Automatic Detection of Hand/Upper Body Movement and Facial Expressions as Cues to Feelings of Exclusion

Marie Postma-Nilsenová (m.nilsenova@tilburguniversity.edu),[1] Eric Postma (e.o.postma@tilburguniversity.edu),[1] Martijn Balsters (m.j.h.balsters@tilburguniversity.edu),[1] Emiel Krahmer (e.j.krahmer@tilburguniversity.edu),[1] Juliette Schaafsma (j.schaafsma@tilburguniversity.edu),[1] Marc Swerts (m.g.j.swerts@tilburguniversity.edu),[1] Ad Vingerhoets (vingerhoets@tilburguniversity.edu)[2] Lisette Mol (l.mol@tilburguniversity.edu)[1]

[1] Tilburg center for Cognition and Communication (TiCC), Tilburg School of Humanities, Tilburg University PO BOX 90153, NL-5000 LE Tilburg, The Netherlands

[2] Tilburg School of Social and Behavioural Sciences, Tilburg University PO BOX 90153, NL-5000 LE Tilburg, The Netherlands

Abstract

We used a modification of the Frame Differencing Method to detect left and right hand/body movement in a corpus of recordings collected in two experimental conditions; a condition in which participants were included in a group decision-making process and one in which they were excluded. The results showed a lower degree of activation in the condition with exclusion, possibly due to withdrawal. An automatic detection of facial expressions indicated a difference with respect to expressions of Joy and Sadness; exclusion from the interaction led to decreased Joy and increased Sadness. Expressions of Joy were also correlated with increased hand/body movement.

Keywords: Exlusion; Hand Activation; Frame-Differencing Methods (FDMs); FACS; Facial Expressions.

Introduction

Past research has shown that qualities of upper body movement, such as amplitude, speed, and fluidity of movement, as well as its energy and spatial extent, can be used to recognize emotions (Wallbott, 1998; Glowinski, et al., 2011, a.o.). Given that interlocutors tend to be less aware of the signals conveyed by their body movement (as opposed to facial expressions, e.g., a smile), they can be used as more reliable cues to emotions and attitudes. In the present study, we explored hand/upper body activation in situations where the interlocutors might feel rejected by others, and compared the detected levels of activations to emotions measured in the face. Apart from the link between body activation and emotions, the goal of the study was to describe behavioral cues to feelings of exclusion. To our best knowledge, research on nonverbal behavior of excluded or ignored persons is sparse. However, in a number of situations, including organizational contexts and medical communication, it may be important to recognize early signs of social exclusion, lack of control and emotional distress. These signs may not always be obvious to others because of coping strategies that may involve detached and emotionally indifferent behavior (Krahmer et al., 2010). Their recognition with the help of automatic tools might help to create affective technologies with a wide range of interactive applications.

Current Study

We made use of an existing collection of recordings created at the University of Tilburg in 2009-201 as a part of a study on the effects of exclusion. In the study, fifty-eight undergraduates students participated in a 8-minute decision task with two confederates. They were randomly divided between two conditions; in one condition (Inclusion), they received the full attention of the confederates and their contributions were rewarded with positive remarks, in the second condition (Exclusion), they were mostly ignored by the confederates. The experiment was presented as a study on group decision-making under time pressure and participants were led to believe that they would be engaging in a discussion with other peers. In reality, they interacted with a pair of actors (one male, one female) who operated on the basis of an elaborate script. The experimental manipulation occurred after the initial 4 minutes of discussion. The contributions of included participants were continuously focused on by the confederates who emphasized how much they appreciated them, e.g., by stating “yes, that's an excellent suggestion”; in the exclusion condition, the confederates only appeared to appreciate each others' contributions and ignored those offered by the participant (for a full description of the experimental procedure, see Krahmer et al., 2010).

The data collected in the study have been previously analyzed with the help of a manual coding for nonverbal cues (Troisi, 2002) and a holistic perceptual study with naïve third-party observers (Krahmer et al., 2010). Both studies used thin slices of behavior extracted from the original recordings (two 30-second fragments and two 8-second fragments, respectively). The manual coding with the Ethological Coding System for Interviews (ECSI) showed a significant effect for the category Affiliation: the included participants showed more affiliative behavior than the excluded ones. The coding of Affiliation is based on features 2-6 of the ECSI scheme, all involving facial expressions and head movements (head to side, sharp upwards movement of the head, quick or slow raising and lowering of the eyebrows and a smile) and this result indicates that facial expressions can reflect participants feelings of inclusion/exclusion. Contrary to the expectations, excluded participants did not engage in more withdrawal-related non-verbal behaviors (in the ECSI scheme, features 10-15 indicating Flight: looking away, looking down, closing eyes, chin drawn towards the chest, crouching, and freezing). In the perceptual study, third-party raters evaluated whether participants were included or excluded. The results showed that in general, it is possible for external observers to detect from thin slices of behavior if a person is being in-/excluded; however, the standard deviations were relatively large suggesting possible individual differences in the use of display rules to mask or neutralize own feelings in the presence of others.

In our analysis, we focused on measures of non-verbal behavior that can be detected with existing computer techniques. Contrary to the previous analyses of the data which were based on thin slices of behavior, we made use of longer recordings segments in order to achieve more representative measurements. Given large individual differences in non-verbal behavior, we also included recordings made before the experimental manipulation in the analysis. Due to the fact that some of these recordings could not be digitally processed, the final dataset consisted of 53 participants (33 female) in two between-participant conditions (Included, Excluded) and two within-participant measurements (before Inclusion/Exclusion and after). We analyzed two types of behavioral cues, namely left and right-hand/upper body part activation (by means of splitting the screen into two, respectively four equal fields) and facial expression, using an implementation of the Frame Differencing Method and automatic Facial Action Coding System Action Unit detection with the Computer Expression Recognition Toolbox. Building on previous work, the outcome of the analysis of facial expressions (the standard behavioral measure of interlocutor's emotions) was compared to the measure of body part activation in the two experimental conditions.

Frame-Differencing Methods

Frame differencing has been used earlier in studies exploring interpersonal synchrony (Ramseyer, & Tschacher, 2008; Paxton, & Dale, 2012) as a way to decrease the time typically needed to analyze interactive data. In particular, manual hand activation coding is assumed to require four times the length of the recording for each hand (H. Lausberg, p.c.), with 20-25% of the data analyzed independently by two coders for reliability checks. Frame-differencing methods are fully automatic and relatively easy to implement. They have been used in the past in a number of domains (Ramseyer & Tschacher, 2011; Paxton & Dale, 2013) with various degrees of success. By tracking the changes in pixels from one frame to the next, given a static background and more or less stable lighting, the methods are able to detect a person's movement in a cost-effective manner. First, the absolute differences of the values of the matching pixels in subsequent frames (images) are computed. For static sequences, frame differencing leads to (absolute) differences of zeros, but for dynamic scenes, subsequent matching pixels may differ in value, yielding a non-zero difference. A measure of visual change is obtained by averaging over the differences for each pair of subsequent frames, see Figure 1. Frame differencing gives a useful measure of the amount of movements, as long as no other motion is present in the sequence.To minimize such interference from the background or irrelevant image regions, frame differencing can be confined to spatial regions of the image.

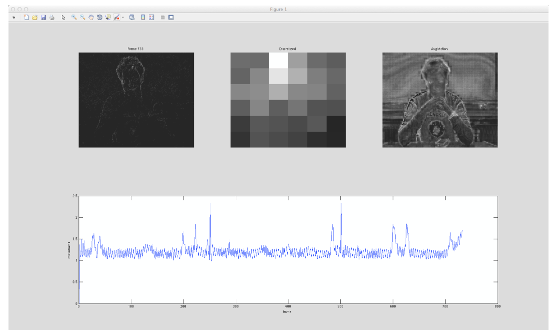

Figure 1: Frame-Differencing Methods.

It is not clear to what extent Frame-Differencing Methods can be compared to manual codings of hand activation, therefore, a comparison of the two approaches was included in the study. Rules to hand activation manual coding have been most explicitly formulated in the NEUROGES Coding System, a research tool designed to annotate and explore hand movement behavior and its anatomy (Lausberg & Slöetjes, 2009). In NEUROGES, Activation refers to the movement of right and left upper limbs, including fingers, hands, arms and shoulders. A movement is defined both by motion and by muscle contraction; a motionless phase may thus be coded as a part of the Activation unit if the hand is held in an anti-gravity position and the phase is framed by phases with movement. This definition is needed in order to include all parts of a single gesture into a single Activation unit; however, the fine degree of analysis may be difficult to achieve with computer-based activation recognition.

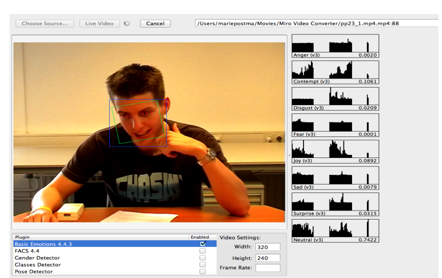

Figure 2: Detection of the face localization serves as the basis for FACS and emotion recognition with CERT.

Automatic FACS detection

With recent developments in the area of facial expression recognition system, a number of computer tools have become available that offer interpretation of basic emotions based on localization of key facial areas (eyes, nose tip, lips, head). Among these tools, the Computer Expression Recognition Toolbox (CERT) offers a fully automatic real-time recognition of the Facial Action Coding System (FACS) Action Units (Littlewort et al., 2011). FACS, originally developed by Ekman & Friesen (1978), is the most precise existing system used to code component movements of the facial muscles. Given a video sequence, CERT localizes the face (see Figure 2) and estimates the presence of facial action units by performing local Gabor transforms at informative facial locations. CERT has been trained on a large database of emotional expressions and has an action-unit recognition accuracy of approximately 80%. In our experiment, we use the estimates of the basic emotions as computed by CERT on the basis of combinations of individual facial action units.

Results

Using a mixed analysis of variance, we first examined the effect of the experimental manipulation (Inclusion vs. Exclusion) as a between-participant factor and the measurement moment (before vs. after the experimental manipulation) as the within-participant factor on the level of average Activation detected in the lower left bottom square of the recordings (LB ~ right hand), the lower right bottom square (RB ~ left hand), the left half of the recordings (L ~ right upper body part), the right half (R ~ left upper body part), and the total activation, viz. Figure 3.

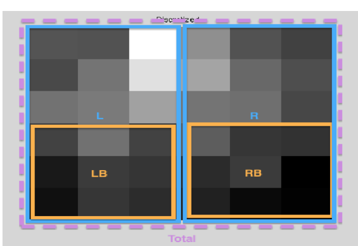

Figure 3: Decomposition of the recording into 5 activation areas.

The analysis revealed no main effect of manipulation and measurement for the LB activation. There was an interaction effect of manipulation and measurement moment for the LB activation, $F(1, 51) = 5.836$, $p = .02$, $\eta_p^2 = .10$. There was a main effect of measurement on the L activation, $F(1, 51) = 4.259$, $p = .044$, $\eta_p^2 = .08$, with no main effect of manipulation but with a significant interaction effect between measurement and manipulation $F(1, 51) = 9.120$, $p = .004$, $\eta_p^2 = .15$. There was a significant effect of manipulation on the RB activation, $F(1, 51) = 7.850$, $p = .007$, $\eta_p^2 = .13$, with no main effect of measurement and no interaction effect. For the R activation, we found no main effect of manipulation, but a significant effect of measurement, $F(1, 51) = 4.608$, $p = .037$, $\eta_p^2 = .08$. There was also an interaction effect for the R activation, $F(1, 51) = 4.844$, $p = .032$, $\eta_p^2 = .09$ and an interaction effect for total activation, $F(1, 51) = 7.557$, $p = .008$, $\eta_p^2 = .13$, with both a main effect of manipulation, $F(1, 51) = 4.804$, $p = .033$, $\eta_p^2 = .09$, and a main effect of measurement, $F(1, 51) = 4.739$, $p = .034$, $\eta_p^2 = .09$. The means are reported in Table 1. The results of the series of mixed ANOVAs indicate that the level of average Activation measured with the FDM can be used to distinguish between Inclusion and Exclusion. In particular, if we take into account inter-personal differences, the level of Activation in the left half of the screen (the location of the right-hand gestures) appears to be indicative of the interlocutors involvement in the interaction.

Table 1: Mean Activation detected by the Frame Differencing Method ($N = 53$).

Category	Included		Excluded	
	Before	After	Before	After
Left Bottom	**1.668**	**1.785**	**1.432**	**1.248**
	(0.734)	(0.639)	(0.864)	(0.865)
Right Bottom	1.738	1.827	1.250	1.224
	(0.697)	(0.766)	(0.766)	(0.783)
Left	**2.165**	**2.208**	**2.043**	**1.815**
	(0.519)	(0.508)	(0.554)	(0.585)
Right	**2.242**	**2.244**	**1.945**	**1.788**
	(0.568)	(0.566)	(0.575)	(0.604)
Total	**2.204**	**2.226**	**1.994**	**1.801**
	(0.525)	(0.518)	(0.549)	(0.591)

In the second part of the analysis, we explored the link between the outcomes of the Frame-Differencing Method and manual annotation of Hand Activation according to the NEUROGES guidelines. A random selection of 50 two-minute recordings was annotated for Left Hand (LH) and Right Hand (RH) Activation in Elan under supervision of a certified NEUROGES coder. The annotator and the supervisor were both blind to the experimental condition. A comparison of the total length of LH and RH manually coded activation and the activation measured with FDM gave no significant correlations, suggesting that the two methods are independent and that the FDM cannot be implemented as a substitute for the manual codings. In the third part of the analysis, in order to determine the effect of the experimental manipulation on participants' facial expressions compared before and after the manipulation, we conducted a series of tests with 8 emotions (Anger, Contempt, Disgust, Fear, Joy, Sadness, Surprise, and Neutral). Given that the measurements for all the emotions were not normally distributed (Shapiro-Wilk's test of normality $p < .001$), we made use of nonparametric statistics for the two conditions tested separately. In the condition where participants were included in the interaction ($N = 28$), the statistical outcomes of the Wilkoxon Signed Rank's test revealed no significant differences between the measurements before and after the experimental manipulation. In the condition with exclusion ($N = 27$), we found a significant effect for Joy, $Z = -2.714$, $p = .007$, and for Sadness, $Z = - 2.114$, $p = .034$, as well as a trend for Neutral expressions, $Z = -1.786$, $p = .074$. The participants exhibited less joyful and more sad expressions when they were excluded, viz. Table 2.

Table 2: Median emotions measured by automatic FACS detection.

Category	Included		Excluded	
	Before	After	Before	After
Anger	1.818	1.583	2.220	2.162
Contempt	36.016	39.323	43.985	35.565
Disgust	1.210	1.378	1.589	1.366
Fear	0.363	0.543	0.351	0.396
Joy	3.517	5.014	**4.846**	**2.123**
Sadness	5.063	4.814	**4.941**	**5.933**
Surprise	0.502	0.669	0.334	0.356
Neutral	40021	39.782	35.370	44.834

Finally, to explore the link between body movement and facial expressions, we conducted nonparametric correlation analyses for the Activation measurements and the facial expression measurements in the measurement after the experimental manipulation.

Table 3: Spearman's rho correlations between handy/body movement and emotions.

	LB	RB	L	R	Total
Joy	.51**	.55**	.58**	.55**	.56**
Neutral	-.30*	-.23	-.19	-.12	-.16
Surprise	.24	.28	.37**	.40**	.41**
Fear	.17	.30*	.25	.27**	.33*

Note: * $p < .05$, ** $p < .01$; *LB* = left bottom, *RB* = right bottom, *L* = left half, *R* = right half.

We found that activation in all the analyzed parts of the recording correlated strongly and positively with facial expressions of Joy. There was also a negative correlation between activation in the lower bottom half (~ right hand) and Neutral expressions and positive correlations between Surprise and general activation and Fear and activations especially in the right half of the screen, see Table 3.

Discussion and Conclusion

We employed recent computer techniques to analyze hand/upper body movement in relation to facial expressions of interlocutors included vs. excluded in the social interaction: the Frame Differencing Method and the automatic FACS detection with CERT. We found significant differences in the behavior of included vs. excluded interlocutors with respect to both body movement (especially in the right part of the upper body), and facial expressions indicating Joy and Sadness. We also observed a link between detected levels of activation and emotions expressed in the face, especially Joy. In future research, we intend to compare the measurements obtained with the two above mentioned automatic tools with more detailed gesture annotations with NEUROGES, as well as with an analysis of vocal data.

Acknowledgments

The authors would like to thank Susan Vermeer for her help with segmentation and annotation work and to Hedda Lausberg for a helpful discussion regarding the experimental procedure.

References

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., & Scherer, K. (2011). Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing, 2*, 106-118.

Krahmer, E., Schaafsma, J., Swerts, M., Balster, M., & Vingerhoets, A. (2010). Non-verbal responses to being ignored: evidence of cognitive deconstruction? *Proceedings of CogSci 2010*.

Lausberg, H., & Slöetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behaviour Research Methods, 41*, 841-849.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). *The Computer Recognition Toolbox* (CERT).

Paxton, A., & Dale, R. (2013). Frame-Differencing Methods for measuring bodily synchrony in conversation. *Behaviour Research Methods*, in press.

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body-movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology, 79*, 284-295.

Wallbott, H.G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28*, 879-896.