4D Cardiff Conversation Database (4D CCDb): A 4D Database of Natural, Dyadic Conversations

Jason Vandeventer, Andrew J. Aubrey, Paul L. Rosin, David Marshall

School of Computer Science and Informatics
Visual Computing Group
Cardiff University
Cardiff, Wales, UK

VandeventerJM@cardiff.ac.uk, ajaubrey@3dmd.com,
Paul.Rosin@cs.cardiff.ac.uk, Dave.Marshall@cs.cardiff.ac.uk

Abstract

The 4D Cardiff Conversation Database (4D CCDb) is the first 4D (3D Video) audio-visual database containing natural conversations between pairs of people. This publicly available database contains 17 conversations which have been fully annotated for speaker and listener activity: conversational facial expressions, head motion, and verbal/non-verbal utterances. It can be accessed at *http://www.cs.cf.ac.uk/CCDb*.

In this paper we describe the data collection and annotation process. We also provide results of a baseline classification experiment distinguishing frontchannel from backchannel smiles, using 3D Active Appearance Models for feature extraction, polynomial fitting for representing the data as 4D sequences, and Support Vector Machines for classification. We believe this expression-rich, audio-visual database of natural conversations will make a useful contribution to the computer vision, affective computing, and cognitive science communities by providing raw data, features, annotations, and baseline comparisons.

Index Terms: 4D Databases, Affective Computing, Face and Gesture Recognition, Speech Analysis

1. Introduction

Face-to-face conversations are a frequent occurrence for most people and are an important part of social communication. These conversations, whether with well-known friends or complete strangers, consist of a variety of verbal and non-verbal signals (e.g., expressions, gestures) which control the tone, content, and flow of a conversation [1, 2, 3, 4].

Given the frequency and importance of these social interactions and the advances of recent technology, it is surprising that little research has focused on analysing and modelling the components of natural, human conversations. Many expression databases focus solely on the so-called *prototypical expressions*, such as anger, fear, and disgust; and not the *conversational expressions* people observe and express on a daily basis, such as agreement, thinking, and confusion [5, 6].

Some previous works have used 2D data for modelling conversational interactions [7, 8]. While 2D data is useful for some cases, 3D data offers the advantage of providing intrinsic geometry which is invariant to pose and lighting. Moreover, 3D dynamic (4D) data is preferred over 3D static data because it includes temporal information, which is very important for modelling and synthesising realistic facial expressions.

No such databases currently exist of 4D conversations, and so we have created the first 4D (3D video) database of natural, dyadic conversations. This publicly available database contains 17 minutes of natural, expression rich, dyadic conversations and was captured on two back-to-back, synchronised, 3dMD 4D (3D video) capture systems at 60 frames per second (FPS) [9]. This setup allowed for an unobstructed line-of-sight between the participants (Figure 3). Four experienced annotators annotated 17 conversations (34 sequences). Here, *sequence* is used to refer to one side of a conversation). Two annotators marked 8 conversations, while two others marked 9 conversations. Due to the amount of data and time required for capturing and processing 4D conversations (which is on the order of terabytes), this dataset is not as large as those which only capture short, specific facial expressions. However, this database allows for the first time the modelling, analysis, and synthesis of conversational interactions in 4D.

Hereafter, *expression periods* refers to specific annotated instances. The annotations consist of 764 Frontchannel/Backchannel expression periods (329 Frontchannel, 435 Backchannel. Note: Multiple annotation types can fall under the same annotated period), 433 rigid expression periods (e.g., head nod), 450 non-rigid expression periods (e.g., smiles), 305 verbal/non-verbal utterance periods, and 307 'Other' expression periods (Full List with Descriptions: 3.3.2).

A baseline experiment classifying speaker from listener smile interactions was performed to show one of the many applications the database can allow.

Understanding the nuanced expressions of conversations will allow for advances in synthesised facial expressions, deception detection, behaviour analysis, animated character interaction and modelling, etc. Thus, the data will be of interest to computer vision, affective computing, and cognitive science researchers alike. The fully annotated database, including 2D videos of the conversations so researchers can easily create their own annotations, can be accessed at *http://www.cs.cf.ac.uk/CCDb*.

The following sections are organised as follows: Section 2 covers related work, Section 3 describes the data collection and annotation process, Section 4 presents a baseline experiment performed using conversational interactions, Section 6 covers the future work that the authors would like to conduct, and Section 5 concludes the paper.

2. Background

Early work on conversational modelling focused on written transcripts of conversations. As a result, traditional models of communication assumed that in any dyadic conversation one person was active (the speaker) and one was passive (the listener). Since at least 1970, however, it has been repeatedly shown that human conversations are very much multimodal. In addition to the words chosen, it has been found that prosody, facial expressions, hand and body gestures, and gaze all convey conversational information. For example, Bridwhistell has shown that speech conveys only about one-third of the information in a conversation [10]. The rest of the information is distributed throughout a number of non-verbal semiotic channels, such as hand or facial motions [11]. It has also been shown that non-verbal information is often given a greater weight than spoken information: when the spoken message conflicts with facial expressions, the information from the face tends to dominate [12, 13].

2.1. Conversational Expressions

Once real conversations (and not just written texts) are examined, it is clear that listeners are in fact not passive. During face-to-face conversations, there is a considerable degree of communication from the listener to the speaker, which often serves to control conversational flow [1, 2, 3, 4, 14, 15, 16]. In [4], Yngve coined the term *backchannel* to describe this exchange of signals from the listener(s) to the speaker (Figure 1). This feedback can indicate comprehension (e.g., a look of confusion), provide an assessment (e.g., saying "correct"), control conversational flow, or even add new content (e.g., sentence completion). For obvious reasons, we use the term *frontchannel* to refer to the speaker's behaviour.

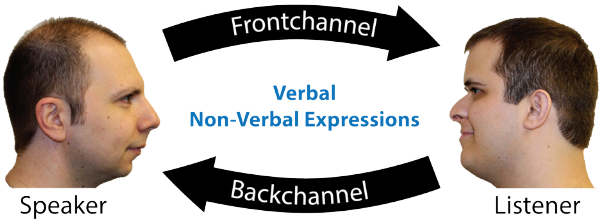

Figure 1: Backchannel signals can have a significant effect on conversational flow. They can be multimodal, including speech and facial expressions.

In most conversations the role of the speaker and listener changes from person to person throughout the conversation. One moment an individual may be the speaker and producing frontchannel expressions, while in the next moment their role has shifted to listener and their expressions are of the backchannel type. This dynamic relationship is what allows for the conversation's path to be altered based on expressed and received conversational expressions.

In order to detect conversational expressions, let alone fully model them, it is necessary to obtain and analyse real-world test data.

2.2. 3D/4D Databases

There are many 3D/4D databases of facial expressions which currently exist and a comprehensive survey of these databases can be found in [17]. Unfortunately, none of these databases contain conversations, and as a result, conversational expressions; those expressions found more commonly in everyday conversation, such as laughing, thinking, confusion, and an expression we have termed *interesting-backchannel* (Figure 2, Descriptions: 3.3.2). While these databases are potentially useful for modelling and synthesis of prototypical facial expressions, they can not be used for our purposes of creating coupled models of conversational expressions.

Figure 2: Conversational Expressions: Laugh, Thinking, Confusion, and Interesting-Backchannel

2.3. Conversational Databases

While some conversational databases exist (e.g., [18, 19, 20, 21, 22, 23]), the general lack of interaction between participants, poor visibility of the face, and lack of 4D data, make these unsuitable for our research. In [18], pre-defined speaker/listener roles are assigned, which constrains the naturalness of the conversation. In [19], one side of the conversation contains an operator-controlled synthesised face. In [20, 22] the subjects are often too far from the camera for the face to be visible. Finally, the works of [21, 23] focus more on the gestures and body movement than the facial expressions of the individuals in the conversations.

It is for these reasons we found it necessary to create our own 4D (3D video) database of natural, dyadic conversations.

3. Database

This paper builds on the previous work of the 2D Cardiff Conversation Database (CCDb) [24]. The 2D CCDb contains 30 videos of 2D annotated, natural dyadic conversations. In this paper we present a new multimodal 4D database of natural conversations, designed specifically to allow analysis, modelling, and synthesis of frontchannel/backchannel signals, conversational facial expressions, head motion, and verbal/non-verbal utterances.

The database presented here contains *natural* conversations. While it was collected in a laboratory, the participants had free rein to discuss whatever subject they wished; the conversations were *not* scripted. Furthermore, the participants did not act in a simulated manner, nor were they prescribed roles to fulfil (i.e. a participant is not given the role of speaker or listener). The conversations were driven by the participant's knowledge (or lack) of the discussion subject, which led to spontaneous behaviour. No equipment was altered between the recording sessions, with the exception of the chair height to ensure the participant's head was clearly visible to the cameras.

3.1. System Setup

Two synchronised, 4D (3D video) 3dMD, capture systems were used for data acquisition (Figure 3). Each system consists of

7 cameras: 4 monochrome and 3 colour. These cameras are 2 megapixel with gigabit Ethernet interfaces, have a resolution of 1200×1600, a bit depth of 14-bit (mono) and 12-bit (colour), and a capture frame rate of 60 FPS. The systems use active stereo to create the 3D models for each frame and the geometric model for each frame typically consists of 30,000 vertices. To capture the speech of each subject a lapel microphone is worn by each speaker. The audio was recorded at 44.1 KHz.

Figure 3: 3dMD Synchronised 4D Capture Systems

3.2. Capturing Process

There were four volunteer participants: two male and two female, all Caucasian, ranging from approximately 20 to 50 years of age (Figure 4). They were recruited from the general public and had no tie to the lab performing the data acquisition. Three of these volunteers had a background in acting, although for the purpose of our experiment we specified they should act naturally. Participants with acting backgrounds were recruited because two separate data captures were occurring on the day. The other experiment required directed and controlled facial expressions and was done after the conversation captures. The participants were unaware of the specifics of the research and only given details at the conclusion of each experiment.

Figure 4: 3D Capture Examples of the 4 Participants

There were 17 one-minute conversation captures (34 sequences) captured over the span of 2 hours There were six capture sessions consisting of each pair of the four subjects. For each session there were three expression-rich, one-minute captures made. The 3dMD systems were placed back-to-back to allow the subjects to sit face-to-face with an unobstructed line-of-sight (Figure 3). Participants were given roughly a minute before each conversation capture session to allow them to become comfortable with the environment and the other participant. They were given an indication of when recording began. As stated above, to ensure natural conversations, the participants were not guided nor given topics to discuss. The main topics they tended to discuss were hobbies, films and television shows, and travel experiences. The participants were swapped after each capture session to allow them resting time in-between sessions, as well as to ensure they were captured on both systems.

3.3. Database Contents

3.3.1. 3D Frames

The database consists of 17 one-minute, conversation captures (34 Sequences). Therefore, each sequence consists of approximately 3500-4000 frames, with 7 camera images for each frame: 4 mono and 3 colour (Figure 5). The 7 images are used with the camera calibration information to create 3D frames. The frames are 3D surface object OBJs, with a 3-image texture map (BMP) (Figure 6, Left). Each OBJ consists of approximately 30,000 vertices, normals, and texture coordinates; and 55,000 faces (polygons). The total size per 3D frame (OBJ and BMP) is typically around 20 MB. A *cleaned* OBJ is then produced using an in-lab tool which removes non-manifold vertices and edges, isolated vertices, and small components, and then produces a unified (single-image) texture map (PNG) (Figure 6, Right). The total size of the new OBJ and PNG is typically around 4.5 MB. Aside from taking up much less space, the single-image texture map resolves texture uv-coordinate issues researchers will have when they make certain modifications to the original 3D object, such as tracking non-vertex feature points through a sequence. In that specific case, new uv texture coordinates will often be located in separate images of the 3-image texture map, resulting in errant texture patches for the affected faces. It is for this reason that researchers wishing to track features or manipulate the 3D objects will want to use the cleaned OBJ data. Other researchers may be happy with the originally captured data. Therefore, both the original and cleaned OBJ data have been made available to the research community.

Figure 5: Mono and Colour Image Examples of a Single Frame

3.3.2. Annotations

Manual annotation of the sequences was carried out in ELAN (Figure 7) [25]. ELAN is a publicly available, easy to use software tool that allows for multiple annotation tracks and hierarchical tracks. It also allows for time-accurate text annotation of speech sections. A variety of facial expressions and head motion (e.g., nodding) were annotated. For the database, two trained annotators were used for each sequence. The annotators were instructed to mark a backchannel signal

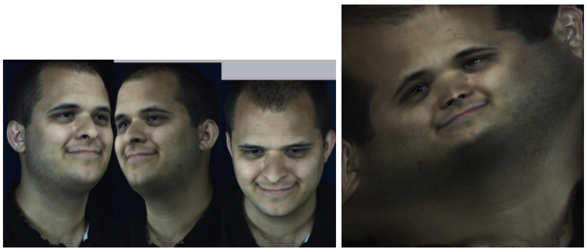

Figure 6: 3-Image and Unified Texture Maps

as any expression or gesture made in response to verbal or non-verbal action from the speaker. These backchannel signals can occur during or after the action. The annotation tracks included are (based on those discussed in [6]):

- **Frontchannel:** Main speaker periods.
- **Backchannel:** Expressions, gestures, and utterances that would be classified as backchannels.
- **Agree:** Up/down rigid head motion and/or vocalisation (e.g., "yeah").
- **Disagree:** Left/right rigid head motion and/or vocalisation (e.g., "no").
- **Utterance:** The periods of speaker activity, including all verbal and non-verbal activity.
  - **Verbal:** Whole or partial words spoken.
  - **Non-Verbal:** Verbal fillers (e.g., "umm", "err") and other non-verbal sounds (e.g., "uh-huh")
- **Happy:** Smile or laugh.
  - **Smile:** Lip corners move upwards.
  - **Laugh:** Spontaneous smile and sound.
- **Interesting-Backchannel:** Eyebrows slightly raised, lip corners move downwards, slight head nod.
- **Surprise-Positive:** Mouth opening and/or raised eyebrows and/or widening of eyes. Upward motion of lip corners.
- **Surprise-Negative:** Mouth opening and/or raised eyebrows and/or widening of eyes. Wrinkled brow. Downward and/or backward-pull of mouth corners.
- **Thinking:** Eye gaze goes up and left/right.
- **Confusion:** Slight squint of the eyes, eyebrows move towards each other.
- **Head Nodding:** Up/down rigid head motion. This can be agreement or motion made during speech.
- **Head Shake:** Left/right rigid head motion. This can be disagreement or motion made during speech.
- **Head Tilt:** In plane left/right rotation of the head.
- **Other:** Expressions not included in the list, but are interesting, such as consistent facial mannerisms of an individual

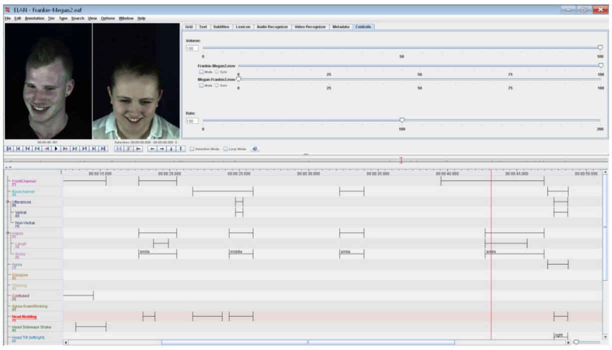

Figure 7: Screenshot of ELAN Software

4. Experiment

In normal everyday conversations, especially involving people with whom we are unfamiliar, it is common to project a friendly demeanour. This is most commonly achieved through the smile expression [26]. Whether due to the conversation topic, expression mimicry, or some other factor, this expression often produces a more comfortable feeling for the individuals in the conversation, as people tend to feel more comfortable when individuals around them reflect their own emotional state. It is unsurprising then that the smile expression is, by far, the most frequent conversational expression annotated in our dataset. Given its importance in conversations, and frequency, smile interactions were chosen for use in our classification experiment. This provides a baseline for comparison.

Using the annotated dataset described in 3.3.2, interactions consisting of a frontchannel (FC) smile expression with a corresponding backchannel (BC) smile expression, within 2 seconds, were selected (Figure 8). This resulted in 22 conversation interactions (44 sequences), which were 4D tracked and inter-subject registered using an in-lab developed approach. An example interaction can be seen in the supplementary materials (ConvoInteraction.mp4).

4.0.3. *Classification Methodology*

In this experiment we attempted to differentiate frontchannel from backchannel smile sequences, using 3D AAMs for feature extraction, polynomial fitting for 4D sequence representation, and Support Vector Machines (SVMs) for classifying the 4D sequences.

Figure 8: Screenshot of Subjects in Conversation - Smile Exchanges

For each subject, Sub_{target}, a 3D Active Appearance Model (AAM) was built using all sequence frames from every other subject, Sub_{others} [27]. 95% of the eigenenergy was kept. For each sequence, bVectors (feature vectors) were calculated by projecting every frame into the AAM. These bVectors describe the shape and texture features for each projected frame.

An n^{th} degree polynomial fit was performed on each sequence of bVectors, for each principal component (Figure 9). A grid search was performed to empirically find an appropriate polynomial degree and number of principal components to use for fitting, for each Sub_{target} AAM model. In the resulting polynomial equation, the coefficients make up the feature vector which is used as input into a Support Vector Machine (SVM) classifier. The main strength of this approach is that it allows sequences of different lengths and characteristics to be represented by the same number of values, which makes subsequent processing (classification, modelling, etc.) more straightforward. As a result of this fitting process, a sequence made up of discrete 3D frames is now represented by a single, continuous, multivariate function.

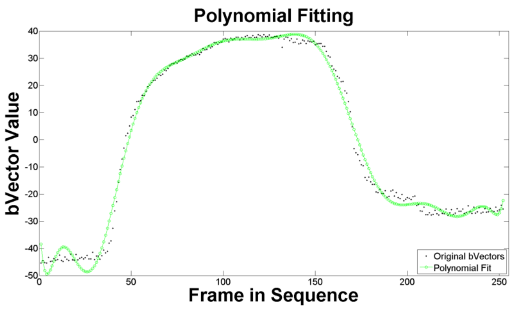

Figure 9: Example of a Polynomial Fit on bVector Sequence Data

In the SVM classifier (*libSVM* [28]) Sub_{others} sequences comprised the training set and Sub_{target} sequences comprised the testing set. A ν-SVM with a Gaussian RBF kernel was used, and a grid search was performed for parameter optimisation, as suggested in [29, 30]. As stated above, these steps were performed for each subject, so as to provide a fully-generalised approach to classification.

4.0.4. Results and Analysis

For classification accuracy, Area Under the ROC Curve (AUC) was chosen as the performance metric because it has been shown to be more reliable and contain more preferable properties than raw classification accuracy, as described in [31, 32, 33]. The average accuracy for all four subjects was 97.54%. Details of the scores, polynomial degrees, number of principal components used, and confusion matrices for each subject can be found in the supplementary materials (ClassificationResultDetails.pdf).

This experiment was able to validate two main points. First, frontchannel and backchannel signals contain characteristics which allow them to be differentiated; this is most likely the vertical movement of the mouth of the speaker (frontchannel signal). Second, the results support the idea of using this database for modelling, analysing, and synthesising conversational interactions.

5. Conclusion

In this paper we presented the first 4D database of natural, dyadic conversations. This publicly available database consists of 17 minutes of expression rich conversations, and manual annotations of frontchannel and backchannel signals, which include conversational facial expressions, head motion, and verbal/non-verbal utterances. A baseline experiment classifying frontchannel and backchannel smile interactions was performed. The results showed a respectable 97.54% classification accuracy across subjects.

Due to the amount of data and time required for capturing and processing 4D conversations (raw data, OBJ data, cleaned data) this dataset is not as large as those which only capture short, specific facial expressions. However, this database allows for the first time the modelling, analysis, and synthesis of conversational interactions in 4D, and once we and the research community better understand the characteristics of interesting conversations, we can capture more data for other uses.

The full database includes the original 3D frames, cleaned 3D frames, manual annotations, and 2D videos of the conversations, and can be accessed at *http://www.cs.cf.ac.uk/CCDb*. It is the hope of the authors that the research community will use this database of 4D conversations to further research in computer vision, affective computing, cognitive science, and related fields.

6. Future Work

While the authors are excited to see what the greater community can produce from this database, our work will continue with building 4D models of appearance, specifically of conversational expressions. Analysis of conversation roles, the effect of conversational expression mimicry, and perceptual experiments using synthesised expressions are just some of the research topics that will be explored using this database.

7. Acknowledgements

The authors would like to thank Ben Barbour, Lukas Gräser, and Abhishek Sunkari for their assistance in the data annotation process, and Dr. Job van der Schalk for recruiting the participants for the data capture.

8. References

[1] J. Bavelas, L. Coates, and T. Johnson, "Listeners as co-narrators," *J Pers Soc Psychol*, vol. 79, no. 6, pp. 941–52, 2000.

[2] P. Bull, "State of the art: Nonverbal communication," *The Psychologist*, vol. 14, pp. 644–647, 2001.

[3] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan, "More than just a pretty face: conversational protocols and the affordances of embodiment," *Knowledge-Based Systems*, vol. 14, no. 1, pp. 55–64, 2001.

[4] V. Yngve, "On getting a word in edgewise," in *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, 1970, pp. 567–578.

[5] K. Kaulard, D. W. Cunningham, H. H. Bülthoff, and C. Wallraven, "The MPI facial expression database – a validated database of emotional and conversational facial expressions," *PloS One*, vol. 7, no. 3, p. e32321, 2012.

[6] M. Nusseck, D. W. Cunningham, C. Wallraven, and H. H. Bülthoff, "The contribution of different facial regions to the recognition of conversational expressions," *Journal of Vision*, vol. 8, no. 8, pp. 1–23, 2008. [Online]. Available: http://www.journalofvision.org/content/8/8/1.abstract

[7] D. Hogg, N. Johnson, R. Morris, D. Buesching, and A. Galata, "Visual models of interaction," in *2nd International Workshop on Cooperative Distributed Vision*, 1998.

[8] A. J. Aubrey, D. W. Cunningham, D. Marshall, P. L. Rosin, A. Shin, and C. Wallraven, "The face speaks: Contextual and temporal sensitivity to backchannel responses." in *ACCV Workshop on Face analysis: The intersection of computer vision and human perception*, 2012, pp. 248–259.

[9] 3dMD http://www.3dmd.com.

[10] R. L. Birdwhistell, *Kinesics and context: Essays on body motion communication*. University of Pennsylvania press, 2010.

[11] J. de Ruiter, S. Rossignol, L. Vuurpijl, D. Cunningham, and W. J. Levelt, "SLOT: A research platform for investigating multimodal communication," *Behavior Research Methods, Instruments, & Computers*, vol. 35, no. 3, pp. 408–419, 2003.

[12] P. Carrera-Levillain and J. Fernandez-Dols, "Neutral faces in context: Their emotional meaning and their function," *Journal of Nonverbal Behavior*, vol. 18, pp. 281–299, 1994.

[13] A. Mehrabian and S. Ferris, "Inference of attitudes from nonverbal communication in two channels," *Journal of Consulting Psychology*, vol. 31, pp. 248–252, 1967.

[14] J. B. Bavelas, A. Black, C. R. Lemery, and J. Mullett, "I show how you feel – motor mimicry as a communicative act," *Journal of Personality and Social Psychology*, vol. 59, pp. 322–329, 1986.

[15] J. Cassell and K. R. Thorisson, "The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents," *Applied Artificial Intelligence*, vol. 13, pp. 519–538, 1999.

[16] R. Vertegaal, "Conversational awareness in multiparty VMC," in *"Extended Abstracts of CHI'97"*. Atlanta: ACM, 1997, pp. 496–503.

[17] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3d facial expression recognition: A comprehensive survey," *Image Vision Comput.*, vol. 30, no. 10, pp. 683–697, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2012.06.005

[18] I. de Kok and D. Heylen, "The MultiLis corpus - dealing with individual differences in nonverbal listening behavior," in *Third COST 2102 International Training School*, A. Esposito, R. Martone, V. Müller, and G. Scarpetta, Eds., vol. 6456. Springer Verlag, 2011, pp. 362–375.

[19] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schröder, "The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," *Affective Computing, IEEE Transactions on*, vol. 3, no. 1, pp. 5–17, 2012.

[20] L.-P. Morency, I. de Kok, and J. Gratch, "Predicting listener backchannels: A probabilistic multimodal approach," in *Intelligent Virtual Agents*. Springer, 2008, pp. 176–190.

[21] C. Oertel, F. Cummins, J. Edlund, P. Wagner, and N. Campbell, "D64–a corpus of richly recorded conversational interaction," *Journal of Multimodal User Interfaces*, vol. 7, no. 1-2, pp. 19–28, 2013. [Online]. Available: http://www.springerlink.com/openurl.asp?genre=article\&id=doi:10.1007/s12193-012-0108-6

[22] A. Vinciarelli, A. Dielmann, S. Favre, and H. Salamin, "Canal9: A database of political debates for analysis of social interactions," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pp. 1–4.

[23] E. P. Volkova, B. J. Mohler, T. J. Dodds, J. Tesch, and H. H. Blthoff, "Emotion categorization of body expressions in narrative scenarios," *Frontiers in Psychology*, vol. 5, 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25071623

[24] A. J. Aubrey, D. Marshall, P. L. Rosin, J. Vandeventer, D. W. Cunningham, and C. Wallraven, "Cardiff conversation database (CCDb): A database of natural dyadic conversations," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 277–282. [Online]. Available: http://dx.doi.org/10.1109/CVPRW.2013.48

[25] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "ELAN: a professional framework for multimodality research." *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 2006, software available at http://tla.mpi.nl/tools/tla-tools/elan/.

[26] J. B. Bavelas and N. Chovil, "Faces in dialogue," in *The psychology of facial expression*, J. A. Russell and J. M. Fernández-Dols, Eds. Cambridge University Press, 1997, pp. 334–346, Cambridge Books Online. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511659911.017

[27] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[28] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011. [Online]. Available: http://doi.acm.org/10.1145/1961189.1961199

[29] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Department of Computer Science, National Taiwan University, Tech. Rep., 2003. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers.html

[30] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000. [Online]. Available: http://dx.doi.org/10.1162/089976600300015565

[31] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recogn.*, vol. 30, no. 7, pp. 1145–1159, 1997. [Online]. Available: http://dx.doi.org/10.1016/S0031-3203(96)00142-2

[32] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully automatic facial action recognition in spontaneous behavior," pp. 223–230, 2006.

[33] C. X. Ling, J. Huang, and H. Zhang, "AUC: A statistically consistent and more discriminating measure than accuracy," in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003, pp. 519–524. [Online]. Available: http://dl.acm.org/citation.cfm?id=1630659.1630736