

# Perception of Facial Emotion in Video Clips for Supporting Connected Presence

Jeremy Thorne, BT Group CTO  
 {jeremy.thorne@bt.com}

## Abstract

*This paper reports the results of an experiment into the human perception of emotional Arousal and Valence in short video clips showing close-ups of the human face. The results suggest that motion is more important than colour for the perception of Arousal. Emotion perception is important in the exchange of messages that build Connected Presence.*

## 1. Introduction

We report the results on an experiment into how the perception of facial emotion as expressed in short video clips is affected by manipulations of those clips.

We can envisage systems that seek to support Connected Presence through pairs of linked displays and cameras. As these displays are glanced at, the viewer is able to gain a ‘thin slice’ [5] from the life of a significant another. Emotional expression, perception and response are important factors in this connection, and motion has been shown to be important for perception of information from ‘thin slices’.

We show that there is evidence to support the recommendation that displays used in Connected Presence [1] type situations, where background awareness supports an ongoing sense of connectedness [2] [3], should prioritise motion over colour and resolution. This also has relevance for the design of affective displays such as avatars or virtual agents [4], and for display technologies.

The competing viewpoint which this paper seeks to argue against is the drive for ever higher resolution, colour depth and fidelity of the still image. While these factors are important, for Connected Presence they are of less value than fidelity of motion.

To meet space constraints this poster-paper is necessarily very brief and readers are encouraged to contact the author for further details.

### 1.1 Related work and novelty

In the closest prior experiment to that described in this paper, Ehrlich et al. [6] performed an experiment to investigate the importance of motion. The face of an actor performing transitions from neutral to one of six basic emotions was filmed. Emotional intensity was rated higher for the moving clips than for stills taken from these clips. Motion also increased the perceived intensity in images where the video had been degraded to a sketch-like presentation.

This paper seeks to extend the Ehrlich’s experiment with a wider range of stimuli that do not fall neatly into categories,

with varying dynamic directions, natural emotions expressed by real people not actors, and a novel ‘comic like’ display form [7].

## 2 Stimuli and procedure

An automatic face tracker was used to detect faces within a single episode of Big Brother. The videos were cropped and scaled around the detected face, the audio track was removed and clips were trimmed to two seconds long. From the 73 clips, 5 versions of each were created (Fig 1).

1. *Movie* - the original clip.
2. *Still* - the last frame from the Movie, displayed for the entire duration of the original clip.
3. *Sketch* - each frame of the movie is passed through a sketch algorithm that converts it to a binary black and white image.
4. *Still-sketch* - the last frame of the movie sketched.
5. *Comic* - the first and last frames of the clip are sketched, and displayed side by side.



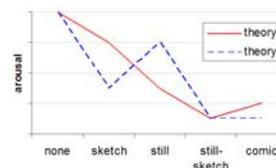
**Figure 1 examples of 'Still', and 'Still-Sketch' display conditions**

Clips are displayed with a field of view of 5-8°. After display, participants are asked to rate the emotion described in the previous clip on the 2 dimensions of emotional Arousal and Valence (unpleasant-pleasant) by clicking or dragging a marker around a box. After 25 training clips, participants proceed at their own pace through the remaining 340 clips.

## 3 Hypothesis

Theory 1: we expect highest Arousal to be rated for full colour original clips, losing motion to have a greater affect than losing colour. A two frame comic-like arrangement will lead people to re-imagine the motion and increase the perceived Arousal. We expect a similar pattern for Valence.

Theory 2: the competing pattern that would be expected if colour and resolution were more important (See Fig 2).



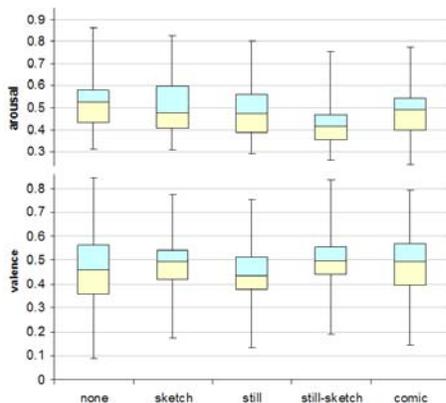
**Figure 2 Expected patterns of results**

### 4 Results

6 female and 4 male participants took part. It took participants about an hour to complete the experiment.

The results are summarised in Figure 3. Treating Arousal and Valence separately, we calculate the mean rating (across participants) for each version of each clip and perform a standard “repeated measures contrast for one group with two patterns” [8] between the two theories.

Arousal  $t(df=67) = 1.82, p=0.04, rcontrast = 0.22$   
 Valence  $t(df=67) = 3.06, p=0.002, rcontrast = 0.35$



**Figure 3 box plots (quartiles) of ratings averaged across participants for each condition. Top Arousal, bottom Valence.**

### 5 Discussion

The consumer electronics industry constantly pushes for ever higher resolutions and sizes for digital cameras and TVs. Previous Presence research [9] has also shown how bigger displays that occupy a greater field of view, lead to a greater sense of Spatial Presence. However this argument may not hold true for Connected Presence.

The contrasts for both Arousal and Valence are statistically significant. That is, the ratings of perceived emotion given by the participants more closely follow ‘Theory 1’ than ‘Theory 2’. This gives evidence to support the hypothesis that motion is more important than colour and resolution in representing emotion, though the effect sizes are not high. Considering Arousal only, this result might suggest that a moving, black and white sketch will be more effective at supporting a relationship than a colour still image; and that it will lead to a greater sense of Connected Presence.

The Valence result is slightly problematic; the median rating for Valence for the original clips is slightly below 0.5 (Figure 3), suggesting on average the clips were of negative emotions (bored, angry, frightened etc) and a reduction in Valence is an *increased* negativity. If we look instead at Valence in terms of errors; how often is an original clip with a positive average Valence given a negative Valence in the

other conditions or vice versa (Figure 4), sketching causes more of a drop in performance than a still image, supporting the importance of colour. Further experiments are needed to fully understand what is happening here.

sketch	still	still-sketch	comic
73.53	76.47	57.35	70.59

**Figure 4 Percentage agreement with original clip in Valence ratings**

### Conclusions

We have shown that motion is more important than colour for the perception of the level of emotional Arousal or intensity in short glances at small displays. This is an important consideration for the design of Connected Presence displays and virtual characters.

The result for emotional Valence, that is how pleasant or unpleasant the emotion is perceived to be, might however argue for colour rather than motion.

Further research should address emotion perception in more naturalistic two-way relationships, and what quantity of motion is important (is 2 fps enough or do we need 25fps?)

### Acknowledgements

This work has been supported by BT, UCL, and The EPSRC.

### References

- [1] Licoppe C., ‘Connected’ presence: the emergence of a new repertoire for managing social relationships in a changing communication technoscape (2004) Environment and Planning D: Society and Space 2004, volume 22, 135 – 156
- [2] Nardi B. A. Beyond Bandwidth: Dimensions of Connection in Interpersonal Communication. CSCW 2005
- [3] Rettie, R., (2003), Connectedness, Awareness and Social Presence, 6th International Presence Workshop, Aalborg.
- [4] Garau M. Selective Fidelity: Investigating Priorities for the Creation of Expressive Avatars in Schroeder R. and Axelsson A-S. Avatars at Work and Play Collaboration and Interaction in Shared Virtual Environments (2006)
- [5] Histology of Social Behavior: Judgmental Accuracy from Thin Slices of Behavior. In M. P. Zanna (Ed.), Advances in Experimental Social Psychology, 201-272. 2000
- [6] Ehrlich, S.M., Schiano, D.J., Sheridan, K.: Communicating facial affect: It’s not the realism, it’s the motion. CHI 2000 (2000) 252-253
- [7] McCloud, S.: Understanding Comics. Kitchen Sink Press. Northampton, Massachusetts (1993)
- [8] Furr R. M. , Rosenthal R. Repeated-Measures Contrasts for “Multiple-Pattern” Hypotheses. Psychological Methods 2003, Vol. 8, No. 3, 275–293
- [9] Hendrix, C., Woodrow B. Presence within virtual environments as a function of visual display parameters. Presence: Teleoperators and Virtual Environments. Vol. 5, no. 3, pp. 274-289. Summer 1996, 1996