The Belfast Storytelling Database

A spontaneous social interaction database with laughter focused annotation

Gary McKeown, William Curran
School of Psychology
Queen's University Belfast
Belfast, UK.
e-mail: g.mckeown@qub.ac.uk; w.curran@qub.ac.uk

Johannes Wagner, Florian Lingenfelser and Elisabeth André
Human Centered Multimedia
University of Augsburg
Augsburg, Germany.
e-mail: wagner@hcm-lab.de; lingenfelser@hcm-lab.de; andre@hcm-lab.de

Abstract—To support the endeavor of creating intelligent interfaces between computers and humans the use of training materials based on realistic human-human interactions has been recognized as a crucial task. One of the effects of the creation of these databases is an increased realization of the importance of often overlooked social signals and behaviours in organizing and orchestrating our interactions. Laughter is one of these key social signals; its importance in maintaining the smooth flow of human interaction has only recently become apparent in the embodied conversational agent domain. In turn, these realizations require training data that focus on these key social signals. This paper presents a database that is well annotated and theoretically constructed with respect to understanding laughter as it is used within human social interaction. Its construction, motivation, annotation and availability are presented in detail in this paper.

Keywords—laughter; emotion; database; conversation; social interaction

I. INTRODUCTION

The importance of laughter within conversational and social interaction has long been recognised within the Conversation Analysis tradition [1], [2]. However, it has taken many years for other academic domains interested in human social interaction to pay sufficient attention to this important social signal. Within the domain of Affective Computing laughter is a particularly important social signal, it signals positive affect [3] and social affiliation [4] and has important conversational functions that are likely to be crucial in creating more human-oriented interactions between computers and humans [5]. It serves as an important regulator of many functional features of human social interaction regulating topics and turn-taking within conversation, and it aids in the repair of conversations [6], [7]. As it is such an important social signal, it is important that data exist upon which laughter in interaction can be modelled if it is to be understood with regard to building intelligently interactive socio-communicative systems. Many of the currently available laughter databases concentrate on laughter generated by watching video clips [8], [9]–often this laughter has particularly high levels of intensity. These databases are an important part of the endeavour of understanding laughter, both laughter on its own and in certain kinds of social interactions; they are also particularly important in the synthesis of the visual and acoustic properties of high intensity laughter. However, they are less important in the goal of understanding laughter's important role within coversational interactions. Additonally, many of the corpera used by conversation analysts do look at the interactional aspects of laughter [10], [11], but they are typically auditory in nature and do not concentrate on the multimodal nature of social interactions. With these limitations in mind the Belfast Story-telling Database was created.

The Belfast Story-telling sessions were designed with the goal of capturing naturalistic audio-video laughter in quasi-natural social interactions with a variety of levels of laughter intensity. High intensity laughs have been shown to be more often related to humour [12], whereas low intensity laughs appear to have many more roles, including important conversational functions [13].

II. RECORDING SCENARIO

To meet these goals we required conversational interactions that generated laughter in reasonable quantities but at the same time did not impose too many laboratory-based constraints on the participants. Previously we had created laughter scenarios that were social in nature but more oriented towards high intensity humour associated laughter [14]. In the current circumstances we wanted the laughter to be generated in a more naturalistic social interaction setting aimed, not only at generating laughter associated with humour, but also at other social-interaction-based laughter. To create this environment we considered a story-telling setting and used the 16 enjoyable emotions induction task [15]. The task was designed to create a scenario conducive to the induction of laughter in a semi-structured story-telling environment that is not dissimilar to conversational interactions and at times becomes more discussion-like and conversational in nature. It involves participants taking turns to recount stories relating their previous experience of 16 enjoyable emotions or sensory pleasures proposed by Ekman [16], these are: Amusement, Auditory, Contentment, Ecstasy, Excitement, Fiero (pride in achievement), Gratitude, Gustatory, Olfactory, Naches (pride of a parent or mentor in the accomplishment of offspring or mentee), Elevation, Relief, Schadenfreude, Tactile, Visual, and Wonder.
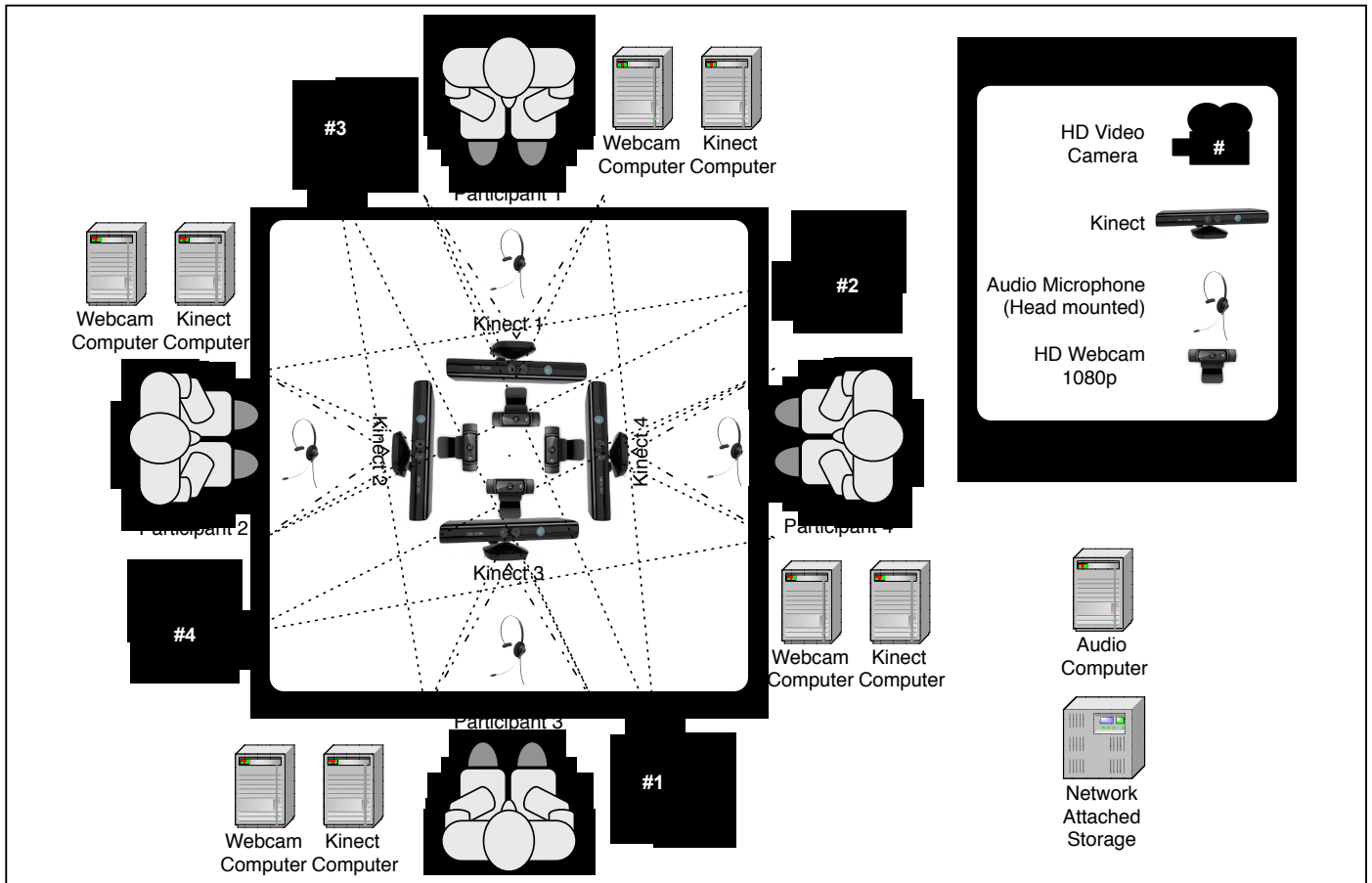
Fig. 1 Schematic diagram displaying the layout of the sensors and data capture equipment used in the Belfast Story-Telling sessions

III. PARTICIPANTS AND PROCEDURE

Native speakers of English and Spanish were recruited, and participants were filmed recounting and listening to stories in their native language. The English speakers were all from Ireland; the Spanish speaking group contained people from Spain and Latin America–the Latin Americans had all been living for several years within the European Union.

There were six sessions, three in English and three in Spanish. Four participants were recruited for each session; however, in each of the English speaking sessions only three participants took part. In contrast each Spanish-speaking session had four participants. While this was an unfortunate occurrence, the major theoretical transition in the group dynamics of laughter interactions occurs in the transition between two party interactions and multi party interactions [10]; thus laughter related phenomena should be comparable in the two groups.

Participants were provided with the list of enjoyable emotions ahead of the task and asked to prepare a story for each emotion or sensory pleasure, making brief notes to remind them of the story. Within each session the participants were seated around a table, Fig. 1 shows a schematic diagram of the system and the placement of cameras and sensors. While microphones and sensors were being adjusted each participant was asked to recite 10 phrases drawn from the TIMIT acoustic-phonetic corpus [17]. Following this the storytelling began. The order of the stories was randomised for each round of stories, and each participant took it in turns to recount their story while the other participants listened. While participants were free to ask questions during the stories, most discussion occurred at the end of the stories where there would commonly be a moment of more involved social interaction between participants. This story-telling in rounds continued until each of the participants had recounted 16 stories, one from each enjoyable emotion.

The Belfast Story-telling sessions produced a large quantity of data. The six sessions recorded 21 participants for over an hour in each session, creating a combined database duration of 25 hours and 40 mins of high quality audio and video material including both speaker and listener laughs. In addition Microsoft Kinect systems provide an additional 25 hours of motion tracking material in various forms. Information regarding the data gathering and synchronisation tools will now be detailed.

IV. DATA COLLECTION

The video was collected using Logitch HD webcams. Three or four webcams were used depending on how many participants were present. Three of the HD Webcams were Logitech c920 webcams streaming video only data to a single computer at 25fps with a resolution of 1024×576 pixels. A fourth webcam used in the situations where there were four participants was a

Logitech c900 streaming video only data to a single computer at 25fps with a resolution of 960x720 pixels. Original plans were to use the lower resolution but it was increased for the three cameras where that was possible. Further video recordings were taken by the Kinect RGB cameras producing video at 25 fps with a resolution of 640x480 compressed with the Microsoft Video 1 CRAM codec (this will play using VLC software). In addition to these we recorded sessions using HD video cameras as a backup, but we do not intend to make these recordings available as part of the database unless requested.

Audio information was gathered through the use of three or four head mounted microphones–two wired and two wireless. The two wired microphones were AKG HC-577-L condenser microphones. The two wireless microphones were Trantec HM-22 Headband Microphone connected to TOA WM-4300 wireless transmitter packs. This results in three or four mono audio channels recorded at a sound rate of 48 kHZ, with 24bit PCM quality. Files are in .wav format. Audio was additionally captured by the Kinect microphone at 16 kHZ, with 24bit PCM quality. The HD video cameras again served as an additional audio backup but again we do not intend to make these recordings available as part of the database unless requested. One audio session did not record properly, (Session 5, Participant 2); in this case the Kinect audio was substituted for the higher quality audio in the final synchronised version.

The Kinect systems (Kinect for Windows, version 1) recorded six streams of differing types of motion data for each of the participants: An Action Unit stream, a face unit stream, a head stream, two skeleton streams and a depth stream. The data collected take the form:

1. Facial Action Units (upper lip raiser, jaw lowerer, lip stretcher, brow lowerer, lip corner depressor, outer brow raiser) [18].

2. Face point tracking: tracking of 100 facial points.

3. Head poses: 3 values - Pitch, Roll and Yaw

4. Skeleton: 20 skeletal joints. These are adjusted for people close up and in the sitting position.

5. Depth: This seeks to capture movement towards and away from the camera.

The various data streams were synchronized using Social Signal Interpretation (SSI), a framework to record, analyse and recognize human behaviour in real-time [19]. Captured signals are stored on a Network Attached Storage device. A summary of the data collected during the Belfast Story-telling session can be found in Table I, Kinect information has been left out of the table but 6 streams were collected for each session and all audio is available as mono wav files.

TABLE I. AUDIO, VIDEO DATA AND PARTICIPANT DETAILS FOR THE BELFAST STORYTELLING DATABASE

Participant	Session/ Language	Video	Audio	Sex
1 (S1P1)	1-English	25fps 1024×576	48 kHZ, 24bit	Male
2 (S1P2)	1-English	25fps 1024×576	48 kHZ, 24bit	Male
3 (S1P3)	1-English	25fps 1024×576	48 kHZ, 24bit	Male
4 (S2P1)	2-Spanish	25fps 1024×576	48 kHZ, 24bit	Male
5 (S2P2)	2-Spanish	25fps 1024×576	48 kHZ, 24bit	Female
6 (S2P3)	2-Spanish	25fps 1024×576	48 kHZ, 24bit	Female
7 (S2P4)	2-Spanish	25fps 960x720	48 kHZ, 24bit	Female
8 (S3P1)	3-English	25fps 1024×576	48 kHZ, 24bit	Female
9 (S3P2)	3-English	25fps 1024×576	48 kHZ, 24bit	Male
10 (S3P3)	3-English	25fps 1024×576	48 kHZ, 24bit	Female
11 (S4P1)	4-Spanish	25fps 1024×576	48 kHZ, 24bit	Male
12 (S4P2)	4-Spanish	25fps 1024×576	48 kHZ, 24bit	Female
13 (S4P3)	4-Spanish	25fps 1024×576	48 kHZ, 24bit	Male
14 (S4P4)	4-Spanish	25fps 960x720	48 kHZ, 24bit	Female
15 (S5P1)	5-English	25fps 1024×576	48 kHZ, 24bit	Male
16 (S5P2)	5-English	25fps 1024×576	16 kHZ, 24bit	Male
17 (S5P3)	5-English	25fps 1024×576	48 kHZ, 24bit	Female
18 (S6P1)	6-Spanish	25fps 1024×576	48 kHZ, 24bit	Male
19 (S6P2)	6-Spanish	25fps 1024×576	48 kHZ, 24bit	Male
20 (S6P3)	6-Spanish	25fps 1024×576	48 kHZ, 24bit	Male
21 (S6P4)	6-Spanish	25fps 960x720	48 kHZ, 24bit	Male

To capture the data we required the use of 9 computers and a network attached storage (NAS) system. Streaming the data from a single participant required a dedicated computer for each HD webcam and Kinect, making a total of 8 computers to capture the data. The audio from each head mounted microphone was fed into a MOTU 8pre FireWire audio interface preamp, and from there into another computer with five Firewire 800 recording hard drives. These sessions generated large quantities of data, making storage and compression the major bottleneck in gathering the data. Streams were compressed using the Huffyuv lossless codec for within project storage, and then further compressed with the H.264 codec to make them available at useable sizes on the ILHAIRE laughter database. Further compression procedures were used to ensure short segmentation clips were playable, and are detailed in the annotation section. The Network Attached Storage device was a QNAP TS659 Pro II which was used to store the approximately 3 Terabytes of data generated by these sessions.

V. ANNOTATION

Annotation took place at a number of levels. Principally structural story-telling annotations, physical laughter annotations, interpretation of laughter annotations and automated laughter annotations.

A. Structural Story-telling Annotations

This annotation level resulted from the particular nature of the data collection methodology. Each participant was recorded for the duration of the session meaning that they were, at different periods in the interactions, the speaker or a listener–the speaker is defined as a participant telling a story and quite clearly holds the "floor" of the interaction. These annotations are concerned with the turn-taking elements of the sessions. Segmentation

occurs as the "floor" passes from one participant to another and the story-telling focus changes. There is often overlap between these sessions as the criterion for terminating a "floor holding session" for segmentation purposes is when the last vestiges of a facial expression associated with a story telling session are no longer visible on the face of the story teller. For example, it may take some time after a new story-teller has begun recounting their story for the smile on the face of the last story-teller to fully return to a neutral face–these segmentation decisions were made by a single certified FACS coder. Annotated segments were made into storytelling and listening audiovisual clips and are available for each participant for each storytelling emotion period; they were also further segmented to distinguish when the story had ended and a period of more interactive social conversation had become established. A distinction is made between speaker (the person telling the story) and listener (the other group members listening to the story). Audiovisual files at this level of segmentation are available on the database and may have broader interest than simply for the study of laughter.

B. Physical Laughter Annotations

The segmentation of laughter episodes involves a multi-stage process of 1) finding the laughter episode, 2) annotating the video frames at which the phenomenon begins and ends in the master file, and 3) extracting the relevant section of audio and video for further annotation purposes as well as making the video clip segments more usable in experiments and for database users. There are two gradations of laughter segmentation annotation used in the Belfast story-telling sessions; these are based on the visual aspects of laughter, in particular the facial expressions, and on the auditory laughter signals. The level of annotation is different for the Spanish and English speaking sessions due to varying constraints and goals of the ILHAIRE project when annotation was taking place. Both the Spanish and English Speaking clips have been segmented based on the visual components of a laughter episode. These annotations take their starting point as the first visual element of the laughter episode until the final visual element of the laughter episode–typically from the start of the AU12 indicated smile associated with the laughter episode until the face returns to a neutral state and AU12 is no longer visible (Based on Action Unit 12 of the Facial Action Coding System [18]). There are, as always, exceptions; on occasions the mouth is obscured and complications arise from multiple laughs in a sequence. Where there are multiple auditory laughs, segmentation aims to find the minima in intensity between laugh peaks–this level of intensity is decided by the annotator on the basis of facial expressions as auditory features of laughs are typically not present in these "between laugh" minima. The English speaking sessions have additonally been segmented based on the auditory features of a laugh, from the first audible sound associated with a laugh until the final sound associated with a laugh. This is a task made more difficult by the addition of speech; however, in most instances auditory laugh annotations should be fairly unambiguous. Laugh particles have also been annotated, with a laugh particle being defined as a very short laugh aspiration that occurs as part of speech. In these cases the laugh is almost always so short that attempting to annotate the start and finish of the laugh is too difficult; in

Story Context	English Total Laughs Duration	Laugh Clips	Spanish Total Laughs Duration	Laugh Clips
Amusement	3:07	64	1:47	22
Auditory	0:51	24	0:58	17
Contentment	1:03	28	0:56	16
Ecstasy	1:25	42	1:59	34
Elevation	1:04	26	0:35	9
Excitement	2:47	66	1:13	15
Fiero	1:46	40	1:03	18
Gratitude	1:25	38	0:20	6
Gustatory	1:47	48	0:55	15
Naches	1:12	40	0:24	7
Olfactory	1:58	50	1:01	15
Relief	1:49	58	0:23	7
Schadenfreude	3:02	72	1:32	27
Tactile	1:50	38	1:05	19
Visual	1:24	36	1:20	21
Wonder	2:14	48	0:53	14
Total	*28:44*	*718*	*16:24*	*262*

these cases the start and end segmentation annotations are broadened a little which incorporates parts of the speech within which the laugh particle occurs. The annotation is only partially based on certain FACS Action Units and the full laugh annotation protocol is available at [20].

The segmentation annotations made for the story-telling aspects of the database are used as the operational basis for the separation of listener and speaker laughter annotations. If a participant laughs while recounting a story (and therefore holds the floor) the laughs are deemed to be speaker laughs. Note that speaker laughs almost always occur within speech. If the laugh occurs while a participant is listening to a story-teller and does not hold the floor then the laugh is deemed to be a listener laugh.

Speaker and listener laughs for the English speaking sessions are segmented at both the visual and auditory levels, providing a total of 2,074 audiovisual laughter clips for the database. The Spanish clips are only currently segmented at the speaker and visual level, adding a further 262 laughs clips to the database.

There are some slightly more informal annotations that were made during the segmentation process. These are rough guides to phenomena and are not intended to be exhaustive or comprehensive. Story-telling sessions where there are examples of smile voices–where the speaker talks in a manner suggestive that a smile or laughter is going to be produced – have been highlighted [10], as have stories that contain a topic terminating laugh [7]. The number of laugh episodes annotated and cumulative amount of laughter time these clips is presented for each story-telling context in Table II. As there are at least two listeners for each story telling session there are typically more laugh segments for the listener annotations. Listener laughs for the English speaking sessions are segmented at both the visual and auditory levels, providing a total of 1356 laugh clips for the database. The number of

TABLE III. LISTENER LAUGH SEGMENTATION ANNOTATION BY STORY-TELLING CONTEXT

Story Context	English Listener Total Laughs Duration	Laugh Clips
Amusement	7:00	120
Auditory	3:41	78
Contentment	3:43	88
Ecstasy	4:26	102
Elevation	4:02	80
Excitement	4:21	100
Fiero	2:04	58
Gratitude	2:34	62
Gustatory	3:56	90
Naches	3:51	108
Olfactory	4:26	94
Relief	3:09	66
Schadenfreude	4:31	84
Tactile	3:24	76
Visual	2:00	46
Wonder	3:50	84
Outside stories	0:39	20
Total	61:37	1356

laugh episodes annotated and cumulative amount of laughter time the listener clips is presented for each story-telling context in Table III.

In total the Belfast Story-telling database currently contains audiovisual clips of 2,336 laugh episodes totaling over 106 minutes of the social signal of laughter drawn from different kinds of story-telling context.

C. Interpretation of Laughter Annotations

The segmented laughs were then placed on an online survey where annotators recruited from Amazon's Mechanical Turk rated them along a number of dimensions

As part of the ILHAIRE project numerous attempts were made to develop annotation schemes that could successfully categorize laughter. These included categorization on the basis of the story-telling context, and categorization based on laughter drawn from previous databases [21] where both categorical and dimensional schemes were drawn up. When each of these schemes was assessed inter-rater reliability was found to be at chance levels. Eventually, we realized that ambiguity may have a functional role to play in the perception of laughter; a detailed exposition of this theoretical account is provided in [13]. Subsequently a more functional annotation scheme addressing key goals of the ILHAIRE project was adopted and a number of the dimensions proved to be particularly useful–especially laughter intensity [12].

The nature of the annotations varied across the duration of the ILHAIRE project, but annotators were asked at different times to rate laughs on levels of intensity, maliciousness, benevolence, humour, their conversational nature, politeness, and whether they seemed genuine or fake. Annotations were unipolar (except for the genuine-fake dimension which was bipolar) and rated on a scale of 1 to 10. Strong correlations

TABLE IV. SPEAKER LAUGH SEGMENTATION ANNOTATION BY STORY-TELLING CONTEXT

Annotation	Number of Annotations	Number of Annotaters	Mean Rating	Stddev
Intensity	9421	748	4.21	2.32
Humor	9421	748	4.34	2.44
Maliciousness	9421	748	2.21	1.87
Benevolence	9421	748	4.8	2.54
Conversational	9421	748	6.05	2.53
Politeness	3731	38	6.07	2.29
Genuine/Fake	3731	22	5.36	2.72

were found between laugh intensity and humour ($r=0.7$, this is addressed in detail in [12]), and between laughter described as conversational and laughter described as polite ($r=0.5$). Details of these annotations are provided in Table IV. Distributions and discussions about the relationships between these variables can be seen in more detail in [12] [13].

For a small amount of the laughs we also asked raters to indicate whether they thought the laugh actually was a laugh or not, and to provide a confidence rating for this decision. The reasoning behind this assessment was that the segmentations were conducted to be inclusive; that is, they included small nasal aspirations and laugh particles that would be considered very low intensity laughs. In these cases we wished to gain some knowledge concerning the level of uncertainty on whether these social signals should be considered laughter or not.

D. Automated Laughter Annotations

It is, of course, desirable to automate some of the extensive manual work that is included in the annotation process. While full automation is currently not capable of producing the same annotation quality as a human expert, we can try to lighten the necessary workload by automatic pre-annotation of laughter occurrences, which can then be refined manually. This is achievable by means of machine learning techniques: having annotated a sufficient amount of laughter episodes manually, these labels can be used to train classification models.

TABLE V. LIST OF FEATURES EXTRACTED FROM THE AUDIO AND VIDEO CHANNELS TO DETECT AUDIBLE AND VISUAL LAUGHTER

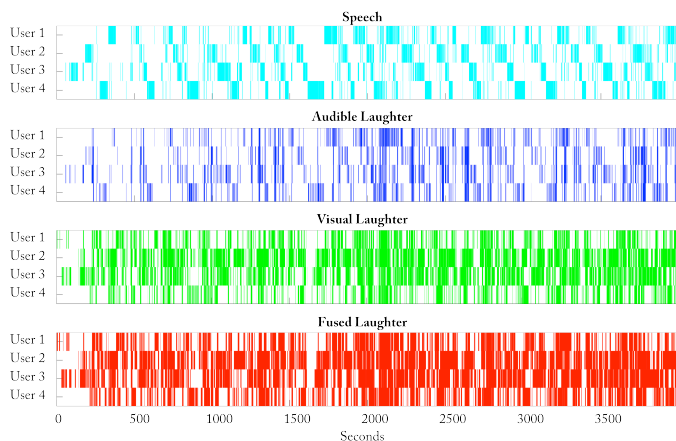| Channel	Short-term Feature	Long-term Statistics	Total
Mono Audio 48kHz	Pitch Energy MFCCs Spectral Voice Quality	Mean Median Maximum Minimum Variance Median Lower/Upper Quartile Absolute/Quartile Range	1451
Action Units 25Hz	Upper Lip Raiser Jaw Lowerer Lip Stretcher Brow Lowerer Lip Corner Depressor Outer Brow Raiser	Mean Energy Stddev Maximum Minimum Range	36

Fig. 2. Automated annotation obtained for session 6 after learning with session 1. Clear alternating periods of "floor holding" are obvious in the first panel as each participant delivers their story. The second panel shows audible laughter which follows the alternating pattern to some extent but interspersed with episodes of shared laughter. The third panel involves smiling which is commonly used as a backchannel and is more common. The fourth panel shows the final fused result.

The feature sets we use to detect audible and visual laughter are listed in Table V. For audio analysis we compute acoustic features related to the paralinguistic message of speech, which means that the features describe "how" something is said. Paralinguistic features are extracted with EmoVoice [22]. Laughter detection on the video channel is trained with 36 features, gained from statistics over the action units provided by the Microsoft Kinect. These features are used to train Support Vector Machines as classifiers for each channel. Finally, the confidence values of respective classification models are fused to gain a multimodal decision if the currently observed window contains a laugh segment or not. This decision determines if the observed window is labeled as a laughter window. A detailed description of the recognition system is found in [23]. Once trained on a subset of the corpus, it can be used to generate automated annotations for the remaining sessions. An example is shown in Figure 2.

VI. AVAILABILITY

The Belfast Story-telling Database is available as part of the ILHAIRE laughter Database, which is a meta-database that gathers together many different resources for the use of laughter researchers. Access to the database is available on completion of an End User License Agreement, which is available at the ILHAIRE laughter Database (www.qub.ac.uk/ilhairelaughter).

VII. CONCLUSION

The Belfast Storytelling database makes a useful contribution to the set of databases that exist for the purposes of understanding laughter. The understanding of laughter within more social settings was its original purpose and most of the annotation has been collected with these goals in mind. However, there is a lot of information available in these interactions that is likely to be of value in understanding other social signals and we would encourage the use of the database beyond laughter research. We would also encourage researchers to provide any further annotations derived from using the database to the authors, who will then make them available to the broader research community.

REFERENCES

[1] G. Jefferson, H. Sacks, and E. Schegloff, "Notes on laughter in the pursuit of intimacy," in *Talk and social organisation*, G. Button and J. R. E. Lee, Eds. Clevedon, England: Multilingual Matters, 1987, pp. 152–205.

[2] P. J. Glenn, "Initiating shared laughter in multi-party conversations," *Western Journal of Speech Communication*, vol. 53, no. 2, pp. 127–149, 1989.

[3] M. J. Owren and J.-A. A. Bachorowski, "The evolution of emotional expression: A' selfish-gene' account of smiling and laughter in early hominids and humans. | Michael Owren - Academia.edu," in *Emotions: Current Issues and Future Directions*, no. 5, T. J. Mayne and G. A. Bonanno, Eds. New York: The Guilford Press, 2001.

[4] M. Smoski and J.-A. A. Bachorowski, "Antiphonal laughter between friends and strangers," *Cognition and Emotion*, vol. 17, no. 2, pp. 327–340, Jan. 2003.

[5] H. J. Griffin, M. H. 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 Transactions on Affective Computing*, pp. 1–1.

[6] E. Holt, "On the nature of 'laughables': laughter as a response to overdone figurative phrases," *Pragmatics*, vol. 21, no. 3, pp. 393–410, 2011.

[7] E. Holt, "The last laugh: Shared laughter and topic termination," *Journal of Pragmatics*, vol. 42, no. 6, pp. 1513–1525, 2010.

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

[9] R. Niewiadomski, M. Mancini, T. Baur, G. Varni, H. Griffin, and M. S. H. Aung, "MMLI: Multimodal Multiperson Corpus of Laughter in Interaction," in *Human Behavior Understanding*, A. A. Salah, H. Hung, O. Aran, and H. Gunes, Eds. 2013, pp. 184–195.

[10] P. J. Glenn, *Laughter in Interaction*. Cambridge: Cambridge University Press, 2003.

[11] P. Glenn and E. Holt, Eds., *Studies of Laughter in Interaction*. London: Bloomsbury Academic, 2013.

[12] G. McKeown and W. Curran, "The Relationship Between Laughter Intensity and Perceived Humour," presented at the The 4th Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech, Enschede, Netherlands, 2015, pp. 27–29.

[13] G. McKeown, I. Sneddon, and W. Curran, "The underdetermined nature of laughter," *In preparation*, 2015.

[14] G. McKeown, W. Curran, C. McLoughlin, H. J. Griffin, and N. Bianchi-Berthouze, "Laughter Induction Techniques Suitable for Generating Motion Capture Data of Laughter Associated Body Movements," presented at the 2nd International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space (EmoSPACE) In conjunction with the IEEE Conference on Automatic Face and Gesture Recognition, 2013, pp. 1–5.

[15] J. Hofmann, F. Stoffel, A. Weber, and T. Platt, "The 16 enjoyable emotions induction task (16-EEIT)." Unpublished Research instrument, Department of Psychology, University of Zurich, Switzerland.

[16] P. Ekman, "Sixteen Enjoyable Emotions," *Emotion Researcher*, vol. 18, pp. 6–7, 2003.

[17] L. F. Lamel, R. H. Kassel, and S. Seneff, "Speech database development: Design and analysis of the acoustic-phonetic corpus," presented at the Proceedings of DARPA Speech Recognition Workshop, 1986.

[18] P. Ekman and W. V. Friesen, *Manual for the facial action coding system.* Consulting Psychologists Press, 1978.

[19] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. Andre, "The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time," presented at the the 21st ACM international conference, New York, New York, USA, 2013, pp. 831–834.

[20] H. J. Griffin, G. McKeown, G. T. Lourido, and N. Bianchi-Berthouze, "ILHAIRE Report 5.2 Model of Cross-Cultural Differences," ILHAIRE Project European Union FP7-ICT Grant No270780., 2014.

[21] G. McKeown, R. Cowie, W. Curran, W. Ruch, and E. Douglas-Cowie, "ILHAIRE Laughter Database," presented at the ES³ 2012 4th International Workshop on Corpora for Research on emotion, sentiment, & social signals at the eighth international conference on Language Resources and Evaluation (LREC), Istanbul, 2012, pp. 1–4.

[22] T. Vogt, E. Andre, and N. Bee, "EmoVoice — A framework for online recognition of emotions from voice," presented at the Proceedings of Workshop on Perception and Interactive Technologies for Speech-Based Systems, 2008.

[23] F. Lingenfelser, J. Wagner, E. Andre, G. McKeown, and W. Curran, "An Event Driven Fusion Approach for Enjoyment Recognition in Real-time," presented at the Multimedia 2014, Orlando, Florida, 2014, pp. 1–10.