Article

Critical Examination of the Parametric Approaches to Analysis of the Non-Verbal Human Behavior: A Case Study in Facial Pre-Touch Interaction

Soheil Keshmiri ^{1,*} , Masahiro Shiomi ¹, Kodai Shatani ², Takashi Minato ¹
and Hiroshi Ishiguro ^{1,2}

¹ Advanced Telecommunications Research Institute International (ATR), Kyoto 619-0237, Japan; m-shiomi@atr.jp (M.S.); minato@atr.jp (T.M.); ishiguro@sys.es.osaka-u.ac.jp (H.I.)

² Graduate School of Engineering Science, Osaka University, Osaka 565-0871, Japan; shatani.kodai@irl.sys.es.osaka-u.ac.jp

* Correspondence: soheil@atr.jp

Received: 21 April 2020; Accepted: 28 May 2020; Published: 30 May 2020

Abstract: A prevailing assumption in many behavioral studies is the underlying normal distribution of the data under investigation. In this regard, although it appears plausible to presume a certain degree of similarity among individuals, this presumption does not necessarily warrant such simplifying assumptions as average or normally distributed human behavioral responses. In the present study, we examine the extent of such assumptions by considering the case of human–human touch interaction in which individuals signal their face area pre-touch distance boundaries. We then use these pre-touch distances along with their respective azimuth and elevation angles around the face area and perform three types of regression-based analyses to estimate a generalized facial pre-touch distance boundary. First, we use a Gaussian processes regression to evaluate whether assumption of normal distribution in participants’ reactions warrants a reliable estimate of this boundary. Second, we apply a support vector regression (SVR) to determine whether estimating this space by minimizing the orthogonal distance between participants’ pre-touch data and its corresponding pre-touch boundary can yield a better result. Third, we use ordinary regression to validate the utility of a non-parametric regressor with a simple regularization criterion in estimating such a pre-touch space. In addition, we compare these models with the scenarios in which a fixed boundary distance (i.e., a spherical boundary) is adopted. We show that within the context of facial pre-touch interaction, normal distribution does not capture the variability that is exhibited by human subjects during such non-verbal interaction. We also provide evidence that such interactions can be more adequately estimated by considering the individuals’ variable behavior and preferences through such estimation strategies as ordinary regression that solely relies on the distribution of their observed behavior which may not necessarily follow a parametric distribution.

Keywords: physical interaction; physical pre-touch distance; parametric analysis; non-parametric analysis; non-verbal behavior

1. Introduction

Social and cognitive psychology aims to unlock the secret of emotional [1] and personality [2,3] traits that humans adopt while navigating the complex world of their social relations. Research suggests that the signatures of these internal states are etched in our brain [4–6] and that such states can determine individuals’ responses to psychological stressors [7]. These findings also point at the correspondence [8] between these internal states and their more immediate manifestations

in the form of personal space [9] and interpersonal distance [10–12]. The significance of personal space is well-documented in the findings that emphasize its positive socioemotional effects on our wellbeing [9,10,13–16] and that associate its invasion to a stressful situation [17].

One important aspect of these studies is to identify the universal dimensions along which human behavior is expressed. In this regards, Goldberg [3] asserts that “the variety of individual differences... are insignificant in people’s daily interactions”; a viewpoint that echoes that of Galton [18]. On the other hand, although it appears plausible to presume a certain degree of similarity among individuals, this presumption does not necessarily warrant such simplifying assumptions as average or normally distributed [19–21] behavioral response. For instance, Micceri [22] analyzed the distributional characteristics of over 440 samples of achievement and psychometric measures. His results revealed that several classes exhibit deviations from the normal distribution. Similarly, analysis of the 693 psychological data by Blanca et al. [23] identified that merely 5.50% of the distributions were close to expected values under normality while 74.40% exhibited slight to moderate and 20.00% more extreme deviations. These findings indicated that the differences in individuals’ psycho-behavioural and personality traits are associated with types of variability whose trend may not readily and necessarily follow a normal distribution.

In the same vein, our recent study [24] also found that such non-normal characteristics in individuals’ personality was also present in their more observable behaviour. Specifically, we observed that these individuals’ facial pre-touch interaction did not follow a specific distribution pattern. Our results further identified that the non-parametric cluster analysis of their facial pre-touch interaction resulted in formation of well-defined facial pre-touch subspaces that significantly correlated with the participants’ openness in the five-factor model (FFM) [3]. Additionally, they indicated that such a non-parametric modeling was able to predict individuals’ openness with a significantly above chance level accuracy.

Considering the above findings in the literature, the present study set to examine the justifiability of such assumptions as average or normally distributed behavioural response in the context of estimating a generalized distance boundary in the case of human–human facial pre-touch interaction. For this purpose, we use the data from our previous study [24] and perform three types of regression-based estimation of such a distance boundary on it. First, we use a Gaussian processes regression to evaluate whether assumption of normal distribution in participants’ pre-touch distances warrants a reliable estimate of this boundary. Second, we apply a support vector regression to determine whether estimating this space by minimizing the orthogonal distance between participants’ pre-touch data and its corresponding pre-touch boundary can yield a better result. Third, we use ordinary regression to validate the utility of a non-parametric regressor with a simple regularization criterion in estimating such a pre-touch space. In addition, we compare these models with the scenarios in which the average pre-touch distance or a fixed boundary distance (i.e., a spherical boundary) is adopted. It is worthy of note that we consider the facial-area touch interaction because people tend to be more sensitive about touching their faces, compared to the other body parts that they may more openly share during their social interaction (e.g., handshakes or a pat on the shoulder). This makes the facial boundary to be a good candidate for understanding the people’s behavioural responses within the context of touch interaction.

Our results indicate that such assumptions as averaged/fixed boundary thresholding or imposing an unjustified distribution on data do not warrant an acceptable estimate of the human subjects’ facial pre-touch behaviour. In a broader sense, they also suggest that the use of such simple mathematical models as ordinary regression to unveil rather than impose underlying similarities can yield more reliable results that reflect the observed human behaviour.

2. Materials and Methods

2.1. Participants

Forty younger adults ($M = 21.83$, $SD = 1.53$) participated in our experiment. These individuals were paired into four distinct toucher-evaluator groups (i.e., 10 pairs per group). The four groups were:

1. “female” touchers and “female” evaluators
2. “female” touchers and “male” evaluators
3. “male” touchers and “female” evaluators
4. “male” touchers and “male” evaluators

In the present study, all the participants were Japanese nationals who were strangers to each other. None of these participants had any (social) relation/connection with the other participants. All experiments were carried out with written informed consents from all subjects.

2.2. Paradigm

We acquired the facial pre-touch distances that were measured between the hand of a toucher and the face of a person who was about to be touched (evaluator). Figure 1a shows an instance of the experiment. The evaluator was seated on a chair in the middle of the experimental room and the toucher stood close to the evaluator in a distance that was adjusted based on the arm’s length of each of the touchers in our experiment. The nine approaching positions from which the toucher reached for the face area of the evaluator are shown in Figure 1b (positions 0 through 8).

Figure 1. Experimental Setup. (a) An instance of toucher-evaluator interaction. The two Kinect V2 sensors that were mounted behind the evaluators’ seat to automatically track the touchers’ hand and the evaluators’ face positions are visible in this figure. (b) Toucher-evaluator interaction settings. The toucher (i.e., T) moves along the positions 0 through 8 and stretches her hand toward the face of the evaluator (i.e., E) who is seated in the middle. The two Kinect V2 sensors mounted behind the evaluator collect the joint and the head positions of the toucher and the evaluator.

In our experimental setup, the touchers slowly stretched their hand toward the evaluators’ face. While doing so, they freely decided their initial hand position and their approaching angle. When the evaluators felt that the touchers’ hand were exceeding their comfort zone and wanted them to stop, they clicked a mouse bottom whose clicking sound was audible to the touchers. We instructed the touchers to immediately stop getting their hand any closer to the evaluators’ face once they heard the mouse clicking sound. We then measured the distance between the touchers’ hand and the evaluators’ face and used these measured distances as the minimum comfortable pre-touch distance of the individuals (i.e., their behavioural-based facial pre-touch boundary). We did not fix the number of pre-touch interactions and allowed the participants to continue as many times as their allocated

interaction time permitted. In total, we collected 13,456 facial-area pre-touch distance data points (Mean (M) individuals' sample = 292.5 and standard deviation (SD) = 81.10).

The total interaction time, per pair, was two hours. However, we did not allow free interaction (e.g., talking to each other, etc.) between them during the experiment session.

2.3. Data Acquisition

We used two Kinect V2 sensors that were mounted behind the evaluators' seat (Figure 1a) to automatically track the touchers' hand and the evaluators' face positions. We collected their data on 3D positions of each joint of the touchers (including the center of their hands) and the 3D head position of the evaluators. Simultaneously, we also recorded the timing of the mouse clicks by the evaluators that signaled the touchers to stop getting their hand any closer to evaluators' face. In order to calculate the minimum comfortable distance between the touchers' hand and the evaluators' face, we subtracted the size of the touchers' hand (measured prior to the commencement of the experiment) from the average Japanese face size (i.e., 9.0 cm for female and 10.0 cm for males) [25].

Giancola et al. [26] suggested that Kinect sensors are suitable for applications in which the joint position accuracy does not exceed a few cm. However, in their study, they focused on the accuracy of a whole body tracking algorithm in an upper-limb rehabilitation scenario. Our experiment differed from their setting in which we considered the interaction space between the touchers' hand and the evaluators' face. Therefore, we employed (unlike Giancola et al. [26]) two Kinect sensors for data acquisition, thereby bypassing the use of markers on touchers' hand and the evaluators' face to prevent their potential confounding effect on participants' pre-touch feelings. To increase the accuracy of the detected joint positions, we further calibrated the relative positions of these two sensors and used their absolute positions to integrate their joint positions data. In an event that one of these Kinect sensors failed to estimate the joint positions, we used the other sensor if its estimates were continuous and stable.

2.4. Analysis

Our earlier results [27] based on the present dataset indicated that such factors as gender and approaching angles did not induce any significant difference on the evaluators' comfortable distances and that although the acclimation effect showed a trend, its difference was less than 2.0 cm. Therefore, we combined the data from all four distinct male/female pairs in the present analyses.

First, we constructed (Figure 2a) the grid of the participants' pre-touch distance (in cm) based on their associated azimuth and elevation angles (both angles within $[-89.00, \dots, +89.00]$ in degree). Next, we modified this grid to eliminate the entire row that included a cell with no pre-touch distance (i.e., grey cells in Figure 2a). The resulting modified grid included 5664 facial-area pre-touch samples whose azimuth and elevations angles were $[-80.00, \dots, +80.00]$ (mean (M) = 0.54, median (Mdn) = 0.10, standard deviation (STD) = 38.99, 95.0% confidence interval (CI_{95.0%}) = $[-0.11, 1.20]$) and $[-32.00, \dots, +53.00]$ (M = 9.83, Mdn = 9.17, STD = 24.17, CI_{95.0%} = $[9.33, 10.31]$). We then used this modified grid to construct the pre-touch surface (Figure 2b). Considering the mean and median of the azimuth angles associated with these data points that are closely centered about zero, the participants' data around azimuth axis was uniformly distributed between the two sides of the face. In the case of elevation angles, we observed that the touchers' data were mostly approached the touchee's face above the lips area. Figure 2c visualizes the distribution of touchers' pre-touch distances around the touchees' face area (i.e., the location at which the touchers were asked by the touchees to stop approaching). The coordinates of these data points correspond to the toucher's hand distance from the touchee's face (x -axis, in cm), azimuth angle (y -axis, in degree), and elevation angle (z -axis, in degree) with respect to the touchees' face. Figure 2d shows the head direction.

Figure 2. Pre-touch distance surface. (a) Full participant's pre-touch distance (cm) grid along the azimuth and elevation angles. Azimuth and elevation angles are within -89.00 and $+89.00$ degrees. The color-bar on the side indicates the pre-touch distance value (in cm) in each cell of the azimuth-elevation grid whose cells are constructed by 1-degree increment along these angles' directions. Grey cells in this subplot show the areas in which no pre-touch distance was available. (b) 3-D heatmap of pre-touch surface around the face using the modified grid that includes 5664 facial-area pre-touch data points. The azimuth and elevations angles of these samples are within $[-80.00, \dots, +80.00]$ and $[-32.00, \dots, +53.00]$, respectively. The z-axis corresponds to the distance of the touchers' hand from the touchees' face (i.e., distance at which the touchers were asked by the touchees to stop approaching hand to get any closer). (c) Distribution of touchers' pre-touch data points around the touchees' face area. Points are plotted in different shades of grey for better visualization purpose. These data points correspond to the locations where the touchers were asked by the touchee to stop approaching their face. Coordinates of these data points correspond to toucher's hand distance (x-axis, in cm), azimuth angle (y-axis, in degree), and elevation angle (z-axis, in degree) with respect to the touchees' face. (d) Schematic diagram of the face's orientation with respect to the azimuth and elevation angles as shown in subplot (b).

We considered two pre-touch boundary estimation scenarios. In the first scenario that was based on our earlier results [27], we used a fixed pre-touch distance of 20.0 cm (i.e., a sphere of diameter 40.0 cm) around the face. In the second scenario, we used the data associated with participants' pre-touch surface (Figure 2b) and used regression models for estimating this boundary. We considered three regression models: (1) Gaussian processes regression (GR) to evaluate whether assumption of normal distribution in participants' pre-touch distances warrants a reliable estimate of this pre-touch boundary, (2) support vector regression (SVR) to determine whether estimating the pre-touch boundary by minimizing the orthogonal distance between participants' pre-touch data and its corresponding pre-touch boundary can yield a better result, and (3) ordinary regression with Lasso regularization (Lasso) to validate the utility of a non-parametric regressor with a simple criterion i.e., L1-regularization (i.e., sum of the model's absolute weight values) for estimating this pre-touch boundary. For all these models, we used the azimuth and elevation angles as input features and their corresponding pre-touch distances as output values that these models needed to estimate. The preprocessing of the models' input features included the scaling of the azimuth and elevation angles within $[0, \dots, 1]$ using $\frac{\alpha - \min(A)}{\max(A) - \min(A)}$ where A is the set of either azimuth or elevation angles and α identifies a specific angle that to be scaled.

We also examined whether the use of polynomial degree on input features (i.e., azimuth and elevation angles) benefited these models' estimate of participants' facial-area pre-touch boundary. The use of polynomial degree was a natural consideration since any such pre-touch distance boundary was analogous to approximation of a three-dimensional volume around the face (e.g., a perfect spherical space in the case of fixed pre-touch distance). We examined polynomial degrees 0 (i.e., no polynomial degree or control condition) through 10. All models attained their best estimates using polynomial degree 4. Therefore, we used 0 (i.e., control setting in the case of the models) and 4 (i.e., feature vectors of length $2^4 = 16$, given the azimuth and elevation angles as original input features) polynomial degrees in our analyses. For each of these two cases, we adopted a cross-validation strategy in which we randomly split the pre-touch surface that included 5664 distance data (Figure 2b) into 80.0% train (4531 pre-touch data) and 20.0% test (i.e., 1133) sets. We repeated this procedure for 1000 simulation runs during which we recorded the root-mean-squared-error (RMSE) and coefficient of determination (R^2) of the models as measures of their goodness of estimates. The RMSE allowed us to determine how well each model was estimating the pre-touch distance values on test set. On the other hand, the R^2 enabled us to evaluate whether the use of averaged pre-touch distance in the train set might yield a better estimates of test set's pre-touch distances than the models' estimates based on their training. Precisely, in the case of regression, coefficient of determination (i.e., $R^2 = 1 - \frac{SS_{residual}}{SS_{total}}$ where $SS_{residual}$ and SS_{total} refer to residual and explained sum of squares, respectively) measures how well the regression predictions approximate the real data points [28,29]. As a result, $R^2 < 0.0$ would mean that the use of the averaged pre-touch distances as their estimated pre-touch boundary was better than the model estimates, $R^2 = 0.0$ would indicate no difference between the two, and $R^2 > 0.0$ would justify the use of a model for estimating this boundary.

For the analysis, we performed 1000 rounds of random sampling in which we selected (without replacement) 100 of these RMSE and R^2 values for our statistical tests of significance. At each round, we first applied pairwise Wilcoxon rank sum between each pair of polynomial degrees (i.e., 0 versus 4) for each model. Next, we applied Wilcoxon rank sum test between models and the fixed pre-touch distance (i.e., 20.0 cm). In the case of models, we considered their pairs of equal degrees (e.g., GP vs. SVR using degree 0). In the case of fixed pre-touch distance, R^2 indicated whether the use of averaged pre-touch distance might yield a better estimate than such a fixed spherical boundary. We also applied Wilcoxon signed rank (i.e., one-sample test) on R^2 values of these models (for both polynomial degrees 0 and 4) to determine how significantly better or worse (in the case of $R^2 < 0$) than the use of averaged pre-touch distance they performed. For each round, we also calculated the effect size $r = \frac{W}{\sqrt{N}}$ [30] with N denoting the sample size ($N = 100$, i.e., RMSE and R^2 associated with each of the 1000 random sampling without replacement rounds) and W denoting the Wilcoxon test-statistics. We Bonferroni-corrected all p -values by multiplying them with N , given the use of non-parametric tests. We reported the average of these 1000 simulation rounds.

To test whether the differences between these models can also be verified in their respective learning process, we further used the full facial pre-touch distance data and trained these models. We then used their estimated pre-touch distances that was based on the full dataset and computed the difference between these estimates (per model) and the actual pre-touch distances and applied Friedman's test on these differences to test for any significant difference. For Friedman's test, we reported the effect size $r = \sqrt{\frac{\chi^2}{N}}$ [31] with N denoting the sample size and χ^2 is the Friedman's test-statistics. We also computed the RMSE and R^2 values of these models on full data.

We used Python 2.7 for models' training and testing. All analyses were carried out in MATLAB 2016a. We used MATLAB and the Python version of Raincloud plots [32] for data visualization. We applied Bonferroni-correction to all our tests. Specifically, our analyses were between four settings: "GP", "SVR", "Lasso", and "Fixed" pre-touch distance boundary. As a result, the Bonferroni-corrected p -value for significance level $p = 0.05$ was $\frac{0.05}{4} = 0.0125$.

2.5. Ethics Statement

This study was carried out in accordance with the recommendations of the ethical committee of the Advanced Telecommunications Research Institute International (ATR) with written informed consent from all subjects in accordance with the Declaration of Helsinki. The protocol was approved by the ATR ethical committee (approval code: 17-601-4).

3. Results

3.1. Rmse

The use of polynomial degree resulted in significantly reduced RMSEs in the case of GP, SVR, and Lasso (Figure 3 horizontal within-subplots asterisks and Table 1). We also observed that (Figure 3 vertical between-subplots asterisks and Table 2) both Lasso and SVR significantly outperformed GP. Their RMSEs were also significantly lower than the fixed pre-touch distance (i.e., fixed 20.0 cm facial pre-touch distance). Although the difference between Lasso and SVR RMSEs was non-significant when these models did not use a polynomial degree (Figure 3a), RMSEs in the case of Lasso were significantly lower than those of SVR when the polynomial degree 4 was used (Figure 3b). Last, RMSEs associated with GP were also significantly lower than the case of fixed pre-touch distance setting.

Figure 3. Root-mean-squared-error (RMSE) when the models trained using (a) polynomial degree 0 (b) polynomial degree 4. The horizontal bars correspond to the difference between the polynomial degrees for each model and the vertical bars are associated with the difference between models in each polynomial degree settings. The asterisks indicate the significance in their differences (***: $p < 0.001$).

Table 1. RMSE values using Lasso, SVR, and GP. Entries M_i and SD_i , $i = 0, 4$ refer to the mean and the standard deviation of RMSEs associated with these models while using polynomial degrees 0 (i.e., no polynomial degree) and 4.

Models	$p <$	W(26)	r	M_0	SD_0	M_4	SD_4
Lasso	0.001	36.57	1.16	8.19	0.85	5.79	1.60
SVR	0.001	12.19	0.39	8.07	0.84	7.69	0.85
GP	0.001	19.64	0.62	16.79	3.64	14.27	2.70

Table 2. Pairwise models' RMSE comparison. Subscripts 0 and 4 correspond to the case of polynomial degree 0 (i.e., no polynomial degree) and 4.

Models	$p_0 <$	$W_0(26)$	r_0	$p_4 <$	$W_4(26)$	r_4
Lasso vs. GP	0.001	12.22	0.86	0.001	12.21	0.86
Lasso vs. SVR	=0.31	1.03	0.07	0.001	8.00	0.57
Lasso vs. Fixed	0.001	14.11	1.00	0.001	14.11	0.86
SVR vs. GP	0.001	12.22	0.86	0.001	12.20	0.86
SVR vs. Fixed	0.001	14.11	1.00	0.001	14.11	1.00
GP vs. Fixed	0.001	14.11	1.00	0.001	14.11	1.00

3.2. Coefficient of Determination R^2

The use of polynomial degree resulted in significant increase R^2 values in the case of GP, SVR, and Lasso (Figure 4 horizontal within-subplots asterisks and Table 3). We also observed that (Figure 4 vertical between-subplots asterisks and Table 4) the R^2 values associated with both Lasso and SVR were significantly higher than those of GP and the fixed pre-touch distance (i.e., fixed 20.0 cm facial pre-touch distance).

Figure 4a subplot shows that without the use of polynomial degree (i.e., degree 0) all these models performed poorly (i.e., lower $R^2 < 0.0$) compared to the fixed pre-touch distance (i.e., fixed 20.0 cm facial pre-touch distance). On the other hand (Table 4), the use of polynomial degree 4 (Figure 4b) resulted in Lasso and SVR significantly increased R^2 values. However, we did not observe this increase in R^2 in the case of GP.

Finally, we observed that (Table 5) whereas GP performed significantly worse than fixed pre-touch distance in the case of both polynomial degrees 0 ($p < 0.001$, $W = 14.11$, $r = 1.00$, $M_{Fixed} = -0.94$, $SD_{Fixed} = 0.21$) and 4 ($p < 0.001$, $W = 14.11$, $r = 1.0$), fixed distance was significantly outperformed by SVR (0: $p < 0.001$, $W = 14.11$, $r = 1.00$ and 4: $p < 0.001$, $W = 14.11$, $r = 1.0$) and Lasso (0: $p < 0.001$, $W = 14.11$, $r = 1.00$ and 4: $p < 0.001$, $W = 14.11$, $r = 1.00$).

Figure 4. Coefficient of Determination (R^2) when the models trained using (a) polynomial degree 0 (b) polynomial degree 4. The horizontal bars correspond to the difference between the polynomial degrees for each model and the vertical bars are associated with the difference between models in each polynomial degree settings. The asterisks indicate the significance in their differences (***: $p < 0.001$).

Table 3. Effect of polynomial degree on improving the coefficient of determination (R^2) in Lasso, SVR, and GP. M_i and SD_i , $i = 0, 4$ refer to the mean and the standard deviation of RMSEs associated with these models while using polynomial degrees 0 (i.e., no polynomial degree) and 4.

Models	$p <$	W(26)	r	M_0	SD_0	M_4	SD_4
Lasso	0.001	36.74	1.16	−0.06	0.08	0.44	0.30
SVR	0.001	29.08	0.92	−0.03	0.09	0.06	0.11
GP	0.001	10.29	0.33	−2.47	0.93	−2.05	0.84

Table 4. Coefficient of determination (R^2). Comparison of the use of average with Lasso, SVR, and GP with and without polynomial degrees. Subscripts 0 and 4 correspond to the case of polynomial degree 0 (i.e., no polynomial degree) and 4.

Models	$p <$	W(26)	r	$p <$	W(26)	r
Lasso	0.001	27.39	0.87	0.001	47.84	1.51
SVR	0.001	4.51	0.14	0.001	37.27	1.18
GP	0.001	22.90	0.72	0.001	52.62	1.66

We also identified that (Table 5) the use of Lasso resulted in significantly higher R^2 values than GP (0: $p < 0.001$, $W = 12.22$, $r = 0.86$ and 4: $p < 0.001$, $W = 12.21$, $r = 0.86$) and SVR (0: $p < 0.001$, $W = 4.95$, $r = 0.35$ and 4: $p < 0.001$, $W = 6.84$, $r = 0.48$). SVR also performed significantly better than GP in the case of both polynomial degrees 0 ($p < 0.001$, $W = 12.22$, $r = 0.86$) and 4 ($p < 0.001$, $W = 12.20$, $r = 0.86$).

Table 5. Pairwise model's R^2 comparison. Subscripts 0 and 4 correspond to the case of polynomial degree 0 (i.e., no polynomial degree) and 4.

Models	$p_0 <$	$W_0(26)$	r_0	$p_4 <$	$W_4(26)$	r_4
Lasso vs. GP	0.001	12.22	0.86	0.001	12.21	0.86
Lasso vs. SVR	0.001	4.95	0.35	0.001	6.84	0.48
Lasso vs. Fixed	0.001	14.11	1.00	0.001	14.11	1.00
SVR vs. GP	0.001	12.22	0.86	0.001	12.20	0.86
SVR vs. Fixed	0.001	14.11	1.00	0.001	14.11	1.00
GP vs. Fixed	0.001	14.11	1.00	0.001	14.11	1.00

3.3. Investigation of the Training Process

Considering these models' respective training phases, we noticed that GP perfectly fitted (Figure 5a) its training data ($RMSE = 0.00$, $R^2 = 1.00$), thereby yielding an overfitted model. On the other hand, we observed that SVR learning process overtly discarded (Figure 5b) the differences between these facial pre-touch distances ($RMSE = 7.69$, $R^2 = 0.10$). The SVR's performance can be explained in light of its optimization criterion that attempts to minimize the orthogonal distances between data and its fitted hyper/plane. Considering the distribution of facial-area pre-touch distances around the face area (i.e., an approximately hemispheric surface), it is foreseeable for SVR's optimization criterion to find one such candidate hemisphere (i.e., analogous to a plane in 3D that is bent along the azimuth angle). On the other hand, Lasso (Figure 5c) whose only criterion was to penalize the growth of its learned weight matrix appeared to achieve a balance between bias and variance ($RMSE = 4.90$, $R^2 = 0.64$). This suggestion is clarified by comparing its $RMSE$ and R^2 in the case of full data and its average performance on test cases in Sections 3.1 and 3.2 ($M_{RMSE} = 5.79$ and $M_{R^2} = 0.44$).

Figure 5. Subplots (a) through (c): estimated 2-D heatmap surfaces by models trained on the entire facial-area pre-touch distances. In each of these subplots, the z-axis corresponds to the facial pre-touch distances that were estimated by these models. The x- and y-axes correspond to the elevation azimuth and elevation angles associated with these distances (i.e., the feature inputs) to these models. (a) GP appears to perfectly fit the data during its learning, thereby resulting in an overfitted model. (b) SVR discards the differences between pre-touch distances and instead finds the hyper/plane that best suites its optimization criterion i.e., minimization of the orthogonal distances between pre-touch data and the candidate boundary. It is apparent that in 3D space of the facial pre-touch distances a hemisphere is the best candidate. (c) Lasso's learning process appears to better balance the balanced bias-variance tradeoff while satisfying its criterion that is to keep the entries of its learned weight matrix small (i.e., potentially as coarse-grained as possible), thereby allowing for a better generalization to novel cases. The middle row gives the top-view of the surfaces that were estimated by each of these models.

Although we observed a high variability in under/overfitting the pre-touch distances by these models, it was interesting to observe that the Friedman's test (Figure 6) identified a non-significant difference between these models based on the difference between their estimations and the actual pre-touch distances ($p = 0.95$, $H(2, 16,991) = 0.10$, $r = 0.02$). However, this failure is clarified once one observes that the difference between the estimations and actual pre-touch distances by GP and Lasso yielded zero-mean differences ($M_{GP} = 0.00$, $SD_{GP} = 0.00$, $M_{Lasso} = 0.00$, $SD_{Lasso} = 4.92$) with a negligibly above zero in the case of SVR ($M_{SVR} = 0.11$, $SD_{SVR} = 7.72$).

Figure 6. Distribution of estimated pre-touch distances by each of these models. This subplot verifies that GP perfectly matched (i.e., overfitted) the data and that SVR's estimate resulted in a uniform boundary in the form of a hemisphere around the face area.

4. Discussion

In this article, we examined the use of fixed or normal distribution for modeling individuals' behavioural responses. In so doing, we contrasted these assumptions to the case that primarily relied on data to uncover its underlying properties as a more tenable choice. Our study was motivated

by the previous findings that identified the shortcoming of such assumptions [19–23] and differed from the viewpoints that advocate the insignificant differences among the individuals' behavioural responses [3,18].

We used individuals' facial pre-touch distance and considered the problem of estimating a generalized facial pre-touch boundary based on it. We considered three strategies for approximating such a generalized facial pre-touch distance: (1) averaged pre-touch distance in which we considered a pre-touch sphere whose radius was equal to the average of all individuals' pre-touch distances; (2) fixed distance in which we considered a fixed (i.e., 20.0 cm) radius for this sphere, and (3) application of estimators to approximate this surface. For the third case, we further used three different estimators: (1) Gaussian processes regression to verify whether assumption of normality among individuals' pre-touch distances was warranted for estimation of such a facial pre-touch surface; (2) SVR to determine whether such criteria as minimization of the orthogonal distances of these pre-touch distances from their estimated common surface could yield a better result, and (3) a simple linear regression model that regularized its weights (i.e., coefficients of the linear model) while learning to estimate this pre-touch surface.

Our results indicated that estimation of the individuals' facial pre-touch space through modeling of such data yielded a significantly better results than opting for such strategies as averaged pre-touch distance or fixed facial boundary. Additionally, these results identified that among the adopted modeling approaches the one with minimal assumptions/criteria while exploring the data (i.e., Lasso regressor) resulted in the estimates that were significantly closer to the actual individuals' pre-touch distances. Further scrutiny of the learning process of these strategies revealed three interesting observations. First, we observed that GP suffered from overfitting during its learning process in which its trained model flawlessly fitted the constructed facial pre-touch surface while falling short on estimating novel data. An immediate implication of this observation was the danger of the reliance on the training estimation, thereby concluding on misleading results [33,34]. Second, we also observed that although SVR performed better than GP, its core optimization criterion (i.e., minimization of the orthogonal distances of the data points from their estimated facial pre-touch surface) led to overtly discarding the differences between these facial pre-touch distances. As a result, it approximated a surface that closely resembled a hemisphere around the face area. In fact, considering the distribution of facial-area pre-touch distances that roughly approximated a hemisphere in conjunction with the SVR's optimization criterion, it was reasonable for this model to find one such candidate hemisphere. Third, we observed that Lasso regressor whose only criterion was to penalize the growth of its learned weight matrix (and hence preventing the overfitting the data as was the case for GP) appeared to achieve a reasonable balance between bias and variance. This observation also indicated that allowing for certain degree of mistake enabled this model to better estimate the future novel/unseen data through higher variability that was accommodated by its learned parameters.

From the human behaviour perspective, we observed that (Figure 2b) the individuals' pre-touch distances were larger around the forehead and the nose areas and that the distances in these areas exhibited rather variable changes than smooth gradients. These observations indicated that these individuals (regardless of which gender they were interacting with) showed higher sensitivity toward a touch around these areas than other facial parts (e.g., their chin). Such a variability in individuals' behaviour that was reflected in their embodied interaction (even in such a limited area as facial touch interaction) further highlighted the importance of the use of a justifiable approach toward modeling of the human behaviour. For instance, if the individuals' preferred pre-touch distances were equally distant around the face (i.e., not affected by different facial areas) or if such distances were rather well-defined by people's blindsight (e.g., uniformly decreasing as we moved away from their field of view and toward the sides of their face) then SVR or GP might have been the better choices.

Our results benefit the human behaviour research by identifying the need for thorough examination of potential approaches prior to their application to ensure the reliability of their conclusions [22,23]. In this respect, an immediate implication of our findings is that the special

care must be given to the danger of interpreting the significance of observations based on estimations that are based on training than the test phase of modeling approaches [33,34]. Human–robot interaction (HRI) is another field of research that can benefit from our findings, considering the recent urge for more robust evaluations that are founded on theoretical than sheer empirical approaches [35] to enable these agents to better meet the grand social challenges [36] (p. 9) while interacting with individuals [37].

It is important to note our study is not meant to be construed as a hard and fast rule for preferring one of the adopted model over and above all the other alternatives but to provide demonstrable results on some of the issues that can arise while opting for unwarranted assumptions [8,19–23,38]. In fact, falling for such simplifications may well explain some of the discrepancies that are present in the field of human behaviour [8,38]. Our results primarily state that unless the underlying properties of the data is well understood (e.g., its distribution, etc.) the better strategy to make sense of such properties is to let the data reveal itself. Moreover, our results re-emphasize the danger of the interpreting the significance of observations based on estimations that are based on training than the test phase of modeling approaches, as it has been noted by previous studies [34].

5. Limitations and Future Direction

An important consideration with regards to the present study is the limited number of participants and the toucher-touchee pairs that they formed. On the one hand, although the number of participants was limited, we collected a moderately large sample (i.e., 5664 facial area pre-touch data points) from them. On the other hand, since these data were all collected from the same individuals, our data acquisition procedure could have been affected by two important factors: habituation and sensitization effects. For instance, in the context of present study the habituation could have taken place if the touchee felt more relaxed after a few trials (due to repeated facial pre-touch attempts by the same toucher) and subsequently allowed the toucher's hand to approach closer to the face area. Conversely, the sensitization effect could have occurred if the repeated attempts by the toucher resulted in higher sensitivity of the touchee, thereby asking the toucher to stop in ever quicker and therefore farther distances from the touchee's face. The occurrence of such effects could also become more possible, considering the limited number of toucher-touchee pairs in the present study whose members were fixed. Specifically, although the participants in our study were all strangers to one another, it would be foreseeable that the toucher-touchee pairs could feel more comfortable (or conversely more sensitive to the presence of a stranger) after the first few trial attempts. Therefore, it is crucial for the future research to further investigate the potential effect of such factors by including a greater number of participants. In this context, it is also interesting to study how swapping the individuals between the pairs may affect the responses of the touchees on the one hand and the confidence of the touchers for approaching new individuals' faces on the other hand.

With regards to the distribution of touchers' data around the face area of the touchees, we observed that the mean and median of their data were close to zero. This indicated that the touchers' data was quite fairly and smoothly distributed along the azimuth axis. On the other hand, we observed that their elevation angles were more elongated towards the above-lips area. Perhaps such a stretch of data towards the area above the lips is to an extent expected when we consider the fact that the touchees were seated and the touchers were standing in front of the touchees' seat. However, this observation also highlights a shortcoming in our setting in which the fixed height of the seat did not allow for the lower portion of the faces to be more conveniently exposed to touchers' reach. This indicated that our consideration for the position of the touchees' seat according to the touchers' arm length needed to be complemented with subsequent adjustment of the seat's height as well. Therefore, future research is necessary to analyze the potential effect of this shortcoming on the data collection and the performance of different regression models that were studied in the present study.

Another issue that requires further consideration is the physical characteristics of the touchers. In the present study, we did not collect information about the participants' height. We also did not acquire touchees' subjective impression of the touchers' appearances on their decision for allowing

or stopping the reaching hand. In this respect, it is worthy of note that we did ask the touchees to constantly look straight ahead (i.e., to avoid eye-contact and/or to look at/follow the approaching hand) to alleviate such potential effects as touchers' appearance. However, these measures would probably be insufficient to eliminate these effects. Therefore, the future research should consider collecting data about participants' subjective feelings (e.g., in the form of questionnaires), thereby allowing for more robust analysis of such non-verbal human behaviour.

Cultural backgrounds and social norms are another important issue that will inevitably affect the outcome of such analyses as human (non-verbal) interaction. In this respect, the scope of our study was also limited in that our participants were not multinationals. Although the fact that these individuals were total strangers to each other help validate the interpretation of our findings, this factor on its own is not sufficient for generalization of our results across cultures. Therefore, future research must take into account such differences, thereby focusing on studies that introduce such cultural variabilities.

From a broader perspective, it is a well-known fact that the performance of the machine learning algorithms is highly dependent on the sample size based on which these algorithms are trained. It is also widely demonstrated that limited data can result in degradation of these models' performance and consequentially bias and/or reduce their accuracy. To (at least partially) control for such undesirable effects, we opted for a strategy in which we repeated our cross-validation for 1000 random split of our 5664 distance data points between train and test sets. We also further carried out our statistical tests of significance on these models' RMSE and R^2 values through 1000 rounds of random sampling (without replacement). These steps enabled us to estimate these models' RMSE more reliably and R^2 values as well as to obtain a more accurate estimate of their differences. However, to draw a more informed conclusion on our results, it is necessary to repeat these analyses with a larger sample sizes and in scenarios in which more comprehensive sets of non-verbal human behaviour are considered. Such analyses will not only let researchers realize the extent to which our results can be generalized but also will help them acquire more thorough understanding of the domains of application for these algorithms.

Last, we limited our analyses of the facial pre-touch distances to GP, SVR, and Lasso regressors. It is apparent that the choice of regression models for approximating non-verbal human behaviour is not limited to these options. Future research can benefit from the use of other sophisticated predictive approaches such as linear mixed-effects models [39] and Bayesian linear regression [40], thereby further advancing the findings that are presented in the present study.

Author Contributions: Conceptualization, M.S. and T.M.; formal analysis, S.K.; investigation, M.S., T.M., and K.S.; resources, H.I.; data curation, M.S.; writing—original draft preparation, S.K.; writing—review and editing, S.K., M.S.; project administration, M.S. and H.I.; funding acquisition, M.S., T.M., and H.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research work was supported in part by Japan Science and Technology Agency (JST) CREST (Grant Number JPMJCR18A1), Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Number JP17K00293), and JST ERATO Ishiguro Symbiotic Human-Robot Interaction Project (Grant Number: JPMJER1401).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nummenmaa, L.; Glerean, E.; Hari, R.; Hietanen, J.L. Bodily maps of emotions. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 646–651. [[CrossRef](#)] [[PubMed](#)]
2. Gerlach, M.; Farb, B.; Revelle, W.; Amaral, L.A.N. A robust data-driven approach identifies four personality types across four large data sets. *Nat. Hum. Behav.* **2018**, *2*, 735–742. [[CrossRef](#)] [[PubMed](#)]
3. Goldberg, L.R. An Alternative "Description of Personality": The Big-Five Factor Structure. *J. Personal. Soc. Psychol.* **1990**, *59*, 1216–1229. [[CrossRef](#)]
4. Saarimäki, H.; Gotsopoulos, A.; Jääskeläinen, I.P.; Lampinen, J.; Vuilleumier, P.; Hari, R.; Sams, M.; Nummenmaa, L. Discrete neural signatures of basic emotions. *Cereb. Cortex* **2015**, *26*, 2563–2573. [[CrossRef](#)]

5. Liu, W.; Kohn, N.; Fernández, G. Intersubject similarity of personality is associated with intersubject similarity of brain connectivity patterns. *NeuroImage* **2019**, *186*, 56–69. [\[CrossRef\]](#)
6. Hu, Y.; Hu, Y.; Li, X.; Pan, Y.; Cheng, X. Brain-to-brain synchronization across two persons predicts mutual prosociality. *Soc. Cogn. Affect. Neurosci.* **2017**, *12*, 1835–1844. [\[CrossRef\]](#)
7. Xin, Y.; Wu, J.; Yao, Z.; Guan, Q.; Aleman, A.; Luo, Y. The relationship between personality and the response to acute psychological stress. *Sci. Rep.* **2017**, *7*, 16906. [\[CrossRef\]](#)
8. Ickinger, W.J.; Morris, S. *Psychological Characteristics and Interpersonal Distance*; Tulane University: New Orleans, LA, USA, 2001.
9. Sommer, R. *Personal Space: The Behavioral Basis of Design*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, USA, 1969.
10. Hall, E.T. *The Silent Language*; Anchor Books: New York, NY, USA, 1959.
11. Hall, E.T. *The Hidden Dimension*; Anchor Books: New York, NY, USA, 1963.
12. Gallace, A.; Spence, C. The science of interpersonal touch: An overview. *Neurosci. Biobehav. Rev.* **2010**, *34*, 246–259. [\[CrossRef\]](#)
13. Field, T. Touch for socioemotional and physical well-being: A review. *Dev. Rev.* **2010**, *30*, 367–383. [\[CrossRef\]](#)
14. Chatel-Goldman, J.; Congedo, M.; Jutten, C.; Schwartz, J.L. Touch increases autonomic coupling between romantic partners. *Front. Behav. Neurosci.* **2014**, *8*, 95. [\[CrossRef\]](#)
15. Yun, K.; Watanabe, K.; Shimojo, S. Interpersonal body and neural synchronization as a marker of implicit social interaction. *Sci. Rep.* **2012**, *2*, 959. [\[CrossRef\]](#)
16. Singh, H.; Bauer, M.; Chowanski, W.; Sui, Y.; Atkinson, D.; Baurley, S.; Fry, M.; Evans, J.; Bianchi-Berthouze, N. The brain's response to pleasant touch: An EEG investigation of tactile caressing. *Front. Hum. Neurosci.* **2014**, *8*, 893. [\[CrossRef\]](#)
17. Duffy, K.G.; DeJulio, S.S. The Relationship of Neuroticism to Proxemic Behavior; Education Resources Information Center, Sponsored by the Institute of Education Sciences (IES) of the U.S. Department of Education. 1974. Available online: <https://files.eric.ed.gov/fulltext/ED136093.pdf> (accessed on 30 May 2020).
18. Galton, F. Measurement of character. *Fortnightly* **1884**, *36*, 179–185.
19. Bono, R.; Blanca, M.J.; Arnau, J.; Gómez-Benito, J. Non-normal distributions commonly used in health, education, and social sciences: A systematic review. *Front. Psychol.* **2017**, *8*, 1602. [\[CrossRef\]](#)
20. Taleb, N.N. *The Black Swan: The Impact of the Highly Improbable*, 1st ed.; Random House: New York, NY, USA, 2007.
21. Boyle, E.O., Jr.; Aguinis, H. The best and the rest: Revisiting the norm of normality of individual performance. *Pers. Psychol.* **2012**, *65*, 79–119.
22. Micceri, T. The unicorn, the normal curve, and other improbable creatures. *Psychol. Bull.* **1989**, *105*, 156–166. [\[CrossRef\]](#)
23. Blanca, M.J.; Arnau, J.; López-Montiel, D.; Bono, R.; Bendayan, R. Skewness and kurtosis in real data samples. *Methodology* **2013**, *9*, 78–84. [\[CrossRef\]](#)
24. Keshmiri, S.; Shiomi, M.; Shatani, K.; Minato, T.; Ishiguro, H. Facial Pre-Touch Space Differentiates the Level of Openness Among Individuals. *Sci. Rep.* **2019**. [\[CrossRef\]](#)
25. Makiko, K.; Mochimaru, M. *Japanese Head Size Database*; AIST: Tokyo, Japan, 2001. (In Japanese)
26. Giancola, S.; Corti, A.; Molteni, F.; Sala, R. Motion Capture: An Evaluation of Kinect V2 Body Tracking for Upper Limb Motion Analysis. In Proceedings of the International Conference on Wireless Mobile Communication and Healthcare, Milan, Italy, 14–16 November 2016; pp. 302–309.
27. Shiomi, M.; Shatani, K.; Minato, T.; Ishiguro, H. How Should a Robot React Before People's Touch? Modeling a Pre-Touch Reaction Distance for a Robot's Face. *IEEE Robot. Autom. Lett.* **2018**, *3*, 3773–3780. [\[CrossRef\]](#)
28. Steel, R.G.; Torrie, J.H. *Principles and Procedures of Statistics with Special Reference to the Biological Sciences*; McGraw-Hill Book Company, Inc.: New York, NY, USA, 1960.
29. Glantz, S.A.; Slinker, B.K.; Neilands, T.B. *Primer of Applied Regression and Analysis of Variance*; McGraw-Hill Book Company, Inc.: New York, NY, USA, 1990.
30. Tomczak, M.; Tomczak, E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sport Sci.* **2014**, *1*, 19–25.
31. Rosenthal, R.; DiMatteo, M.R. Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annu. Rev. Psychol.* **2001**, *52*, 59–82. [\[CrossRef\]](#) [\[PubMed\]](#)

32. Allen, M.; Poggiali, D.; Whitaker, K.; Marshall, T.R.; Kievit, R. Raincloud plots: A multi-platform tool for robust data visualization. *PeerJ Prepr.* **2018**, *6*, e27137v1. [[CrossRef](#)]
33. Oishi, H.; Takemura, H.; Aoki, S.C.; Fujita, I.; Amano, K. Microstructural properties of the vertical occipital fasciculus explain the variability in human stereoacuity. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12289–12294. [[CrossRef](#)] [[PubMed](#)]
34. Kriegeskorte, N.; Simmons, W.K.; Bellgowan, P.S.; Baker, C.I. Circular analysis in systems neuroscience: The dangers of double dipping. *Nat. Neurosci.* **2009**, *12*, 535–540. [[CrossRef](#)] [[PubMed](#)]
35. Jung, M.; Hinds, P. Robots in the wild: A time for more robust theories of human-robot interaction. *ACM Trans. Hum. Robot Interact. (THRI)* **2018**, *7*, 2. [[CrossRef](#)]
36. Yang, G.Z.; Bellingham, J.; Dupont, P.E.; Fischer, P.; Floridi, L.; Full, R.; Jacobstein, N.; Kumar, V.; McNutt, M.; Merrifield, R.; et al. The grand challenges of Science Robotics. *Sci. Robot.* **2018**, *3*, eaar7650. [[CrossRef](#)]
37. Clabaugh, C.; Matarić, M. Robots for the people, by the people: Personalizing human-machine interaction. *Sci. Robot.* **2018**, *3*, eaat7451. [[CrossRef](#)]
38. Hayduk, L.A. Personal space: Where we now stand. *Psychol. Bull.* **1983**, *94*, 293–335. [[CrossRef](#)]
39. Bates, D.; Mächler, M.; Bolker, B.M.; Walker, S.C. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1. [[CrossRef](#)]
40. Box, G.E.; Tiao, G.C. *Bayesian Inference in Statistical Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 1973.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).