems have been built to detect laughter. It started with audio-only classification. Kennedy and Ellis [8] obtained 87% accuracy with Support Vector Machines fed with 6 MFCCs; Truong and van Leeuwen [9] reached slightly better results (equal error rate of 11%) with Neural Networks fed with Perceptual Linear Prediction features; Knox and Mirghafori [10] obtained better performance (around 5% error) by using temporal feature windows.

In 2008, Petridis and Pantic started to enrich the so far mainly audio-based work in laughter detection by consulting audio-visual cues for decision level fusion approaches [11, 12]. They combined spectral and prosodic features from the audio modality with head movement and facial expressions from the video channel. They reported a classification accuracy of 74.7% to distinguish three classes, namely unvoiced laughter, voiced laughter and speech.

Since laughter detection robustness increases when combining audio and facial features [12], including other modalities can probably further improve the performance. First, the production of audible laughter is, in essence, a respiratory act since it requires the exhalation of air to produce distinctive laughter sounds ("Ha") or less obvious sigh- or hiss-like verbalizations. The respiratory patterns of laughter have been extensively researched as Ruch & Ekman [13] summarize. A distinctive respiration pattern has emerged of a rapid exhalation followed by a period of smaller exhalations at close-to-minimum lung volume. This pattern is reflected by changes in the volume of the thoracic and abdominal cavities, which rapidly decrease to reach a minimum value within approximately 1s [14]. These volumetric changes can be seen through the simpler measure of thoracic circumference, noted almost a century ago by Feleky [15]. Automatic detection of laughter from respiratory actions has previously been investigated using electromyography (EMG). Fukushima et al. [16] analyzed the frequency characteristics of diaphragmatic muscle activity to distinguish laughter, which contained a large high-frequency component, from rest periods, which contained mostly low-frequency components. In this paper, we will explore automatic laughter detection from the measure of the thoracic circumference (Section 4).

Second, intense laughter can be accompanied by changes in postures and body movements, as summarized by Ruch [17] and Ruch & Ekman [13]. Throwing the head backwards will ease powerful exhalations. The forced expiration movements can cause visible vibrations of the trunk and shoulders. This is why we propose features characterizing such laughter-related body movements, that are presented in Section 3.

3 Body analysis

The EyesWeb XMI platform is a modular system that allows both expert (*e.g.*, researchers in computer engineering) and non-expert users (*e.g.*, artists) to create multimodal installations in a visual way [18]. The platform provides modules, that can be assembled intuitively (*i.e.*, by operating only with the mouse) to create programs, called patches, that exploit system resources such as multimodal files, webcams, sound cards or multiple displays. The body analysis input component consists of an EyesWeb XMI patch performing analysis of the user's body movements in real-time. The computation performed by the patch can be split into a sequence of distinct steps, described in the following paragraphs.

Currently, the task of the body analysis module is to track the user's shoulders and characterize the variation of their positions in real-time. To this aim we could use a sensor like Kinect to provide the user's shoulders data as input to our component. However, we observed that the Kinect shoulders' position do not consistently follow the user's actual shoulder movement: in the Kinect skeleton, shoulders' position is extracted via a statistical algorithm on the user's silhouette and depth map and usually this computation cannot track subtle shoulder movement, for example, small upward/downward movements.

This is why in this paper we present a different type of shoulder movement detection technique: two small and lightweight green polystyrene spheres have been fixed on top of the user's shoulders. The EyesWeb patch separates the green channel of the input video signal to isolate the position of the two spheres. Then a tracking algorithm is performed to follow the motion of the sphere frame by frame, as shown in Figure 3. However, the above technique can be used only in controlled environments, i.e., it can not be used in real situations when users are free to move in the environment. So we plan to perform experiments to compare the two shoulder movement detection techniques: the one based on Kinect and the one based on markers. Results will guide us in developing algorithms for approximating user's shoulder movement from Kinect data.

Fig. 3. Two green spheres placed on the user's shoulders are tracked in real-time (red and blue trajectories)

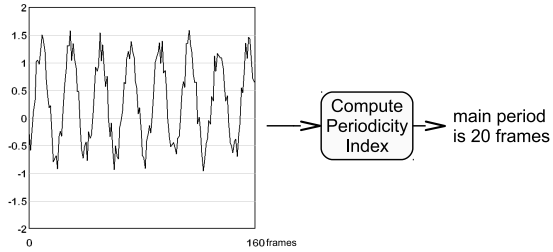

Fig. 4. An example of Periodicity Index computation: the input time-series (on the left) has a periodicity of 20 frames.

The position of each user's shoulder is associated to the barycenter of each sphere, which can be computed in two ways. The first consists in the computation of the graphical barycenter of each sphere, that is, the mean of the pixels of each sphere's silhouette is computed. The second option includes some additional steps: after computing the barycenter like in the first case, we consider a square region around it and we apply a Lukas-Kanade [19] algorithm to this area. The result is a set of 3 points on which we compute the mean: the resulting point is taken as the position of the shoulder. From this shoulder tracking, several laughter-related features can be computed:

- **Correlation:** The correlation ρ is computed as the Pearson correlation coefficient between the vertical positions of the user's shoulders. Vertical positions are approximated by the y coordinate of each shoulder's barycenter.
- **Kinetic energy:** The kinetic energy is computed from the speed of user's shoulders and their percentage mass as referred by [20]:

$$E = \frac{1}{2}(m_1 v_1 + m_2 v_2) \tag{1}$$

- **Periodicity:** Kinetic energy is serialized in a sliding window time-series having a fixed length. Periodicity is then computed, using *Periodicity Transforms* [21]. The time-series is decomposed into a sum of its periodic components by projecting data onto periodic subspaces. Periodicity Transforms also output the relative contribution of each periodic signal to the original one. Among many algorithms for computing Periodicity Transforms, we chose *mbest*. It determines the m periodic components that, subtracted from the original signal, minimize the residual energy. With respect to the other algorithms, it provides a better accuracy and does not need the definition of a threshold. Figure 4 shows an example of computation of the Periodicity Index in EyesWeb for a sinusoidal signal affected by a uniform noise in the range $[0, 0.6]$.
- **Body Laughter Index:** Body Laughter Index (BLI) stems from the combination of the averages of shoulders' correlation and kinetic energy, 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 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:

$$BLI = \alpha\bar{\rho} + \beta\bar{E} \tag{2}$$

As reported in [13], rhythmical patterns produced during laughter usually have frequencies around 5 Hz. 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), 25 in our case.

Figure 5 displays an example of analysis of user's laugh. A previously segmented video is provided as input to the EyesWeb XMI body analysis module. The green plot represents the variation of the BLI in time. When the BLI is acknowledged by the Periodicity Index value the plot becomes red. In [22] we present a preliminary study in which BLI is validated on a corpus of laughter videos. A demonstration of the Body Laughter Index can be watched on `http://www.ilhaire.eu/demo`.

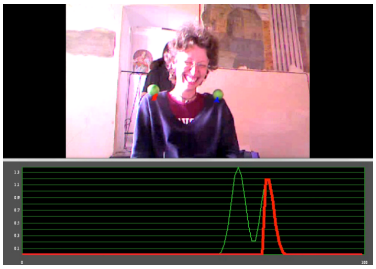

Fig. 5. An example of Body Laughter Index computation

4 Respiration

In order to capture the laughter-related changes in thoracic circumference (see Section 2), we constructed a respiration sensor based on the design of commercially available sensors: the active component is a length of extensible conductive fabric within an otherwise inextensible band that is fitted around the upper thorax. Expansions and contraction of the thorax change the length of the conductive fabric causing changes in its resistance. These changes in resistance are used to modulate an output voltage that is monitored by the Arduino prototyping platform[2]. A custom written code on the Arduino converts the voltage to a 1-byte serial signal, linear with respect to actual circumference, which is passed to a PC over a USB connection at a rate of approximately 120Hz.

While Fukushima et al. [16] designed a frequency-based laughter detection module (from EMG signals), our approach is time-based. Laughter onset is identified through the appearance of 3 respiration events (see Figure 6):

[2] `http://www.arduino.cc/`

1. A sharp change in current respiration state (inhalation, pause, standard exhalation) to rapid exhalation.
2. A period of rapid exhalation resulting in rapid decrease in lung volume.
3. A period of very low lung volume.

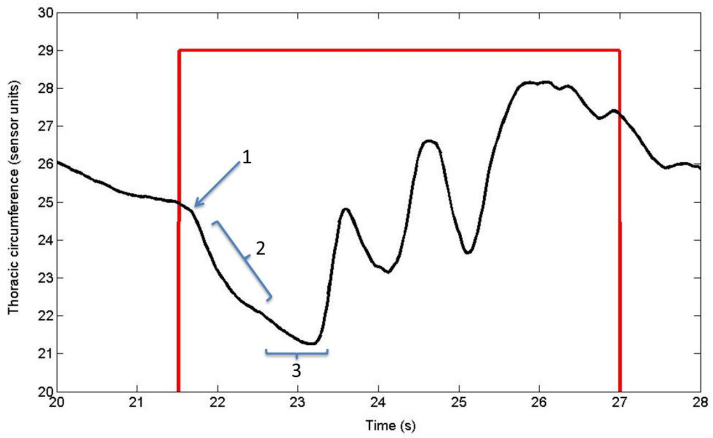

Fig. 6. Example of thoracic circumference, with laughter episode marked in red, and notable features of laughter initiation. Feature 1 - a sharp change in current respiration state to rapid exhalation; feature 2 - a period of rapid exhalation; feature 3 - a period of very low lung volume.

These appear as distinctive events in the thoracic circumference measure and its derivatives:

1. A negative spike in the second derivative of thoracic circumference.
2. A negative period in the first derivative of thoracic circumference.
3. A period of very low thoracic circumference.

These were identified by calculating a running mean (λ_f) and standard deviation (σ_f) for each measure. A running threshold (T_f) for each measure was calculated as: $T_f = \lambda_f - \alpha_f \sigma_f$, where α_f is a coefficient for that measure, empirically determined to optimise the sensitivity/specificity trade-off. Each feature was determined to be present if the value of the measure fell below the threshold at that sample. Laughter onset was identified by the presence of all three features in the relevant order (1 before 2 before 3) in a 1s sliding window. This approach restricts the number of parameters to 3 ($\alpha_{1\text{—}3}$) but does introduce lag necessary for calculating valid derivatives from potentially noisy data. It also requires a period for the running means and standard deviations, and so the thresholds, to stabilise. However, this process would be jeopardised by the presence of large, rapid respiratory event such as coughs and sneezes. The robustness of this detection module remains to be investigated, as well as what it can bring in multimodal detection.

5 Conclusion and future work

In this paper we have focused on the development of two new modalities to detect and characterize laughs that are integrated in a broader, fully functional, interactive application. These two modalities are affordable to include in multimodal systems and offer real-time monitoring. The proposed features are related to laughter behavior and will provide useful information to classify laughs and measure their intensity.

This is ongoing work. We will go on developing robust laughter detection. For example, the rules for laughter detection from respiration features, currently determined empirically, will be optimized in a larger study. In addition, other modalities will be included, for example facial tracking. For this purpose we plan to include another sensor, *i.e.* a Kinect camera. The latest version of the Microsoft Kinect SDK not only offers full 3D body tracking, but also a real-time 3D mesh of facial features tracking the head position, location of eyebrows, shape of the mouth, etc. Action units of laughter could thus be detected in real-time.

Secondly, our analysis components need formal evaluation. For this purpose we have recently captured using our analysis components the data of more than 20 people participating in laughter-eliciting interactions. The collected data will now be used to validate these components. In the future, we will also perform a methodical study of multimodal laughter detection and classification (*i.e.*, distinguishing different types of laughter), to evaluate the performance of each modality (audio, face, body, respiration) and measure the improvements that can be achieved by fusing modalities. The long term aim is to develop an intelligent adaptive fusion algorithm. For example, in a noisy environment audio detection should receive a lower importance.

This additional information will allow our decision components to better tune the virtual character reactions to the input, and hence enhance the interactions between the participant and the virtual agent.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement $n°270780$. H. Çakmak receives a Ph.D. grant from the Fonds de la Recherche pour l'Industrie et l'Agriculture (F.R.I.A.), Belgium.

References

1. Vettin, J., Todt, D.: Laughter in conversation: Features of occurrence and acoustic structure. Journal of Nonverbal Behavior **28**(2) (2004) 93–115
2. Niewiadomski, R., Hofmann, J., Urbain, J., Platt, T., Wagner, J., Piot, B., Cakmak, H., Pammi, S., Baur, T., Dupont, S., Geist, M., Lingenfelser, F., McKeown, G., Pietquin, O., Ruch, W.: Laugh-aware virtual agent and its impact on user amusement, Saint Paul, Minnesota, USA (May 2013)

3. Niewiadomski, R., Bevacqua, E., Le, Q.A., Obaid, M., Looser, J., Pelachaud, C.: Cross-media agent platform. In: Web3D ACM Conference, Paris, France (2011) 11–19
4. Urbain, J., Niewiadomski, R., Bevacqua, E., Dutoit, T., Moinet, A., Pelachaud, C., Picart, B., Tilmanne, J., Wagner, J.: AVLaughterCycle: Enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audiovisual laughter animation. JMUI **4**(1) (2010) 47–58
5. Urbain, J., Dutoit, T.: A phonetic analysis of natural laughter, for use in automatic laughter processing systems. In: ACII'11, Memphis, Tennesse (2011) 397–406
6. Niewiadomski, R., Pammi, S., Sharma, A., Hofmann, J., Platt, T., Cruz, R., Qu, B.: Visual laughter synthesis: Initial approaches. In: Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech, Dublin, Ireland (2012)
7. Ekman, P., Friesen, W., Hager, J.: Facial action coding system: A technique for the measurement of facial movement (2002)
8. Kennedy, L., Ellis, D.: Laughter detection in meetings. In: NIST ICASSP 2004 Meeting Recognition Workshop, Montreal (May 2004) 118–121
9. Truong, K.P., van Leeuwen, D.A.: Automatic discrimination between laughter and speech. Speech Communication **49** (2007) 144–158
10. Knox, M.T., Mirghafori, N.: Automatic laughter detection using neural networks. In: Proceedings of Interspeech 2007, Antwerp, Belgium (August 2007) 2973–2976
11. Petridis, S., Pantic, M.: Fusion of audio and visual cues for laughter detection. In: Proceedings of the 2008 international conference on Content-based image and video retrieval, ACM (2008) 329–338
12. Petridis, S., Pantic, M.: Audiovisual discrimination between speech and laughter: Why and when visual information might help. IEEE Transactions on Multimedia **13**(2) (2011) 216–234
13. Ruch, W., Ekman, P.: The expressive pattern of laughter. In Kaszniak, A., ed.: Emotion, qualia and consciousness. World Scientific Publishers, Tokyo (2001) 426–443
14. Filippelli, M., Pellegrino, R., Iandelli, I., Misuri, G., Rodarte, J., Duranti, R., Brusasco, V., Scano, G.: Respiratory dynamics during laughter. Journal of Applied Physiology **90**(4) (2001) 1441
15. Feleky, A.: The influence of the emotions on respiration. Journal of Experimental Psychology **1**(3) (1916) 218–241
16. Fukushima, S., Hashimoto, Y., Nozawa, T., Kajimoto, H.: Laugh enhancer using laugh track synchronized with the user's laugh motion. In: Proceedings of CHI'10. (2010) 3613–3618
17. Ruch, W.: Exhilaration and humor. Handbook of emotions **1** (1993) 605–616
18. Camurri, A., Coletta, P., Varni, G., Ghisio, S.: Developing multimodal interactive systems with eyesweb xmi. In: NIME'07. (2007) 302305
19. Lukas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th IJCAI. (1981)
20. Winter, D.: Biomechanics and motor control of human movement (1990)
21. Sethares, W., Staley, T.: Periodicity transforms. IEEE Transactions on Signal Processing **47**(11) (1999) 2953–2964
22. Mancini, M., Varni, G., Glowinski, D., Volpe, G.: Computing and evaluating the body laughter index. In Salah, A., Ruiz-del Solar, J., Merili, ., Oudeyer, P.Y., eds.: Human Behavior Understanding. Volume 7559 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2012) 90–98