MMLI: Multimodal Multiperson Corpus of Laughter in Interaction

Radoslaw Niewiadomski[1], Maurizio Mancini[1], Tobias Baur[2], Giovanna Varni[1], Harry Griffin[3], and Min S.H. Aung[3]

[1] Università degli Studi di Genova, Viale Francesco Causa, 13, 16145 Genova, Italy
{radek,giovanna}@infomus.org, maurizio.mancini@unige.it
[2] Augsburg University, Universitätsstr. 6a, 86159 Augsburg, Germany
baur@hcm-lab.de
[3] University College London, Gower Street, London, WC1E 6BT, United Kingdom
{harry.griffin,m.aung}@ucl.ac.uk

Abstract. The aim of the Multimodal and Multiperson Corpus of Laughter in Interaction (MMLI) was to collect multimodal data of laughter with the focus on full body movements and different laughter types. It contains both induced and interactive laughs from human triads. In total we collected 500 laugh episodes of 16 participants. The data consists of 3D body position information, facial tracking, multiple audio and video channels as well as physiological data.

In this paper we discuss methodological and technical issues related to this data collection including techniques for laughter elicitation and synchronization between different independent sources of data. We also present the enhanced visualization and segmentation tool used to segment captured data. Finally we present data annotation as well as preliminary results of the analysis of the nonverbal behavior patterns in laughter.

1 Introduction

Laughter is one of the most commonly appearing human communicative signals [1]. Despite its high incidence, knowledge about the multimodal expressive pattern of laughter is rather limited. Laughter is a very complex behavior that includes the majority of expressive modalities. Most research to date has focused on acoustic and facial cues of laughter. However, they are often accompanied by body movements and changes in posture[2] including, among others, head backwards movements and trunk/shoulders vibrations caused by forced exhalations. We argue that special attention should be paid to these body movements, as they are important in both laughter detection and synthesis. Laughter synthesis, for example, may benefit from the analysis of laughter episodes' body movements, as realistic body movements are crucial in distinguishing between laughter and smile visual patterns.

It was shown that many different types of laughter e.g. happy, embarrassed, contemptuous or schadenfreude laughter can be differentiated on a linguistic basis [3]. However it is still not clear if these laughter types also exhibit distinctive

A.A. Salah et al. (Eds.) : HBU 2013, LNCS 8212, pp. 184–195, 2013.

expressive patterns. Existence of distinct expressive patterns for two laughter types i.e. *Duchenne laughter* (i.e. a sign of enjoyment) and *social laughter* was shown (see [4]). It is also suggested that voluntary down/up regulation of laughter can be detected from acoustic and facial cues [2].

Building a multimodal laughter corpus is a challenging task because laughter mainly occurs during social interaction. This important aspect of laughter is often neglected and existing corpora usually contain data captured in induced, non-interactive setups e.g. data of people watching a video containing funny stimuli (e.g., [5,6]). Another approach consists of capturing the behaviors of the participants of multi-party meetings (e.g., [7,8]). These corpora allow one to capture the dynamics of the whole interaction but often contain only audio modality, and to our knowledge there is no database that focuses on capturing body movements.

In this paper we describe a new corpus - Multimodal and Multiperson Corpus of Laughter in Interaction[1] (MMLI). Capturing this data we focus on laughter full body movements in different contexts, and for different laughter types. The MMLI corpus will be made freely available for research purposes on the ILHAIRE database website: `http://qub.ac.uk/ilhairelaughter`.

2 State of the Art

Only a few corpora exist that were explicitly created with the purpose of studying laughter. More often the data collected for other aims e.g. multi-participants meetings analysis (such as AMI Corpus [8] or ICSI Meeting corpus [7]) is used. For instance, Truong and Leeuwen [9] selected audible laugh episodes from the ICSI corpus to build an acoustic laughter detector. Existing corpora often focus on one modality only. They are created ad-hoc with a concrete aim e.g. acoustic or visual laughter detection. Aiming at emotion differentiation in laughter Szameitat et al. [10] recorded 8 actors performing four types of laughter i.e., joyous, tickling, schadenfreude, and taunting. The actors were instructed to put themselves into emotional states with the help of self-induction techniques. The database contains 429 audio episodes and it was used to investigate the acoustical correlates of laughter expressing four emotions. Scherer et al. [11] collected the FreeTalk corpus consisting of 90 minutes of multiparty conversations of four participants. The corpus includes audio and video recordings (with 360 degree camera) of about 300 laugh episodes and was used for audio visual laughter detection.

The AudioVisualLaughterCycle corpus [5] contains multimodal data of about 1000 spontaneous laughter episodes recorded from 24 subjects. Each subject was recorded while watching a 10-minute comedy video. Each episode was captured with one motion capture system (either Optitrack or Zigntrack) and synchronized with the corresponding audiovisual sample. The material was manually segmented

[1] The research leading to the results presented in this paper has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 270780 (ILHAIRE project).

into single laugh episodes. The number of episodes for a subject ranges from 4 to 82.

Aiming to collect multimodal data for automatic audiovisual laughter detection Petridis et al. created the MAHNOB database [6]. It contains nearly 4 hours of recordings with the participation of 22 subjects who were recorded while watching funny video clips. The collected data is publicly available and consists of audio, the upper body video, as well as recordings of a thermal camera. It mostly contains induced laughter of enjoyment (563 episodes) but also posed smiles and laughs as well as speech. The database was primarily used for laughter vs. speech discrimination showing the advantage of a multimodal approach over audio-only detection in noisy environments.

Suarez et al. [12] created the PinoyLaughter audiovisual laughter corpus containing about 500 spontaneous and acted laugh episodes. The aim was to display various emotions that can be transmitted with laughter such as happiness, giddiness, excitement, embarrassment and hurtful laughter. While professional actors acted emotions, spontaneous laughter was collected from volunteers. Both acted and spontaneous data was pre-processed, manually segmented, and then annotated both with discrete and dimensional labels of emotions. Different emotional states were recognized mainly from the audio modality and the context.

With the aim of collecting samples of different laughter types such as conversational laughter McKeown et al. [13] reviewed existing six databases and built the ILHAIRE laughter database. This corpus contains audio and video of more than 1000 laugh episodes in various contexts: laughter induced by watching funny clips; laughter in conversations or when performing some engaging tasks.

Table 1. Comparison of main laughter corpora

	Modality			Interaction type			Laughter type
Name	Audio	Face/Head	Body	Posed	Induced	Interactive	(e.g., social)
MAHNOB	yes	yes	no	yes	yes	no	enjoyment, posed
AVLC	yes	yes	no	no	yes	no	enjoyment
PinoyLaughter	yes	yes	no	yes	yes	no	different emotions
Szameitat	yes	no	no	yes	no	no	different types
FreeTalk	yes	yes	no	no	no	yes	enjoyment
ILHAIRE	yes	yes	no	yes	yes	yes	enjoyment, social/conversational

As shown in Table 1 most of the existing corpora consist of audio and eventually facial cues of laughter. They often contain only posed or induced (i.e. non interactive) laughter. Finally, none of them offer high quality data of full body movements.

3 MMLI Data Collection

By capturing data for the MMLI corpus of laughter we mainly aimed at:

- capturing full body movements with special attention to shoulders, torso and respiration,

– capturing laughter in different contexts, and in particular, during unsupervised "free" interactions,
– capturing different types of laughter.

To collect multimodal data we built a complex setup that allowed us to collect the information from different sources. First of all, three high precision inertial motion capture systems were used to collect high quality data of body movements. These systems were complemented by Microsoft Kinect sensors, high frame rate cameras and a respiration sensor. All the data is synchronized through a freely available software called SSI (see Section 4). This allows one to analyze not only synchronization between different modalities in a laughter episode but also intra-subject synchronization.

To capture laughter in different contexts and various laughter types we invited groups of friends and asked them to perform six enjoyable tasks (T1 - T6). Beside classical laughter inducing tasks such as watching funny clips we proposed participants to play several "simple" social games, i.e. games regulated by one simple general rule in which participants are free to improvise. We supposed that a lack of detailed rules could encourage easy-going spontaneous behaviors that may include reactions such as commenting, joking, irony, or even embarrassment or schadenfreude. Additionally, some tasks were expected to cause down or up regulation of laughter. Thus, we expected that the resulting data could consist not only of enjoyment laughter, but also of some other laughter categories.

3.1 Techical Setup

In the data collection we captured the behavior of up to three interacting participants at the same time. For this purpose, as shown in Figure 1, we used 2 types of different inertial mocap systems:

– (XS1, XS2) 2 out of 3 participants were captured using Xsens MVN Biomech system[2] that is composed of 17 inertial sensors placed on Velcro straps. Data is captured at 120 frames per second; each frame consists of 22 joints' location and rotation in a 3D reference space. Data can be exported to common formats like C3D or BVH,
– (AZ) the third participant was recorded with the Animazoo IGS-190 system, equipped with 19 inertial sensors. Recorded data consists of 3D body joints' rotation in BVH format.

Additionally audio and video were recorded:

– 2 (M1 - M2) personal wireless microphones (Mono, 16 kHz) placed close to the participants' mouth,
– 1 (A1) microphone (Mono, 16 kHz) placed in between the participants to record the room ambient sound,
– 4 (W1 - W4) webcams Logitech Webcam Pro 9000 (640x480, 30fps),

[2] www.xsens.com

- 2 (C1, C2) high-frame rate cameras Philips PC webcam SPZ5000 (640x480, 60fps),
- 2 (K1, K2) Kinect cameras. Kinect cameras were used to collect video (640x480, 30fps) as well as additional data of face, head and body movements. The Kinect SDK allows tracking and extraction of 100 facial points, 20 body points, as well as 6 high-level facial actions such as: smiling or frowning; all these data are extracted with the frequency of 30Hz,
- 1 (RS) Respiration sensor (ProComp Infiniti, Thought Technology) to capture thoracic and abdominal circumference of one participant at 256 samples/second.

Two different setups (S1 and S2) were used. In the main setup, S1, used for tasks T1, T2, T5 and T6 (see Section 3.2) the cameras were placed as in Figure 1.

Fig. 1. a) Setup S1; b) Setup S2 - Tasks T3 and T4

Tasks T3 and T4 consist of two social games expected to trigger different types of laughter (see Section 3.2). According to the literature (e.g. [2]) face allows one distinguish between different laughter types, so we decided to record close-ups of the participants' faces. For this reason, setup S2 with cameras C1-C2 and K1-K2 placed closely to the participants faces was used for these task (see Figure 1).

3.2 Scenarios

In order to record spontaneous as well as controlled laugh reactions we asked participants to perform the following tasks: T1 - watching funny videos together, T2 - watching funny videos separately, T3 - "Yes/no" game, T4 - "Barbichette" game, T5 - pictionary game, and T6 - tongue twisters.

T1 and T2 consist of classic laughter inducing task, i.e., watching funny videos selected by experimenters and participants. Differently to other laughter corpora (e.g. [5]) in our data collection participants were not left alone; they could talk freely (e.g. comment videos) and hear each other. In more details, in task T1, all the participants, as well as a part of the technical stuff watched a 9 minute video. In T2 one participant was separated from everyone else by a curtain, which completely obscured her view of the others participants while still allowing her to hear (see Figure 2).

T3 and T4 are two social games that were expected to trigger different laughter types such as enjoyment or social laughter as well as down and up regulation.

Fig. 2. The views from cameras W3 and W4 in task T2

These were carried out in turns with all participants taking each role or competing against every other participant in the triad. In T3 one of the participants must respond quickly to questions from the other participants without saying "yes", "no" or any variation of these. The role of the other two participants is to ask questions and distract him in an attempt to provoke the use of the prohibited words. T4 is a classic French children's game that we included in order to elicit down-regulated laughter. Two participants face each other, make eye contact and hold the other's chin. The aim of the game is to avoid laughing; the one who laughs first is the one who loses. Players may do anything (talk, move, pull faces etc.) but must maintain physical and eye contact. In this scenario the third person acts as a judge and plays against the winner in the next round.

In T5 one participant drew words printed on a piece of paper extracted from an envelope. His task was to convey the word to the other participants by drawing on a large board. Each participant had 2 minutes to convey as many words as possible, each correct answer worth one point. The groups of participants competed against each other with the promise of the highest scoring group receiving a special prize.

T6 involved participants in pronouncing tongue twisters in four different languages (French, Polish, Italian, and English). One participant read the tongue twisters which were printed on a piece of paper held by another participant. The other two participants were encouraged to distract and ridicule the speaker e.g. by using a fake laughter, in order to make him laugh.

3.3 Protocol

For the purpose of the data collection it was important that participants knew each other well, so we recruited groups of friends.

Data collection consisted of recording all inter-subject interactions. Thus, there was no clear start and end of recordings between tasks. In particular we also recorded the participants between tasks. Experimenters were present most of the time in the recording room, they talked with participants and commented the events; they even participated in some of the tasks. This procedure was chosen because much of the laughter interaction occurs between the tasks, i.e. when participants comment on the previous events, or they discuss freely. We did not wanted to interrupt these spontaneous interactions.

Each person could participate only in one session. One session consisted of six tasks T1 - T6. It lasted about 2 hours and included from 31 to 51 minutes of recording. The participants were first asked to complete additional questionnaires. Next, they were fitted with mocap sensors and microphones. Additionally, three small green markers were placed on the shoulder and chest of each participant to track body movements with vision processing algorithms. The fully equipped participants can be seen in Figure 2. The session usually started with a task T1. The order of the remaining tasks was variable. At the end participants were debriefed and given small gifts.

We recorded 6 sessions with 16 participants: 4 triads & 2 dyads (groups G3 and G5), age 20 - 35; 3 females; 8 French, 2 Polish, 2 Vietnamese, 1 German, 1 Austrian, 1 Chinese and 1 Tunisian. Participants were not obliged to speak French during the sessions. They could use the language they usually speak with each other. The instructions were given in one of 3 languages: English, French or Polish.

4 Data Synchronization

For collecting, processing and synchronizing multimodal data between different signals we used the open-source Social Signal Interpretation framework[3] (SSI) [14]. SSI allows synchronization of data from different sensor devices in real time. For our extended scenario with a high number of sensor devices SSI was extended to support not only local synchronization on a single computer, but also on multiple machines. As SSI guarantees local synchronization, the multi-computer approach is realized by a host-client architecture where multiple clients wait for a host to send a start command. To ensure all network delays can be detected at a later point we first processed a clock synchronization between all machines. For our data set we found a negligible maximum delay of only a few milliseconds.

We ran up to 8 computers simultaneously. Each of the two most powerful machines, that is, machines 1 and 2, performed the recording of an Xsens motion capture suit, a Kinect sensor, an additional webcam, as well as high quality audio recordings (XS1, M1, W1, K1 and XS2, M2, W2, K2 resp.). Machine 3 performed the recording of the Animazoo motion capture suit (AZ) and machine 4 recorded both rear cameras: W3 and W4, as well as ambient audio of the room (A1). Machines 5 and 6 ran two instances of the EyesWeb platform (see Section 5.1 for further details) used to record high frequency cameras (C1,C2, 60Hz). These two instances were synchronized to start recording when the start command, sent by the host, was received. Machine 7 was attached in some of the sessions involving video playback, to also ensure that the playback position of the video watched by participants was synced with all the other data. Finally, for some of the sessions, we also recorded respiration data (RS) on machine 8, which had to

[3] `http://openssi.net`

be synchronized by hand as the used sensor device was not yet integrated in the framework.

5 Data Playback, Annotation, Segmentation and Processing

The EyesWeb XMI platform[4] 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 [15]. We developed some EyesWeb tools to play back, annotate, segment, and process the corpus data.

5.1 Playback

The playback tool, illustrated in Figure 3, allows one to select a recording session and play back the multimodal data in a synchronized way. A synchronization counter/clock signal is generated at $120Hz$. This signal provides two different information: (i) it encodes a *frame number* that is increased by 1 each time the signal is generated; (ii) it is a clock, that is, when the signal is received, the receiving module generates its output. Separate file reading modules are associated with each recorded stream: Xsens, Animazoo, cameras, respiration. In the current version of the playback tool audio is not yet supported. Each reading module receives the same number of synchronization signals but the encoded number is scaled depending on the original frame rate of the recorded stream. For example, the Animazoo data stream is recorded at $60Hz$, so the value encoded in the synchronization signal is divided by 2; the webcam video is recorded at $30Hz$, so the synchronization signal value is divided by 4.

The stream data and video frames corresponding to the synchronization signal are then read from the stream and video files and the result is a multimodal synchronized output consisting of a 3D visualization of the participants' bodies, videos corresponding to the setup camera and the respiration sensor's graph.

5.2 Annotation and Segmentation

Another tool, based on the playback tool, for annotating and segmenting the recorded sessions has been developed using EyesWeb XMI.

The annotation phase (label A in Figure 4) consists of determining the start and end frame of each laughter event in which a particular participants' behavior can be observed (e.g., at least one of the participants laughs). During this phase, the user sets up starting and ending frame number via a GUI. Each time these numbers are modified a synchronization signal is generated (as described in Section 5.1) and the corresponding video frames coming from the recorded video streams are shown, providing a feedback to the user. Once the segment's frame interval is decided by the user, the pair $(startframe, endframe)$ is stored into the session's annotation file.

[4] `http://www.eyesweb.infomus.org`

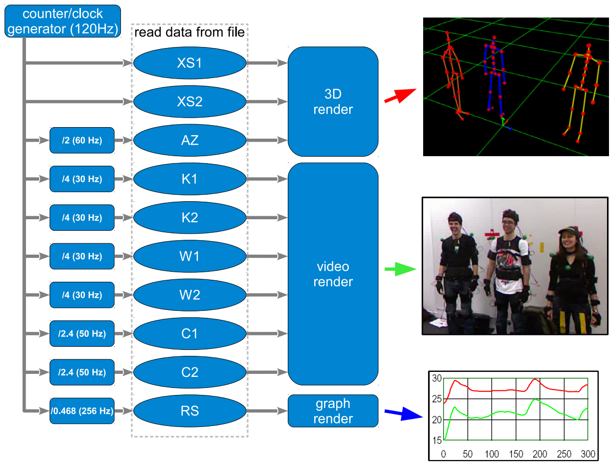

Fig. 3. The EyesWeb XMI playback tool

In the segmentation phase (label S in Figure 4), the annotation file is read and the starting and ending frame of each segment are provided as input to the counter/clock generator: that is, a sequence of consecutive frame numbers in the interval $[startframe, endframe]$ is provided as input to the playback tool. Then, the corresponding video and 3D frames coming from the recorded video and 3D streams are written to separate files.

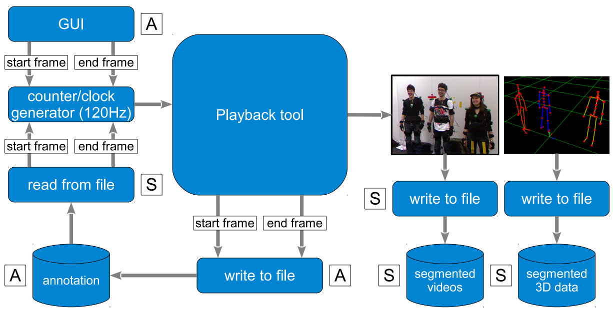

Fig. 4. The EyesWeb XMI annotation and segmentation tool

6 Results

During 6 sessions we collected nearly 4 hours and 16 minutes of data. Due to some technical problems we were not able to record 4 tasks: namely task T2 for group G1, G2 and task T6 for group G1, G3. We annotated *laughter events* and *laughter episodes*. A *laughter event* is a time interval in which at least one of the participants laughs. Annotation of laughter events is useful to analyze the overall laughter dynamics and to measure how successful respective task was in producing the laugh. A *laughter episode* corresponds to a single laugh generated by one participant. Thus one laughter event can be composed of several laughter episodes that correspond to different people that laugh at the same time. The annotation contains the following information: participants' task (T1-T6), start and end time of laughter event and camera (W1, W4, C1, C2, K1, K2) that captured the event.

We annotated 439 laughter events[5] , corresponding to 31 minutes of laughter, that is, 12% of total recording time (4 hours and 16 minutes). The rates obtained in other laughter data collections are not much different: $5-8\%$ in meeting recordings (FreeTalk, [11]), 18% in laughter inducing study (AVLC, [5]). We observed variability in the laughter frequency between the tasks (6% - 22%) and between groups (4% - 16%). The details are reported in Tables 2 and 3, and in Figure 5. Interactive games based tasks appear to elicit more laughter than laughter inducing tasks with the highest rate of laughter events in barbichette game (T4). Results of Tasks T1 and T2 cannot be compared as different videos were shown to the participants in T1 and T2. One out of two dyads, group G5, was laughing particularly rarely. The quantity of laughter in other groups is comparable.

Table 2. Laughter per group

		Laughter events			Laughter episodes		
Group	Task duration	Number of events	Duration	Percentage of laughter	Number episodes	Average duration	Std
G1	00:37:35	74	00:03:37	9.62%	91	3.09s	2.29s
G2	00:48:40	47	00:07:55	16.27%	49	8.02s	6.52s
G3	00:31:40	68	00:04:22	13.79%	76	3.93s	3.36s
G4	00:51:31	99	00.08.16	16.05%	137	4.90s	4.64s
G5	00:42:00	59	00:01:43	4.09%	58	1.89s	1.04s
G6	00:44:55	92	00:04:58	11.06%	109	3.28s	2.16s

Table 3. Laughter per task

		Laughter events			Laughter episodes		
Task	Task duration	Number of events	Duration	Peecentage of laughter	Number episodes	Average duration	Std
T1	00:56:24	83	00:05:31	9.78%	101	4.23s	5.09s
T2	00:13:45	24	00:01:48	13.09%	29	4.93s	2.88s
T3	00:43:05	86	00:04:54	11.37%	98	4.05s	3.16s
T4	00:46:52	94	00:10:21	22.08%	110	3.86s	5.79s
T5	00:58:05	64	00:03:30	6.03%	71	4.67s	3.37s
T6	00:38:10	88	00:04:47	12.53%	111	2.99s	3.22s

[5] We consider only laughs occurred when certain task was performed. Laughs between tasks were not included.

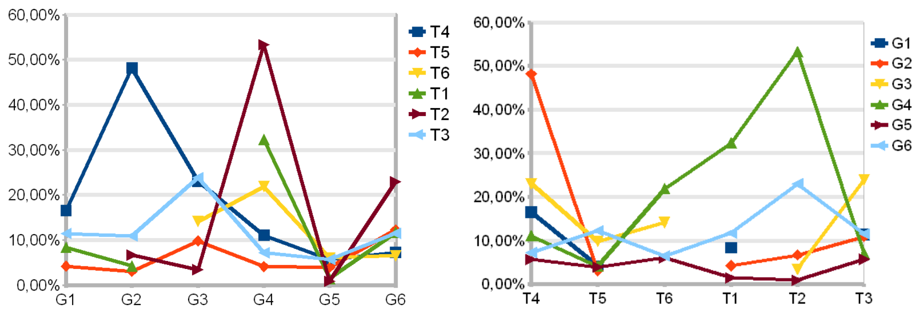

Fig. 5. Percentage of laughter events per task and group

7 Conclusion

We presented collection and analysis of the MMLI corpus. To our knowledge, this is the first corpus of this richness, dedicated to different laughter contexts, containing various data sources (mocap, audio, video, physiological), a large spectrum of captured modalities and that is synchronized across multiple participants. We proposed different scenarios that were successful in eliciting the laughter in our participants. We presented tools to synchronize and visualize the multimodal data from many cameras, various inertial mocap systems as well as Kinects. In the future, we will conclude the annotation work. Through FACS coding and perceptive studies we aim to validate whether our corpus does contain different laughter types. We will work on laughter detection from body cues as well as fusion algorithms that take into account more than one modality. We also plan to build a model of laughter mimicry and contagion in the interaction. Our corpus will be made freely available for research purposes on the ILHAIRE database website: `http://qub.ac.uk/ilhairelaughter`.

References

1. Chapman, A.: Humor and laughter in social interaction and some implications for humor research. In: McGhee, P., Goldstein, J. (eds.) Handbook of Humor Research, vol. 1, pp. 135–157 (1983)
2. Ruch, W., Ekman, P.: The expressive pattern of laughter. In: Kaszniak, A. (ed.) Emotion, Qualia and Consciousness, pp. 426–443. World Scientific Pub., Tokyo (2001)
3. Huber, T., Ruch, W.: Laughter as a uniform category? a historic analysis of different types of laughter. In: 10th Congress of the Swiss Society of Psychology, University of Zurich, Switzerland (2007)
4. Huber, T., Drack, P., Ruch, W.: Sulky and angry laughter: The search for distinct facial displays. In: Banninger-Huber, E., Peham, D. (eds.) Current and Future Perspectives in Facial Expression Research: Topics and Methodical Questions, Universitat Innsbruck, pp. 38–44 (2009)

5. 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. Journal on Multimodal User Interfaces 4(1), 47–58 (2010)
6. Petridis, S., Martinez, B., Pantic, M.: The mahnob laughter database. Image Vision Comput. 31(2), 186–202 (2013)
7. Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., Wooters, C.: The ICSI Meeting Corpus. In: Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003), vol. 1, pp. 364–367 (2003)
8. Carletta, J.: Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus. Language Resources and Evaluation 41(2), 181–190 (2007)
9. Truong, K.P., van Leeuwen, D.A.: Automatic discrimination between laughter and speech. Speech Communication 49(2), 144–158 (2007)
10. Szameitat, D.P., Alter, K., Szameitat, A.J., Wildgruber, D., Sterr, A., Darwin, C.J.: Acoustic profiles of distinct emotional expressions in laughter. Journal of The Acoustical Society of America 126 (2009)
11. Scherer, S., Schwenker, F., Campbell, N., Palm, G.: Multimodal laughter detection in natural discourses. In: Ritter, H., Sagerer, G., Dillmann, R., Buss, M. (eds.) Human Centered Robot Systems. Cognitive Systems Monographs, vol. 6, pp. 111–120. Springer, Heidelberg (2009)
12. Suarez, M.T., Cu, J., Maria, M.S.: Building a multimodal laughter database for emotion recognition. In: Calzolari, N. (ed.) Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012). European Language Resources Association (ELRA), Istanbul (2012)
13. McKeown, G., Cowie, R., Curran, W., Ruch, W., Douglas-Cowie, E.: ILHAIRE laughter database. In: Proceedings of the LREC Workshop on Corpora for Research on Emotion Sentiment and Social Signals (ES 2012). European Language Resources Association (ELRA), Istanbul (2012)
14. Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., André, E.: The Social Signal Interpretation (SSI) Framework - Multimodal Signal Processing and Recognition in Real-Time. In: Proceedings of The 21st ACM International Conference on Multimedia, Spain (2013)
15. Camurri, A., Coletta, P., Varni, G., Ghisio, S.: Developing multimodal interactive systems with EyesWeb XMI. In: Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME 2007), pp. 302–305 (2007)