On the Influence of Emotional Feedback on Emotion Awareness and Gaze Behavior

Fabien Ringeval*, Andreas Sonderreger†, Basilio Noris‡, Aude Billard‡ Juergen Sauer† and Denis Lalanne*
*Document Image and Voice Analysis group, Department of Informatics
University of Fribourg, Fribourg, Switzerland
Email: {fabien.ringeval,denis.lalanne}@unifr.ch
†Cognitive Ergonomic group, Department of Psychology
University of Fribourg, Fribourg, Switzerland
‡Learning Algorithms and Systems Laboratory
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

***Abstract*—This paper examines how emotion feedback influences emotion awareness and gaze behavior. Simulating a videoconference setup, 36 participants watched 12 emotional video sequences that were selected from the SEMAINE database. All participants wore an eye-tracker to measure gaze behavior and were asked to rate the perceived emotion for each video sequence. 3 conditions were tested: (c1) no feedback, i.e., the original video-sequences, (c2) correct feedback, i.e., an emoticon is integrated in the video to show the emotion depicted by the person in the video and (c3) random feedback, i.e., the emoticon displays at random an emotional state that may or may not correspond to the one of the person. The results showed that emotion feedback had a significant influence on gaze behavior, e.g., over time random feedback led to a decrease in the frequency of episodes of gaze. No effect of emotion display was observed for emotion recognition. However, experiments on the automatic emotion recognition using gaze behavior provided good performance, with better score on arousal than valence, and a very good performance was obtained in the automatic recognition of the correctness of the emotion feedback.**

I. Introduction

Previous research in the domain of work psychology has indicated that mood, emotion and team members empathy may influence team processes and the outcomes of teamwork, such as performance, cohesion and satisfaction [1]. This suggests that emotion awareness of each team members is important for efficient and satisfactory teamwork. However, a gap may exist between felt emotions and the corresponding teammate's understanding of them, especially in computer-mediated interaction, where the medium may introduce communication bias. This gap, which indicates the level of emotion awareness between collaborators, impacts the collaboration by either impairing mutual understanding or maintaining a good relationship between participants [2].

Since research in the domain of affective computing has made significant progress in automatically detecting emotional states of humans [3], we plan to develop a tool, the EmotiBoard, for providing automatic affective feedback in teamwork during video-conference [4]. This tool aims to reduce the emotion awareness's gap between collaborators, by providing an emotion feedback based on real-time automatic emotion recognizers. A preliminary study on the EmotiBoard has shown that the evaluation of the other team members' emotional state is more accurate when a mood feedback is presented, using self-reported data [5].

Before using the EmotiBoard with an automatic emotion feedback setting, we first wanted to estimate how such feedback might influence the behavior of a participant, regarding both the perceived emotion and the gaze toward the emotion feedback. Simulating a video-conference setup, participants were asked to watch emotional sequences that were selected from the SEMAINE database, and to evaluate the emotional state expressed by the person seen in the video. Participants also wore an eye-tracker to measure their gaze behavior on the emotion feedback, which was incrusted into the video data using an adaptive emoticon. 3 conditions were tested to study the impact of the emotional feedback: (c1) no feedback, i.e., the original video-sequences, (c2) correct feedback, i.e., the emoticon displays the mean arousal-valence annotation values and (c3) random feedback, i.e., the emoticon shows random emotion representations. The main question as well as contributions of this paper are thus the followings:

(i) does the gap in emotion awareness, i.e., the difference between perceived and expressed emotion, is reduced when participants receive a correct emotional feedback, compared to a random emotional feedback, or not any feedback?
(ii) is it possible to use gaze data to automatically recognize the emotion reported by participants, or the one depicted in the video-sequence?
(iii) are there features of gaze behavior that depend on the condition of feedback, and make it possible to recognize this condition, as a model of the quality of feedback?

In the remainder of this paper we introduce some related works on emotional feedbacks for remote collaborative interactions (Sec. II), then present the experimental setup (Sec. III), the methodology (Sec. IV) used for the analysis of emotion feedback's influence on both emotion perception and gaze behavior (Sec. V), with some automatic recognition experiments using gaze data (Sec. VI), before concluding (Sec. VII).

II. Related Works

There are many research topics that might be related to the EmotiBoard, such as the model to quantify the emotion, the system to recognize automatically emotion from sensors' data, the type of communication medium used to convey the recognized emotion and the user evaluation scheme used to quantify the benefits of the system on remote collaborative interactions.

Several models have been defined in the literature to quantify emotional behaviors of humans, such as, categorical [6], dimensional [7] and appraisal-based models [8]. Concerning emotion recognition, there is a trend for using realistic recordings of spontaneous human's behaviors [3], [4]; see also the series of AVEC 2011-2013 ICMI Grand Challenges [9]. Such authentic behaviors are difficult to collect and recognize, because they are relatively rare, short lived, and filled with subtle context-based changes [3]. Emotion recognition performance reported by the actual systems on spontaneous interactions are usually lower on valence than for arousal [9].

Several strategies can be used for enhancing remote interaction through an emotion feedback, using different communication medium, e.g., visual displays [5], [10], taste and smell [11], or touch [12]. Even though haptic systems can reduce sadness emotion and general negative mood significantly in remote interaction [12], they are more intrusive and less easy to use compared to a visual feedback display. For example, Sánchez et al. have shown that it is possible to improve existing functionality of an instant messaging system for conveying emotion, as we did in a preliminary study for video-conferencing [5].

Concerning the user's evaluation schemes, systems can be used to provide objective measurements of the user behavior, to quantify the influences of contextual variations. For example, eye-tracking systems can be used to measure information processing during emotion perception of facial expressions [13], or effectiveness of emotion feedback for remote interaction [14]. Even though many studies focused on the automatic recognition of emotion from facial expressions, we did not find any study, at the time of writing, that explicitly used gaze data to perform this automatic recognition, especially with different conditions of emotion, e.g., perceived or expressed, as well as emotion feedback, e.g., correct or random.

III. Experimental Setup

The main setup of this study consists in the simulation of a video-conference setup, in which participants were asked to watch audio-visual sequences of different persons. The setup includes furthermore an adaptive emotion feedback in the form of an emoticon which is incrusted into the video data. A review of the audio-visual databases that include time-continuous annotations of natural emotions was made, to provide stimuli for this study [4]. We chose the SEMAINE database [15], because its interaction scenario is close to the one we target for the EmotiBoard, i.e., two persons talking remotely to each other.

A. Annotation data pre-processing

Audio-visual sequences of the SEMAINE database were filtered out to keep only those that were annotated by at least 3 raters, to provide a more precise estimation of the inter-rater agreement. A zero-mean normalization was performed locally on the annotation data, as in [4]. The inter-rater agreement was quantified by the standard-deviation of the annotations provided by the N available raters on a given video frame; std_ann. To ease the selection using the inter-rater agreement as criterion, we normalized the std_ann measure to get a value in percentage, with 0% meaning no agreement and 100% full agreement. We considered that a uniform distribution in the annotation data, i.e., $[-1, 1]$ for $N = 2$, $[-1, 0, 1]$ for $N = 3$, etc., would provide the lowest inter-rater agreement with the standard-deviation measure; std_uni. Whereas a perfect agreement will always provide a std_ann value equal to 0 whatever the number of raters. The inter-rater agreement IRA was computed as follows on each annotation frame (20ms):

$$IRA = \left(\frac{std\_uni - std\_ann}{std\_uni} \right) \% \quad (1)$$

B. Video-sequences selection

Software including a graphical user interface and a searching algorithm was designed using Matlab, to ease the selection of video-sequences that will be used as stimuli in this study. The software tool allows for the identification of video-sequences from the SEMAINE database according to the following criteria: minimum / maximum value of arousal, valence, inter-rater agreement and sequence duration. The video-sequences were selected based on the two following criteria: a mean IRA higher than 70% and a sequence duration longer than 30s. This is supposed to ensure that the displayed emotions are not ambiguous, and the duration of the emotional episode is long enough to be well perceived. Two emotion classes were selected for each affective dimension, using the mean annotation value from the SEMAINE database: passive / active for arousal, and negative / positive for valence, i.e., a mean value inferior or superior to 0, respectively. Sequences were then automatically selected by our software to have at least one significant variation (e.g., from positive to negative) in the emotional dimension that is not targeted, while the targeted one did not change significantly. Finally, sequences for which the verbal information was correlated to the produced emotion have been rejected and 3 sequences were kept for each of the 4 emotional classes, representing data from 5 different persons (3 females and 2 males), cf. Table I.

One may notice that the first two sequences of low arousal were annotated by only 2 raters and that the mean annotation value is positively close to zero, whereas it should be negative. This is because only one sequence matched all our criteria for a passive arousal condition. The other two sequences we found had a duration less than 30s. But we yet decided to keep these two sequences, by moving back the start time index to reach the desired 30s, because the duration of the low arousal episode lasts for at least the 2/3 of the overall duration.

TABLE I

INFORMATION REGARDING THE SELECTED SEQUENCES FROM THE SEMAINE DATABASE: SESSION INDEX / GENDER ID (FEMALE, MALE), NUMBER OF ANNOTATORS, START TIME INDEX, MEAN VALUE OF ANNOTATION AND INTER-RATER AGREEMENT (IRA) OF AROUSAL AND VALENCE.

	AROUSAL						VALENCE					
	passive			active			negative			positive		
Session / Gender_ID	4 / F1	29 / F1	72 / M1	4 / M2	9 / M1	47 / F2	46 / F1	73 / F1	78 / M2	78 / F2	79 / F2	95 / F3
N annotators	2	2	6	6	6	6	6	6	6	6	6	8
Start time in s.	98.64	0.02	69.88	6.68	17.0	78.34	31.66	73.44	34.96	4.42	21.66	19.58
Annotation - Arousal	0.01	0.03	-0.22	0.21	0.47	0.24	-0.05	-0.01	0.07	-0.14	-0.02	0.05
IRA - Arousal	0.82	0.76	0.77	0.78	0.81	0.71	0.73	0.75	0.48	0.58	0.55	0.41
Annotation - Valence	-0.03	0.45	-0.26	0.09	-0.28	-0.14	-0.21	-0.08	-0.06	0.17	0.45	0.34
IRA - Valence	0.90	0.77	0.82	0.89	0.69	0.54	0.85	0.79	0.72	0.74	0.88	0.77

C. Emotional feedbacks

The 12 selected video-sequences were processed to mix the stereo soundtrack with the video data; participants could thus hear the two persons engaged in the discussion. According to the EmotiBoard-setup, an emoticon was incrusted in the top-left corner of the video frames to show an emotion feedback to the participants of this study. The size of the emoticon was adapted with regard to arousal and a dashed-circle was drawn to represent a "neutral" state, whereas valence was represented by adapting the smile of the emoticon: a smiling face as indicator for positive valence and a frowning face as indicator for negative valence, cf. Fig. 1.

Two versions of the EmotiBoard were developed for this study: (c2) correct-feedback, using the mean annotation value from the SEMAINE database and (c3) random-feedback, using a dedicated algorithm. The random behavior of the emoticon was defined with 4 variables for each sequence. The first two are initialized once and are used to define the first emotional state of the emoticon, whereas the last two variables are used to set the duration of the emotional episodes, with values ranging in $[1-30]s$. These last two variables were computed iteratively until the overall duration of the episodes exceeds 30s for each affective dimension, and the duration was then fixed to 30s.

Fig. 1. Incrustation of an emotion feedback into the selected video-sequences using an emoticon; left: active arousal and positive valence; right: passive arousal and negative valence.

IV. METHODOLOGY

A. Participants, experimental design and instruments

36 participants (28 females, 8 males), aged between 19 and 41 years ($M = 24.5, SD = 5$) were recruited for this study. All of them were students at University of Fribourg, Switzerland. As English is the spoken language in the SEMAINE database, participants had to rate their English comprehension skills. 20 participants reported to have basic skills, 14 had good skills and 2 had very good skills.

A 3x4 mixed design was chosen for this experiment. Emotion feedback was varied as a between-subjects variable at 3 levels: (c1) the first group of participants received no feedback, (c2) the second group received continuous emotion feedback with the emoticon and (c3) the third group of participants received continuous emotion feedback that was generated randomly. As a within-subjects variable, participants were asked to rate the affective states expressed by the persons in the video-sequence for each of the 4 emotional conditions, i.e., passive / active for arousal and negative / positive for valence. A dimensional scale [16], which provided scores for arousal, valence and tension, based on the circumflex model of affect [7], was used for this purpose. Participants in the c2 and c3 feedback conditions furthermore rated the usefulness of the emotion feedback on a $[1-100]$ visual analogue scale; "*To what extend do you think the mood display was helpful for the assessment of the person's mood?*".

B. Materials

A wearable gaze-tracking system, WearCam, cf. Fig. 2, was used to estimate the gaze behavior of participants toward the emotion feedback [17]. The WearCam uses two miniature CCD cameras, which provide 768 x 576 pixel images at 25 frames per second with an angle of 96° x 96°. A 2.7cm x 0.7cm mirror is mounted between the two cameras to reflect the eyes of the wearer back into the bottom camera, which will serve to estimate the direction of the gaze, i.e., looking or not to the emotion feedback. The gaze estimation accuracy of the system with adult participants is around 1.6°.

In order to maintain synchronization between the cameras, because we used USB video converters for data transfer, the two images are blended into a single vertical image of 384 x 576 pixels. Because the estimation of the gaze direction is performed later using a semi-supervised classification framework, no calibration step was required during data recording. We just had to optimize the angle of the mirror before launching the gaze data recording, to have all the eyes well visible in the bottom camera, by using a remote controller rotating this mirror with a DC motor.

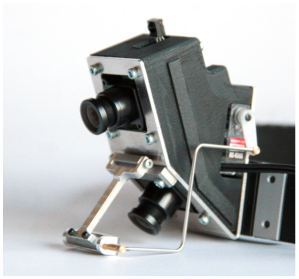

Fig. 2. WearCam: gaze-tracking system worn by participants [17].

C. Procedure

Each participant was welcomed in a room at the department of psychology of the University of Fribourg, Switzerland, that is devoted to data acquisition, i.e. isolated from external noises [4]. First, participants rated their actual emotional state using the dimensional scale, and filled in a questionnaire about their identity, i.e., age, gender and skills in English understanding. Then, the WearCam was mounted and the 12 video-sequences were back-projected on a large display surface. After each sequence, participants were asked to rate the emotional state of the person seen in the video-sequence, as well as the usefulness of the emotion feedback for the c2 and c3 feedback conditions. In order to reduce the effect of participants having similar rating behaviors at specific timings of the experiment, e.g., first or last played sequence, we randomly permuted the order of the video-sequences and iteratively shifted the result by one for each consecutive participants.

V. Influences of Emotion Feedback on Emotion Perception and Gaze Behavior

A. Data processing: estimation of gaze direction

Data collected with the WearCam were processed to estimate the timings when the participant was looking or not to the emotion feedback on each video-sequence, cf. Fig. 3. All pre-processing steps used to extract the cropped image of the mirrored region of the eyes are fully detailed in [17]. Spatial movements of the eyes are characterized by Gabor filters, and the gaze direction of the participant, i.e., looking or not to the emoticon, is estimated as a two-class problem by a C-SVM classifier; gaussian kernel width: $\gamma = 0.01$; penalty factor: $C = 1000$. For the learning phase, we manually provided to the system an equal number of samples when the participant was looking or not to the emoticon for each video-sequence; roughly 50 samples in total per participant. Then the system returned for each video frame a binary estimation of the gaze direction of the participant, i.e., looking or not to the emoticon.

B. Data processing: features extraction

Discontinuities in the data were filtered out frame-wisely to remove single positive cases, which are considered as noise. 6 different low-level descriptors (LLD) were then computed on the data for each sequence: (i) the gaze count, i.e., the number of times the participant looked at the emotion feedback, (ii) the gaze duration, i.e., the duration of the episodes when

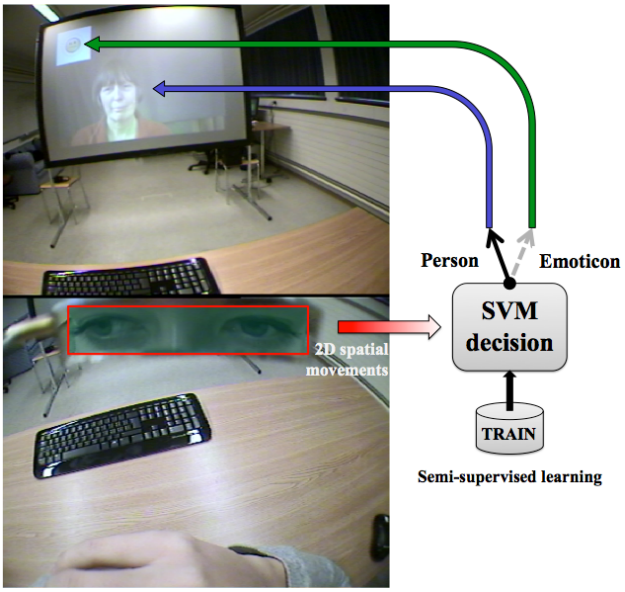

Fig. 3. Estimation of gaze direction as a bi-class problem using 2D spatial movements of the eyes and SVM based semi-supervised learning.

gaze is toward the emoticon, (iii) the gaze interval, i.e., the duration between the middle of the episodes of gaze toward the emoticon and (iv) the horizontal and vertical movements of the eyes used in the C-SVM classification task, as well as their euclidean distance for the diagonal movements. A series of 29 statistics (cf. Table II) were finally computed on all LLDs excepted gaze count, for each video-sequence and for all, to provide a detailed description of the gaze data.

C. Influence of emotion feedback on emotion perception

Because different scales were used to quantify the emotion depicted in the video-sequence and those reported by the participants, we normalized all these values in the range $[1-5]$. For each statistical test we performed on the data, including emotion perception and gaze behavior (Sec. V-D), we did not use the self-reporting of participants on age, gender and English level as co-variables, because they did not change the results significantly. Difference between the emotion reported by the participant and the one shown in the video-sequence, i.e., the bias in emotion awareness, was computed on each condition of feedback and emotion. Unfortunately, we did not find any statistical difference in this bias between the conditions of feedback.

It is interesting to note that the rating of usefulness of the emotion feedback does not differ significantly between the c2 ($M = 31.23$, $SD = 25.83$), and the c3 conditions ($M = 19.63$, $SD = 14.96$), and that values are rather low because the scale is $[1 - 100]$. There is a significant effect between these two conditions when considering the evolution over the trials; $F = 1.89; df = 11, 242; p < .05$. The usefulness of the emotion feedback is for the first trials similar for the c2 and c3 conditions. However, the longer the experiment takes, the less useful the feedback is rated in the c3 condition. Whereas in c2, the usefulness-ratings increase in the end.

TABLE II
LIST OF STATISTICAL MEASURES COMPUTED ON THE LLDS

Measure	Description	Measure	Description
H	maximum	IQR_Std	|IQR−Std|
Rpos_H	relative position (max)	Reg_slope	regression slope
L	minimum	Onset	1st value
Rpos_L	relative position (min)	Target	middle value
Rpos_D	|Rpos_H−Rpos_L|	Offset	last value
Range	H−L	Tar_On	|Target−Onset|
Range_n	(H−L)/Rpos_D	Off_On	|Offset−Onset|
Mean	mean value	Off_Tar	|Offset−Target|
Std	standard-deviation	%raising	%raising values
Skewness	skewness	%falling	%falling values
Kurtosis	kurtosis	mean_raising	mean(raising)
Q1	1st quartile	std_raising	std(raising)
Q2	median value	mean_falling	mean(falling)
Q3	3rd quartile	std_falling	std(falling)
IQR	inter quartile range		

D. Influence of emotion feedback on gaze behavior

A statistical analysis was performed on the LLDs of gaze, to evaluate which feature might be related to the condition of feedback. We used as inputs the mean value and the regression slope of the LLDs. In the c2 condition, participants looked on an average 3.77 times on the emotion feedback; $SD = 2.05$. This value is slightly higher in the c3 condition ($M = 4.64, SD = 3.31$), but does not differ significantly with regard to the feedback condition. When looking at the development of the gaze behavior over the video-sequences (cf. Fig. 4), it can be observed that the number of eye-movements towards the emoticon remains more or less stable during the 12 trials of the c2 condition; Reg_slope: $M = 0.025, SD = 0.25$. Whereas the number of eye-movements towards the emotion feedback decreases over time in the c3 condition; Reg_slope: $M = -0.24, SD = 0.27$. This difference is significant: $F(1, 22) = 6.08, p < 0.05$.

E. Discussion

There are many possible explanations for the non-expected results of the influence of the different variations of feedback on the accuracy of emotion evaluation. Could it be that the emoticon was not understood correctly? Results regarding the helpfulness of the emotion feedback seem to confirm this misunderstanding because values are rather low. Furthermore, we used in this study verbal representations for the questionnaire data, whereas a graphical representation was used for the feedback. It could also be argued that the sequences that have been chosen were emotionally so easy to discern, as we fixed the IRA criterion value to be higher than 70%, that participants were pretty sure about the emotion of the person in the film scene; the maximum bias in the emotional rating is 25.9%. Moreover, it is possible that because participants did not have to be highly concentrated on the subject of discussion in the video-sequence, compared to a real interaction, they thus did have all their cognitive load available for the emotion recognition task.

Concerning the gaze behavior, we found that there is a significant difference in the development of the number of eye-movements towards the emoticon according to the condition of feedback. Indeed, participants seek to minimize discongruencies regarding the emotion feedback and their own estimate: they check whether this difference will change over time when the feedback is correct, whereas they abandon the feedback when they realize it has no value for them since it is random.

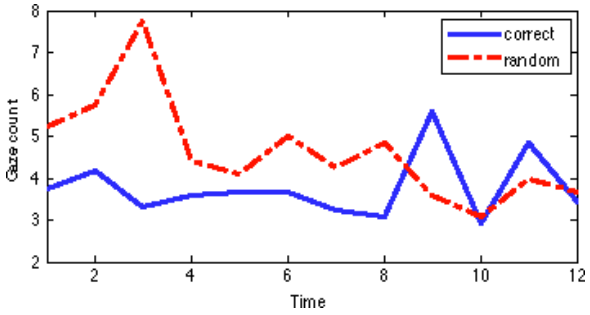

Fig. 4. Comparison of the evolution of gaze count (mean value) toward the emotion feedback between correct- and random-feedback conditions.

VI. AUTOMATIC EMOTION / FEEDBACK RECOGNITION FROM GAZE BEHAVIOR

We focused here on the last two conditions of feedback c2 and c3 to investigate whether gaze data could be used to estimate automatically the emotion reported by the participant, or the one depicted in the video-sequence, as well as the correctness of the feedback, i.e., correct or random.

A. Reported / depicted emotion recognition

An SVM classifier with RBF kernel was trained on the gaze data to estimate whether the emotion reported by the participant on arousal and valence, as well as the one shown in the video-sequence, could be automatically recognized. We used as input the 146 features set, cf. Sec. V-B, and as output, the emotion reported by the participant, or the one depicted in the video-sequence. Positive and negative labels used for the classification task were obtained by thresholding the mean annotation value of sequences from Table I; we obtained 5 sequences for a passive arousal and 7 sequences for an active, and 7 sequences for a negative valence and 5 for a positive. A grid-search procedure was used to estimate the optimal kernel width and penalty factor on a subset of the data; $\gamma = [1e^{-4} - 1.0]$ in 20 steps; $C = [0.1 - 1000]$ log-scale in 10 steps. For each participant, the 2/3 of sequences were used for training and optimization of the classifier's parameters, whereas the remaining four (1/3) was used for testing. Performance of the system was estimated with a 12-fold cross-validation (i.e., leave-one-sequence-out), and results are reported with F-score in % for each condition of feedback in Table III.

Overall, i.e., for both correct and random feedbacks, the system is able to correctly recognize the reported arousal 82% of the time, whereas the actual, i.e., ground-truth, arousal is recognized 74% of the time. Conversely, the system is able to correctly recognize the reported emotional valence 70% of the time, whereas the performance slightly increases to 75%

when the ground-truth value of the valence is used as output. In addition, there are almost no variation in the performance between the conditions of feedback.

B. Feedback recognition

Because there are many possibilities to display an emotional feedback [10], the automatic estimation of its quality could help at choosing the most appropriate visualization technique to use. To this end, we trained a decision tree classifier with AdaBoost on the LLDs' features, using as output the type of emotion feedback, i.e., correct vs random. Optimal number of tree branches were optimized in the same as for the previous experiment, and performance was estimated in the same 12-folds cross-validation scheme. Results show that the system is able to successfully recognize the type of emotion feedback as correct or random 81.6% of the time.

C. Discussion

One may notice that the automatic recognition of the emotion reported by the participant performs lower on valence than for arousal, which is a result pretty well known in the emotion-oriented speech and image processing communities [9], [18], [19]. Furthermore, we found that there is a difference in how the performance evolves when considering the emotion depicted by the person in the video-sequence as output of the system, instead of the one reported by the participant, for all conditions of feedback. This would mean that our understanding of arousal impacts more the way we are doing gaze, whereas valence would have a more unconscious impact.

VII. Conclusion

This paper examined how emotion feedback influences emotion recognition and gaze behavior. Results show that emotion perception is not influenced by emotion feedback, as we choose sequences where the emotion was too easily identifiable. However, participants looked at the correct-feedback constantly all over the experiment, whereas they abandonned little by little the random-feedback. Automatic emotion recognition experiments using gaze behavior provided better performance on arousal than valence, and a very good performance was obtained in the automatic recognition of the correctness of the emotion feedback, which is a promising contribution toward the development of automatic feedback quality estimation.

Different strategies of display will be studied in the near future to provide emotion feedback during remote collaborative interactions. We first plan to re-use the methodology of this study but with video-sequences where the emotion is perceived as being ambiguous, i.e., with a low IRA. The setup will then include live emotion-recognition for the automatic estimation of arousal and valence from speech and physiological data using electro-dermal activity sensors.

TABLE III
Emotion recognition performance using gaze data on arousal and valence for the emotion reported by the participant, or the one depicted by the person in the video-sequence, and according to the type of feedback.

F-score$\pm\sigma$	AROUSAL		VALENCE	
	Reported	Depicted	Reported	Depicted
Correct	82.1 $\pm$ 2.4	74.4 $\pm$ 2.4	70.2 $\pm$ 1.5	74.7 $\pm$ 2.4
Random	83.3 $\pm$ 2.2	74.7 $\pm$ 3.2	72.0 $\pm$ 2.5	75.0 $\pm$ 2.6
Both	82.2 $\pm$ 1.7	74.3 $\pm$ 2.0	69.9 $\pm$ 1.8	74.8 $\pm$ 1.8

References

[1] J. R. Kelly and S. G. Barsade, "Mood and emotions in small groups and work teams," *Organizational Behavior and Human Decision Processes*, vol. 86, no. 1, pp. 99–130, 2001.

[2] V. U. Druskat and S. B. Wolff, "Building the emotional intelligence of groups," *Harvard Business Review*, vol. 79, no. 3, pp. 80–90, 2001.

[3] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: current trends and future directions," *Image and Vision Computing*, vol. 31, no. 2, pp. 120–136, 2012.

[4] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, "Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions," in *EmoSPACE, proc. of IEEE Face & Gestures*, Shanghai, China, 2013.

[5] A. Sonderegger, D. Lalanne, L. Bergholz, F. Ringeval, and J. Sauer, "Computer-supported work in distributed and co-located teams: the influence of mood feedback," in *Interact*, Cape Town, South-Africa, 2013.

[6] P. Ekman, *Emotion in the human face*. Cambridge, UK: Cambridge University Press, 1982.

[7] J. Russel, "A circumplex model of affect," *Journal of Personality and Social Psychology*, pp. 1161–1178, 1980.

[8] D. Grandjean, D. Sander, and K. R. Scherer, "Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization," *Consciousness and Cognition*, vol. 17, pp. 484–495, 2008.

[9] B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic, "AVEC 2011 — The first international audio/visual emotion challenge," *D'Mello, S. et al. (eds.), ACII 2011, LNCS 6975*, pp. 415–424, 2011.

[10] J. A. Sánchez, N. P. Hernández, J. C. Penagos, and Y. Ostróvskaya, "Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states," in *Anais do IHC*, Natal, Brasil, 2008, pp. 66–72.

[11] N. Ranasinghe, K. Karunanayaka, A. D. Cheok, O. N. Newton Fernandao, H. Nii, and P. Gopalakrishnakone, "Digital taste and smell communication," in *BodyNets*, Boston (MA), USA, 2011, pp. 78–84.

[12] R. Wang and F. Queck, "Touch & talk: contextualizing remote touch for affective interaction," in *Int. Conf. on Tangible, Embedded, and Embodied Interaction*, Cambridge (MA), USA, 2010, pp. 13–20.

[13] P. Schmid, M. Mast, D. Bombari, F. Mast, and J. Lobmaier, "How mood states affect information processing during facial emotion recognition: an eye tracking study," *Swiss J. of Psychology*, vol. 70, no. 4, pp. 223–231, 2012.

[14] D. K. Kim, J. Kim, E. C. Lee, M. Whang, and Y. Cho, "Interactive emotional content communications system using portable wireless biofeedback device," *IEEE Trans. on Consumer Electronics*, vol. 57, no. 4, pp. 1929–1936, 2011.

[15] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder, "The SEMAINE database: annotated multimodal records of emotionally colored conversations between a person and a limited agent," *IEEE Trans. on Affective Computing*, vol. 3, no. 1, pp. 5–17, 2012.

[16] P. Wilhelm and D. Schoebi, "Assessing mood in daily life," *Eur. J. of Psychological Assessment*, vol. 43, no. 4, pp. 258–267, 2007.

[17] B. Noris, J.-B. Keller, and A. Billard, "A wearable gaze tracking system for children in unconstrained environments," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 476–486, 2011.

[18] F. Ringeval, M. Chetouani, and B. Schuller, "Novel metrics of speech rhythm for the automatic assessment of emotion," in *Interpseech*, Portland (OR), USA, 2012.

[19] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, and E. Marchi et al., "The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism," in *Interspeech*, Lyon, France, 2013.