

# Communication Skills Training Intervention Based on Automated Recognition of Nonverbal Signals

Monica Pereira

London Metropolitan University  
monica.pereira@londonmet.ac.uk

Kate Hone

Brunel University London  
kate.hone@brunel.ac.uk

## ABSTRACT

There have been promising studies that show a potential of providing social signal feedback to improve communication skills. However, these studies have primarily focused on unimodal methods of feedback. In addition to this, studies do not assess whether skills are maintained after a given time. With a sample size of 22 this paper investigates whether multimodal social signal feedback is an effective method of improving communication in the context of media interviews. A pre-post experimental evaluation of media skills training intervention is presented which compares standard feedback with augmented feedback based on automated recognition of multimodal social signals. Results revealed significantly different training effects between the two conditions. However, the initial experiment study failed to show significant differences in human judgement of performance. A 6-month follow-up study revealed human judgement ratings were higher for the experiment group. This study suggests that augmented selective multimodal social signal feedback is an effective method for communication skills training.

## CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); Empirical studies in HCI.

## KEYWORDS

Social Signals, Communication skills training, Media interviews, Off-the-shelf emotion recognition technology

## ACM Reference Format:

Monica Pereira and Kate Hone. 2021. Communication Skills Training Intervention Based on Automated Recognition of Nonverbal Signals. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, May 08–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3411764.3445324>

## 1 INTRODUCTION

Effective communication is important in everyday social interactions. Within many organisations, large investments are made by training employees to communicate better. It is also well recognized that non-verbal communication plays a significant role in the competency of communication styles in a wide range of contexts [63]. Most training interventions rely on human trainers to provide

feedback to learners, but this is costly, labor intensive, subjective and heavily reliant on the skill and experience of individual trainers. Research within the domain of affective computing and social signals processing have started to explore the potential of augmenting or replacing human trainers through the use of automated recognition of nonverbal signals with promising results [7, 14, 30, 31, 59]. However, most studies have focused on a narrow range of channels. They also tend to focus on evaluating the performance of one individual, rather than considering the interplay of signals between communicators.

Our previous work has explored which multimodal signals best predict human ratings of communication skills in the context of TV interviews and developed a usable feedback display to provide participants with information about their performance [47]. Effective media skills are important for many organisations including commercial companies, political parties and non-profits, since performance in this context can have huge implications for organizational reputation and outcomes. To our knowledge, none of the previous affective computing interventions for communication skills have focused on this specific domain.

In the current paper, we report an experiment which was conducted to assess the impact of the feedback intervention we had developed on the effectiveness of training to improve media interview performance. We present a controlled between-groups pre-post experiment study where half of the trainees received standard media skills training and half received the standard training augmented with tailored feedback based on automated recognition of facial expression, vocal signals, hand movements and 'honest signals' [45]. We compared performance before and after training across both groups using subjective measures of performance and using measurements of the participants' displayed social signals. The methods used here have potential to be adapted to support real world training interventions for media skills. Longer-term, the results are relevant to the development of an automatic training feedback system to help learners self-reflect upon their performance.

## 2 RELATED WORK

The social signal processing (SSP) domain aims to understand and interpret social interactions using nonverbal cues [65]. Signal expression depends highly on context. To recognize the significance of an expression researchers must note *where* an expression is displayed, *when* it is displayed and *who* the presenter is [65]. Later, researchers included the significance of *why* and *how* a cue is expressed [49]. Research in this field has been successful in capturing postures [55], gestures [11], vocal behaviour [15] and inferring emotions from facial expression and eye movements [68]. The contexts which have been investigated includes job interviews [42], health-care [26], public speaking [53, 54] and in the classroom [3]. Earlier

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

*CHI '21*, May 08–13, 2021, Yokohama, Japan

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8096-6/21/05...\$15.00

<https://doi.org/10.1145/3411764.3445324>

research investigated social signals in isolation (i.e. facial expression only during an interaction); however, research has demonstrated that multimodal analysis is more informative of understanding naturalistic interactions. Van den Stock and colleagues (2007) investigated emotions associated with body expression and found that when investigated in isolation, the recognition of emotions is incorrectly recognized and cannot be interpreted [57]. It is noted that this is a result of visual integration of such cues which are necessary for adaptive behavior when responding to others [2].

Augmentation of social interactions requires the use of sensor and visual displays that provide trainees with real-time feedback on nonverbal behaviours. The purpose of this is to increase trainees awareness of their use of nonverbal signals and improve the quality of their behaviour in any given context. The behavioral feedback method has been used to provide the user with real-time feedback that is suited to the user, the context and the scenario [17]. This method of feedback provision includes observational learning, operant conditioning, social cognitive theory, perception, reflection and action [18]. Several studies have investigated the efficacy of this method [6, 52, 53]. Even though these studies found promising results, researchers did not investigate whether this method was distracting. In contrast, a study found that visual displays during an interaction was not distracting [18]. However, research in cognitive functioning postulates that an increase in visual load is cognitively taxing [1] and could impact trainees overall performance.

Studies using technology enhanced training have been successful in improving social skills or communication skills [31]. Researchers [27] developed My Automated Conversation Coach (MACH). MACH is a social skills training platform which allows users to communicate with a virtual character. MACH captures facial expression and speech and generates information on the users use of nonverbal cues. Similarly, a study by Damian and colleagues [16] found that this technique was useful in improving job interview training in underprivileged adolescents. Another group of researchers developed a feedback system called ROC Speaking Framework [68][70]. Researchers found that feedback of social signals during job interview training significantly improved in comparison to traditional methods of training. Similar results were found when attempting to improve communication skills in those with social impairments [60], public speaking [13, 52, 53], medical students [39], job interviews [3, 9, 31, 32, 41–43]. However, no studies investigate whether social skills can be improved using automated feedback in the context of media skills training.

### 3 THE CURRENT STUDY

A review of the literature revealed that research in augmentation of training feedback does not include media skills training. This context of communication requires a unique technique to communicate to a wider audience. Topics such as job interviews and public speaking where there is a vast quantity of literature are different to media interview communication. The concept of context in communication is important as noted in section 2.

Furthermore, research is typically limited to investigating unimodal (facial expression alone) or bimodal (e.g., facial expression and hand gestures) signals and no research exist which investigates

the retention of skills over time when augmenting training. The current research aims to investigate whether communication skills can be improved using automated feedback to improve self-awareness. Based on the literature we aim to address the following research questions:

1. Can communication skills be improved by providing automated feedback in the context of media interview training?
2. Are there differential training effects for social signal feedback when tested after 6 months?

Based on previous research it is predicted that trainees who receive augmented feedback during media skills training will be rated as better communicators, even after 6 months

The remainder of this paper is organized in the following manner: section 4 details the design of this experiment, population sample, off-the-shelf recognition technology used to capture social signals, characteristics of the judges of communication performance and the communication rating scale used to rate interviewees performance, training intervention design, method of feedback methodology and procedure of data collection. Section 5 discusses how the data will be analysed. Section 6 presents the results for the initial experiment / training intervention and 6-month follow-up. Section 7 discusses the results considering previous research, provides some possible research limitations. Section 8 provides a conclusion and future research recommendations.

## 4 METHOD

### 4.1 Study Design

An experimental study was conducted to compare standard media skills interview training to technology-enhanced training. A between-subjects design was used with participants randomly assigned to either the experimental group (social signal training) or the control group (standard training). Participant communication skills were measured before training (pre-test), after training (post-test) and again after six months (follow-up study). Media skills training consisted of a taught session followed by a series of practice interviews. Participant communication skills were measured by the journalist in the study and neutral observers who were both blind to the experimental condition. More details about the judges of performance can be seen in section 4.4. Moreover, participants were asked to self-rate their improvement in skill and confidence following training after the initial training day was complete.

*4.1.1 Initial Training Intervention Design – Assessing Immediate Training Effects.* The initial study was a 2x2 mixed factorial experiment. The between-participants independent variable was feedback type which has two levels: social signals feedback and traditional feedback. The within-participants independent variable was session with two levels: pre-training and post-training tests. Detailed information about the design of the feedback intervention can be found in section 4.5 and details of the experimental procedure can be found in section 4.6. The dependent variables were subjective judgements of training skills (trainer rating, neutral observer rating) and automatically detected social signals. Detailed information about the subjective measures used in the study can be found in section 4.4. The study was designed to investigate immediate training effects achieved on the day of training with the following hypotheses:

H1: There will be significant improvements in subjective ratings of improvement communication skills from baseline interview to post-training interview in both training conditions (main effect of training)

In addition, a two-condition comparison was conducted between the control (standard feedback) and experimental (social signal feedback) training groups with the following hypotheses:

H2: There will be significant changes in observed social signals detected between baseline and post training interview in both conditions (main effect of training)

H3: Subjective training gains from baseline interview to post-training interview will be greater for the experimental condition (interaction effect)

H4: Greater changes in social signals will be detected between baseline interview and post-training interview for the experimental condition (interaction effect)

**4.1.2 Follow-up Study Design – Skill Maintenance after 6 Months.** The follow-up study was a between participants' design. The between participants independent variable was feedback type which included two levels: traditional feedback and social signal feedback and the dependent variables included the social signals captured during interviews. In this follow-up study, details about the studies procedure can be found in section 4.6. The study was designed to investigate the maintenance of skill after 6 months with the following hypotheses:

H5: Subjective ratings of observed interview performance will be greater for the experiment group after 6 months

## 4.2 Participants

A total of 22 research staff and students recruited (age ranged from 18 – 55 years old; 6 male and 16 female) for this stage. Experience in public speaking ranged from no experience to extensive and from none to some experience in media interviews. The roles that participants had within the university included taught students (3), research staff (1) and research students (18). Participants were from different cultural backgrounds including 6 native English speakers (participants who stated that English was their first language) and 16 non-native English speakers. Participants who took part in initial training were recalled using participant recruitment posters.

Of the 22 participants trained, a total of 16 participants (age ranged from 18 – 55 years old; 13 females and 3 males) were included in this follow-up study. The roles that participants had within the university included research students (12), research staff (1) and taught students (3). A total of 14 participants were non-native English speakers and 2 participants were native English speakers. Ethical approval was granted by the Brunel Research Ethics Office and Ministry of Defence Research Ethics Council. All participants were not affiliated with the lab in which this study was conducted.

## 4.3 Off-the-shelf Nonverbal Signal Detection Technology

Nonverbal signals detected throughout interview included facial expressions, vocal signals, honest signals and hand movements. This section describes the technology used to capture these signals. Receiver Operating Characteristic (ROC) measures will be reported for each technology. This demonstrates the diagnostic ability of a system which is based on a curve that is created by its true positive rate against the false positive rate. A ROC score ranges from 0 to 1 where a score of 1 demonstrates a perfect classifier [12]. All recordings of communication channels were synchronized to one second timestamp.

**4.3.1 Facial Expression Recognition.** Capturing facial expressions and hand movements were synchronized on iMotions which is a Biometric Research Platform. Facial expressions were detected using AFFDEX by Affectiva a Facial Action Coding Unit System (FACS) which detects Action Units (AU) which are derived from facial muscles, associated with basic emotions [20]. The signals captured for this system were the 7 basic emotions: sad, happy, contempt, anger, fear, surprise and disgust. Additionally, all the AU associated with those facial expressions were also captured. The ROC score for AFFDEX is 0,8 for joy, disgust, contempt and surprise ([www.developer.affectiva.com/determining-accuracy/](http://www.developer.affectiva.com/determining-accuracy/)). To ensure high quality recordings, interviews were conducted in rooms with natural light and where this was not possible a Lishuai LED 312 Panel light diffuser was used. Hand movements signify hand use in the form of gestures during interviews were captured using Shimmer Unit+ 3 which was attached to the participants non-dominant wrist.

**4.3.2 Vocal Behaviour Detection.** Vocal signals were detected using Nemesysco's QA 5 Evaluation System which is a voice recognition technology used to capture vocal signals. This software uses proprietary signal processing algorithms to extract vocal parameters which are classified according to a range of signals in real time or recordings can be imported and post-processed. For the current study, vocal behaviour was captured in real time. The signals captured by Nemesysco can be seen in Table 1.

The area under the ROC curve score for Nemesysco ranged from 0.53 – 0.71 [36, 37]. However, this study did not specify which system was investigated. However, the certain signals captured by the QA 5 system used in this study has been validated such as 'embarrassment' [25] and 'arousal' [35]. In addition to these validations, the QA 5 has been used in the development of a conversational robot [27, 62] which suggest the application of Nemesysco in social interactions. Signals detected by QA 5 can be seen here [46]. To validate the signals captured by Nemesysco using an open source software, signals were correlated with prosodic features extracted from PRAAT which is a voice extraction software which can be used to analyze, synthesis and manipulate speech [8]. A correlation analysis was conducted to validate the features collected by QA 5. Features extracted from PRAAT included intensity (mean, energy, minimum and maximum), fundamental frequency / pitch (mean and maximum). All results produced were all significantly correlated ( $p < .001$  to  $p = .04$ ).

**Table 1: Definitions of social signal labels produced by Nemesysco**

Signals	Description
Energy	Indicates if speaker is sad, tired, boredom, comfortable or highly energetic.
Content	Indicates how pleased or happy a person is
Upset	Indicates how displeased or sad a person is
Angry	Indicates how angry a person is
Stressed	Indicates how nervous a person is
Embarrassment	Indicates how uncomfortable a person is
Intensive thinking	Indicates thinking intensity while speaking
Imagination Activity	Indicates whether the person is recalling information or visualizing something
Hesitation	Indicates how comfortable a person is when making the statement
Uncertainty	Indicates how certain or uncertain a person is
Excitement	Indicates how positively or negatively excited a person is
Concentration	Indicates how concentrated the person is
Arousal	Indicates deep and profound interest in the conversation
Extreme Emotion	Indicates overall emotional activity
Cognitive Activity	Overall cognitive activity
Rationality	Indicates rationality

**4.3.3 Honest Signals.** Honest Signals are the subtle nonverbal patterns of behaviour which reveal our emotions and attitudes towards others in a social interaction. Pentland [45] proposed four main honest signals: activity (indicates increased activity often indicates interest and excitement), consistency (consistently paced speech indicates mental focus, whereas irregular or jerky speech suggests lack of focus, and can signal an openness to influence others), influence (extent to which a person's action causes another person's actions to match their own) and mimicry (*chameleon effect* - extent to which interlocutors reflexively copy each other).

Sociometric badges were developed by Pentland and used to detect a range of signals hypothesised by Pentland to relate to 'honest signals'. The sociometric badges were developed to capture 'honest signals' in a multiperson interaction containing two or more individuals and worn around the neck. Badges have also been used to detect signals in dyadic interactions [29]. These signals are detected by four sensors: a microphone, an infrared sensor, a Bluetooth detector and a motion detector [45]. These signals are detected by four sensors: a microphone, an infrared sensor, a Bluetooth detector and a motion detector [45]. Microphone captures vocal tones and does not capture content (Table 2. Feature L – U). The infrared sensor captures movement in relation to other interlocutors (Table 2. Features – E, F, J, K). The Bluetooth detector captures other badge wearing interlocutors. The signals produced by Sociometric Badges during TV interviews for the current study can be seen in Table 2.

The ROC score for these badges have been reported at 0.8 [69]. After the interview, the data stored locally on the badges were exported as structured meetings (as participants were facing each other in a single meeting) with a resolution of 1 second intervals. Badges worn by the trainer and the participant were synced Socio-metric DataLab Enterprise Edition 3.1.2824.

#### 4.4 Journalist and Neutral Observer Characteristics and Subjective Ratings of Communication

This study included a total of four judges, a journalist (interviewer) and three neutral observers. The interviewer for the initial study was a trained journalist. This journalist was a female with 4 years field experience and had conducted all the interviews. The journalist recruited for the follow-up study was a male with 4 years field experience and had conducted all the interviews. As part of a journalists training, they are required to maintain objectivity. The neutral observers recruited for this study acted as a general audience and were not trained on what is considered 'effective communication' and were treated as a member of the general population. The three neutral observers recruited for the initial study were different to the neutral observers in the follow-up study.

Both the journalist and the neutral observers were asked to fill in a communication skill rating scale to measure each interviewees communication ability. Subjective ratings of performance were collected using the Conversation Skill Rating Scale (CSRS) [56]; which includes 25 conversational feature scale ratings and a composite measure of 5-items on communication performance called molar scores. The molar scores were used as a measurement of communication performance. Raters rated performance using a Likert scale from 1-7 on the following 5 items: 1) poor conversationalist to good conversationalist, 2) socially unskilled to socially skilled, 3) incompetent communicator to competent communicator, 4) inappropriate communicator to appropriate communicator and 5) ineffective communicator to effective communicator. The CSRS 5-item molar scores have produced an internal reliability between .85 and .90 and an acceptable inter-rater reliability of .75 [56]. All judges were given instructions on how to fill in the molar scores.

**Table 2: Signals Produced by Sociometric Badges**

Feature	Description
A) Body movement	Normalized acceleration magnitude over 3 movement axes
B) Body movement activity	Absolute value of the first derivative of the accelerometer's energy
C) Body movement rate	Indicates the direction of change in activity level (compared to first derivative)
D) Body movement consistency	Movement consistency throughout interaction
E) Body movement mirroring	Mimicking of other badge wearers body movement
F) Body movement mirror lag	Delay in mimicking of body movement
G) Posture front back	Orientation of front back panel
H) Posture activity	Absolute angular velocity
I) Posture rate	Angular acceleration
J) Posture mirroring	Mimicking of other badge wearers posture
K) Posture mirror lag	Delay in mimicking of posture
L) Successful Interruptions	Number of successful interruptions made by the badge wearer
M) Unsuccessful Interruptions	Number of unsuccessful interruptions made by the badge wearer
N) Speed of turn-taking	Indicates speed of turn-taking in a conversation
O) Overlap	Total amount of speaking whilst someone else is also speaking
P) Total speaking	Total amount of combined speaking (speaking and overlap combined)
Q) Volume Front	Average absolute value of amplitude of the front microphone
R) Volume consistency front	Measurement of change in speech volume
S) Front pitch	Pitch of the voice from the front mic correlated with the fundamental frequency of the voice signal
T) Volume mirroring	Mimicking of other badge wearers volume
U) Volume mirroring lag	Delay in mimicking of other badge wearers volume

Two further questions were administered to participants aimed at obtaining a rating of their confidence and skills based on the training they received:

- 1) Please indicate on the scale below how much you estimate your skills in giving a media interview has improved over the course of today's training event.
- 2) Please indicate on the scale below the extent to which your confidence in giving a media interview has improved over the course of today's training event.

Participants in both conditions had rated their skills and confidence after training on a scale of 'no improvement', 'slight improvement', 'some improvement', 'considerable improvement' and 'great improvement'.

#### 4.5 Social Signal Feedback Method and Visualization

Participants in the control condition received formative feedback following practice interviews while participants in the experimental condition received formative and summative feedback following practice interviews. After each interview, participants' videos in both groups were played back their videos for reflection to improve self-awareness as those who are self-aware are known to be effective communicators [28, 52, 66]. The group that received social signal feedback received technology enhanced feedback with an emotion dashboard [47] and those who received standard training feedback were presented with the same video playback visualization but without the emotional dashboard. Both groups received verbal feedback and an opportunity to discuss their feedback with the journalist. The method of feedback applied was the sandwich

feedback model [19]. Participants would receive positive feedback, then negative feedback on which elements to improve on and then followed by positive feedback again.

Figure 1 shows how participant interviews were played back in a formative structure which allowed participants to view their overall interview performance for self-reflection. An emotional dashboard below the video display was used to observe expression and expression range only for the experiment condition. The emotional dashboard was removed for the control condition.

A summative feedback method was presented to the experiment group after each interview formative feedback session. Nonverbal signal feedback visualization presented using a bar chart template (see Figure 2) which contains a 'good performance' element based on thin slices of behaviour in the first 30 seconds of the interview [47]. The 'good performance' element is similar to the performance threshold developed in [18] where participants were given feedback based on the level of behaviour displayed above the threshold, suggesting a behaviour that needs modifying for improved performance. Another study which used a bar chart method of visualizing feedback was also presented to participants in [59] which also included a behaviour component. For this study, signals were selected for feedback by the trainer based on whether participants' performance exceeded the minimum and maximum threshold, similar to [18]. Moreover, selective feedback was done to diminish an increase in cognitive overload as an increase in cognitive load has negative effects on task performance as information held in our working memory that transfers into long term memory is affected [22]. Furthermore, by feeding back additional information that is irrelevant for improvement, comprehension will be reduced [50], otherwise known as the seductive details effect [23].

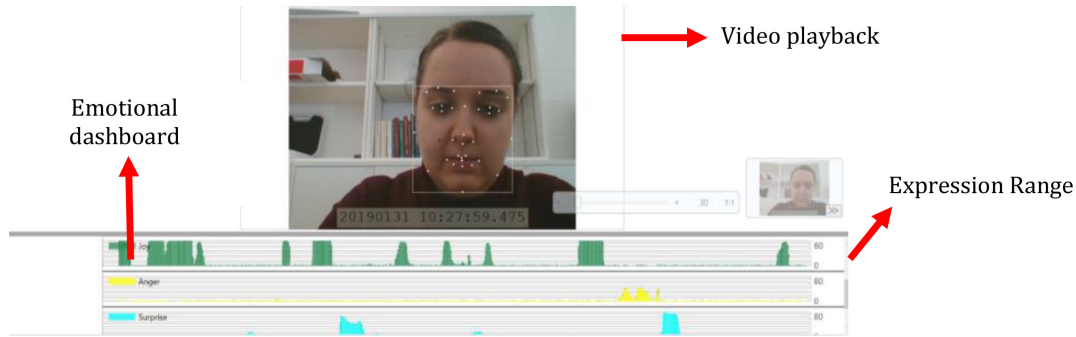


Figure 1: Formative Feedback for Facial Expression and Hand Movement

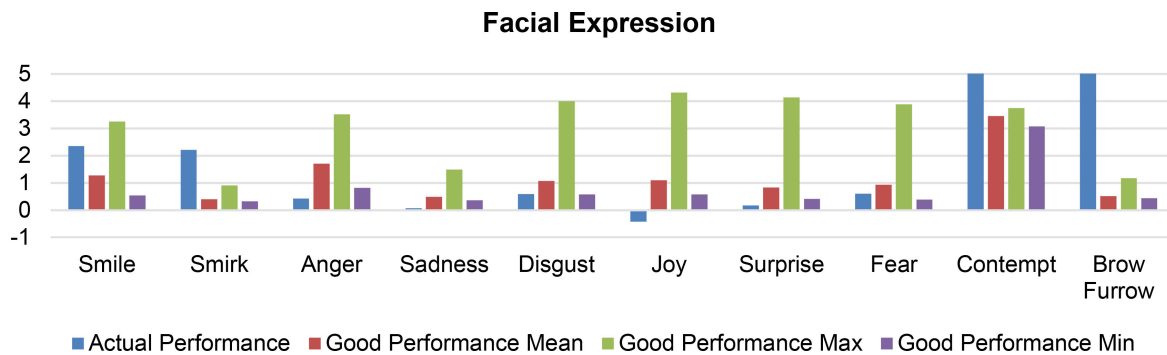


Figure 2: Summative Feedback Template for Social Signal Feedback Group

Figure 2 shows the summative feedback method. This bar chart template was produced for each channel communication (vocal signals, honest signals, hand movements and facial expression). The signals included in the facial expression template were smile, smirk, anger, sadness, disgust, joy, surprise, fear, contempt and brow furrow (frown). The signals included in the honest signals template were movement rate, movement mirroring, posture, posture mirroring, turn taking speed, volume and volume mirroring. The signals included in the vocal signal feedback included content, upset, hesitation and extreme emotion. Vocal signals, honest signals and hand movements were the same as the template demonstrated in Figure 2. Our previous study was conducted to identify which social signals are relevant for effective communication in the context of media interviews and the signals identified in the previous study were feedback to participants in this study. The justification for feedback of these signals can be seen in our previous paper [47].

**4.5.1 Interview Description and Task Difficulty.** Interview questions were designed prior to the interviews. Participants were asked to provide information about their research prior to the training day to allow the journalist to understand their research and create questions. This is similar to what a Press Officer would be tasked with in a real-world setting [61]. Example questions included “Please provide a short summary of your research in non-specialist/ non-academic language”, “Why is your research important to a general audience?”, “What is the worst possible question you fear being asked by a member of the public or a journalist?”. Some examples of research topics of

participants included immunology, positive psychology, creativity, engineering and design.

On the initial training day participants took part in 5 interviews: 1) a pretraining interview, 2) a radio interview, 3) a face-face TV interview, 4) a down-the-line interview and 5) a post-training interview. The radio interview, face-face interview and down-the-line interviews were practice interviews, each of which increased in task difficulty. The pre-training and post training interview were face-face interviews and were a moderate level of task difficulty. Difficulty was assessed by what the participants labelled as a fearful question. Example questions in pre-training and post training interviews included “Tell me about your research”, “Why is your research important to the general public”, the more difficult questions were related to the ethics surrounding participants research. Examples of difficult questions in practice interviews included participants were asked to justify why funding their research was important if it is not as impactful as other research in the media.

The interview that participants took part in 6 months following their initial training were the similar to the pre-training and post-training interview task difficulty.

## 4.6 Procedure

For the initial training study, participants were recruited via email to take part in research in media skills training workshop. Participants who had expressed interest in taking part in the study were given a participant information sheet. On the day of the media

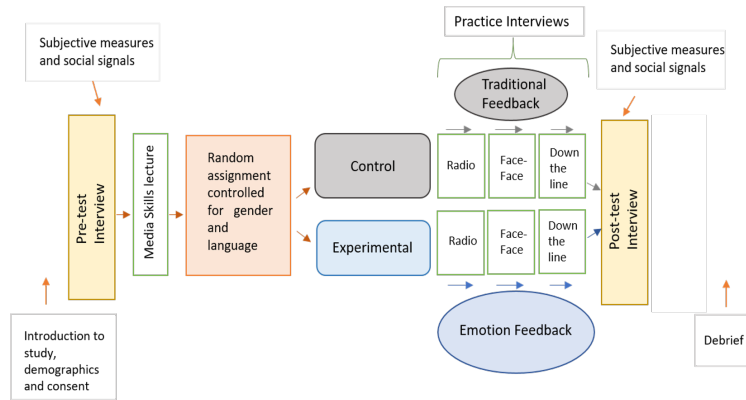


Figure 3: Schematic Presentation of Experimental Procedure for the Initial Study

skills training participants were introduced to the study, the social signal technology and were informed of their right to withdraw their participation without reason. Participants were also given the option to take part in the study without social signal recognition technology turned on; however, none had chosen this option. At this point demographics and consent were collected. Participants then engaged in a pre-training interview where subjective ratings gathered from the journalist and social signals were recorded. Participants were then split into pairs matched as closely as possible by gender, native language (English as first language), and on average pre-test communication skill ratings based on journalist ratings of communication skills (molar scores). One member of each pair was allocated to the experimental or control condition at random (by the toss of a coin). This procedure aims to reduce the variability between conditions which may be attributable to extraneous variables. Participants then engaged in a 30-minute pre-recorded lecture which introduced them to effective communication (verbal and nonverbal) in media interviews and were given examples of good and bad media interviews.

Next, participants individually engaged in their training session which comprised of a radio, face-face and down-the line interview and lasted a total of 3 hours. Both the control and the experimental conditions received feedback from the journalist about their interview. Journalist feedback was given after the playback of videos and were given the opportunity to discuss their feedback with the journalist. Following this, the journalist was asked to leave the room to mask which group the participants were assigned to. The duration of the post interview was 10 minutes for the control group and 15 minutes for the experiment group. This was because the experiment condition received additional social signal feedback as described in section 4.5.

All participants then engaged in a post-training interview where dependent variables were recorded exactly like the pre-training interview. Social signals were collected from both groups during their pre and post-training interviews. Overall, all interviews for both conditions lasted 5-6 minutes. Participant interview recordings were later rated by three neutral observers. Both the trainer and neutral observers were blind to the participant feedback condition. Neutral observers were also blind to the recording order (i.e., pre vs post-test were randomized). Participants were reimbursed £5 /

hour in recognition of their time. A summary of the experimental procedure can be seen in Figure 3

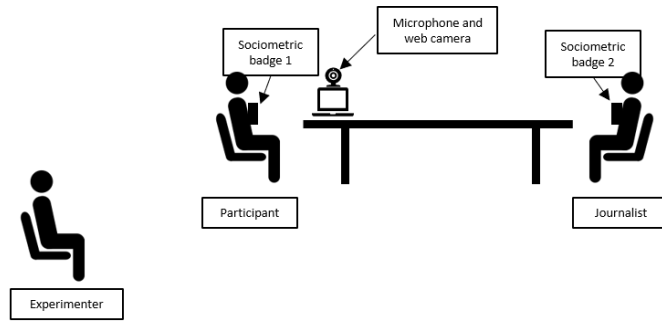
Six months later, participants were asked to take part in a final media interview. Once participants arrived, they were briefed on the study and the recording equipment, formal consent and demographics were collected. Permission was requested for the researchers to access participants' data from the previous study and all participants consented to this. Participants then took part in a 7 – 10-minutes media interview with a journalist. After the media interview was completed, participants' interviews were played back to them. The journalist, as a conversational partner, was asked to rate the participants communication skills performance (molar scores). Participants' performance was also later rated by three neutral observers who were blind to condition assignment. The study layout (test scene) can be seen in Figure 4

## 5 DATA ANALYSIS

A 2x2 mixed factorial analysis of variance will be conducted to assess differences in subjective ratings of communication performance provided by three neutral observers. The between-participants independent variable is feedback type which has two levels: social signals feedback and traditional feedback. The within-participants independent variable was session which has two levels: pre-training and post-training interviews. The dependent variables were subjective judgements of communication skills. To meet the assumption of normal distribution the data was normalised using the minimum and maximum values of the dataset for each signal using the following formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Data was averaged 10s intervals which introduced 24 trials (4 minutes). Four minutes were considered due to missing data after this time point. Additionally, 10s intervals were used as it is enough time to account for question and answers in chunks. Data which was not collected by the technology were considered missing data. A tabulated pattern analyses of the missing data for 'pitch', 'volume mirroring', 'volume mirror lag' and 'speed of turn-taking' as missing data was in abundance but produced no cohesive pattern in missing data and was then removed [58]. Content detected by voice



**Figure 4: Study Layout / Test scene.** The microphone was to ensure that vocal analysis was being analysed throughout the interview. The camera was to ensure facial expression recognition was recorded throughout. Both the participant and journalist were wearing sociometric badges. The experimenter was in the room to ensure the equipment was working throughout and to produce the bar chart for the experiment group once the journalist left the room.

analysis was also removed as all values produced by Nemesysco were mostly '0' which significantly skewed the data. The resulting total number of valid cases produced were 646 cases.

We did not compare the 6-month interview with previous subjective ratings of communication performance because the journalist recruited for the 6-month was different so we could not measure like for like in interview setting. This could also introduce an extraneous variable (i.e., the first journalist was a female and the second was a male with different experience in training).

### 5.1 Assumption Testing

A principal component analysis (PCA) will be conducted to obtain components of social signals which are correlated and linearly uncorrelated variables [21]. A PCA will be conducted to reduce and merge similar variables to conduct a multivariate analysis (Cheng Li & Wang, 2014). The reason for this was that the total number of variables included in the analysis was large and included multiple emotions across communication channels. Additionally, this will be done to identify whether the signals which were fed back to participants were highlighted in the PCA analysis. Assumptions for running a PCA was met. The data was normally distributed, no extreme outliers were observed, adequate sample size as demonstrated by KMO, the variables were unrelated and suitable for reduction [21, 44].

A multivariate analysis will be conducted where the between-participants independent variable was *feedback type* with two levels: social signals feedback and traditional feedback. The within-participants independent variable was the session which has two levels: pre-training and post-training interviews. The dependent variables were the components extracted from the PCA analysis. Assumptions for running a multivariate analysis on PCA components were explored and most were met. Homogeneity of variance was not assumed, as results a stricter alpha level was adopted at  $p = 0.1$  as advised in [58]. Each component was also explored, and an outlier for component 7 was identified and removed. A Shapiro Wilks test was observed for normality and when the  $p$  value was less than .05. Histograms were eyeballed. Each signal contained a pre-training histogram which was similar in the post training histogram demonstrating consistency in signals which is acceptable to run an

ANOVA test [51]. This enabled an analysis using the PCA-ANOVA method. A Levene's test was interpreted for homogeneity of variance and if not assumed then univariate ANOVAs will be conducted with interpretation of the alpha level at a stricter level;  $p < .01$  [4]. Welch's statistic will be interpreted for post hoc analyses to adjust for unequal variances [21]. Box's M was significant ( $p < .001$ ); however, when groups are larger than 30, ANOVA is robust to this.

## 6 RESULTS

### 6.1 Subjective Ratings of Performance

The degree to which raters agreed on the participant's performance was moderate. The average measure intraclass correlation was .696 with a 95% confidence interval from .394 to .863 ( $F(21, 42) = 3.79, p < .001$ ) in the pre-training interview. Neutral observers were in moderate agreement in the post-training interview with an average measure of .603 with a 95% confidence interval from .146 to .827 ( $F(21, 42) = 3.868, p < .001$ ). On average neutral observers rated the control group as improving by 15% on the subjective rating scale for overall communication skill and the social signal feedback group as improving by 20% on the subjective rating scale. Descriptive statistics can be seen in Table 3

A 2x2 mixed ANOVA was conducted on communication skills rated by neutral observers. Results revealed no significant main effect for session ( $F(1, 20) = 34.269, p < .001, \eta^2 = .631$ ) and for feedback type ( $F(1, 20) = 1.127, p = .301, \eta^2 = .053$ ). There was a significant interaction between session x feedback type ( $F(1, 615) = 1.022, p = .324, \eta^2 = .049$ ). A follow-up analysis of the main effect illustrates that participants who received social signal feedback significantly improved in communication from pre-test to post-test ( $t = -4.509, p = .001$ ). Furthermore, participants who did not receive social signal feedback significantly improved in communication ratings from pre-test to post-test ( $t = -3.735, p = .004$ ).

**6.1.1 Confidence and Skills Ratings.** The mean and standard deviations of the self-evaluation scores between the two groups can be seen in Table 4

[Mean (Standard Deviation)]

Participants who received social signal feedback reported higher confidence post-training than participants who did not ( $t = 2.500$ ,



**Table 3: Descriptive statistics for neutral observer ratings of performance**

	Pre-test		Post-test	
	Social signal Feedback Group	Traditional Feedback Group	Social signal Feedback Group	Traditional Feedback Group
Neutral Observers Communication Score	4.44 (.89)	4.18 (.95)	5.30 (.73)	4.79 (1.03)

[Mean (Standard Deviation)]

**Table 4: Descriptive statistics for self-evaluations of confidence and skill post-test**

Item	Social signal Feedback Group	Traditional Feedback Group
Confidence	4.36 (.67)	3.36 (.81)
Skill	4.09 (.83)	3.18 (.87)

**Table 5: PCA Components**

Component	Signals
Confidence	Rationality, Stressed (-) <sup>*</sup> , Excited, Energy, Imaginative Think (-), Upset (-), Cognitive Activity, Imagination (-) and Uncertain (-)
Disgust	Disgust, Jaw Drop, Upper Lip Raise and Nose Wrinkle
Frowning	Brow Furrow, Lid Tighten, Anger and Sadness
Eagerness	Movement Activity, Volume, Movement Consistency (-), Unsuccessful Interruptions and Posture Activity
Expression	Engagement, Surprise and Brow Raise
Engagement	
Posed Expression	Dimpler, Lip Stretch, Lip Press and Lip Suck
Posture	Movement and Posture

\* (-) is negatively loaded

$p = .021$ ). This was also observed for skill ratings post-training ( $t = 19.370$ ,  $p = .005$ ).

## 6.2 Principal Component Analysis

A PCA was run on the data using a Varimax Rotation. Analysis of the scree plot revealed the second elbow would include 7 Components accounting for 42.85% of the total variance explained (Li & Wang, 2014). Variables retained in the rotation matrices were above .6 [21]. Component labels and items can be seen in Table 5

## 6.3 Multivariate Analysis

A multivariate analysis revealed a significant main effect for session,  $F(5, 611) = 8.097$ ,  $p < .001$ ,  $\eta^2 = .082$  and a significant main effect for feedback type,  $F(5, 611) = 44.782$ ,  $p < .001$ ,  $\eta^2 = .330$ . There was also significant interaction effect for session x feedback type,  $F(5, 611) = 4.701$ ,  $p < .001$ ,  $\eta^2 = .049$ .

**6.3.1 Confidence.** A 2x2 ANOVA revealed no significant main effect for session ( $F(1, 642) = 1.470$ ,  $p = .226$ ,  $\eta^2 = .002$ ) and for feedback type ( $F(1, 642) = 5.568$ ,  $p = .019$ ,  $\eta^2 = .009$ ). However, there was a significant interaction between session x feedback type ( $F(1, 615) = 8.433$ ,  $p = .004$ ,  $\eta^2 = .013$ ). Follow-up analysis using a Welsh's

statistic revealed that there was a significant difference in the pre-training interview between social signal feedback and traditional feedback types ( $F(1, 231.961) = 14.891$ ,  $p < .001$ ,  $\eta^2 = 0.05$ ,  $d = 0.475$ ) and no differences in groups in the post-test ( $F(1, 342.956) = .148$ ,  $p = .701$ ,  $\eta^2 = .00$ ,  $d = .040$ ). There was a significant reduction in confidence from pretraining to post-training was observed the control group, ( $F(1, 339.549) = 12.265$ ,  $p = .001$ ,  $\eta^2 = 0.033$ ,  $d = 0.376$ ). Whereas those who were given social signal feedback had not significantly improved in confidence from the pre-training interview to the post-training interview, ( $F(1, 293.599) = 1.156$ ,  $p = .283$ ,  $\eta^2 = 0.004$ ,  $d = 0.123$ ).

**6.3.2 Disgust.** Results revealed a significant main effect for session ( $F(1, 642) = 25.737$ ,  $p < .001$ ,  $\eta^2 = .039$ ) and feedback type ( $F(1, 642) = 69.691$ ,  $p < .001$ ,  $\eta^2 = .098$ ). However, there was no significant interaction between session x feedback type ( $F(1, 642) = 2.333$ ,  $p = .127$ ,  $\eta^2 = .004$ ). Follow-up analysis using a Welsh's statistic revealed that the control group significantly reduced in demonstrations of disgust from the pre to post-training interview ( $F(1, 264.137) = 15.095$ ,  $p < .001$ ,  $\eta^2 = 0.046$ ,  $d = 0.431$ ). Those who had received social signal feedback reduced displays of disgust from pre-training to post training ( $F(1, 220.177) = 10.438$ ,  $p = .001$ ,  $\eta^2 = 0.037$ ,  $d = 0.386$ ).

**6.3.3 Frowning.** A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158, p < .001, \eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780, p < .001, \eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460, p < .001, \eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449, p = .001, \eta^2 = 0.046, d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007, p = .933, \eta^2 = 0.000, d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899, p = .049, \eta^2 = 0.012, d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582, p < .001, \eta^2 = 0.067, d = 0.496$ ).

**6.3.4 Eagerness to Speak.** A 2x2 ANOVA revealed no significant main effect for session ( $F_{(1, 642)} = .186, p = .666, \eta^2 = .000$ ), but a significant main effect for feedback type ( $F_{(1, 642)} = 64.358, p < .001, \eta^2 = .091$ ). There was also no significant interaction between session x feedback type ( $F_{(1, 642)} = .078, p = .780, \eta^2 = .000$ ). Follow-up analysis using a Welch's statistic revealed that there was a significant difference between social signal feedback and traditional feedback types in the pre-training interview ( $F_{(1, 230.506)} = 23.423, p < .001, \eta^2 = 0.083$ ). A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158, p < .001, \eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780, p < .001, \eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460, p < .001, \eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449, p = .001, \eta^2 = 0.046, d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007, p = .933, \eta^2 = 0.000, d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899, p = .049, \eta^2 = 0.012, d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582, p < .001, \eta^2 = 0.067, d = 0.496$ ). 0.593) as well as in the post training interview ( $F_{(1, 341.316)} = 30.218, p < .001, \eta^2 = 0.077$ ). A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158, p < .001, \eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780, p < .001, \eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460, p < .001, \eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449, p = .001, \eta^2 = 0.046, d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007, p = .933, \eta^2 = 0.000, d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899, p = .049, \eta^2 = 0.012, d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582, p < .001, \eta^2 = 0.067, d = 0.496$ ). 0.573).

**6.3.5 Expression Engagement.** A 2x2 ANOVA revealed no significant main effect for session ( $F_{(1, 642)} = .719, p = .397, \eta^2 = .001$ ) but there was a main effect for feedback type ( $F_{(1, 642)} = 55.381, p < .001, \eta^2 = .079$ ).

There was no significant interaction between session x feedback type ( $F_{(1, 642)} = .544, p = .461, \eta^2 = .001$ ). Follow-up analysis using Welch's statistic revealed that there was a significant difference between social signal feedback and traditional feedback types in the pre-training interview ( $F_{(1, 230.506)} = 23.423, p < .001, \eta^2 = 0.083$ ). A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158, p < .001, \eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780, p < .001, \eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460, p < .001, \eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449, p = .001, \eta^2 = 0.046, d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007, p = .933, \eta^2 = 0.000, d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899, p = .049, \eta^2 = 0.012, d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582, p < .001, \eta^2 = 0.067, d = 0.496$ ). 0.593) as well as in the post training interview ( $F_{(1, 341.316)} = 30.218, p < .001, \eta^2 = 0.077$ ). A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158, p < .001, \eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780, p < .001, \eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460, p < .001, \eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449, p = .001, \eta^2 = 0.046, d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007, p = .933, \eta^2 = 0.000, d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899, p = .049, \eta^2 = 0.012, d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582, p < .001, \eta^2 = 0.067, d = 0.496$ ). 0.573).

**6.3.6 Posed Expression.** A 2x2 ANOVA revealed no significant main effect for session ( $F_{(1, 642)} = .819, p = .366, \eta^2 = .001$ ), but a significant main effect for feedback type ( $F_{(1, 642)} = 14.805, p < .001, \eta^2 = .023$ ). There was also no significant interaction between session x feedback type ( $F_{(1, 642)} = 3.732, p = .054, \eta^2 = .000$ ). Follow-up analysis revealed that there was no significant difference between social signal feedback and traditional feedback types in the pre-training interview ( $F_{(1, 210.101)} = 1.735, p = .189, \eta^2 = 0.007$ ). A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158, p < .001, \eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780, p < .001, \eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460, p < .001, \eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449, p = .001, \eta^2 = 0.046, d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007, p = .933, \eta^2 = 0.000, d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899, p = .049, \eta^2 = 0.012, d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582, p < .001, \eta^2 = 0.067, d = 0.496$ ).

$\eta^2 = 0.067$ ,  $d = 0.496$ ). 0.163) but there was a significant difference observed in the post training interview ( $F_{(1, 370.988)} = 18.130$ ,  $p < .001$ ,  $\eta^2 = 0.046$ ). A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158$ ,  $p < .001$ ,  $\eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780$ ,  $p < .001$ ,  $\eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460$ ,  $p < .001$ ,  $\eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449$ ,  $p = .001$ ,  $\eta^2 = 0.046$ ,  $d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007$ ,  $p = .933$ ,  $\eta^2 = 0.000$ ,  $d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899$ ,  $p = .049$ ,  $\eta^2 = 0.012$ ,  $d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582$ ,  $p < .001$ ,  $\eta^2 = 0.067$ ,  $d = 0.496$ ). 0.441).

**6.3.7 Posture.** A 2x2 ANOVA (feedback type [social signal feedback vs traditional feedback] vs session [pretraining interview vs post-training interview]) revealed a no significant main effect for session ( $F_{(1, 641)} = 3.214$ ,  $p = .073$ ,  $\eta^2 = .005$ ) but a significant main effect for feedback type ( $F_{(1, 641)} = 5.703$ ,  $p = .017$ ,  $\eta^2 = .009$ ). There was no significant interaction between session x feedback type ( $F_{(1, 641)} = .415$ ,  $p = .520$ ,  $\eta^2 = .001$ ). Follow-up analysis using a Welch's statistic revealed that there was no significant difference between social signal feedback and traditional feedback types in the pre-training interview ( $F_{(1, 241.169)} = .970$ ,  $p = .396$ ,  $\eta^2 = 0.004$ ). A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158$ ,  $p < .001$ ,  $\eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780$ ,  $p < .001$ ,  $\eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460$ ,  $p < .001$ ,  $\eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449$ ,  $p = .001$ ,  $\eta^2 = 0.046$ ,  $d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007$ ,  $p = .933$ ,  $\eta^2 = 0.000$ ,  $d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899$ ,  $p = .049$ ,  $\eta^2 = 0.012$ ,  $d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582$ ,  $p < .001$ ,  $\eta^2 = 0.067$ ,  $d = 0.496$ ). 0.120) but there was a significant difference observed in the post training interview ( $F_{(1, 348.603)} = 6.976$ ,  $p = .009$ ,  $\eta^2 = 0.019$ ). A 2x2 ANOVA revealed a significant main effect for session ( $F_{(1, 642)} = 28.158$ ,  $p < .001$ ,  $\eta^2 = .042$ ) and feedback type ( $F_{(1, 642)} = 16.780$ ,  $p < .001$ ,  $\eta^2 = .025$ ). There was also a significant interaction between session x feedback type ( $F_{(1, 642)} = 16.460$ ,  $p < .001$ ,  $\eta^2 = .025$ ). Follow-up analysis using a Welch's statistic revealed a difference between feedback types in the pre-training interview ( $F_{(1, 137.151)} = 11.449$ ,  $p = .001$ ,  $\eta^2 = 0.046$ ,  $d = 0.425$ ) but none in the post training interview between groups ( $F_{(1, 362.912)} = .007$ ,  $p = .933$ ,  $\eta^2 = 0.000$ ,  $d = 0.009$ ). The control group reduced displays of frowning / anger from the pre-training to the post-training interview ( $F_{(1, 274.480)} = 3.899$ ,  $p = .049$ ,  $\eta^2 = 0.012$ ,  $d = 0.219$ ). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ( $F_{(1, 129.012)} = 15.582$ ,  $p < .001$ ,  $\eta^2 = 0.067$ ,  $d = 0.496$ ). 0.276).

**6.3.8 6-Month Follow-up Subjective Ratings of Performance.** The interclass correlation was conducted to assess the amount of rating agreement between the trainer and three neutral observers. The average measure intraclass correlation was .972 with a 95% confidence interval from .936 to .989 ( $F_{(15, 30)} = 34.148$ ,  $p < .001$ ,  $\eta^2 = 0.245$ ) which suggests high level of agreement of trainee performance. Neutral observers rated participants who had received social signal feedback ( $M = 6.458$ ;  $SD = .460$ ) as significantly better social communicators than those who had received traditional feedback ( $M = 5.333$ ,  $SD = 1.419$ ),  $F_{(1,14)} = 4.553$ ,  $p = .05$ ,  $\eta^2 = .245$ .

## 7 DISCUSSION

The aim of this study was to investigate whether nonverbal feedback using commercial automated affect recognition technology is more effective in improving communication skills in media interview training than traditional methods of media interview training. The results obtained are in support of H1, H2, H4 and H5 where H1 proposed that there will be significant improvements in subjective ratings of performance from pre-test to post-test interview in both training conditions. H2 predicted that there will be significant changes in observed social signals detected between pre-test and post-test interview in both training conditions. H4 predicted that there will be greater changes in social signals detected between pre-test and post-test interview for the experimental condition. Finally, the H5 predicted that there would be better subjective ratings of performance in favor of the experiment group. Subjective ratings of trainee performance and behavioral modifications by feedback are discussed considering previous research.

### 7.1 Subjective Ratings of Performance

Subjective judgements of communication skills provided by neutral observers illustrate a clear improvement in both feedback groups, supporting H1. According to subjective ratings neutral observers rated the improvement of social signal feedback participants at 20%, compared to 15% for standard training; however, the interaction effect was not statistically significant. The current data set does not provide statistical evidence for H3 in the initial study. This could be due to the subjective nature of the ratings used by the neutral observers, even though both received training on how to answer the scale. However, even though there were differences in how each condition displayed signals during the interview, it could be that this did not affect how an audience and a journalist perceive their effectiveness of communication. However, a larger effect size was found for improvements from pre-training to post-training interview for trainees who received social signal training ( $d = 1.064$ ) than for those who received traditional feedback ( $d = 0.614$ ).

When participants were asked to score how much they had improved directly, it was found that participants who received social signal feedback rated their skills and confidence higher post-training (on average citing 'considerable improvement') compared to participants that did not receive social signal feedback (on average citing 'some improvement' according to the scale labels). This provides partial support of hypothesis 3. Previous research suggests that communication training increases confidence as well as perception of its effectiveness [23]. In the 6-month follow-up interview,

the experiment group was rated more effective communicators by neutral observers. This result is in support of H5. This suggests that the skills gained in the original training session were long lasting as they were observed in the social signal group after 6 months.

## 7.2 Social Signals

The first component extracted by a PCA were derived from vocal behaviour analysis and included scores for 'rationality', 'stressed', 'excited', 'energy', 'logic' (loading positively) and 'imaginative thinking', 'upset', 'imagination' and 'uncertain' (loading negatively). These have been labelled as confidence based on the included signals. The results revealed that those who received traditional feedback had a significant reduction in confidence from pre-training interview to post training interview. This finding is partially supportive of H4, the control group were not as confident as social signal group in the pre-training interview. Those that received social signal feedback (experiment group) did not change in confidence score from pre to post-test resulting a relatively positive outcome compared to the control group. The result could suggest that the control group were not able to reflect on their feedback as much as the social signal group who received additional, more in-depth feedback due to the nature of the summative feedback. However, trainees who received social signal feedback had rated their skills and confidence higher than the control group. This finding was also found in [70].

The second component extracted was disgust. The features included both EMFACS and facial expression classifiers resulting from Affdex expression recognition software. This included disgust, jaw drop, upper lip raises and nose wrinkle (AU9, with some AU4 and AU10). Both groups displayed a reduction in disgust from pre-training to post-training interview which is partially in support of H2. This could be a consequence of the AU involved in displays of disgust, which include nose wrinkler, lip corner depressor (AU15) (AU26 with some AU25) and low lip depressor (AU16 with some AU25). Jaw drop (AU26 with some AU25) could have been captured due to its role in the act of verbal communication during conversation. Disgust is often confused with fear [34], while fear could be presented in this context, it may be present in micro expressions. There was a reduction which could suggest that participants became less fearful of the situation from the pre-training interview to the post-training interview. Empirically, disgust has also been confused with anger and contempt [27]. It has been suggested that the confusion resulting from facial expression of disgust could be a consequence of culture-specific [10]. Overall, the expression of disgust is revealing of a negative emotion suggesting that both groups were aware of their facial expressions which could be a result of watching videos of their interviews.

The third component extracted was anger / frowning. The signals included in this component included both facial expressions classifiers and EMFACS captured by Affdex which included brow furrow (AU4), lid tighten (AU7), anger and sadness. Both groups had significantly reduced displays of frowning in post-training. The experiment group displayed more frowning than the control group in the pre-training interview suggesting scope for improvement. This result is in support of H4. Furthermore, trainees that received social signal feedback did not frown as much in the post

training interview compared to the control groups. This suggests that the summative feedback provided to the experiment group was effective in altering the behaviour. Media skills training guides suggest that frowning during an interview is negatively perceived by the audience [61]. From this it can be implied that social signal feedback helped trainees to improve their social skills in how they are perceived which may have been acquired from watching their interviews back for reflection and summative feedback that emphasizes a threshold of good and bad performance. This was a key finding as previous research found no significance for the implementation of thresholds during feedback because trainees did not cross the threshold [18].

The fourth component extracted using PCA was engagement. The features included are made up facial expression classifiers and EMFACS that include engagement, surprise and brow raise (AU1). Trainees who received social signal feedback displayed more facial expression engagement in the pre-training interview and in the post-training interview compared to trainees that received standard feedback training. This result is consistent with H4. It suggests that social signal feedback provision encouraged trainees to use more facial expressions in their conversations with the journalist. Previous research has shown that engagement is required for effective communication [42, 48, 67].

The fifth component extracted was 'posed expression'. This component was derived from facial expression classifiers captured by Affdex. Included were dimpler (AU14), lip stretcher (AU20), lip press (AU24) and lip suck (AU28). Those who received traditional training feedback had displayed more dimpler (AU14) expression in the post-training interview. This result is partially supportive of H4. However, the presentation of dimpler (AU14) has been found to be over exaggerated during posed expressions facial actions [38], suggesting that expressions shown in the post-training interview were posed. This may signpost boredom or fatigue after training [61]. Similarly, the expression of dimpler (AU14) is the main AU that makes up contempt which is often confused with disgust which was a PCA component. This may have been noticed by subjective raters and could be an explanation for the non-significant results between groups obtained using subjective ratings of performance.

Signals included in the posture component were honest signals resulting from sociometric badges detection of the interaction. Signals included were movement and posture. Those who received standard training feedback displayed a relaxed posture compared to trainees who received social signal feedback in the post-training interview. This result supports H4. Research suggests that a relaxed posture is directly linked to attitude and signifies boredom or aloofness [61, 64]. This could imply that those who received traditional feedback may have been bored after training session or that they were comfortable enough in the final interview. This is like the finding where those who received traditional feedback displayed more posed expressions, suggesting that they were bored. Research has also found that a straightened posture is associated with higher cognitive engagement and an increase in self-awareness [33, 64], similar findings were found in [40]. This suggests that trainees in the experiment group were more attentive and engaged after in the post training interview.

It is noted that this study includes a small sample size. Future work could increase the sample size. Furthermore, it is possible

that ceiling effects were present where trainees had displayed effective communication skills in the pre-training interview; thereby receiving a high score (such as scores 6 – 7 out of 7 for communication skills). Receiving a high score does not leave much room for improvement which impacts on the results obtained. Research suggests that trainees who perform very poor in the pre-training interview / baseline are more likely to gain skills from training than those who scored high baseline scores [5]. Future research could control for high scores obtained in the pre-training interview by excluding them. Additionally, it would be interesting in future work to include expert as well as novice participants. Even though this population may be representative of a professional role, it would be interesting to confirm these findings with trainees in other types of organisations.

## 8 CONCLUSION

This study allowed researchers to gain details surrounding the provision of social signal feedback using a combination of feedback techniques. The results of this study highlight key points when implementing a technology enhanced method of communication skills training. The use of COTS technology in training is effective and is consistent with the literature. It proposes an alternative method to an already practical method, the behavioral feedback loop. Rather than providing real-time feedback, which could be distracting, the technique used in this study provides a summary of the behaviours displayed and improves self-awareness through formative feedback (video playback). This permits discussion of performance with the trainer and reflection of behaviour. Even though there were limitations surrounding re-test timing there were signs of habituation due to the differences between the two groups in social signal displays. Future work could improve this training by implementing a longer timeframe between practice interviews, excluding trainees that initially displayed good levels of communication skills performance and provide a concrete behavioral threshold for what is classified as a 'good performance'.

## ACKNOWLEDGMENTS

This work was jointly supported by the Defence, Science and Defence Lab and the Engineering and Physical Science Research Council.

## REFERENCES

- [1] Jans Aasman, Gijsbertus Mulder, and Lambertus J.M. M Mulder. 1987. Operator effort and the measurement of heart-rate variability. *Human Factors* 29, 2: 161–170.
- [2] Reginald B. Adams and Kestutis Kveraga. 2015. Social Vision: Functional Forecasting and the Integration of Compound Social Cues. *Review of Philosophy and Psychology* 6, 4: 591–610.
- [3] Mohammad Rafayet Ali and Ehsan Hoque. 2017. Social skills training with virtual assistant and real-time feedback. 325–329.
- [4] P. Allen and K. Bennett. 2007. *SPSS for the Health and Behavioural Sciences*. Cengage Learning Australia.
- [5] Knut Aspegren. 1999. BEME Guide No. 2: Teaching and learning communication skills in medicine - A review with quality grading of articles. *Medical Teacher* 21, 6: 564–570.
- [6] Roghayeh Barmaki. 2016. Improving social communication skills using kinesics feedback. *Conference on Human Factors in Computing Systems - Proceedings 07-12-May-*: 86–91.
- [7] Tobias Baur, Gregor Mehlmann, Ionut Damian, et al. 2015. Context-Aware Automated Analysis and Annotation of Social Human-Agent Interactions. *ACM Transactions on Interactive Intelligent Systems (TiIS)* 5, 2: 11.
- [8] Paul Boersma and Vincent van Heuven. 2001. Speak and unSpeak with Praat. *Glott International* 5, 9–10: 341–347.
- [9] Willem Paul Brinkman, Joost Broekens, Dirk Heylen, et al. 2015. Towards a Socially Adaptive Virtual Agent. *International Conference on Intelligent Virtual Agents*, 3–16.
- [10] Andrew J. Calder and Andrew W. Young. 2005. Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience* 6, 8: 641–651.
- [11] Rafael Calvo, Sidney D’Mello, Jonathan Gratch, et al. 2015. Expressing Emotion Through Posture and Gesture. In *The Oxford Handbook of Affective Computing*. 3–10.
- [12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16: 321–357.
- [13] Mathieu Chollet, Pranav Ghate, and Stefan Scherer. 2018. A generic platform for training social skills with adaptive virtual agents. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*.
- [14] Mathieu Chollet, Torsten Wörtwein, Louis-Philippe Morency, Ari Shapiro, and Stefan Scherer. 2016. Exploring feedback strategies to improve public speaking. September 2016: 1143–1154.
- [15] Marco Cristani, R Raghavendra, Alessio Del Bue, and Vittorio Murino. 2013. Human behavior analysis in video surveillance: A Social Signal Processing perspective. *Special issue: Behaviours in video* 100: 86–97.
- [16] Ionut Damian, Tobias Baur, Birgit Lugin, Patrick Gebhard, Gregor Mehlmann, and Elisabeth André. 2015. Games are better than books: In-situ comparison of an interactive job interview game with conventional training. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9112: 84–94.
- [17] Ionut Damian, Michael Dietz, and Elisabeth André. 2018. The SSJ Framework: Augmenting Social Interactions Using Mobile Signal Processing and Live Feedback. *Frontiers in ICT* 5: 13.
- [18] Ionut Damian, Chiew Seng (Sean) Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting Social Interactions: Realtime Behavioural Feedback Using Social Signal Processing Techniques. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*: 565–574.
- [19] A. C. Daniels. 1989. *Performance Management*. Performance Management Publications, Atlanta, GA.
- [20] Paul Ekman. 2003. Sixteen enjoyable emotions. *Emotion Researcher* 18, 2: 6–7.
- [21] Andy Field. 2013. *Discovering statistics using IBM SPSS statistics*. Sage Publications.
- [22] Scott T. Frein, Samantha L. Jones, and Jennifer E. Gerow. 2013. When it comes to Facebook there may be more to bad memory than just multitasking. *Computers in Human Behavior* 29, 6: 2179–2182.
- [23] Sakiko Fukui, Keiko Ogawa, and Naoshi Fukui. 2010. Communication skills training on how to break bad news for Japanese nurses in oncology: Effects of training on nurses’ confidence and perceived effectiveness. *Journal of Cancer Education* 25, 1: 116–119.
- [24] Ruth Garner, Mark G. Gillingham, and C. Stephen White. 1989. Effects of &quot;Seductive Details&quot; on Macroprocessing and Microprocessing in Adults and Children. *Cognition and Instruction* 6, 1: 41–57.
- [25] Young Han and Joseph Nunes. 2010. Read the Signal But Don’t Mention It: How Consumption Embarrasses the Signaler. *Advances in Consumer Research* 37: 81–84.
- [26] Yuval Hart, Efrat Czerniak, Orit Karnieli-Miller, et al. 2016. Automated video analysis of non-verbal communication in a medical setting. *Frontiers in Psychology* 7, AUG: 1130.
- [27] Minoru Hashimoto, Misaki Yamano, and Tatsuya Usui. 2009. Effects of emotional synchronization in human-robot KANSEI communications. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*: 52–57.
- [28] R. Glen Hass and Donna Eisenstadt. 1990. The effects of self-focused attention on perspective-taking and anxiety. *Anxiety Research* 2, 3: 165–176.
- [29] Benjamin C. Holding, Tina Sundelin, Mats Lekander, and John Axelsson. 2019. Sleep deprivation and its effects on communication during individual and collaborative tasks. *Scientific Reports* 9, 1: 3131.
- [30] Me Hoque, Matthieu Courgeon, and Jc Martin. 2013. Mach: My Automated Conversation coach. *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*, 697–706.
- [31] Mohammed Hoque and Rosalind W Picard. *Automated Coach to Practice Conversations*. .
- [32] Hazaël Jones and Nicolas Sabouret. 2013. *TARDIS - A simulation platform with an affective virtual recruiter for job interviews*. .
- [33] Johanna K. Kaakinen, Ugo Ballenghein, Geoffrey Tissier, and Thierry Baccino. 2018. Fluctuation in cognitive engagement during reading: Evidence From Concurrent Recordings of Postural and Eye Movements. *Journal of Experimental Psychology: Learning Memory and Cognition* 44, 10: 1671–1677.
- [34] Douglas M Klieger and Kimberly K Siejak. 1997. Disgust as the source of false positive effects in the measurement of ophidiophobia. *Journal of Psychology: Interdisciplinary and Applied* 131, 4: 371–382.
- [35] L. Konopka, J. Duffecy, and K. Hur. 2010. Technology assisted PTSD screening (TAPS) for the Vietnam Veterans of America (VVA) study. *Pilot Study Abstract*,

- The Chicago School of Professional Psychology, Northwestern University, and The University of Illinois at Chicago.*
- [36] Francisco Lacerda. 2009. LVA-technology-The illusion of “lie detection.” *Proceedings, FONETIK 2009*: 220–226.
- [37] Francisco Lacerda. 2012. Money Talks: The Power of Voice: A critical review of Mayew and Ventachalam’s The Power of Voice: Managerial Affective States and Future Firm Performance. *PERILUS*: 1–10.
- [38] G Littlewort, M S Bartlett, and K Lee. 2008. *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction*. Springer Verlag.
- [39] Chunfeng Liu, Karen M Scott, Renee L Lim, Silas Taylor, and Rafael A Calvo. 2016. EQClinic: a platform for learning communication skills in clinical consultations. *Medical Education Online* 21: 1–7.
- [40] Markus Muehlhan, Michael Marxen, Julia Landsiedel, Hagen Malberg, and Sebastian Zaunseder. 2014. The effect of body posture on cognitive performance: a question of sleep quality. *Frontiers in Human Neuroscience* 8: 171.
- [41] Skanda Muralidhar. 2017. On Job Training: Automated Interpersonal Behavior Assessment & Real-Time Feedback. .
- [42] Iftekhar Naim, Md. Iftekhar Tanveer, Daniel Gildea, et al. 2016. Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing*: 1–14.
- [43] Laurent Son Nguyen and Daniel Gatica-Perez. 2015. I Would Hire You in a Minute. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15*, 51–58.
- [44] S.Gopal Krishna Patro and Kishore Kumar Sahu. 2015. Normalization: A Preprocessing Stage. *IARJSET*: 20–22.
- [45] Alex Pentland and Tracy Heibeck. 2010. *Honest signals: how they shape our world*. MIT press, Cambridge, Massachusetts.
- [46] Monica Pereira. 2018. Media Skills Training Intervention Based on Automated Recognition of Human Emotion and Non-Verbal Behaviour. 1: 1.
- [47] Monica Pereira, Federico Colecchia, and Kate Hone. 2018. User Centred Design of Social Signals Feedback for Communication Skills Training. *Proceedings of British HCI, 2018*: 1–5.
- [48] Namita Mediratta Rana el Kaliouby, Evan Kodra, Pankaj Jha. 2014. *Does Facial Coding Generalize Across Cultures? A Spotlight on Asia*. .
- [49] Ognjen Rudovic, Mihalisa A. Nicolaou, and Vladimir Pavlovic. 2017. Machine learning methods for social signal processing. *Social Signal Processing*: 234–254.
- [50] Christopher A. Sanchez and Jennifer Wiley. 2006. An examination of the seductive details effect in terms of working memory capacity. *Memory and Cognition* 34, 2: 344–355.
- [51] Emanuel Schmider, Matthias Ziegler, Erik Danay, Luzi Beyer, and Markus Bühner. 2010. Is It Really Robust?: Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology* 6, 4: 147–151.
- [52] Jan Schneider, Dirk Börner, Peter Van Rosmalen, and Marcus Specht. 2016. Can You Help Me with My Pitch? Studying a Tool for Real-Time Automated Feedback. *IEEE Transactions on Learning Technologies* 9, 4: 318–327.
- [53] Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht. 2015. Presentation Trainer, your Public Speaking Multimodal Coach. *ICMI '15: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*: 539–546.
- [54] Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht. 2016. Enhancing public speaking skills - An evaluation of the presentation trainer in the wild. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer, Cham, 263–276.
- [55] By Jamie Shotton, Toby Sharp, Alex Kipman, et al. 2011. *Real-Time Human Pose Recognition in Parts from Single Depth Images*. .
- [56] Brian H Spitzberg and Thomas W Adams. 2007. CSRS, the Conversational Skills Rating Scale: An Instructional Assessment of Interpersonal Competence. *NCA, National Communication Association*.
- [57] Jan Van den Stock, Ruthger Righart, and Beatrice de Gelder. 2007. Body Expressions Influence Recognition of Emotions in the Face and Voice. *Emotion* 7, 3: 487–494.
- [58] B. G. Tabacknick, L. S. Fidell, and J. B. Ullman. 2007. *Using Multivariate Statistics*. Pearson, Boston, MA.
- [59] Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2016. Automated social skills training with audiovisual information. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, IEEE*, 2262–2265.
- [60] Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, and Satoshi Nakamura. 2015. NOCOA+: Multimodal computer-based training for social and communication skills. *IEEE Transactions on Information and Systems* E98D, 8: 1536–1544.
- [61] Robert Taylor. 2015. *Media Interview Techniques: A complete guide of media training*. Kogan Page, London.
- [62] Tatsuya Usui, Kazuomi Kume, Misaki Yamano, and Minoru Hashimoto. 2008. A robotic KANSEI communication system based on emotional synchronization. *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE*, 3344–3349.
- [63] Alessandro Vinciarelli, Maja Pantic, and H Bourlard. 2008. Social signals, their function, and automatic analysis: a survey. *Proceedings of the 10th international conference on Multimodal interfaces (ICMI '08)*: 61–68.
- [64] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and Vision Computing* 27, 12: 1743–1759.
- [65] Alessandro Vinciarelli, Hugues Salamin, and Maja Pantic. 2009. Social signal processing: Understanding social interactions through nonverbal behavior analysis. *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, IEEE, 42–49.
- [66] Robert A. Wicklund. 1979. The Influence of Self-Awareness on Human Behavior: The person who becomes self-aware is more likely to act consistently, be faithful to societal norms, and give accurate reports about himself. *American Scientists* 67, 2: 187–193.
- [67] World Health Organization and WHO. 2005. Effective media communication during public health emergencies - A WHO Handbook. *World Health July*.
- [68] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1: 39–58.
- [69] Yanxia Zhang, FX Palo Alto Laboratory, Usa Jeffrey Olenick, et al. 2018. TeamSense: Assessing Personal Affect and Group Cohesion in Small Teams through Dyadic Interaction and Behavior Analysis with Wearable Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol* 2, 150: 150.
- [70] Ru Zhao, Vivian Li, Hugo Barbosa, Gourab Ghoshal, and Mohammed Ehsan Hoque. 2017. Semi-Automated 8 Collaborative Online Training Module for Improving Communication Skills. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2: 1–20.