Frustration in the Face of the Driver: A Simulator Study on Facial Muscle Activity during Frustrated Driving

Klas Ihme[1*], Christina Dömeland[1], Maria Freese[2] & Meike Jipp[1]

[1]Institute of Transportation Systems and [2]Institute of Flight Guidance, German Aerospace Center, Braunschweig, Germany.

Correspondence concerning the manuscript should be addressed to Klas Ihme, German Aerospace Center, Lilienthalplatz 7, 38102 Braunschweig, Germany, Email: klas.ihme@dlr.de, Phone: +49 531 295 3497

This work was (partially) supported by the funding initiative Niedersaechsisches Vorab of the Volkswagen Foundation and the Ministry of Science and Culture of Lower Saxony (as part of the Interdisciplinary Research Center on Critical Systems Engineering for Socio-Technical Systems). The funding source had no influence on any aspect of the study's content and its publication.

Abstract

Frustration in traffic is one of the causes of aggressive driving. Knowledge whether a driver is frustrated may be utilized by future advanced driver assistance systems to counteract this source of crashes. One possibility to achieve this is to automatically recognize facial expressions of drivers. However, only little is known about the facial expressions of frustrated drivers. Here, we report the results of a driving simulator study investigating the facial muscle activity that comes along with frustration. Twenty-eight participants were video-taped during frustrated and non-frustrated driving situations. Their facial muscle activity was manually coded according to the Facial Action Coding System. Participants showed significantly more facial muscle activity in the mouth region. Thus, recording facial muscle behavior potentially provides traffic researchers and assistance system developers with the possibility to recognize frustration while driving.

Keywords: Frustration, Facial Action Coding System, driving simulator, facial expressions

Frustration in the face of the driver: A simulator study on facial muscle activity during frustrated driving

Frustration can be seen as an aversive emotional state when goal-directed behavior is blocked (Lazarus, 1991) associated with negative valence and slightly elevated arousal (Russell, 1980; Scherer, 2005). Increasing or persisting frustration can result in anger and aggressive behaviors (Ekman & Friesen, 2003). During everyday driving, many events and situations can lead to frustration. Frustration ranges from confrontations with red lights to interactions with other road participants blocking the way. Frustration can lead to aggressive behaviors, a contributing factor to many crashes (Deffenbacher, Lynch, Oetting, & Swaim, 2002). Thus, reducing frustration behind the wheel is an important step towards reducing crashes on the road. One possibility to decrease frustration is to design intelligent driver assistance systems capable of recognizing and mitigating negative emotions of car drivers. This requires an unobtrusive measurement of frustration in the car.

Humans communicate emotions by changing the activation of the muscles in the face which is fundamental to understand each other in social interaction (Ekman & Friesen, 2003). This is why it is envisioned to equip machines with the capability to read facial expressions and herewith increase an understanding of users and make human-machine interaction more natural (e.g., Bruce, 1992). With the developments in video processing and machine learning, cars and advanced driver assistance systems are also envisioned to become capable of interpreting and properly reacting to drivers' emotions (Tews, Oehl, Siebert, Höger, & Faasch, 2011). In previous studies frustration (e.g. Lee, 2010; Lee & LaVoie, 2014; Malta, Miyajima, Kitaoka, & Takeda, 2011) and other emotions (e.g. Gao, Yuce, & Thiran, 2014; Healey & Picard, 2005) of drivers in certain driving situations have been examined. Other authors used facial expressions as method comparing affective reactions between driving conditions (e.g. Donkor, Burnett, & Sharples, 2014). However, none of these studies reported the facial signs of spontaneously experienced frustration during (simulated) driving. Thus, while the recognition of facial expressions appears feasible in cars (Gao, Yuce, & Thiran, 2014), to the best of our knowledge no study has reported the changes in facial muscle activity that come along with frustrated driving so far.

Ekman and colleagues (Ekman, Friesen, & Hager, 2002) proposed the Facial Action Coding System (FACS) for describing facial muscle activity in terms of so-called *action units* (AUs). The FACS provides researchers with the possibility to rate the expressivity of human faces on a single muscle activity level. According to Ekman (1992), emotions coming from the anger family, with which frustration is associated, can come along with activity in the AUs 4 (brow lowerer), 23 (lip tightener) and 24 (lip pressor). Studies investigating emotions during interaction with computer-mediated learning systems reported that frustration came along with activity in the AUs 1 (inner brow raiser), 2 (outer brow raiser), 4 and 14 (dimpler) (D'Mello et al., 2005; Grafsgaard, Wiggins, Boyer, Wiebe, & Lester, 2013). Of these, however, AUs 1 and 2 were rather linked to surprise (Gosselin, Perron, & Beaupré, 2010). Accordingly, we assume that facial activity related to frustrated driving happens in the AUs 4, 14, 23 and 24. We tested this in a driving simulator study with frustrating situations. Afterwards participants' facial activity was coded using the FACS.

Methods

Participants

Twenty-eight participants (12 females, age mean [M] = 26 years, standard deviation [SD] = 3 years) completed the experiment without technical problems (of 31 in total). All of them were German native speakers with normal or corrected-to-normal vision and held a valid driver's license. Informed consent was obtained from all participants before the conduct of this study.

Experimental Set-Up

The study was accomplished in a driving simulator consisting of a 46 inch screen and a G27 Logitech Racing steering wheel with pedals that controlled a virtual car in a driving simulation (Virtual Test Drive, Vires). Participants' faces were filmed using an IP-Camera with a resolution of 1280 x 720 pixels at a sampling rate of 10 frames per sec.

Experimental Design and Cover Story

The experiment consisted of six experimental drives with time pressure and six baseline drives (see Figure 1). A cover story was used to create time pressure in the drives: Participants had to imagine that they work at a parcel delivery service, have to fetch parcels from the parcel service's headquarter (baseline drive), and deliver these to customers (experimental drives). Each parcel had to be delivered within six minutes and participants were told that they receive 15 € for the experiment plus a bonus of 2 € for each parcel delivered within the given time (the experiment was programmed so that all participants were successful in three out of six drives and thus received 21 €).

Participants were instructed to respect traffic rules and the speed limit of 50 km/h at all times. All drives took place on an urban single carriageway. During driving, a visual message on the screen warned participants in case they drove 7 km/h or more above the speed limit.

In the six *baseline* drives, participants were told to fetch the parcel from headquarters without time pressure. No other traffic and no pedestrians were present during these drives. The baseline drives took between 70 to 120 seconds and ended with a message on the screen confirming that the parcel was picked up from headquarters. Baseline drives were included to minimize carry-over effects from the different experimental drives.

Moreover, there were six different experimental drives with moderate traffic on the opposite lane. In these, participants had six minutes to deliver a package. In three of the six experimental drives, little traffic occurred on the ego lane, so that driving at the maximal allowed speed was mostly possible (*noFrust* condition). In two of these, after a fixed amount of time below six minutes (5:41 and 5:36 minutes), participants were told that they delivered the parcel successfully and gained 2 € extra. These drives served as a condition without frustration (noFrust1 and noFrust2). In order not to make the structure of the experiment too obvious, the third non-frustrated drive ended after six minutes with the message that the time was over and no extra money was gained (*noFrustDummy*).

The other three experimental drives contained traffic and road conditions inducing frustration (*Frust* condition). In two of these drives, after six minutes, participants were told that the parcel could not be delivered in time and that they did not gain extra money (*Frust1* and *Frust2*). In the third drive, participants were told that they successfully delivered the parcel after 5:40 minutes. This drive served as dummy drive preventing drivers from thinking that they cannot reach the goal whenever there are frustrating elements (*FrustDummy*). Frustrating events included red lights, construction sites, and slow or standing lead vehicles that could not be overtaken. The design of the frustrating drives was similar to experimental manipulations of earlier studies on driver frustration. These showed that frustration induced in simulator studies with relatively short driving periods could lead to aggressive behaviors such as speeding (Lee, 2010; Lee & LaVoie, 2014).

The order of the six baseline drives was the same for all participants. The experimental drives were presented in six different pseudorandom orders. Before the first baseline drive, participants completed a four minute training to become acquainted with the steering dynamics of the driving simulator. After the experiment, participants were debriefed. The total experimental procedure took roughly two hours.

Subjective Ratings

Participants completed the self-assessment manikin (SAM, Bradley & Lang, 1994) after each baseline and experimental drive assessing valence (from -4 [negative] to +4 [positive]) and arousal (from 1 [low] to 9 [high]). Moreover, participants completed the frustration scale of the NASA task-load index (NASA-TLX, Hart & Staveland, 1988) after each experimental drive. We considered the rating data from the two standard Frust scenarios (Frust1 and Frust2) and the two standard noFrust scenarios (noFrust1 and noFrust2). Data of the baseline and the dummy scenarios were not taken into account. Each of the scores was pooled across the two Frust and the two noFrust drives. The individual scales of the questionnaires were compared to each other (noFrust vs. Frust) with one-way ANOVAs. One participant did not fill in the NASA-TLX, so that only 27 datasets could be analyzed for the frustration scale.

Video Coding

Videos were analyzed using the FACS. One minute from one of the Frust and one minute from one of the noFrust drives were used. For this, we selected the drives of the Frust condition with the highest rating on the frustration scale and compared it to the noFrust drive with the lowest score of each participant (for the participant without NASA-TLX, the selection was based on the SAM valence rating). The fifth minute of each selected video was coded, because participants were expected to experience high frustration (because they already drove for 4 min), but are not anticipating the end of the drive yet (since about 1 min was left). Thus, in total, 2 (conditions) x 28 (participants) x 1 minute = 56 minutes of video material were coded by a certified FACS coder (author CD). To check the reliability of this coder, a second trained FACS coder (author MF) co-rated three randomly chosen Frust videos (the 5th minute) to calculate an inter-observer agreement (i.e., the number of AUs coded by both observers divided by the total number of coded AUs, see Ekman et al., 2002). The inter-observer agreement was .85.

For the four AUs from our hypothesis (AU4, 14, 23, 24), the frequency of occurrences was counted in the selected minute of the Frust and the noFrust conditions and compared to each other by means of non-parametric Wilcoxon tests (significance threshold of $p < .05$, one-tailed). The non-parametric tests were used, as most AU frequency distributions were not normally distributed.

Finally, we examined whether there was a difference in frequency between the Frust and noFrust condition for any other AU in an explorative way. Significance of these comparisons was evaluated using a series of Wilcoxon tests (significance threshold of $p < .05$, two-tailed). Effect sizes r for all Wilcoxon tests were calculated by dividing the Z score by the square root of the number of observations (which is twice the number of participants).

Results

Ratings

Participants rated the Frust condition as more negative, more arousing, and more frustrating than the noFrust condition (see Table 1).

Activity in Facial AUs

On average, action unit 4 (brow lowerer) occurred 36.4 times ($SD = 25.6$) in the Frust and 34.6 times ($SD = 26.5$) in the noFrust condition. There was no difference between Frust and NoFrust ($Z = 1.3$, $p = .09$, $r = .17$) as tested with the Wilcoxon test.

For AU14 (dimpler), the mean frequency of occurrence was 9.1 ($SD = 8.4$) in the Frust and 7.8 ($SD = 12.5$) in the noFrust condition. The difference between Frust and noFrust frequency was not significant ($Z = 1.5$, $p = .07$, $r = .20$).

Regarding AU23 (lip tightener), a mean frequency of occurrence of 2.0 ($SD = 6.6$) was revealed in the Frust condition. In the noFrust condition, this was .21 ($SD = .7$). The Wilcoxon test indicated a significant difference between conditions ($Z = 2.74$, $p < .01$, $r = .37$).

On average, participants showed AU24 (lip pressor) 4.9 times (*SD* = 7.3) in the Frust condition compared to 1.4 times (*SD* = 2.2) in the noFrust conditions. The difference in frequency of occurrence in AU 24 between the Frust and noFrust condition was significant (Z =-3.34, $p < .01$, $r = .45$) as tested with a Wilcoxon test (Figure 2).

The exploratory analysis of the remaining AUs revealed that AU 10 (upper lip raiser, Frust: 2.6±8.5, noFrust: .1±.4, $Z = 2.2$, $r = .29$), AU 12 (lip corner puller, Frust: 17.5±18.2, noFrust: 10.8±16.2, $Z = 2.7$, $r = .36$), AU 17 (chin raiser, Frust: 5.9±9.9, noFrust: 2.7±6.7, $Z = 2.7$, $r = .36$), and AU 20 (lip stretcher, Frust: 2.6±3.5, noFrust: .8±1.3, $Z = 2.9$, $r = .39$) were shown significantly more often in the Frust compared to the noFrust condition (Figure 3).

Discussion

The current study revealed that activity in the AUs 23 (lip tightener) and 24 (lip pressor) occurred significantly more often (with medium effect sizes) when participants were frustrated during driving compared to when they were not. In addition, we hypothesized that AU 4 (brow lowerer) and AU 14 (dimpler) are more frequently shown when frustrated. However, although these occurred descriptively more often, the differences had only small effect sizes and were not significant. An exploratory analysis additionally revealed that the AUs 10 (upper lip raiser), 12 (lip corner puller), 17 (chin raiser), and 20 (lip stretcher) were shown more often by frustrated than by non-frustrated drivers. The subjective reports of participants indicated that the experimental manipulation successfully induced frustration supporting the validity of the results of research on driver frustration. Previous studies on driver frustration have mostly studied behavioral effects of frustration (e.g. Lee, 2010; Lee & LaVoie, 2014) or used a rather data-driven approach (Malta et al., 2011) and did not focus on the facial signs that come along with frustration. Thus, this study is, to the best of our knowledge, the first to report facial signs accompanying spontaneously experienced frustration during simulated driving.

Our results are in line with previous fundamental and applied research on the expression of emotions of frustration and similar negative emotions (D'Mello et al., 2005; Ekman & Friesen, 2003; Grafsgaard et al., 2013). The small effect size of AU4 may be explained by the fact that it can be a sign of concentration necessary during driving (Ekman & Friesen, 2003). An additional exploratory analysis revealed that the AUs 10, 12, 17, and 20 were shown significantly more often during frustrated compared to non-frustrated driving. Although AU 12 refers to the lip corner puller, or zygomaticus major, generally attributed to a happy facial expression (smiling, e.g. Gosselin et al., 2010), research has also linked non-enjoyment smiles to frustration in human-computer interaction (Hoque, McDuff, & Picard, 2012), rendering its occurrence also likely during frustrated driving. To sum up, this study revealed that muscle activity especially in the mouth region appears to be indicative for frustration.

This study gives a first glance on the facial signs that come along with frustration; however, more research is needed to provide the full picture on how to measure frustration. In this study, we concentrated on the activity of AUs within a small time period. Thus, future studies need to address a one-to-one mapping between the revealed AU activities and frustration as well as consider the time course of the facial signs of frustration to build up reliable models that can be used to automatically recognize the degree of frustration of drivers. Moreover, in order to use AU activity as indicator for frustration in intelligent assistance systems for the mass market, it has to be investigated whether facial signs of frustration are comparable across individuals and contexts. Therefore, future research on frustration could benefit from realistic study conditions in real traffic or on test tracks and software packages for automated recognition of AU activity from videos to reduce the amount of time necessary for data analysis (e.g. Gehrig & Ekenel, 2011; Hamm, Kohler, Gur, & Verma, 2011). As affective states, such as frustration, are multi-component phenomena not only involving changes in facial expressions, but also in behavior, posture, or physiology (Scherer, 2005), a multimodal approach studying frustration should be employed. It has to be acknowledged though, that unlike basic emotion theories (Ekman, 1992) or appraisal theories of emotion (Scherer, 2005), the theory of constructed emotion postulates that there is no unique set of facial (and physiological) markers related to the experience of a particular emotion (e.g. see Barrett, 2016). In this line of argumentation, finding a one-to-one mapping between facial markers and experienced frustration would be impossible. Therefore, it has to be investigated whether the revealed facial action unit activity is related only to frustration, to negative affect in general, or even accompanies other emotional or cognitive states.

In sum, this article reports facial muscle activity that comes along with spontaneously experienced frustration during simulated driving. Results indicate that drivers show more activity in the mouth area when they are frustrated compared to when they are not.

Acknowledgement

The authors thank Dirk Assmann, Björn Hendriks and Michael Scholz for their technical assistance.

References

Barrett, L. F. (2016). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, nsw154. doi:10.1093/scan/nsw154

Bruce, V. (1992). What the human face tells the human mind: some challenges for the robot-human interface. In *Proceedings of IEEE International Workshop on Robot and Human Communication* (pp. 44-51).

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.

D'Mello, S., Craig, S., Gholson, B., Franklin, S., Picard, R., & Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. In *Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International Conference on Intelligent User Interfaces*. New York: AMC Press.

Deffenbacher, J. L., Lynch, R. S., Oetting, E. R., & Swaim, R. C. (2002). The Driving Anger Expression Inventory: a measure of how people express their anger on the road. *Behaviour Research and Therapy*, 40(6), 717-737. doi:10.1016/S0005-7967(01)00063-8

Donkor, R., Burnett, G., & Sharples, S. (2014). Measuring the emotional validity of driving simulators. *Advances in Transportation Studies*, (Special, Issue Special Vol1), 51–64.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3), 169-200. doi:10.1080/02699939208411068

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Cambridge, MA: Malor Books.

Ekman, P., Friesen, W. V., & Hager, J. (2002). *The Investigator´s Guide for the Facial Action Coding System*. Salt Lake City: A Human face.

Gao, H., Yuce, A., & Thiran, J.-P. (2014). Detecting emotional stress from facial expressions for driving safety. In *IEEE International Conference on Image Processing (ICIP)* (pp. 5961-5965).

Gehrig, T., & Ekenel, H. K. (2011). A common framework for real-time emotion recognition and facial action unit detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)* (pp. 1–6).

Gosselin, P., Perron, M., & Beaupré, M. (2010). The voluntary control of facial action units in adults. *Emotion*, 10(2), 266-271. doi:10.1037/a0017748

Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2013). Automatically recognizing facial indicators of frustration: A learning-centric analysis. In *Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 159-165).

Hamm, J., Kohler, C. G., Gur, R. C., & Verma, R. (2011). Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods*, 200(2), 237–256. doi:10.1016/j.jneumeth.2011.06.023

Hart, S. G., & Staveland, L. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139-183.

Healey, J., & Picard, R. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2), 156–166. doi:10.1109/TITS.2005.848368

Hoque, M. E., McDuff, D. J., & Picard, R. W. (2012). Exploring temporal patterns in classifying frustrated and delighted smiles. *IEEE Transactions on Affective Computing*, 3(3), 323-334. doi:10.1109/T-AFFC.2012.11

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist*, 46(8), 819–834. doi:10.1037/0003-066X.46.8.819

Lee, Y.-C. (2010). Measuring drivers' frustration in a driving simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(19), 1531–1535. doi:10.1177/154193121005401937

Lee, Y.-C., & LaVoie, N. (2014). Relationship between frustration justification and vehicle control behaviors ? A simulator study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 2235–2239. doi:10.1177/1541931214581429

Malta, L., Miyajima, C., Kitaoka, N., & Takeda, K. (2011). Analysis of real-world driver's frustration. *IEEE Transactions on Intelligent Transportation Systems*, 12(1), 109–118. doi:10.1109/TITS.2010.2070839

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. doi:10.1037/h0077714

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729. doi:10.1177/0539018405058216

Tews, T.-K., Oehl, M., Siebert, F. W., Höger, R., & Faasch, H. (2011). Emotional human-machine interaction: cues from facial expressions. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, . . . (Eds.), *Lecture Notes in Computer Science. Human Interface and the Management of Information. Interacting with Information* (Vol. 6771, pp. 641-650). Berlin: Springer.

Table 1

Mean and SD for Self-Assessment Manikin (SAM) and Frustration scale of NASA task-load index including the results of ANOVAs comparing the two conditions.

	Frust		noFrust		Frust vs. noFrust	
	M	*SD*	*M*	*SD*	*F*	*p*
SAM valence	-0.7	1.4	2.1	1.2	99.3[1]	<.001
SAM arousal	5.0	1.7	4.1	1.8	9.1[1]	<.01
Frustration	7.2	2.4	4.1	2.1	51.2[2]	<.001

Note. [1]$N = 28$, thus $F(1,27)$, [2]$N = 27$, thus $F(1,26)$.

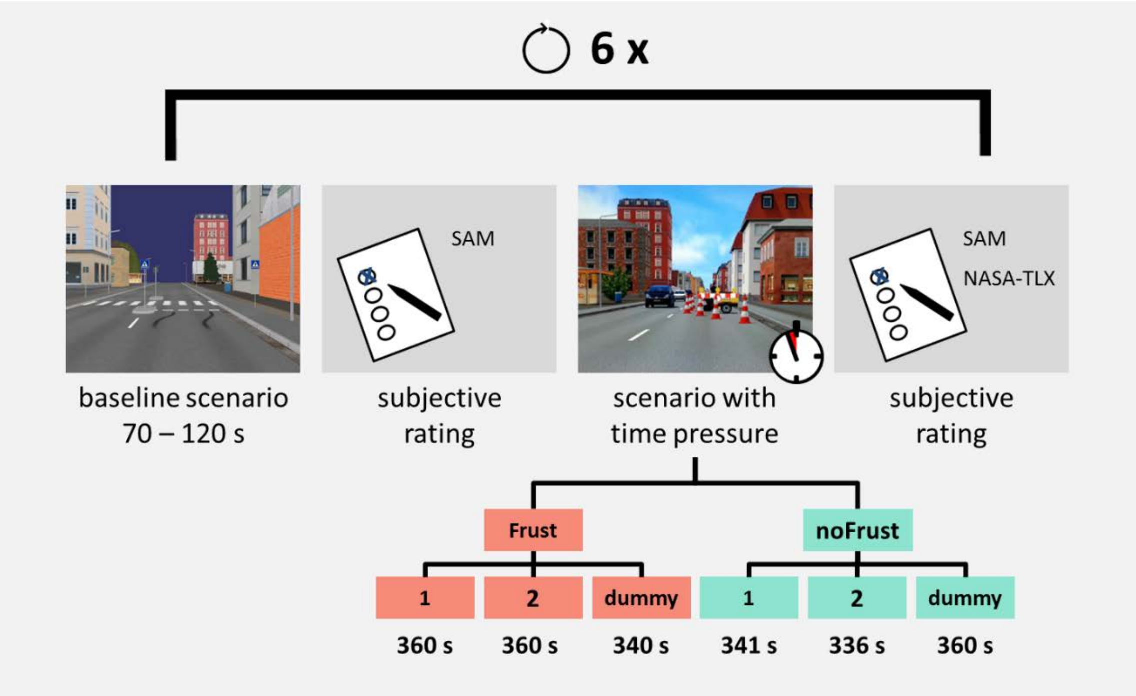

Figure 1. Sketch of the procedure of the experiment.

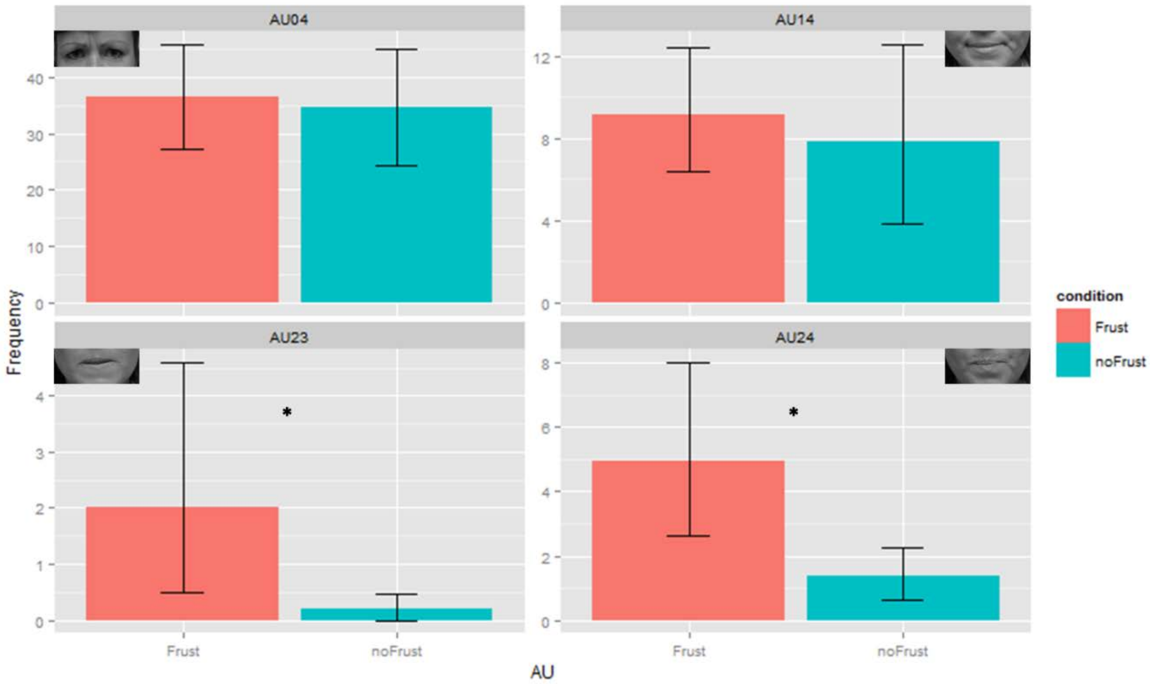

Figure 2. Mean (+ 95% confidence interval) frequency of occurrence of activity for the Actions Units (AUs) (subplots) of interest in the two conditions (color). Significant differences are marked with *.

Figure 3. Results of the exploratory analysis of the remaining Action Units (AUs) (subplots) stratified by condition (color). Significant differences between conditions are marked with *.