Towards 4D Coupled Models of Conversational Facial Expression Interactions

Jason Vandeventer[1]
VandeventerJM@Cardiff.ac.uk
Lukas Gräser[3]
Lukas.Graeser@Campus.tu-berlin.de
Magdalena Rychlowska[2]
Rychlowska@Cardiff.ac.uk
Paul L. Rosin[1]
RosinPL@Cardiff.ac.uk
David Marshall[1]
MarshallAD@Cardiff.ac.uk

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

[2] School of Psychology
Cardiff University
Cardiff, Wales, UK

[3] School of Electrical Engineering and Computer Sciences
Berlin Institute of Technology
Berlin, Germany

Abstract: In this paper we introduce a novel approach for building 4D coupled statistical models of conversational facial expression interactions. To build these coupled models we use 3D AAMs for feature extraction, 4D polynomial fitting for sequence representation, and concatenated feature vectors of frontchannel-backchannel interactions. Using a coupled model of conversation smile interactions, we predicted each sequence's backchannel signal. In a subsequent experiment, human observers rated predicted backchannel sequences as highly similar to the originals. Our results demonstrate the usefulness of coupled models as powerful tools to analyse and synthesise key aspects of conversational interactions, including conversation timings, backchannel responses to frontchannel signals, and the spatial and temporal dynamics of conversational facial expression interactions.

Methodology: Using a 4D database of natural, dyadic conversations [3], conversational interactions were manually annotated for conversational expressions. The sequences were tracked using a 4D sparse tracking approach, which uses 3D shape and texture.These tracked points are used as control points in a dense correspondence method. This method uses a Thin Plate Spline (TPS) based algorithm, with an additional "snapping" step, to modify the geometry of one mesh (reference mesh) so that it matches that of another mesh (target mesh). The tracking and inter-subject registration methods were developed in-lab and details for these approaches can be found in [2]. Statistical modelling of these sequences was performed using 3D Active Appearance Models (AAMs) [1] and a polynomial regression technique for sequence representation.

Experiments: In Experiment 1, individual sequences were classified as either frontchannel or backchannel. In Experiment 2, these sequences were also modified and used in a perceptual experiment that evaluated the realism of the synthesised sequences. For Experiments 3 and 4, a coupled statistical model of conversation interactions was built by concatenating the frontchannel sequence and corresponding backchannel sequence feature vectors (Table 1).

	Sequences			
FC	PC 1	PC 1	PC 1	PC 1
	PC 2	PC 2	PC 2	PC 2
	PC 3	PC 3	PC 3	PC 3
	Offset	Offset	Offset	Offset
BC	PC 1	PC 1	PC 1	PC 1
	PC 2	PC 2	PC 2	PC 2
	PC 3	PC 3	PC 3	PC 3

Table 1: Example of the coupled model's feature vectors.

The *offset* value is the number of frames between the beginning of the frontchannel expression and the beginning of the backchannel expressions. In this work, we used the coupled model to predict the characteristics of one side of the interaction given the information about the other side. In Experiment 3, the predicted sequences were used in a classification experiment which used the original sequences for training. In Experiment 4, predicted backchannel sequences were synthesised and used in a perceptual study, where human observers evaluated the similarity between the original and predicted sequences.

Experiment Results Overview:

- **Experiment 1:** Frontchannel and backchannel sequences were classified with 97.54% accuracy.
- **Experiment 2:** Modified sequences were perceived as realistic, for both expression realism and image quality.
  - Figure 1 shows the realism ratings for each modification level. Any values above the red line (realism midpoint) represent stimuli that were perceived as realistic.

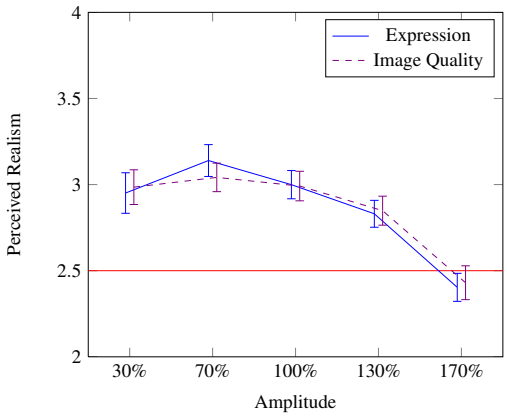

Figure 1: Experiment 2 Results

- **Experiment 3:** Predicted frontchannel and backchannel sequences were classified with 95.45% accuracy.
- **Experiment 4:** Similarity ratings, averaged within participants, were significantly higher than the scale midpoint (2.5), $M = 2.90$, $SD = 0.31$, $t(27) = 7.00$, $p < .001$, suggesting that participants consistently perceived the predicted videos as similar to the original versions.

Conclusion: We introduce a novel method for modelling 4D conversation facial expression interactions. This approach allows for the synthesis of highly-realistic expression sequences, which when modified are still perceived as realistic. Our techniques also allow for the analysis of conversation interactions, as well as the prediction and synthesis of interaction characteristics. The methods described in our paper are supported by the results of two classification experiments and two perceptual studies. Coupled statistical models of conversational interactions will allow for advances in many areas, such as behaviour analysis, perceptual psychology, and modelling and synthesising facial expressions of digital characters.

[1] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. *Pattern Analysis & Machine Intelligence (PAMI), IEEE Transactions on*, 23(6):681–685, 2001.

[2] Lukas Gräser, Jason Vandeventer, Job van der Schalk, Paul L. Rosin, and David Marshall. 4D tracking and inter-subject registration for the synthesis of realistic facial expression sequences. *Under Review*, 2015.

[3] Jason Vandeventer, Andrew J. Aubrey, Paul L. Rosin, and David Marshall. 4D Cardiff Conversation Database (4D CCDb): A 4D database of natural, dyadic conversations. In *Proceedings of the 1st Joint Conference on Facial Analysis, Animation and Auditory-Visual Speech Processing (FAAVSP 2015)*, 2015.