IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 12, 2021, accepted April 30, 2021, date of publication May 7, 2021, date of current version May 25, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3078195*

# Analyzing the Direction of Emotional Influence in Nonverbal Dyadic Communication: A Facial-Expression Study

**MAHA SHADAYDEH[1,*], LEA MÜLLER[2,*], DANA SCHNEIDER[3,*], MARTIN THÜMMEL[1], THOMAS KESSLER[3], AND JOACHIM DENZLER[1,4], (Member, IEEE)**

[1]Computer Vision Group, Friedrich Schiller University of Jena, 07743 Jena, Germany
[2]Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany
[3]Department of Social Psychology, Friedrich Schiller University of Jena, 07743 Jena, Germany
[4]DLR Institute of Data Science, 07743 Jena, Germany

Corresponding author: Maha Shadaydeh (maha.shadaydeh@uni-jena.de)

This work was supported by the Carl Zeiss Foundation for the financial support within the scope of the program line ''Breakthroughs: Exploring Intelligent Systems'' for ''Digitization—explore the basics, use applications.'' The work of Lea Müller was supported by the Computer Vision Group, Friedrich Schiller University of Jena. The work of Dana Schneider was supported by the DFG Scientific Network ''Understanding Others,'' SCHN 1481/2-1. The authors acknowledge support by the German Research Foundation and the Open Access Publication Fund of the Thueringer Universitaets- und Landesbibliothek Jena Projekt-Nr. 433052568.

[*]Maha Shadaydeh, Lea Müller, and Dana Schneider are co-first authors.

**ABSTRACT** Identifying the direction of emotional influence in a dyadic dialogue is of increasing interest in the psychological sciences with applications in psychotherapy, analysis of political interactions, or interpersonal conflict behavior. Facial expressions are widely described as being automatic and thus hard to be overtly influenced. As such, they are a perfect measure for a better understanding of unintentional behavior cues about socio-emotional cognitive processes. With this view, this study is concerned with the analysis of the direction of emotional influence in dyadic dialogues based on facial expressions only. We exploit computer vision capabilities along with causal inference theory for quantitative verification of hypotheses on the direction of emotional influence, i.e., cause-effect relationships, in dyadic dialogues. We address two main issues. First, in a dyadic dialogue, emotional influence occurs over transient time intervals and with intensity and direction that are variant over time. To this end, we propose a relevant interval selection approach that we use prior to causal inference to identify those transient intervals where causal inference should be applied. Second, we propose to use fine-grained facial expressions that are present when strong distinct facial emotions are not visible. To specify the direction of influence, we apply the concept of Granger causality to the time-series of facial expressions over selected relevant intervals. We tested our approach on newly, experimentally obtained data. Based on quantitative verification of hypotheses on the direction of emotional influence, we were able to show that the proposed approach is promising to reveal the cause-effect pattern in various instructed interaction conditions.

**INDEX TERMS** Direction of emotional influence, nonverbal human communication, dyadic interactions, facial action units, fine-grained facial expression, granger causality.

## I. INTRODUCTION

Analyzing the direction of influence between participants in dyadic interactions (i.e., human communication between two people) is of increasing interest in the psychological sciences with applications in psychotherapy [1], critical political interactions [2], and interpersonal conflict behavior [3], among others. In any interpersonal communication, attitudes towards the interaction partner play a crucial role in the outcome of this interaction (e.g., liking or disliking, engagement or disengagement, trust or no trust towards the interaction partner) [4]–[6]. Often self-report measures or verbal cues are used for getting at attitudes towards the interaction partner. However, most interaction partners are very adapted towards socially desirable behavior and thus would not express negative attitudes in an overt manner. The concept of embodiment [1], [7], [8] denotes the theoretical perspective that mental processes are not isolated from bodily processes.

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Ali Babar.

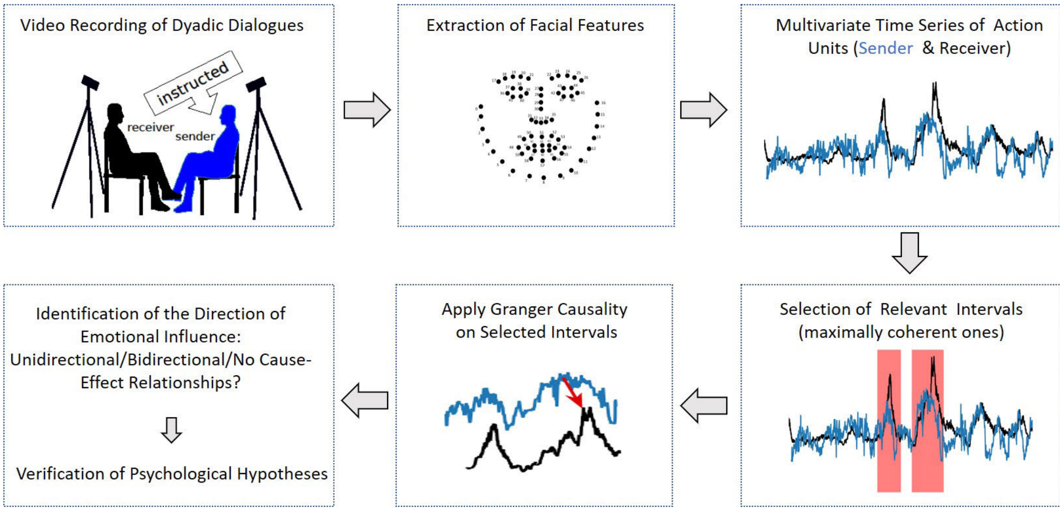



**FIGURE 1. The workflow of the proposed concept for analyzing the direction of emotional influence in dyadic dialogues.**

With this view, psychology is becoming increasingly sensitized to investigate the close association between mental and bodily parameters. In this paper, we are interested in the analysis of the direction of emotional influence in dyadic interaction pairs using non-verbal facial expressions.

Facial expressions, especially basic emotion expressions, are widely described as being automatic, thus hard to be overtly influenced and independent of culture [9]. As such, they seem a perfect measure for a better understanding of unintentional behavioral cues about socio-emotional cognitive processes. Furthermore, there is recent evidence that non-verbal emotion processing is much more relevant than verbal emotion processing [10] and that language plays a marginal role in the perception of emotion [11].

Novel developments in computer vision yielded accurate, open-source and real-time tools to easily extract bodily parameters from images and videos. In this paper, a complete concept for analyzing the direction of emotional influence in dyadic face-to-face dialogues when starting with raw video material is presented (cf. Figure 1). We combine computer vision capabilities with causal inference theory for quantitative verification of psychological hypotheses on the direction of emotional influence in experimentally recorded dyadic interaction dialogues using facial expressions, i.e., facial muscle activation (i.e., Action Units- AUs) [12].

Human nonverbal communication is a process of continual two-sided influence. In a dyadic interaction, such influence occurs over transient time intervals and with intensity and direction that are variant over time. To approach the problem of analyzing the direction of influence, we first present each person's facial emotional expressions in the form of time-series. To investigate who is influencing whom and how, we perform a cause-effect analysis on the obtained time-series using Granger causality (GC) [13], [14], the most widely used method for causal inference in diverse fields [15]–[18]. GC is based on the idea that causes both precede and help predict their effects.

The contributions of this work can be summarized as follows.

1. The use of GC for causal inference in nonverbal communication data has been addressed by several authors using different bodily parameters. However, to the best of our knowledge, no other work has used GC to identify the direction of emotional influence when it comes to facial expressions in dyadic dialogues.
2. Using facial expressions in the upper and lower face regions, we were able to extract fine-grained facial features from the six basic emotions [19] that are present when strong distinct facial emotions are not visible.
3. We present a relevant interval selection approach that we use prior to causal inference to identify those transient intervals where causal inference should be applied. The validation of the relevant interval selection on synthetic data for improved causal inference has been presented in [20] along with the initial results of this study on a small data set. Here we show-based on a larger real data set-the superiority of such an approach in detecting the direction of emotional influence when compared to applying the GC test on the entire time-series.
4. The current work is a study of covert attitudes in interaction partners from a 2nd person perspective [21], meaning in a truly interactive manner. In the past, studying real-time social encounters in an interactive manner was described as the "dark matter" of social (neuro-) science as it is typically explored from a 3rd person perspective [21]. Most studies in the field of social cognition [22] are conducted in a manner that interaction partners observe other interaction partners from a distance (e.g., by judging facial expressions of another person from a picture). Our work goes a step towards the "dark matter" and is looking at interaction partners

that are directly involved, thus continuously responding to the others' actions.

The remainder of this paper is organized as follows. Related work is summarized in Section II. We introduce the experimental setup and psychological hypotheses in Section III. The methodological details are presented in Section IV, followed by experimental results and discussion in Section V. Finally, conclusions are given in Section VI.

## II. RELATED WORK

The topic of finding causal structures in nonverbal communication data is addressed by Kalimeri *et al.* [23]. In their paper, the authors used GC for modeling the effects that dominant people might induce on others when it comes to nonverbal behavior (e.g., speech energy and body motion). Besides audio cues, motion vectors and residual coding bit rate features from skin-colored regions were extracted. In two systems, one for body movement and another for speaking activity, a small GC-based causal network was used to identify the participants with high or low causal influence. Unlike our approach, the authors did not use facial expressions and did not identify relevant intervals in a previous step, but used the entire time-series instead.

Kaliouby and Robinson [24] provided the first classification system for agreement and disagreement as well as other mental states based on nonverbal cues only. They used head motion and facial AUs together with a dynamic Bayesian network for classification. Also, a survey on cues, databases, and tools related to the detection of spontaneous agreement and disagreement was conducted by Bousmalis *et al.* [25]. Despite their ingenious methods, these approaches did not investigate cause-effect relations in a social interaction situation.

Postma and Postma-Nilsenova [26] used the convergent cross mapping method (a causal inference method originating from dynamical systems theory) to study nonverbal interactions. They concluded that there exists bidirectional causal coupling in facial dynamics. Unlike our study, their method focuses on bidirectional causality only rather than identifying all possible directions of emotional influences. Further, their approach was applied to two time-series of a single AU over the entire time-series rather than a combination of two or more AUs presenting emotions in selected intervals. Beyond this, Sheerman-Chase *et al.* [27] used visual cues to distinguish between states such as thinking, understanding, agreeing, and questioning to recognize the agreement. Most intriguingly, Matsuyama *et al.* [28] developed a socially-aware robot assistant responding to visual and vocal cues. For visual features, the robot extracted facial cues (based on OpenFace [29]) using landmarks, head pose, gaze, and facial AUs. Conversational strategies that build, maintain, or destroy connecting relationships were classified. The researchers' approach investigates the building of a social relationship between a human and a robot; however, this study does not deal with the time-variant direction of cause-effect relations.

Joo *et al.* [30] recently presented a motion capture data set and neural network architecture to analyze the direction of influence in triadic interactions. Their approach focuses on predicting social signals, such as speaking status, social formation, and body gestures. The face is represented using a deformable face model with five facial expression parameters. In contrast to Joo *et al.*, we are only interested in facial expressions. This requires a more detailed and supervised representation of facial motion.

Several methods, based on temporal causality analysis, have been proposed for categorizing interactions of individuals in video sequences [31]–[33]. These methods use Granger causality, but they are based on motion analysis of different individuals and are not concerned with facial emotional influences.

## III. EXPERIMENTAL SETUP AND PSYCHOLOGICAL HYPOTHESES

We created an experimental setup in which two participants sat face to face while talking about their personal weaknesses. One participant was in the assigned role of the receiver (R), the other was in the assigned role of the sender (S) (first block in Figure 1). S was instructed to take on a certain attitude (i.e., respectful, objective, contemptuous), whereas R was unaware of that and acted spontaneously.

In total, participants were asked to talk to each other about their personal weaknesses three times, either in circumstances of a respectful, contemptuous, or objective/neutral situation. Both partners were given specific times when to speak and when to listen. In all three experimental conditions, each participant kept their initially assigned role of acting as an S or R. The experimental conditions (i.e., respectful, contemptuous, objective) were conducted in various orders for all pairs to avoid the possible confounding effect of the order of these conditions. The objective/neutral condition consisted of the instruction that S should behave as neutral as possible (i.e., trying not to react readily towards the interaction partner). Thus, S was asked to take time to consider the world through the eyes of the other in order to understand their perspective in an uninvolved manner. As only S had the active experimental interaction attitude task (i.e., to behave either respectfully, contemptuously, or objectively), we expected S to influence R in relevant facial expressions. In order to avoid flirtatious situations, that may overwrite the instructed condition, interaction partners were always of the same gender.

Induction of the experimental conditions (i.e., respectful, contemptuous, objective situation) was secured by the fact that after each conversation both interaction partners filled in a self-report questionnaire. Within that, they reported on their positive affect during the interaction (i.e., on 4 items using a 7-point Likert scale). As expected, in the respectful condition, both interacting partners indicated the highest positive affect (mean value: $M = 5.61$ and standard deviation: $SD = 0.91$), followed by the objective condition ($M = 5.40$, $SD = .98$), followed by the contemptuous condition ($M = 4.15$, $SD = 1.36$).

To capture nonverbal facial behavior, we positioned two frontal perspective cameras (first block in Figure 1), recording at 25 frames per second. Camera positions and lighting conditions were optimized during a test session before the study started. This ensured high video quality in terms of a plain frontal view of the faces and two-sided illumination. Motion blur rarely occurred, but could not be prevented entirely, especially in cases of faster movements like head turns. Except for the label of the experimental condition no other information (e.g., expression annotation per frame) was available for image analysis. The entire data set consisted of 34 pairs (mean age = 20.72, 24 female pairs, German-speaking participants), three conditions per pair, and about four minutes of video per condition for each of the interaction partners, thus about 13 hours of video material. All participants gave written informed consent. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Friedrich Schiller University of Jena. The ethical clearance number is FSV 18/04. Further, the study was pre-registered with aspredicted.org under the title "Social Communication" #13078.

The psychological research question was, whether and how S and R influence each other given different attitude situations. In particular, we were interested in testing the following psychological hypotheses (**H.1-H.4**).

**H.1.** More harmonic expressions (i.e., happiness) present when both interaction partners are confronted with medium to high levels of respect (i.e., respectful and objective/neutral vs. contemptuous).

**H.2.** The strongest activation of negative expressions (e.g., sadness) presents in the disrespectful condition (i.e., contemptuous vs. respectful and objective/neutral).

**H.3.** For all emotional facial actions, S causes the effects and influences R.

**H.4.** In terms of the different facial expressions, the strongest GC causality from S to R occurs for positive expressions (i.e., happiness), followed by negative expressions (e.g., sadness).

## IV. METHODOLOGY

Multiple steps were necessary to get from raw video materials of dyadic interaction dialogues to the detection of the direction of emotional influence between interaction partners. Figure 1 shows the workflow and main steps of the proposed method. In the following subsections, we introduce these steps starting with the feature extraction procedure, the relevant interval selection approach, and the causal inference method. Then, we combine all these steps to elucidate our entire approach.

### A. FACIAL FEATURE EXTRACTION

According to Ekman and Rosenberg [34], facial expressions are the most important nonverbal signals when it comes to human interaction. The Facial Action Coding System (FACS) was developed by Ekman and Friesen [12], [35]. It specifies facial AUs, based on facial muscle activation. Examples of AUs are the *inner brow raiser*, the *nose wrinkler*, or the *lip corner puller*. Any facial expression is a combination of facial muscles being activated, and thus, can be described by a combination of AUs. Hence, the six basic emotions (*anger*, *fear*, *sadness*, *disgust*, *surprise*, and *happiness*) can also be represented via AUs. When for example AU 6 (*cheek raiser*), 12 (*lip corner puller*), and 25 (*lips part*) are activated, the facial expression of *happiness* is visible [36].

**TABLE 1. Percentage of frames where emotions were detected across all experimental conditions.**

| Emotion | Occurrence (in %) |
|---|---|
| Happiness | 12.9 |
| Surprise | 1.00 |
| Anger | .15 |
| Disgust | 3.07 |
| Fear | .06 |
| Sadness | 1.55 |

In general, all dynamic facial expressions are visual nonverbal communication cues transferable to time-series. Regarding our real experimental data, this approach is reasonable for positive emotions like *happiness*, which is frequently visible throughout dyadic interactions. Yet, it is not applicable for negatively associated emotions such as *anger, disgust, fear,* or *sadness* because of social masking (i.e., these emotional expressions are often only visible in a subtle manner). This is emphasized in Table 1, where we show the computed visibility of emotions as defined in [36] in all recordings.

Wegrzyn *et al.* [37] studied the relevance of facial areas for emotion classification and found differences in the importance of the eye and mouth regions. Accordingly, AUs can be divided into upper and lower AUs [38]. Upper AUs belong to the upper half of the face and cover the eye region, whereas AUs in the lower half of the face cover the mouth region.

We decided to split facial expressions of basic emotions into upper-face and lower-face emotion expressions, according to the affiliation of AUs to upper and lower face regions. For example, instead of viewing *sadness* as a combination of AU1, AU4, AU15, and AU17, we used *sadness upper* (AU1 and AU4) and *sadness lower* (AU15 and AU17). All other emotions were split according to their AUs belonging to the upper or lower facial half. This procedure ensured that subtle facial expressions were also detectable. Table 2 shows for all relevant emotional expressions the corresponding upper-face and lower-face AUs. We used the mapping between AUs and the emotional expressions *happiness, surprise, disgust, fear, sadness, and anger* as proposed in [34]. We then split the AUs per expression into upper-face and lower-face AUs, following the correspondence between AUs and upper/lower face regions as provided in [37]. For *anger lower*, we only used AU17 and AU23, as AU24 is not detected by OpenFace2.

In Table 3 the detection percentage of upper-face and lower-face expressions is illustrated. Splitting the facial areas improved our detection algorithm significantly (see Table 1 versus Table 3). For example, emotional expressions *anger*

**TABLE 2. The table shows the used emotional expressions with the corresponding AUs. In this mapping, we follow the relationship between AUs and emotional expressions in [34] and the correspondence between AUs and upper/lower face regions in [37].**

| Expression | Active AUs |
|---|---|
| Happiness upper | 6 |
| Happiness lower | 12, 25 |
| Surprise upper | 1, 2, 5 |
| Surprise lower | 26 |
| Disgust lower | 9, 10, 25 |
| Fear upper | 1, 2, 4, 5 |
| Fear lower | 20, 25 |
| Sadness upper | 1, 4 |
| Sadness lower | 15, 17 |
| Anger upper | 4, 5, 7 |
| Anger lower | 17,23,24 |

**TABLE 3. Percentage of detected emotional expressions in the upper and lower face parts visible throughout the experiment, across all conditions.**

| Emotion | Detection (in %) |
|---|---|
| *Anger lower* | 8.20 |
| *Anger upper* | .67 |
| *Disgust lower* | 3.07 |
| *Fear lower* | 5.94 |
| *Fear upper* | .97 |
| *Happiness lower* | 16.24 |
| *Happiness upper* | 26.32 |
| *Sadness lower* | 9.28 |
| *Sadness upper* | 7.05 |
| *Surprise lower* | 26.91 |
| *Surprise upper* | 2.48 |

**FIGURE 2. Participant with different facial expressions. From left to right: *happiness, sadness lower*, and *sadness upper*.**

*lower*, *sadness lower*, *sadness upper*, and *surprise lower* could be detected in over 7% of the video material on average. Figure 2 illustrates a participant with different facial expressions.

We evaluated the detection accuracy by mapping AUs to emotions based on the AffectNet dataset [39]. We reached an accuracy of 25.8% for the six basic emotions including contempt and neutral. Especially happiness was detected very well, with 77% accuracy, whereas fear and surprise classification seems to be more complex. For feature extraction, we used OpenFace 2.0 [29], [40] which is a state-of-the-art, open-source tool for landmark detection; it estimates AUs based on landmark positions. OpenFace preserves much of the information by regressing AUs instead of only classifying them and is capable of extracting 17 different AUs (1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 45) with an

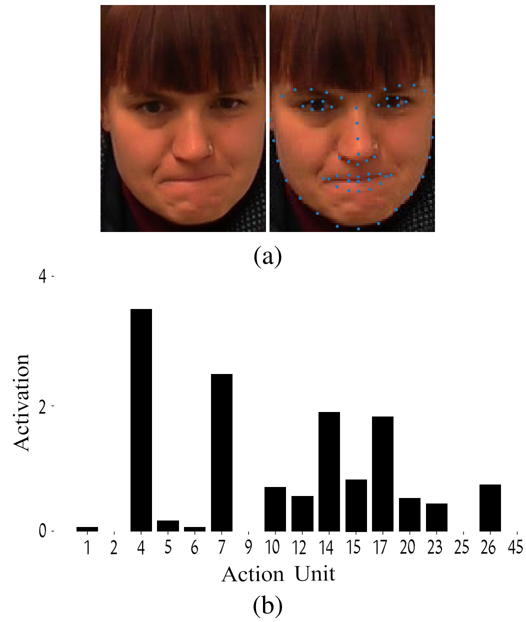



**FIGURE 3. Detection of facial expression *angry*. (a) Landmarks of the facial expression; (b) AUs detected by OpenFace: Strong activation of AU4 (*brow lowerer*), 7 (*lip tightener*), 14 (*dimpler*), and 17 (*chin raiser*).**

intensity ranging from 0 to 5. Figure 3 illustrates the detection of landmarks and AUs for an example image.

Let AUI denote the action unit I, e.g., AU15 is action unit 15. From the OpenFace detections, we computed each participant's average face across the three conditions. That is, for each participant we computed the mean $\mathrm{AUI}_{mean}$ and standard deviation $\mathrm{AUI}_{std}$ of the activation of AUI over all three experimental conditions. An AUI is considered as activated in time frame k, if $\mathrm{AUI}_k \geq \mathrm{AUI}_{mean}+0.5*\mathrm{AUI}_{std}$. Then an emotional expression is counted as activated when all its corresponding AUs are activated. For example, *sadness lower* is considered to be visible in frame $k$ if $\mathrm{AU15}_k \geq \mathrm{AU15}_{mean} + .5 * \mathrm{AU15}_{std}$ and $\mathrm{AU17}_k \geq \mathrm{AU17}_{mean} + .5 * \mathrm{AU17}_{std}$ hold. Following this step, the number of activations per person and per expression was counted for each experimental condition, and normalized by the video length and maximum count of the expression.

### B. CAUSAL INFERENCE WITH GRANGER CAUSALITY

Let $X$ and $Y$ be two stationary time-series with zero mean and length $L$. It is said that a time-series $X$ Granger causes a time-series $Y$ if the inclusion of past observations of $X$ beside $Y$ improves the prediction of $Y$ significantly when compared to the prediction using only past values of $Y$. These two time-series can be represented by the following two vector autoregressive models.

$$X_t = \sum_{j=1}^{M} a_j X_{t-j} + \sum_{j=1}^{M} b_j Y_{t-j} + \varepsilon_t \tag{1}$$

$$Y_t = \sum_{j=1}^{M} c_j X_{t-j} + \sum_{j=1}^{M} d_j Y_{t-j} + \vartheta_t \tag{2}$$

where $t = 1 \dots L$ denotes the time index, $\varepsilon_t$ and $\vartheta_t$ being two independent noise processes. The model order $M$ defines the maximum lag used to estimate causal interactions. It can be estimated using either Akaike [41] or Bayesian Criterion [42]. The model parameters $a_j, b_j, c_j, d_j, j = 1, \dots, M$ can then be estimated using, for example, the method of Least Squares (LS) [43].

To test whether $Y$ Granger causes $X$, two vector autoregressive models are compared. The first model, in which $Y$ is included for predicting $X$, as in (1). The second model is

$$X_t = \sum_{j=1}^{M} a'_j X_{t-j} + \varepsilon'_t \tag{3}$$

where $Y$ is not included. Those models are then compared against each other via a statistical significance test, where the null hypothesis $H_0$ is tested against the alternative $H_1$, with

$$H_0 : b_1 = \cdots = b_M = 0, \tag{4}$$

$$H_1 : \exists b_k \neq 0 \quad k \in \{1 \dots M\}. \tag{5}$$

These hypotheses are equal to the variable selection problem in linear regression. Hence, an F-Test is applicable with

$$F = \frac{(|\Sigma_{\varepsilon'}| - |\Sigma_{\varepsilon}|)(T - 2M - 1)}{|\Sigma_{\varepsilon'}|\, M} \tag{6}$$

where $\Sigma_{\varepsilon'}$ is the covariance matrix of the residual $\varepsilon'_t$ of the simple model in (3), and $\Sigma_\varepsilon$ is the covariance matrix of the residual $\varepsilon_t$ of the enriched model in (1). $F$ follows an F-distribution, with $(M, T - 2M - 1)$ degrees of freedom. The null hypothesis ($Y$ does not Granger cause $X$) can be rejected at a level of significance $\alpha$, if $F > F_{1-\alpha}(x; M, T - 2M - 1)$ where $F_{1-\alpha}(x; M, T - 2M - 1)$ denotes the value $x$, where the $F(x; M, T - 2M - 1) = 1 - \alpha$.

To test whether $X$ Granger causes $Y$, the above steps can be applied, but with the first model as in (2), the second model being $Y_t = \sum_{j=1}^{M} d'_j Y_{t-j} + \vartheta'_t$. When testing for GC, three different cases regarding the direction of influence can occur [44]:

1. If $c_k = 0$ for $k = 1 \dots M$ and $\exists b_k \neq 0$ for $1 \leq k \leq M$ then $Y$ Granger causes $X$.
2. If $b_k = 0$ for $k = 1 \dots M$ and $\exists c_k \neq 0$ for $1 \leq k \leq M$ then $X$ Granger causes $Y$.
3. If for both $\exists b_k \neq 0$ for $1 \leq k \leq M$ and $\exists c_k \neq 0$ for $1 \leq k \leq M$ then a bidirectional (feedback) relation exists.

If none of the above cases holds, $X$ and $Y$ are not Granger causing each other.

### *C. RELEVANT INTERVAL SELECTION*

Considering the experimental setup, we had to expect multiple temporal scenes, further referred to as subintervals, in which the participants influenced each other. The time spans where causality is visible might range from half a second to half a minute. This may occur several times and can be interrupted by irrelevant scenes that differ in the length of time. As outlined above, the direction of influence in a subinterval can either be bidirectional or unidirectional driven by either S or R. This implies that three unwanted effects can occur if the full-time span is analyzed. First, temporal relations are not found at all; second, bidirectional relations mask temporal unidirectional relations and; third, unidirectional relations from $X$ to $Y$ mask temporal bidirectional influences or unidirectional influences from $Y$ to $X$ and vice versa.

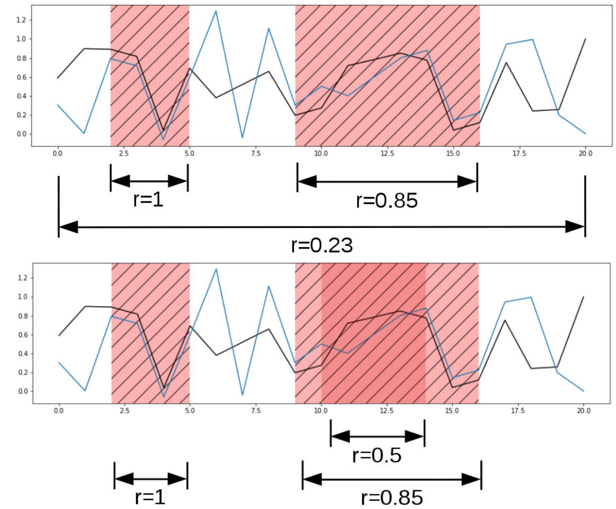
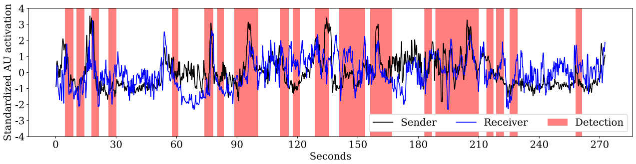



**FIGURE 4. Synthetic example of two time-series showing wrong reasoning of correlation due to surrounding intervals. Upper graph shows an example of highly correlated subintervals of these time-series (Pearson correlation parameter, r = 0.85 and r = 1) within low correlated interval (r = 0.23). While the lower graph shows an example of a weakly correlated subinterval (r = 0.5) within a highly correlated interval (r = 0.85).**

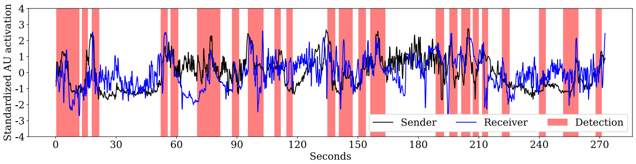
The estimation accuracy of the cause-effect intensity parameters $a_j, b_j, c_j, d_j, j = 1, \dots, M$ is mainly influenced by the accuracy of estimating the correlation of the two time-series $X$ and $Y$ at different time shifts. When the time-series contain several intervals of irrelevant information, the transient similarity in $X$ and $Y$ may not be captured. Figure 4 illustrates that two-time-series can have highly correlated subintervals within low correlated full-time span interval or low correlated subintervals embedded in high correlated full-time span interval. As such, similarity measures that are applied to the entire time range would fail to capture transient similarities.

Our central idea is to apply GC only to time-series obtained by concatenating highly coherent (e.g., in terms of Pearson correlation) subintervals of raw time-series. Instead of using a brute force algorithm, we suggest using a bottom-up approach for finding the longest set of maximal, non-overlapping, correlated intervals in time-series as proposed by Atluri *et al.* [45]. The authors applied their approach to fMRI data where they achieved good results for clustering coherent working brain regions.

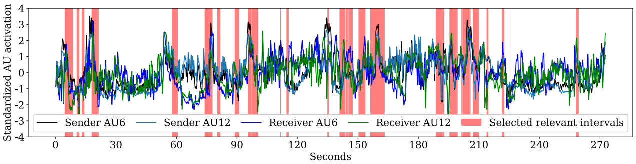
Let $X$ and $Y$ be two time-series of length $L$. An interval is called *correlated interval* for a threshold $\beta$ when all its subintervals up to a lower interval length $L_{min}$ are correlated as well. An interval $I_{(a,b)}$ from $a$ to $b$ is called maximal, when



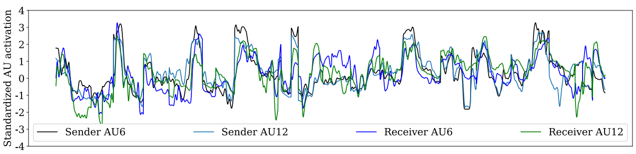

(a) Relevant interval selection for AU6.




(b) Relevant interval selection for AU12.




(c) Selected intervals for a facial expression with AU6 and AU12 being activated ((a) ∩ (b)).




(d) Concatenation of the intervals highlighted in (c).

**FIGURE 5. Process of relevant interval selection in multivariate case, exemplified by AU6 and AU12.**

$I_{(a,b)}$ is a *correlated interval*, but $I_{(a-1,b)}$ and $I_{(a,b+1)}$ are not. And two intervals $I_{(a,b)}$ and $I_{(c,d)}$ are called non-overlapping, when $I_{(a,b)} \cap I_{(c,d)} = \emptyset$. From all intervals fulfilling these conditions the longest set (total length of intervals) is computed.

In the multivariate case, e.g., when multiple AUs define an expression, we propose to compute the set of relevant intervals for each AU. Using the detection of each AU pair, the intersection overall AUs can be used as the set of selected relevant intervals for the expression. It is important to note that relevant intervals are computed for different time shifts to take a possible delay in emotional influence into consideration.

Figure 5 illustrates the selection process, exemplified by AU6 and AU12 of the Sender and Reciever. First, the relevant intervals, i.e., maximally correlated intervals between S and R are computed for AU6 and AU12 separately. We used time shifts of $s \in \{0, 4, 8, 12\}$ frames to take into consideration possible delay in emotional influence. We used a minimum interval length of $L_{min} = 75$ frames. After that step, additional processing steps, such as median filtering can be applied. Figure 5(a) shows the relevant intervals of AU6 and Figure 5 (b) shows the relevant intervals for AU12, where both sets of relevant intervals were median filtered with kernel size 51. Thereafter, the selected relevant intervals are obtained by computing the intersection between the relevant intervals of AU6 and AU12, as visualized in Figure 5 (c). Finally, the selected relevant intervals can be concatenated for further processing, as illustrated in Figure 5 (d).

### D. CAUSAL INFERENCE WITH RELEVANT INTERVAL SELECTION AND GRANGER CAUSALITY

There are two major challenges in the analysis of the emotional cause-effect relation in dyadic dialogues. First, due to the constructed situations, strong distinct emotions, computed by using traditional AU combinations, were barely visible. Second, the time-variant and situation-dependent communication, resulted in high variety and volatility of time spans in which cause-effect behavior between interaction partners is visible. To tackle these difficulties, we use a combination of the time-series of the facial features described in (IV-A) and the proposed relevant interval selection approach (IV-C), as detailed in the following steps.

1. We applied the relevant interval selection approach pairwise to the time-series of the identified AUs for all of the relevant facial expressions as illustrated in Figure 5 (a) and (b), with a minimum interval length of 75 frames and a threshold of 0.8 for Pearson correlation. Based on known average human reaction times (ca. 200 ms or 6 frames [46]), we shifted one time-series by 0, 4, 8, and 12 frames both, back and forth in time, and computed relevant intervals. The grid selected for shifting does cover quicker and slower reactions of participants. Afterward, we computed the longest set of the list of relevant intervals obtained from the different shifts. Before computing GC, we median filtered the selected intervals with a filter length of 51 (2 seconds) and extended the intervals by 12 frames on each side. We removed frames for both, S and R, when for either S or R the OpenFace confidence value was below.89. The confidence score gives a rough orientation about how reliable the AU score computed by OpenFace is. A low confidence score might occur when, for example, the face is occluded or the person moved quickly, which can lead to blurred images.
2. We calculated the average GC on the set of selected intervals of the standardized time-series. The results were counted according to the possible outcomes of the GC test as described in Section IV-B, as either unidirectional caused by S, unidirectional caused by R, bidirectional, or no causality.

## V. EXPERIMENTAL RESULTS AND DISCUSSION

### A. APPEARANCE OF FACIAL EXPRESSIONS IN THE DIFFERENT EXPERIMENTAL CONDITIONS

We first applied a Wilcoxon signed-rank test to compare how often in percent of video length a specific expression was visible between the different experimental conditions (i.e., respectfully, contemptuously, and objectively). We permuted all conditions and tested all expressions reported in Table 2 against each other, summing up to 33 tests in

**TABLE 4.** Results of the Wilcoxon signed-rank (W) test. The table shows all expressions that were significantly different in their occurrence between different conditions after Benjamini-Hochberg p-value correction (p =.05 and Q =.05). The mean activation reports the visibility of an expression in percent of video length across all pairs. From the mean activation value we can conclude in which (First or Second) condition the expression occurred more frequently. These conditions are printed in bold. We permuted all conditions and tested all expressions.

| Expression | First condition | Second condition | p-value | W test statistics | First condition's mean activation | Second condition's mean activation |
|---|---|---|---|---|---|---|
| Happiness upper | **respectful** | objective | .0001 | 511 | **0.4** | 0.29 |
| | contempt | **respectful** | 0.0035 | 695 | 0.32 | **0.4** |
| Happiness lower | contempt | **respectful** | 0.0003 | 580 | 0.29 | **0.41** |
| | **respectful** | objective | 0.0017 | 660 | **0.41** | 0.31 |
| Sadness upper | **contempt** | objective | 0.0002 | 571 | **0.39** | 0.28 |
| Sadness lower | **respectful** | objective | 0.0005 | 606 | **0.38** | 0.29 |
| | **contempt** | objective | 0.0088 | 744 | **0.34** | 0.29 |
| Fear lower | **respectful** | objective | 0.0068 | 730 | **0.36** | 0.32 |
| Anger lower | **contempt** | objective | 0.0020 | 667 | **0.36** | 0.31 |

**TABLE 5.** Number of pairs for which the Granger causality (GC) test, with p-value = 0.05, showed a specific direction of influence in the respectful condition. Average count is the average count of pairs across all expressions listed in the table. Dominant causal direction (the higher value between Sender GC Receiver (S GC R) and Receiver GC Sender (R GC S)) is shown in bold font.

| Expression | Full Time Span | | | | Relevant Interval Selection | | | |
|---|---|---|---|---|---|---|---|---|
| | S GC R | R GC S | Bidirectional | No causality | S GC R | R GC S | Bidirectional | No causality |
| Happiness lower | **4** | 1 | 22 | 7 | **8** | 4 | 14 | 8 |
| Happiness upper | 1 | **3** | 24 | 6 | **6** | 4 | 18 | 6 |
| Sadness lower | 6 | 6 | 4 | 18 | **9** | 3 | 8 | 14 |
| Sadness upper | **6** | 0 | 7 | 19 | 7 | **8** | 6 | 11 |
| Average count | **4**.25 | 2.5 | 14.25 | 12.5 | **7**.5 | 4.75 | 11.5 | 9.75 |

total. In order to decrease the false discovery rate, we used a Benjamini-Hochberg p-value correction [47] with a false discovery rate of Q =.05. Table 4 summarizes the results of the Wilcoxon signed-rank test. In line with the psychological hypothesis **H.1**, participants showed significantly more *happiness upper* and *happiness lower* in the respectful condition than in the contempt and objective condition. Furthermore, we found more *sadness lower* and *sadness upper* expressions in the contempt compared to the objective condition. This supports our psychological hypothesis **H.2**, and demonstrates that our instructed attitude manipulations for the sender had an effective influence on both interaction partners (i.e., sender and receiver). Note that the expressions for *fear lower* occurred against expectations (i.e., those were found more in the respectful compared to the objective condition). This is highly valuable information as it suggests that facial happiness and sadness expressions are the most indicative when comparing negative and positive interaction attitudes.

### B. DIRECTION OF EMOTIONAL INFLUENCE

In order to study the direction of emotional influence, we compared the results of the GC test on the relevant interval selection approach versus the results of the GC test on the full-time span approach. The comparison is represented by the count of pairs for which the Granger causality (GC) test, with p-value = 0.05, showed a specific direction of influence, under the three experimental conditions (Tables 5 (i.e., respectful), 6 (i.e., contemptuous), and 7 (i.e., objective)), for the expressions: happiness upper/lower and sadness upper/lower. These results clearly indicate that the use of the relevant interval selection approach prior to causal inference resulted in considerably more pairs showing unidirectional causation as well as less bidirectional or no causation.

Most interestingly, S influences R particularly in the respectful and objective/neutral compared to the contemptuous condition when using the relevant interval selection approach which partially supports **H.3**. However, in the contemptuous condition, the pattern of dominant influence changes (i.e., on average R and S influence each other similarly). Psychologically this indicates that the receiver in this condition is actively trying to repair the overall negative interaction quality, for example by inducing empathic concern in the sender. This result indicates that the use of facial expressions only has the potential to reveal covert attitudes and behaviors that would easily be missed and overlooked when working with verbal behavioral cues. That is, most likely in situations like these (i.e., an interaction partner acting ignorant and dismissive), a receiver of such contemptuous information would simply produce less speech content.

While we expected a considerable reduction in emotional influence from S to R for negative compared to positive expressions (**H.4**), our results did not confirm that but indicated only a slight reduction in the influence from S to R for negative expressions. Note that in the full-time span approach, the influence of S on R is even higher for negative expressions than positive ones, which is opposite to the initial hypotheses. However, beyond initial hypotheses, we noticed (see Tables 5, 6, and 7) that significantly higher bidirectional influences were observed for positive compared to negative emotions across all experimental conditions and in both approaches. This specifies our initial hypotheses and is once again highly valuable psychological insight as

**TABLE 6. Number of pairs for which the Granger causality (GC) test, with p-value = 0.05, showed a specific direction of influence in the contempt condition. Average count is the average count of pairs across all expressions listed in the table. Dominant causal direction (the higher value between Sender GC Receiver (S GC R) and Reciever GC Sender (R GC S)) is shown in bold font.**

| Expression | Full Time Span | | | | Relevant Interval Selection | | | |
|---|---|---|---|---|---|---|---|---|
| | S GC R | R GC S | Bidirectional | No causality | S GC R | R GC S | Bidirectional | No causality |
| Happiness lower | **3** | 2 | 17 | 12 | **8** | 4 | 10 | 12 |
| Happiness upper | **4** | 1 | 24 | 5 | 3 | **7** | 17 | 7 |
| Sadness lower | **5** | 3 | 10 | 16 | 7 | **8** | 1 | 18 |
| Sadness upper | **5** | 3 | 2 | 24 | 4 | **5** | 6 | 19 |
| Average count | **4**.25 | 2.25 | 13.25 | 14.25 | 5.5 | **6** | 8.5 | 14 |

**TABLE 7. Number of pairs for which the Granger causality (GC) test, with p-value = 0.05, showed a specific direction of influence in the neutral/objective condition. Average count is the average count of pairs across all expressions listed in the table. Dominant causal direction (the higher value between Sender GC Receiver (S GC R) and Reciever GC Sender (R GC S)) is shown in bold font.**

| Expression | Full Time Span | | | | Relevant Interval Selection | | | |
|---|---|---|---|---|---|---|---|---|
| | S GC R | R GC S | Bidirectional | No causality | S GC R | R GC S | Bidirectional | No causality |
| Happiness lower | 0 | **5** | 23 | 4 | **9** | 7 | 7 | 9 |
| Happiness upper | **5** | 0 | 24 | 5 | **9** | 4 | 21 | 0 |
| Sadness lower | 2 | **3** | 7 | 22 | **6** | 7 | 9 | 12 |
| Sadness upper | **4** | 3 | 3 | 24 | **10** | 5 | 2 | 17 |
| Average count | 2.75 | 2.75 | 14.25 | 13.75 | **8**.5 | 5.75 | 9.75 | 9.5 |

**TABLE 8. Comparision of experimental findings and psychological hypotheses.**

| Psychological Hypotheses | Experimental Findings |
|---|---|
| **H.1.** More harmonic expressions (i.e., happiness) present when both interaction partners are confronted with medium to high levels of respect (i.e., respectful and objective/neutral vs. contemptuous) | Wilcoxon signed-rank test (Table 4) showed higher activation of *happiness upper* and *happiness lower* in the respectful condition than in the contempt and objective condition. |
| **H.2.** Stronger activation of negative expressions presents in the disrespectful condition (i.e., contemptuous vs. respectful and objective/neutral) | Wilcoxon signed-rank test (Table 4) showed higher activation of *sadness lower* and *sadness upper* expressions in the contempt compared to the objective condition |
| **H.3.** For all emotional facial actions, S causes the effects and influences R. | GC test on the selected relevant intervals showed that S influences R particularly in the respectful and objective/neutral (Table 5 and 7) compared to the contemptuous condition. In the contemptuous condition (Table 6), the pattern differs as R and S on average influence each other similarly. |
| **H.4.** The strongest GC causality from S to R occurs for positive expressions, followed by negative expressions | GC test on the selected relevant intervals showed a slight reduction in the emotional influence from S to R for negative compared to positive expressions (this is not the case for the full-time span approach). However, across all experimental conditions, results (Tables 5-7) showed significantly higher bidirectional influences for positive compared to negative emotions suggesting that positive facial expressions are in general more "infectious" than negative ones. |

it indicates that positive facial expressions are in general much more "infectious" than negative ones. Interestingly this seems independent of the role a person takes or the attitude/atmosphere of an interaction situation.

The main results of this study in comparison with the psychological hypotheses (given in III) are summarised in Table 8.

## VI. CONCLUSION

In this paper, we have presented a complete concept for identifying the direction of emotional influence in nonverbal dyadic communication when starting with raw video materials using facial expressions only. To this end, we presented an algorithm for the extraction of emotional facial features, capable of capturing emotional expressions even when strong distinct emotions are not visible. To improve causal inference, we proposed an intelligent interval selection approach for filtering relevant information in dyadic dialogues. Subsequently, we were able to apply Granger causality to the set of selected relevant intervals and compute the direction of influence. We applied our approach to real data obtained from a psychological experimental setup. The obtained results revealed that the use of the relevant interval selection approach when combined with the proposed facial features improved the detection of the direction of emotional influence for dyadic communication in various instructed interaction conditions. Such an approach can be used in different applications, i.e., climate science, to improve causal inference in time-series with transient changes. Based on quantitative verification of hypotheses on the direction of emotional influence, we were able to show that the proposed approach is promising to reveal the cause-effect pattern in various instructed interaction conditions. This work also allowed a major step forward in the 2nd person social sciences as we were able to study social emotions in a truly interactive manner. Further, the experimental findings of this study indicate that the use of facial expressions only has the potential to support implicit attitude and behavior research. Overall, we identify our contribution as an important step towards interdisciplinary research that combines the potential of computer vision, psychological experiments, and theoretical knowledge of causality methods, to gain novel insights into emotions in real-time social encounters. For further research, it would be interesting to tear apart the contributions of various social information features (i.e., from speech, non-verbal speech, facial expressions) towards outcomes of the interaction experience (e.g., enjoyment, engagement, and liking of interaction partner).

## REFERENCES

[1] W. Tschacher, G. M. Rees, and F. Ramseyer, "Nonverbal synchrony and affect in dyadic interactions," *Frontiers Psychol.*, vol. 5, p. 1323, Nov. 2014.

[2] B. Renoust, D.-D. Le, and S. Satoh, "Visual analytics of political networks from face-tracking of news video," *IEEE Trans. Multimedia*, vol. 18, no. 11, pp. 2184–2195, Nov. 2016.

[3] C. Reeck, D. R. Ames, and K. N. Ochsner, "The social regulation of emotion: An integrative, cross-disciplinary model," *Trends Cognit. Sci.*, vol. 20, no. 1, pp. 47–63, Jan. 2016.

[4] N. Ambady, F. Bernieri, and J. Richeson, "Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream," *Adv. Exp. Social Psychol.*, vol. 32, pp. 201–271, Dec. 2000.

[5] A. Todorov, "Inferences of competence from faces predict election outcomes," *Science*, vol. 308, no. 5728, pp. 1623–1626, Jun. 2005.

[6] S. T. Fiske, A. J. C. Cuddy, and P. Glick, "Universal dimensions of social cognition: Warmth and competence," *Trends Cognit. Sci.*, vol. 11, no. 2, pp. 77–83, Feb. 2007.

[7] V. Gallese, "Embodied simulation: From neurons to phenomenal experience," *Phenomenol. Cognit. Sci.*, vol. 4, no. 1, pp. 23–48, Mar. 2005.

[8] T. Fuchs and H. De Jaegher, "Enactive intersubjectivity: Participatory sense-making and mutual incorporation," *Phenomenol. Cognit. Sci.*, vol. 8, no. 4, pp. 465–486, Dec. 2009.

[9] P. Ekman and W. V. Friesen, "Measuring facial movement," *Environ. Psychol. Nonverbal Behav.*, vol. 1, no. 1, pp. 56–75, 1976.

[10] P. Filippi, S. Ocklenburg, D. L. Bowling, L. Heege, O. Güntürkün, A. Newen, and B. de Boer, "More than words (and faces): Evidence for a stroop effect of prosody in emotion word processing," *Cognition Emotion*, vol. 31, no. 5, pp. 879–891, Jul. 2017.

[11] D. A. Sauter, "The nonverbal communication of positive emotions: An emotion family approach," *Emotion Rev.*, vol. 9, no. 3, pp. 222–234, Jul. 2017.

[12] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System: The Manual on CD ROM*. Salt Lake City, UT, USA: A Human Face, 2002.

[13] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *J. Amer. Stat. Assoc.*, vol. 77, no. 378, pp. 304–313, Jun. 1982.

[14] C. W. J. Granger, "Testing for causality: A personal viewpoint," *J. Econ. Dyn. Control*, vol. 2, no. 1, pp. 329–352, 1980.

[15] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference-Foundations and Learning Algorithms* (Adaptive Computation and Machine Learning Series). The Cambridge, MA, USA: MIT Press, 2017.

[16] A. Attanasio, A. Pasini, and U. Triacca, "Granger causality analyses for climatic attribution," *Atmos. Climate Sci.*, vol. 03, no. 04, pp. 515–522, 2013.

[17] A. K. Seth, A. B. Barrett, and L. Barnett, "Granger causality analysis in neuroscience and neuroimaging," *J. Neurosci.*, vol. 35, no. 8, pp. 3293–3297, Feb. 2015.

[18] M. Shadaydeh, J. Denzler, Y. G. Garcia, and M. Mahecha, "Time-frequency causal inference uncovers anomalous events in environmental systems," in *Proc. German Conf. Pattern Recognit.* (Lecture Notes in Computer Science), vol. 11824. Cham, Switzerland: Springer, 2019, pp. 499–512.

[19] P. Ekman, "An argument for basic emotions," *Cognition Emotion*, vol. 6, nos. 3–4, pp. 169–200, May 1992.

[20] L. Müller, M. Shadaydeh, M. Thümmel, T. Kessler, D. Schneider, and J. Denzler, "Causal inference in nonverbal dyadic communication with relevant interval selection and granger causality," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 490–497.

[21] L. Schilbach, B. Timmermans, V. Reddy, A. Costall, G. Bente, T. Schlicht, and K. Vogeley, "Toward a second-person neuroscience," *Behav. Brain Sci.*, vol. 36, no. 4, pp. 393–414, Aug. 2013.

[22] F. Happé, J. L. Cook, and G. Bird, "The structure of social cognition: In(ter)dependence of sociocognitive processes," *Annu. Rev. Psychol.*, vol. 68, no. 1, pp. 243–267, Jan. 2017.

[23] K. Kalimeri, B. Lepri, O. Aran, D. B. Jayagopi, D. Gatica-Perez, and F. Pianesi, "Modeling dominance effects on nonverbal behaviors using granger causality," in *Proc. 14th ACM Int. Conf. Multimodal Interact. (ICMI)*, 2012, pp. 23–26.

[24] B. Kisačanin, V. Pavlović, and T. S. Huang, Eds., *Real-Time Vision for Human-Computer Interaction*. Boston, MA, USA: Springer, doi: 10.1007/0-387-27890-7_11.

[25] K. Bousmalis, M. Mehu, and M. Pantic, "Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools," *Image Vis. Comput.*, vol. 31, no. 2, pp. 203–221, Feb. 2013.

[26] E. O. Postma and M. Postma, "Measuring the causal dynamics of facial interaction with convergent cross mapping," in *Proc. CogSci*, 2016, pp. 1–6.

[27] T. Sheerman-Chase, E.-J. Ong, and R. Bowden, "Feature selection of facial displays for detection of non verbal communication in natural conversation," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Sep. 2009, pp. 1985–1992.

[28] Y. Matsuyama, A. Bhardwaj, R. Zhao, O. Romeo, S. Akoju, and J. Cassell, "Socially-aware animated intelligent personal assistant agent," in *Proc. 17th Annu. Meeting Special Interest Group Discourse Dialogue*, 2016, pp. 224–227.

[29] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 59–66.

[30] H. Joo, T. Simon, M. Cikara, and Y. Sheikh, "Towards social artificial intelligence: Nonverbal social signal prediction in a triadic interaction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10873–10883.

[31] K. Prabhakar, S. Oh, P. Wang, G. D. Abowd, and J. M. Rehg, "Temporal causality for the analysis of visual events," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1967–1974.

[32] K. Prabhakar and J. M. Rehg, "Categorizing turn-taking interactions," in *Computer Vision–(ECCV)*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Germany: Springer, 2012, pp. 383–396.

[33] M. Ayazoglu, B. Yilmaz, M. Sznaier, and O. Camps, "Finding causal interactions in video sequences," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3575–3582.

[34] P. Ekman and E. L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. New York, NY, USA: Oxford Univ. Press, 1997.

[35] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[36] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition Emotion*, vol. 24, no. 8, pp. 1377–1388, Dec. 2010.

[37] M. Wegrzyn, M. Vogt, B. Kireclioglu, J. Schneider, and J. Kissler, "Mapping the emotional face. How individual face parts contribute to successful emotion recognition," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0177239.

[38] J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the facial action coding system," in *The Handbook of Emotion Elicitation and Assessment*, vol. 1, no. 3. London, U.K.: Oxford Univ. Press, pp. 203–221.

[39] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," 2017, *arXiv:1708.03985*. [Online]. Available: http://arxiv.org/abs/1708.03985

[40] T. Baltrusaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–6.

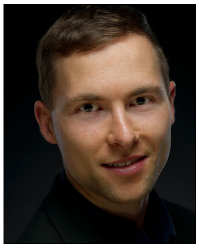
[41] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[42] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.
[43] S. Haykin, *Adapt. Filter Theory*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
[44] P. M. Schulze, "Granger-kausalitätsprüfung: Eine anwendungsorientierte darstellung," Institut für Statistik und Ökonometrie, Johannes-Gutenberg-Universität Mainz, Mainz, Germany, 2004.
[45] G. Atluri, M. Steinbach, K. O. Lim, A. M. Iii, and V. Kumar, "Discovering groups of time series with similar behavior in multiple small intervals of time," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2014, pp. 1001–1009.
[46] A. Jain, R. Bansal, A. Kumar, and K. Singh, "A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students," *Int. J. Appl. Basic Med. Res.*, vol. 5, no. 2, p. 124, 2015.
[47] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 57, no. 1, pp. 289–300, Jan. 1995.


**MAHA SHADAYDEH** received the Ph.D. degree in electrical communications engineering from Tohoku University, Sendai, Japan, in 1999. She held the position of an Associate Professor of Signal Processing at the Electronic Engineering Department, Damascus University, Syria, from 2000 to 2011. She was then a Researcher at the Machine Perception Research Laboratory, Institute for Computer Science and Control, Hungarian Academy of Sciences, from 2012 to 2016. Since 2017, she has been working as a Researcher and a Team Leader with the Computer Vision Group, Department of Mathematics and Computer Science, Friedrich Schiller University Jena, Germany. Her research interests include dynamical system analysis, remote sensing and biomedical signal processing, event detection, and causal reasoning.

**LEA MÜLLER** received the B.Sc. degree in mathematics, with a psychology as an applied subject, from the University of Heidelberg, Germany, and the master's degree (Hons.) in computational and data science from the University of Jena, Germany, in 2018. She is currently pursuing the Ph.D. degree with the Perceiving Systems Department, Max Planck Institute for Intelligent Systems. Her research interests include 3D human pose and shape estimation and the intersection of computer vision and psychology.

**DANA SCHNEIDER** was born in Magdeburg, Germany, in 1982. She studied psychology at Freie Universität Berlin, Germany, and The University of Queensland, Brisbane, QLD, Australia. She received the B.A. (Hons.) and Ph.D. degrees in psychology from The University of Queensland, in 2009 and 2013, respectively. Since 2013, she has been working a Researcher, a Lecturer, and a Group Leader with Friedrich Schiller University Jena, Germany. Her research interests include cognitive, affective, social, and behavioral aspects of human interaction and communication. She is currently an Associate Editor of *Frontiers in Psychiatry: Social Cognition* and the Leading Action Editor of the Special Issue "Understanding Others in Moments of Crisis" in social psychology. She is also the Speaker of the DFG Scientific Network "Understanding Others," SCHN 1481/2-1.



**MARTIN THÜMMEL** received the B.Sc. degree in electrical engineering from the University of Applied Sciences Jena, Germany, in 2014, and the M.Sc. degree (Hons.) in computer science from Friedrich Schiller University Jena, in 2017. From 2017 to 2020, he was a Research Associate with the Computer Vision Group, University of Jena. His research interests include facial expression and behavior analysis, 4D face presentation attack detection, and 3D computer vision.

**THOMAS KESSLER** studied psychology, in 1991. He received the Ph.D. degree from University Osnabrück, Germany, in 1997. He habilitated at the University of Jena, Germany, in 2004. He was a Professor of Social Psychology with the University of Exeter, U.K., from 2007 to 2010. Since 2010, he has been a Professor of Social Psychology with Friedrich-Schiller University, Jena, Germany. His research addresses determinants of tolerance and conflict within and between social groups. This includes research on prejudice and social discrimination, in-group favoritism, and the explicit negative treatment of outgroups or their members. His research interests include motivational and emotional processes reflecting and determining social change as well as evolutionary approaches, including authoritarian, group formation, and social identity processes.

**JOACHIM DENZLER** (Member, IEEE) received the Diplom-Informatiker, Dr.-Ing., and Habilitation degrees from the University of Erlangen, Germany, in 1992, 1997, and 2003, respectively. He currently holds a position as a Full Professor of Computer Science and the Head of the Computer Vision Group, Friedrich Schiller University Jena, Germany. He is also the Director of the DLR Institute of Data Science, Jena. He is the author and coauthor of over 300 journal articles and conference papers as well as technical articles. His research interests include automatic analysis, fusion, and understanding of sensor data, especially the development of methods for visual recognition tasks and dynamic scene analysis. He contributed in the area of active vision, 3D reconstruction, and object recognition and tracking. He is also a member of IEEE Computer Society, DAGM, and GI.