

User Centred Design of Social Signals Feedback for Communication Skills Training

Monica Pereira
Brunel University London
Kingston Ln, London Uxbridge
UB8 3PH
monica.pereira@brunel.ac.uk

Federico Colecchia
Brunel University London
Kingston Ln, London Uxbridge
UB8 3PH
Federico.colecchia@brunel.ac.uk

Kate Hone
Brunel University London
Kingston Ln, London Uxbridge
UB8 3PH
kate.hone@brunel.ac.uk

Affective technologies enable the automatic recognition of human emotional expressions and non-verbal signals which play an important part in effective communication. This paper describes the use of user-centred design techniques to establish display designs suitable for feeding back recognised emotional and social signals to trainees during communication skills training. The channels of communication investigated are facial expression, hand gestures, voice emotion recognition and Pentland's 'honest signals'. Ease and understanding of the different feedback methods was assessed using System Usability Scale (SUS) and obtaining qualitative views using semi structured interviews. The SUS revealed that participants preferred consistency of feedback method over ease of understanding the method on initial exposure. The themes identified across all communication channels were *comparison to good performance, scale considerations, value of visual display, guidance, explanations and temporal behaviour*.

Display design. Affective Computing. Training. Social Signals. Communication.

1. INTRODUCTION

Communication skills are recognised as crucially important across the majority of work contexts and are particularly important for public-facing roles, especially media related work such as giving radio and TV interviews. Media skills courses are common in a range of fields and typically emphasise the importance of non-verbal cues as much as spoken content. This reflects understanding of the importance of non-verbal cues on factors such as impression management, trust and persuasiveness; indeed sometimes non-verbal signals are the most important part of the message (Knapp, Hall, & Horgan, 2013).

Automatic recognition of human emotional expression and social signals has the potential to enhance communication skills training by providing trainees with objective feedback on their non-verbal behaviour. In the short term this could supplement the feedback typically provided by a communication skills trainer; in the longer term it could support the development of automated training systems or avatars.

This paper considers the specific question of how information on recognised emotional and social signals behaviour is best fed back to a trainee in the context of media skills training. It describes a user-centred design process comparing alternative feedback designs and seeking user views on the formats that they would find most easy to understand and act upon in order to improve their communication skills.

2. BACKGROUND

Social signals are defined as emotionally informative non-verbal signals that directly or indirectly provide meaning around social interactions (Poggi & D'errico, 2011). These signals are defined as a perceivable stimulus produced by a sender and recognised by a receiver (Fiske, 2010). They are communicated in different, non-verbal, channels; intonations, gestures, prosody (Kendon, 2004; McNeill, 1992), posture (Condon & Ogston, 1966), head movements (Cerrato, 2005), facial expression (Ekman & Friesen, 1972), gaze (Marzillier, 1976) as well as physical contact and spatial behaviour (Hall, 1969).

Increasingly, off-the-shelf tools are becoming available to automatically detect social signals. These have been developed with a range of different types of display designs for a range of intended user groups (for example researchers, customer service personnel, end users). Some examples are described further below (see section 3.2). However, there is limited research exploring the usability of such interfaces and none which collectively considers feedback across a broad range of different types of emotional and social signals.

There are several elements to consider when developing visual displays in a learning context. The most important one is a requirement not to increase cognitive load, as this can result in divided attention and taxation of working memory (Vicente & Rasmussen, 2014). Suggestions have been made to overcome this, one of which is to map

multiple process variables into a single geometric form that provides high-level visual properties, such as symmetry, i.e an octagon (Bennett & Flach, 1992). This utilises the humans' perceptual capability of processing of symmetrical patterns. However, the method of visual display is dependent upon the context in which the display is utilised.

The current research therefore explores the best method of providing feedback to trainees about their performance in an understandable and actionable way. The research took place within the context of a research programme exploring social signals interventions for media skills training. During the preliminary phase of this research, 17 participants completed standard media skills training while data was collected on their emotional and social signals behaviour using commercial automated emotion recognition technology (see section 3.2 for details). Human subjective judgements of quality of media skills performance was also collected using the Communication Skills Rating Scale (Spitzberg & Adams, 2007) allowing characterisation of the emotional/social signals associated with better performance (as judged by neutral observers and trainers). Participants from this preliminary study were given the option to provide contact details to return to take part in the design study described here.

3. METHOD

3.1 Participants

Participants who took part in a previous standard media skills training course with emotions detected that expressed interest in taking part were recalled. A total number of five participants were recruited (4 female and 1 male, age range; 18 – 35 years old) (Virzi, 1992).

3.2 Materials: Feedback Display Methods

Several alternative feedback presentation formats were explored including those provided by the affective software providers as well as bespoke designs created specifically for this project.

The data fed into performance summary display was presented only for the first 30 seconds of the whole interaction.

3.2.1 Facial Expression

During initial data gathering, facial expressions were detected using Afdex by Affectiva which uses a facial Action Coding System (Ekman, Friesen, & Hager, 2002). This was used within the context of the iMotions Biometric Research Platform which provides one way of visualising the results.

This has the format of a simple video playback with emotions displayed on an emotion dashboard

provided by iMotions (see *Figure 1*). Included emotions were the 6 basic emotions; fear, joy, sadness, disgust, surprise and anger (Paul Ekman, 1992).

The second option was a custom developed software that was implemented using MATLAB www.mathworks.com, (release R2015a8. 5.0.197613, 64 bit) as a means of processing matrix-based data structures (see *Figure 2*). The code corresponds to a basic implementation of an emotion dashboard.

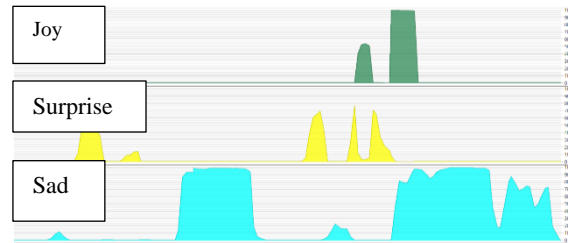


Figure 1: Facial expression feedback using iMotions emotion dashboard (Method 1)

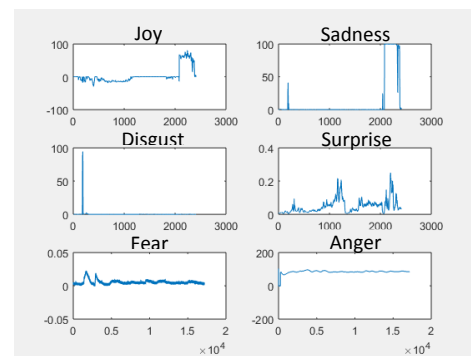


Figure 2: Facial expression feedback using bespoke emotion dashboard (Method 2)

The final method for facial expression feedback was a bar chart style template (see *Figure 3*). The participant's actual performance was plotted alongside a 'good performance' value. All data obtained was normalised to the mean and standard deviation of each feature. This method of presentation is in a numerical format.

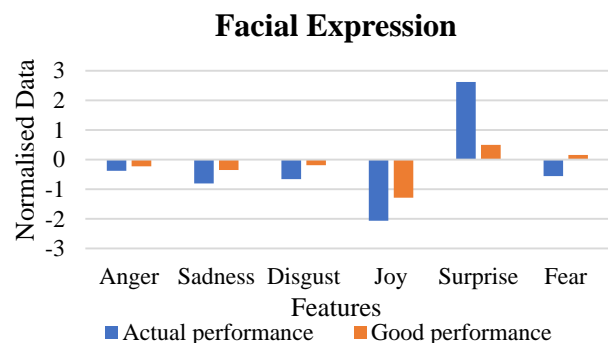


Figure 3: Facial expression feedback using a bar chart template (Method 3)

3.2.2. Voice Emotion Recognition

Voice emotion recognition data was initially gathered using the Layered Voice Analysis developed by Nemesysco^{Ltd}. This provides an output defining a number of variables including energy, content, upset, angry, stressed, uncertain, excited, concentrated, emotion-cognitive ratio, hesitation, brain power, embarrassment, thinking intensity, imagination, extreme emotion and arousal.

Two methods of presenting voice emotion feedback to participants were explored in this study. The first was the 'Emotional Diamond' offered by Nemesysco^{Ltd} (see Figure 4).

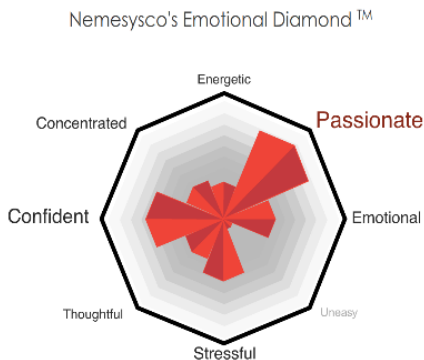


Figure 4: Voice emotion recognition feedback using Emotional Diamond (Method 1)

The second option was a bar chart template layout as presented for facial expression covering the variables listed above.

3.2.3 'Honest signals'

Participants (and interviewers) wore sociometric badges from Humanyze during the data collection phase. These are small recording devices worn around the neck which collect a number of signals relevant to Pentland's concept of 'honest signals' (Pentland & Heibeck, 2010). These include movement, movement activity, movement rate, movement consistency, movement mirroring, posture, posture activity, posture rate, posture mirroring, successful interruptions, unsuccessful interruptions, speed of turn taking, overlap, total speaking, volume, volume consistency, pitch and volume mirroring.

Raw data files can be exported from the sociometric badges, but no user interface is available. For the purposes of this study, a feedback display was developed using the bar chart method (as described above).

3.2.4. Gestures

Gestures were detected using a Shimmer 3 wristband that contains an accelerometer. This

signal could also be integrated into the iMotions Biometric Research Platform providing one way of visualising the results where the video playback was shown alongside the accelerometer values in the iMotions dashboard. Bar chart feedback was also developed to provide an overall summary of gestural movement level.

3.3 Procedure and measures

On arrival participants were introduced to the study and were asked to give informed consent. They were shown feedback on their own previous media interview performance presented with the different formats described above. Each participant was exposed to all the designs described above with the same order of presentation: (1) facial feedback – iMotions, bespoke dashboard, and bar chart, (2) voice feedback – Diamond and bar chart, (3) sociometric badges – bar chart, and (4) gesture – iMotions and bar chart.

Participants were asked to rate each method of feedback using the System Usability Scale (SUS) (Brooke, 1996). The SUS is a reliable 10 item questionnaire with 5 responses for each ranging from 'Strongly agree' to 'Strongly disagree'. It has been used to assess the usability of a range of systems (see Kortum & Bangor, 2013). A high numeric value produced by SUS ratings indicates that the system is easily understood by participants. The benchmark for usability of a system is 68 – 70 and a rating below 50 is cause for concern (Bangor, Kortum, & Miller, 2008; Kortum & Bangor, 2013).

Qualitative interviews were conducted in a semi-structured manner which allowed for more discussion of data. Probe questions explored which designs participants liked best and why.

Participants were compensated with £5 per hour of their participation. The duration of the study was 5 hours. Ethical approval for the study was granted by the Ministry of Defence Research Ethics Committee and Brunel University London Research Ethics Committee.

4. RESULTS AND DISCUSSION

4.1 System Usability Scale

SUS ratings were used to establish ease and understanding of each method (usability). Table 1 summarises the mean SUS scores for the alternative feedback displays. The arrows in the table indicate the SUS scores relative to the benchmark of 68-70 from the literature ((Bangor et al., 2008; Kortum & Bangor, 2013).

Table 1: Mean SUS scores for feedback displays

Channel	Display	SUS (SD)
Facial expression	iMotions	70 ↑ (12.9)
	Bespoke	74 ↑ (10.1)
	Bar chart	58 ↓ (19.8)
Voice	Diamond	84 ↑ (9.1)
	Bar chart	71 ↑ (19.7)
Sociometer	Bar chart	72 ↑ (16.5)
Gesture	iMotions & bar chart	68 → (20)

The results indicate that most of the display formats were considered usable by our sample in the sense that the mean SUS scores obtained were at or above benchmark. None fell below the ‘cause for concern’ level of 50 specified in the literature (Bangor et al., 2008; Kortum & Bangor, 2013). The Emotion Diamond for voice emotion recognition was the highest rated feedback display overall with a SUS score of 84. The lowest score was obtained for the first presentation of the bar chart format (for facial expressions) which received a mean SUS score of 58. However, the bar chart ratings improved on subsequent exposure, suggesting that they were easy to learn.

4.2 Qualitative Interview Findings

Qualitative interviews were conducted to investigate the usability of each method in depth.

4.2.1 Overarching Themes

Two overarching themes were identified from the interviews which applied across all channels. One was the importance of *Comparison to Good Performance*. Participants indicated that they liked to be able to compare their performance to that of others who performed well. This was a feature provided by the bar chart format. The second overarching theme was *Consistency of Presentation*. Participants appreciated consistency across different modalities.

4.2.2 Channel-specific Themes

4.2.2.1 Gestures and Voice Emotion Recognition

The theme identified for gestures and voice emotion recognition is *Scale Concerns* within the bar chart template. Participants had some concerns over the scale on the Y-axis as they had some difficulties understanding what the numeric values represented.

4.2.2.2 Gestures and Facial Expression

The theme identified for gestures and facial expression is *Temporal Behaviour*. Participants preferred the notion of observing gestures and

facial expressions in context of their interview within iMotions.

4.2.2.3 Gestures and Sociometric Badges

The theme identified for gestures and sociometric badges is *Initial Guidance and Explanation*. All participants expressed that they would require an explanation for each channel and what they mean when using the bar chart template for feedback.

4.2.2.4 Facial Expression

The theme identified for facial expression was *Traditional Feedback*. The majority of participants preferred the iMotions video playback. This is a result of learning from video playback and serves as the primary basis for traditional feedback.

4.2.2.5 Voice Emotion Recognition

The theme identified for voice emotion recognition methods of feedback was *Value of Visual Display*. Participants illustrated that the colours of both feedback options enabled them to find key elements of their performance and focus on those points for improvement.

5. Conclusions

The findings of this study suggest a number of design features that will be beneficial to users of affective and social signals technology for communication skills feedback and training. Participants liked visualisations that allowed them to compare their performance to good performance. They also liked consistency across different channels where possible. The findings also showed the importance of feedback designed to provide temporal context, especially for gesture and facial recognition. From these findings we suggest that a combination of emotion recognition augmented video playback, along with consistently formatted summary data showing comparison to ‘good’ performance, most aligns with the views expressed by our participants. In the context of this study, the bar chart was the consistent format used to provide comparative feedback but given the high perceived usability of the Diamond display it may be useful for future work to explore whether geometric forms of this type could be further developed to encompass other feedback channels (face, gesture, sociometrics) as well as the comparative element.

The current study is limited by a small sample size and display options. Future work could evaluate more display options with more potential users.

6. Acknowledgements

The author acknowledges support of the UK Defence Science and Technology Laboratory (Dstl).

7. References

- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *Journal of Human-Computer Interaction Intl. Journal of Human-Computer Interaction*, 24(6), 574–594.
- Bennett, K. B., & Flach, J. M. (1992). Graphical Displays: Implications for Divided Attention, Focused Attention, and Problem Solving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 34(5), 513–533.
- Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4–7.
- Cerrato, L. (2005). Linguistic functions of head nods. In *Proc. Conf. Multi-modal Communication* (pp. 137–152).
- Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, 143(4), 338–347.
- Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*.
- Ekman, P., & Friesen, W. V. (1972). Hand Movements. *Journal of Communication*, 22(4), 353–374.
- Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System*. 577 College Avenue, Palo Alto, California, 94306: Consulting Psychologists Press Inc.
- Fiske, J. (2010). *Introduction to Communication Studies*. Routledge.
- Hall, E. (1969). The Hidden Dimension: man's use of space in public and in private. *Anchor Books*, 217.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.
- Kortum, P. T., & Bangor, A. (2013). Usability Ratings for Everyday Products Measured with the System Usability Scale. *International Journal of Human-Computer Interaction*, 29(2), 67–76.
- Marzillier, J. S. (1976). Gaze and mutual gaze. *Behaviour Research and Therapy*, 14(6), 486.
- McNeill, D. (1992). *Hand and Mind*. Chicago: Chicago Press.
- Pentland, A., & Heibeck, T. (2010). *Honest signals: how they shape our world*. MIT press.
- Poggi, I., & D'errico, F. (2011). Social signals: A psychological perspective. In *Computer Analysis of Human Behaviour* (pp. 185–225). London: Springer.
- Spitzberg, B. H., & Adams, T. W. (2007). CSRS, the Conversational Skills Rating Scale: An Instructional Assessment of Interpersonal Competence. *NCA, National Communication Association*.
- Vicente, K. J., & Rasmussen, J. (2014). The Ecology of Human-Machine Systems 11: Mediating "Direct Perception" in Complex Work Domains. *Ecological Psychology*, 2(3), 207–249.
- Virzi, R. A. (1992). Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 34(4), 457–468.