Facing Employers and Customers: What Do Gaze and Expressions Tell About Soft Skills?

Skanda Muralidhar
Idiap, EPFL
Switzerland
smuralidhar@idiap.ch

Rémy Siegfried
Idiap, EPFL
Switzerland
remy.siegfried@idiap.ch

Jean-Marc Odobez
Idiap, EPFL
Switzerland
odobez@idiap.ch

Daniel Gatica-Perez
Idiap, EPFL
Switzerland
gatica@idiap.ch

ABSTRACT
Eye gaze and facial expressions are central to face-to-face social interactions. These behavioral cues and their connections to first impressions have been widely studied in psychology and computing literature, but limited to a single situation. Utilizing ubiquitous multimodal sensors coupled with advances in computer vision and machine learning, we investigate the connections between these behavioral cues and perceived soft skills in two diverse workplace situations (job interviews and reception desk). Pearson's correlation analysis shows a moderate connection between certain facial expressions, eye gaze cues and perceived soft skills in job interviews ($r \in [-30, 30]$) and desk ($r \in [20, 36]$) situations. Results of our computational framework to infer perceived soft skills indicates a low predictive power of eye gaze, facial expressions, and their combination in both interviews ($R^2 \in [0.02, 0.21]$) and desk ($R^2 \in [0.05, 0.15]$) situations. Our work has important implications for employee training and behavioral feedback systems.

ACM Classification Keywords
I.4.m. Image Processing and Computer Vision: Miscellaneous; J.4 Social and Behavioral Sciences

Author Keywords
First impressions;hirability;job performance; eye gaze; facial expressions;social computing; hospitality; multimodal interaction;

INTRODUCTION
Eye gaze and facial expressions are the foundations of communication and social interaction in humans. They are also the easiest nonverbal cues to signal interest; a person who avoids eye contact doesn't want to be bothered while if he/she smiles and makes eye contact, it indicates his/her willingness to interact. Previous works have explored the interconnections between eye gaze, facial expressions, and various social variables including from hirability [15, 21], sales performance [26], dominance [20] and leadership [34] in the context of a single situation in workplace. In this work, using ubiquitous multimodal sensors and advances in machine learning, we investigate the connections between these nonverbal behavioral cues and first impressions across two diverse workplace settings; job interviews and reception desk.

Literature in psychology has shown that eye gaze and facial expressions play a role in the formation of first impressions in workplaces [24]. In the context of employment interviews, Forbes et al. [15] reported that successful applicants in job interviews made more direct eye contact, produced more facial expressions, smiled and nodded more than unsuccessful ones. In the context of customer-centric service too, these behavioral cues have been reported to be important. In particular, Tsai et al. [45] found that employees' smiling and eye contact were related to the customers' greater willingness to revisit the store. Other studies have shown links between frequent and longer eye contact and increased likability [5], satisfaction [23], credibility and trust [18].

Psychology literature traditionally relies on manual coding of eye gaze and facial expressions, which is an expensive, time consuming and labor-intensive process. With advances in mobile and ubiquitous sensors, capturing eye gaze and facial expressions are becoming inexpensive and common (e.g. Apple's Animoji). Using such advances, eye gaze and facial expressions have been investigated in computing literature. In the context of workplace settings, various social variables including dominance [22, 19], leadership [35], personality [3] have been investigated from the perspective of these cues. In the context of job interviews, Chen et al. [9] investigated the eye gaze, facial expressions and head postures. Using a dataset of 36 videos, the authors reported a correlation of $r = 0.36$ between expert hirability ratings and visual features represented as "visual words". 7 classes of facial expressions

MUM '18, November 25–28, 2018, Cairo, Egypt

ISBN 978-1-4503-6594-9/18/11…$15.00
DOI: https://doi.org/10.1145/3282894.3282925

(a) (b)

Figure 1: Snapshot of the situations studied in this work: (a) job interview, (b) reception desk

and their connections to interview performance was explored Chen et al. [10] using 20 structures interviews. The authors reported a positive correlation ($r = 0.52$) between joy and hirability, while anger was negatively correlated ($r = -0.44$).

Due to the challenges of collecting and analyzing multimodal data, most existing automated prediction frameworks focus on a single setting. In this work, we study the connections between eye gaze, facial expressions and first impressions in the context of two workplace settings; job interviews and hotel reception desk. Specifically, we investigate the connections between eye gaze, facial expressions and perceived hirabilty in interviews and perceived performance on the job. Towards this, we utilize a previously collected data corpus consisting of 169 videos each from job interviews and desk interactions.

The contribution of this work are: (1) We extract eye gaze and facial expressions using state of the art computer vision and machine learning. (2) In the job interview setting, a Pearson's correlation analysis revealed (a) moderate positive correlation between eye gaze and perceived hirabilty ($r = 0.28$), (b) negative correlation between neutral face and perceived hirabilty ($r = -0.28$). (3) In the reception desk setting, a Pearson's correlation analysis revealed moderate positive correlation between (a) eye gaze and perceived performance ($r = 0.36$), (b) anger and perceived performance ($r = 0.24$). (4) Our computational framework to automatically infer perceived soft skills showed low to moderate inference performance for interviews ($R^2 \in [0.02, 0.21]$) and desk ($R^2 \in [0.05, 0.15]$) situations.

Our results demonstrate the feasibility of quantifying such behavioral cues and their contributions to perceived soft skills in various settings, which have previously been reported mainly qualitatively in psychology literature using advances in ubiquitous multimodal senors and machine learning.

DATASET

Data Collection

This dataset consists of 169 interactions previously collected by us [27], role-played by students of a hospitality school in two workplace settings; job interview and reception desk (total of 338 videos). Both the settings were recorded synchronously using (1) two Kinect cameras (one for each protagonist) at 30 frames per second; (2) a microphone array placed at the center of the table recording audio at 48kHz. Due to challenges in feature extraction due to non-frontal body pose, this work uses 161 videos for interview and 153 for desk. This dataset was annotated by us previously [27] for various social variables (Table 1). The annotations for interview videos was done by 5 raters, while for desk there were 3 raters. Both groups were mutually exclusive and consisted of students, who were paid approximately $20USD$ per hour for this work. The raters watched the first two minutes of the videos and annotated the variables on a seven-point Likert scale. Use of two-minute segments, also known as thin slices, follows existing literature [2, 33]. Details of the data collection can be found in [27].

Table 1: List of perceived variables manually annotated for both situations, along with their $ICC(2,k)$ and means.

	Job Interview		Reception Desk	
Variable	$ICC(2,k)$	**Mean**	$ICC(2,k)$	**Mean**
Professional Skills				
Competent (compe)	0.56	6.01	0.69	4.24
Motivated (motiv)	0.52	5.89	0.63	4.80
Social Skills				
Positive (posit)	0.60	5.70	0.60	4.34
Sociable (socia)	0.57	5.67	0.64	4.46
Communication Skills				
Clear (clear)	0.67	5.89	0.66	4.56
Persuasive (persu)	0.69	5.57	0.72	4.01
Overall				
Performance (peImp)	–	–	0.77	4.11
Hirability (hire)	0.69	5.54	–	–

Table 1 presents the inter-rater agreement and their descriptive statistics. Agreement between raters was measured using Intraclass Correlation Coefficient (ICC) [37], a common metric used in psychology. Since annotations were done using a sample of annotators and each annotator judged all videos, we use $ICC(2,k)$. The tables shows that the inter-rater agreement was acceptable with $ICC(2,k) > 0.5$. Specifically, we observe moderate to high agreement between the raters with $ICC(2,k) \in [0.52, 0.77]$ for interview videos and $ICC(2,k) \in [0.60, 0.77]$ for reception desk. Further, we observe that the mean values of the annotated variables are centered on the positive side of the Likert scales (Mean ≥ 4), suggesting that the participants were generally perceived positively by both groups of annotators.

FEATURES EXTRACTED

In the section we outline the extraction of eye gaze and facial expression cues. A number of statistics, including count, mean, median, standard deviation, minimum, and maximum were computed from these cues and used as features.

Eye Gaze

We extract eye gaze using an off-the-shelf method to compute participants' gaze (i.e. line of sight in 3D space) from a Kinect v2 sensor [16]. Taking advantage of the context (i.e. a human-human interaction), we improved its outputs using a subject adaptation method that computes the gaze estimation bias based on speaking turns [38]. An overview of this method is presented in Figure 2 and follows the intuition that people are more likely to look at the person who is speaking during a conversation. The gaze errors corresponding to the frames where the other person is speaking and where the estimated gaze and head pose make a glance at the speaker possible are stored. Then, the bias is estimated using the Least Median Square (LMedS) estimator, so that remaining outliers frames (i.e. frames where the subject is not looking at the other person despite the fact that he/she is talking) are removed. Finally, the bias can be subtracted from the gaze signal, which allows us to compute the Visual Focus of Attention (VFOA) with an

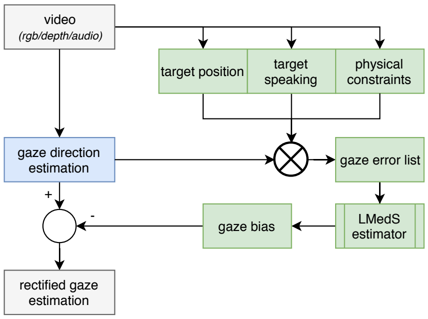

Figure 2: Gaze bias correction method using speaking turns.

higher accuracy without the need to for manual calibration or annotation. This method allowed to reach a mean error of 7.64°on annotated subsets of the interview and desk datasets.

Using this data, we extract the various eye gaze cues, defined as number and duration of events when participant was gazing at the protagonist, from both client and participant. These include gaze while speaking (*GWS*; number and duration of looking at the protagonist and speaking), gazing while listening (*GWL*; number and duration of looking at the protagonist and not speaking), visual dominance ratio (*VDR* defined as ratio of GWS and GWL [14]), and mutual gaze.

Facial Expressions

Literature in psychology indicates connections between facial expressions and impressions of dominance [30, 43], warmth [4] and emotional distance [7]. We extract facial expressions displayed from using "off-the-shelf" Emotion API [12] of Microsoft Azure cognitive services [42, 41]. Such cloud-based services have been used in literature to study diverse social, and political issues [36] including cyber-bullying [39] and public health images [17].

Microsoft Azure Emotion API recognizes 8 facial expressions of emotions (happiness, sadness, surprise, anger, fear, contempt, disgust and neutral). As first step, we identify segments in the videos where the protagonist is not speaking so as to avoid mis-classifications of emotion classes (due to movement of facial muscles and lips). From these segments, we extracted frames at 5 fps, which are then input sequentially to the Emotion API to maintain temporal continuity. If the API detected a face in the frame, its returns confidence values across the 8 emotion types normalized to 1. Else if a face was not found, the API returns 0 for all values. Frames with no faces were filtered before computing various statistics was computed from this 8-dimensional vector and used as features.

Other Features

To further understand the impact of gaze and facial expressions, we combine them with other visual and auditory cues previously extracted by us in [27, 28]. **Visual features** consists of (1) *Overall visual motion* captures the total amount of visual movement displayed [6]; and (2) *Head nods* were extracted using a 3D face centered method [11]. Apart from simple head nods, nodding while speaking was extracted by synchronizing speaking activity with head nod. **Auditory features** include (1) *Speaking activity* includes speaking time, speaking turns, pauses, short utterances, and silence; (2) *Prosody* includes pitch, speaking rate, spectral entropy, energy, voicing rate, and time derivative of energy, extracted using [8, 32].

Table 2: Range of Pearsons correlation between eye gaze, facial expressions and social variables in the interview ($N = 161$) and reception desk ($N = 153$) dataset. $^{***}p < 0.001$;$^{**}p < 0.01$;$^{*}p < 0.05$.

	Professional	Social	Communication	Hirability / Performance
Job Interview				
Applicant				
Std Duration (GWS)	$.15$	$[.15,.18^{*}]$	$[.18,.22]^{**}$	$.17^{*}$
Max Duration (GWS)	$[.16,.17]^{*}$	$[.16,.19^{*}]$	$[.18,.22]^{**}$	$.18^{*}$
Num of GWL	$[-.18^{*},-.11]$	$[-.25,-.28]^{***}$	$[-.30,-.23]^{**}$	$-.22^{**}$
Mean Duration (GWL)	$[.15,.23^{**}]$	$[.12,.17^{*}]$	$[.21,.23]^{**}$	$.16^{*}$
Std Duration (GWL)	$[.13,.16]$	$[.17,.20]^{*}$	$.22^{**}$	$.20^{*}$
VDR	$[.22,.28]^{***}$	$.27^{***}$	$[.17,.25]^{**}$	$.28^{***}$
Facial Expression				
Mean Neutral	$[-.32,-.22]^{**}$	$[-.43,-.40]^{***}$	$[-.24,-.22]^{**}$	$-.30^{***}$
Median Neutral	$[-.27,-.20]^{**}$	$[-.33,-.30]^{***}$	$-.18^{*}$	$-.24^{**}$
Std Neutral	$[.15,.28^{***}]$	$.43^{***}$	$[.19,.24]^{*}$	$.28^{***}$
Var Neutral	$.12,.27^{***}$	$.42^{***}$	$[.18,.23]^{*}$	$.26^{***}$
Mean Happy	$.31^{***}$	$[.41,.42]^{***}$	$[.22,.26]^{**}$	$.30^{***}$
Median Happy	$.25^{**}$	$[.27,.28]^{***}$	$.19^{*}$	$.22^{**}$
Std Happy	$.25^{**}$	$.41^{***}$	$[.16,.22]^{*}$	$.25^{**}$
Var Happy	$.25^{**}$	$[.40,.41]^{***}$	$[.16,.22]^{*}$	$.25^{**}$
Reception Desk				
Receptionist				
Num of GWS	$[.27,.34]^{***}$	$[.26,.32]^{***}$	$[.24,.28]^{***}$	$.30^{***}$
Max Duration (GWS)	$[.14,.21^{*}]$	$[.21,.22]^{**}$	$[.22,.24]^{**}$	$.20^{*}$
Min Duration (GWS)	$[.17,.18^{*}]$	$[.17,.19]^{*}$	$[.15,.20^{*}]$	$.21^{*}$
VDR	$[.31,.32]^{***}$	$[.28,.32]^{***}$	$[.24,.31]^{***}$	$.36^{***}$
Facial Expression				
Std Anger	$[.21,.27]^{**}$	$[.16,.20^{*}]$	$[.20,.21]^{*}$	$.22^{**}$
Var Anger	$.21^{*}$	$[.16,.19^{*}]$	$[.19,.19]^{*}$	$.22^{**}$
Max Anger	$[.22,.26]^{**}$	$[.18,.20]^{*}$	$[.20,.22]^{**}$	$.24^{**}$

CORRELATION ANALYSIS

In the first step, we conduct a Pearson's correlation analysis between perceived soft skills scores, gaze and facial expressions for the two settings (Table 2). In the interview setting, we observe low to moderate correlations between eye gaze, facial expressions and perceived soft skills. Specifically, we observe that applicants who had greater VDR (i.e amount of visual dominance) and presented a happy expression were perceived to be more hirable than applicants who presented a neutral expression and displayed lower VDR (i.e looked less at the interviewer while speaking). Our results concur with literature, which report positive correlation between eye gaze and hirabilty [15, 21, 1], and happy and hirability [21, 10].

Similarly, in the desk setting moderate correlations between eye gaze, facial expressions and perceived soft skills were observed. In particular, we observe that participants who held client's gaze for longer duration while speaking, displayed greater VDR and nonverbal immediacy (by mirroring anger the clients displayed), were perceived to be better performing then participants who did not display these cues. These connections are supported in literature. It has been shown that even in situations when the client is dissatisfied, greater eye contact leads to enhanced perception of credibility [44]. Soderlund et al. [40] reported that dissatisfied client's assessment of employee's emotional state affects their own emotional state, which, in turn, impacts customer's level of satisfaction.

Interestingly, the protagonists' (interviewer and client respectively) gaze and expression cues were not correlated to perceived hirability and performance. This is in contrast to other nonverbal cues (like speaking time) reported in literature for interviews [29, 27] and reception desk [28].

INFERENCE ANALYSIS

The inference of perceived variables was defined as a regression task and was evaluated using random forest (RF) algorithm from the CARET package [25] in R. Hyper-parameters (i.e., number of trees) were automatically tuned by using an inner 10-fold cross-validation (CV) on the training set. The final machine-inferred scores was obtained by leave-one-video-out

Table 3: Regression analysis using various feature set for job interviews ($N = 161$) and desk ($N = 153$) setting. The baseline performance was obtained for interview [27] and desk settings [28]. It must be noted that these numbers cannot be directly compared due to different N values.

	Clear		Persuasive		Positive		Social		Competent		Motivated		Hirability	Performance
	Interview	Desk	Interview	Desk	Interview	Desk	Interview	Desk	Interview	Desk	Interview	Desk	Interview	Desk
Gaze	0.06	0.00	0.08	0.02	0.06	0.10	0.02	0.01	0.00	0.06	0.04	0.09	0.10	0.08
Facial Expressions	0.0	0.00	0.01	0.04	0.13	0.08	0.17	0.05	0.0	0.00	0.04	0.03	0.04	0.01
Gaze + Expressions	0.06	0.05	0.11	0.12	0.21	0.15	0.18	0.10	0.02	0.07	0.10	0.13	0.13	0.15
All Visual	0.04	0.21	0.15	0.31	0.22	0.32	0.22	0.30	0.06	0.31	0.11	0.25	0.14	0.30
All Features	0.19	0.24	0.30	0.33	0.39	0.34	0.32	0.32	0.20	0.31	0.30	0.29	0.34	0.32
Baseline	0.14	0.22	0.20	0.32	0.30	0.32	0.19	0.33	0.18	0.29	0.29	0.30	0.29	0.30

CV repeated 10 times. It must be noted that each setting was evaluated individually.

Job Interviews

Results of the inference task using the interview data is tabulated in Table 3. We observe that eye gaze individually has low inference performance with best of $R^2 = 0.1$ for *Hirability*. Similarly, the performance of facial expressions is low for *Positive* ($R^2 = 0.13$) and *Social* ($R^2 = 0.17$) and poor for other variables. Combining these two cues, we observe an improved inference performance for some variables. Specifically, for *Positive* ($R^2 = 0.21$) followed by *Social* ($R^2 = 0.18$), *Hirability* ($R^2 = 0.13$), and *Persuasive* ($R^2 = 0.11$).

To further understand the impact of eye gaze and facial expressions, we fuse these features with other visual and auditory features. When fused with other visual cues, we observe further improvement in inference performance for *Positive* ($R^2 = 0.22$) followed by *Hirability* ($R^2 = 0.14$), *Social* ($R^2 = 0.22$), and *Persuasive* ($R^2 = 0.15$). Similarly, fusing gaze and expression features with all the previously extracted features (visual + auditory) shows a moderate improvement in inference performance. Particularly, the improvement was highest for *Social* ($R^2 = 0.32$), followed by *Persuasive* ($R^2 = 0.30$), *Positive* ($R^2 = 0.39$), and *Hirability* ($R^2 = 0.32$).

Our results are in agreement with those in psychology and computing literature. Parsons et al. [31] and Forbes et al. [15] showed that high levels of eye contact had a positive effect on interview outcomes. Chen et al. [9], using an off-the-shelf emotion detection toolkit extracted eye gazes, facial expression, and head postures. The authors, using a visual doc2vec representation for features, reported a correlation of $r = 0.36$. For comparison, we convert r to R^2 (by squaring) obtain $R^2 = 0.13$ and is similar to the results in this work. Overall, we see that gaze and facial expressions have a low-moderate effect on inference performance individually and contribute to improved inference when fused with other features.

Reception Desk

In the desk setting, we observe that eye gaze has low inference performance with the best $R^2 = 0.10$ for *Positive* (Table 3). Facial expressions too had low inference capability with the best $R^2 = 0.08$ for *Positive*. Combing these two cues showed improved inference performance. In particular, the combined set of features had a best performance with $R^2 = 0.15$ for *Positive*, *Performance*, while we also observed improved performance for *Motivated* ($R^2 = 0.13$) and *Persuasive* ($R^2 = 0.12$).

Inference performance improve further when gaze and facial expressions were fused with other visual features. The best performance was obtained for *Positive* ($R^2 = 0.32$) followed by *Persuasive* and *Competent* ($R^2 = 0.31$). This fused set of features obtained $R^2 = 0.30$ for *Performance* and *Social*. As a last step, we fused all the visual features with auditory behavioral cues. We observe a moderate improvement in inference performance (as compared to the baseline) for all the variables except *Social* and *Motivated*. Specifically, we observe improved inference for *Positive* ($R^2 = 0.34$), *Persuasive* ($R^2 = 0.33$), *Social* and *Performance* (both $R^2 = 0.32$).

These results are in accordance with those reported in literature. Leigh et al. [26] evaluated impact of eye gaze and other nonverbal behavior on perceived performance of salespersons. They reported significant effect of eye gaze on perceived believability,tactfulness and empathy. DeGroot et al. [13] investigating perceived performance of 110 managers in a news-publishing company, reported a correlation of $r = 0.14$ between visual cues displayed and performance

Our experimental framework shows that eye gaze and facial expressions have an impact on perceived soft skills, especially in customer-centric situations. We believe that this could be due to job interviews being a more predictable setting, with the applicants seated and hence restrictive in terms of behavioral display. In comparison, reception desk is a more unpredictable and spontaneous interaction with greater need for nonverbal communication to handle an unpleasant client.

CONCLUSION

This work studied the links between eye gaze, facial expressions and perceived soft skills in two diverse workplace settings. Towards this, we utilized a previous dataset of 338 interactions in job interview and reception desk settings. First, using advances in multimodal sensor, computer vision and machine learning, we automatically extracted eye gaze and facial expressions. A Pearson's correlation analysis showed low-to-moderate correlations between eye gaze, facial expressions and perceived soft skills in interviews and moderate correlations in the desk setting. A computation framework to infer perceived soft skills (formulated as a regression task) using these features showed low performance for both the settings. Furthermore, fusion of these cues with other visual and auditory features showed improved inference performance with $R^2 = 0.34$ for *Hirability* in interview setting and $R^2 = 0.32$ for *Performance* in the desk setting. Overall, our results parallel those in literature, and show that eye gaze and facial expressions have an impact in reception desk interaction and to a lesser extent in job interviews. In future, we will investigate same-gender and across-gender differences in gaze and facial expressions. Our work has important implications for employee training and for real-time behavioral feedback systems.

ACKNOWLEDGMENTS

This work was partially funded by the Swiss National Science Foundation (SNSF) through the UBImpressed project.

REFERENCES

1. Joseph G Amalfitano and Neil C Kalt. 1977. Effects of eye contact on the evaluation of job applicants. *Journal of Employment Counseling* 14, 1 (1977), 46–48.
2. Nalini Ambady and Robert Rosenthal. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 111, 2 (1992).
3. Ligia Maria Batrinca, Nadia Mana, Bruno Lepri, Fabio Pianesi, and Nicu Sebe. 2011. Please, tell me about yourself: automatic personality assessment using short self-presentations. In *Proc. ACM ICMI*.
4. Marjorie A Bayes. 1972. Behavioral cues of interpersonal warmth. *J. Consulting and Clinical Psychology* 39, 2 (1972), 333.
5. Steven A Beebe. 1976. Effects of Eye Contact, Posture and Vocal Inflection upon Credibility and Comprehension. (1976).
6. Joan-Isaac Biel and Daniel Gatica-Perez. 2013. The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Trans. on Multimedia* 15, 1 (2013).
7. Judee K Burgoon, Thomas Birk, and Michael Pfau. 1990. Nonverbal behaviors, persuasion, and credibility. *Human Communication Research* 17, 1 (1990), 140–169.
8. Ron Caneel. 2005. *Social signaling in decision making*. Ph.D. Dissertation. Massachusetts Institute of Technology.
9. Lei Chen, Gary Feng, Chee Wee Leong, Blair Lehman, Michelle Martin-Raugh, Harrison Kell, Chong Min Lee, and Su-Youn Yoon. 2016. Automated scoring of interview videos using Doc2Vec multimodal feature extraction paradigm. In *Proc. 18th ACM ICMI*. ACM, 161–168.
10. Lei Chen, Su-Youn Yoon, Chee Wee Leong, Michelle Martin, and Min Ma. 2014. An Initial Analysis of Structured Video Interviews by Using Multimodal Emotion Detection. In *Proc. Workshop Emotion Recognition in the Wild Challenge and Workshop*. ACM, 1–6.
11. Yiqiang Chen, Yu Yu, and Jean-Marc Odobez. 2015. Head Nod Detection from a Full 3D Model. In *Proc. IEEE ICCV Workshops*.
12. Microsoft Cognitive Services. Azure Emotion API. (????). `https://azure.microsoft.com/en-us/services/cognitive-services/emotion/`
13. Timothy DeGroot and Stephan J Motowidlo. 1999. Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *J. Applied Psychology* 84, 6 (1999), 986.
14. John F Dovidio and Steve L Ellyson. 1982. Decoding visual dominance: Attributions of power based on relative percentages of looking while speaking and looking while listening. *Social Psychology Quarterly* (1982), 106–113.
15. Ray J Forbes and Paul R Jackson. 1980. nonverbal behaviour and the outcome of selection interviews. *J. Occupational Psychology* 53, 1 (1980), 65–72.
16. Kenneth A Funes-Mora and Jean-Marc Odobez. 2016. Gaze estimation in the 3d space using rgb-d sensors. *International Journal of Computer Vision* 118, 2 (2016), 194–216.
17. Venkata Rama Kiran Garimella, Abdulrahman Alfayad, and Ingmar Weber. 2016. Social media image analysis for public health. In *Proc. 2016 CHI Conf. on Human Factors in Computing Systems*. ACM, 5543–5547.
18. Gordon D Hemsley and Anthony N Doob. 1978. The Effect of Looking Behavior on Perceptions of a Communicator's Credibility 1. *Journal of Applied Social Psychology* 8, 2 (1978), 136–142.
19. Hayley Hung, Yan Huang, Gerald Friedland, and Daniel Gatica-Perez. 2011. Estimating dominance in multi-party meetings using speaker diarization. *IEEE Transactions on Audio, Speech, and Language Processing* 19, 4 (2011), 847–860.
20. Hayley Hung, Dinesh Babu Jayagopi, Sileye Ba, Jean-Marc Odobez, and Daniel Gatica-Perez. 2008. Investigating automatic dominance estimation in groups from visual attention and speaking activity. In *Proceedings of the 10th international conference on Multimodal interfaces*. ACM, 233–236.
21. Andrew S Imada and Milton D Hakel. 1977. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *J. Applied Psychology* 62, 3 (1977).
22. Dinesh Babu Jayagopi, Hayley Hung, Chuohao Yeo, and Daniel Gatica-Perez. 2009. Modeling dominance in group conversations using nonverbal activity cues. *Audio, Speech, and Language Processing, IEEE Trans. on* 17, 3 (2009), 501–513.
23. SM Ketrow and K Perkins. 1986. Operator eye contact and client satisfaction in computer-assisted interactions. In *Annual Convention of the Speech Communication Association, Chicago, IL*.
24. Mark Knapp, Judith Hall, and Terrence Horgan. 2013. *Nonverbal communication in human interaction*. Cengage Learning.
25. Max Kuhn. 2016. A Short Introduction to the caret Package. (2016).
26. Thomas W Leigh and John O Summers. 2002. An initial evaluation of industrial buyers' impressions of salespersons' nonverbal cues. *Journal of Personal Selling & Sales Management* 22, 1 (2002), 41–53.
27. Skanda Muralidhar, Laurent Son Nguyen, Denise Frauendorfer, Jean-Marc Odobez, Marianne Schimd-Mast, and Daniel Gatica-Perez. 2016. Training on the Job: Behavioral Analysis of Job Interviews in Hospitality. In *Proc. 18th ACM ICMI*. 84–91.

28. Skanda Muralidhar, Marianne Schimd-Mast, and Daniel Gatica-Perez. 2017. How May I Help You? Behavior and Impressions in Hospitality Service Encounters. In *Proc. 19th ACM ICMI*.
29. Laurent Son Nguyen, Denise Frauendorfer, Marianne Schmid Mast, and Daniel Gatica-Perez. 2014. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Trans. on Multimedia* 16, 4 (2014).
30. Christopher Y Olivola, Dawn L Eubanks, and Jeffrey B Lovelace. 2014. The many (distinctive) faces of leadership: Inferring leadership domain from facial appearance. *The Leadership Quarterly* 25, 5 (2014), 817–834.
31. Charles K Parsons and Robert C Liden. 1984. Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues. *J. Applied Psychology* 69, 4 (1984), 557.
32. Alex Pentland. 2004. Social dynamics: Signals and behavior. In *Int. Conf. on Developmental Learning*, Vol. 5.
33. Fabio Pianesi, Nadia Mana, Alessandro Cappelletti, Bruno Lepri, and Massimo Zancanaro. 2008. Multimodal recognition of personality traits in social interactions. In *Proc. ACM ICMI*.
34. Dairazalia Sanchez-Cortes, Oya Aran, Dinesh Babu Jayagopi, Marianne Schmid Mast, and Daniel Gatica-Perez. 2013. Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces* 7, 1-2 (2013), 39–53.
35. Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2010. Identifying emergent leadership in small groups using nonverbal communicative cues. In *Int Conf. on Multimodal interfaces and the workshop on machine learning for Multimodal Interaction*. ACM, 39.
36. Albrecht Schmidt. 2016. Cloud-based AI for pervasive applications. *IEEE Pervasive Computing* 15, 1 (2016), 14–18.
37. Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86, 2 (1979).
38. Rémy Siegfried, Yu Yu, and Jean-Marc Odobez. 2017. Towards the use of social interaction conventions as prior for gaze model adaptation. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ACM, 154–162.
39. Vivek K Singh, Souvick Ghosh, and Christin Jose. 2017. Toward Multimodal Cyberbullying Detection. In *Proc. 2017 CHI Conf. Extended Abstracts on Human Factors in Computing Systems*. ACM, 2090–2099.
40. Magnus Söderlund and Sara Rosengren. 2004. Dismantling" positive affect" and its effects on customer satisfaction: an empirical examination of customer joy in a service encounter. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior* 17 (2004), 27.
41. Del A Sole. 2018a. *Getting Started with the Computer Vision API*. Apress, Berkeley, CA.
42. Del A Sole. 2018b. *Introducing Microsoft Cognitive Services*. Apress, Berkeley, CA.
43. Brian R Spisak, Allen E Grabo, Richard D Arvey, and Mark van Vugt. 2014. The age of exploration and exploitation: Younger-looking leaders endorsed for change and older-looking leaders endorsed for stability. *The Leadership Quarterly* 25, 5 (2014), 805–816.
44. DS Sundaram and Cynthia Webster. 2000. The role of nonverbal communication in service encounters. *J. Services Marketing* 14, 5 (2000), 378–391.
45. Wei-Chi Tsai. 2001. Determinants and consequences of employee displayed positive emotions. *Journal of management* 27, 4 (2001), 497–512.