Affective features drive human representations of social touch expressions during naturalistic viewing

Haemy Lee Masson (haemy.lee-masson@durham.ac.uk)
Psychology Department, Durham University, Stockton Rd
Durham, DH1 3LE, UK

Yaxuan Zhai (yaxuan.zhai@durham.ac.uk)
Psychology Department, Durham University, Stockton Rd
Durham, DH1 3LE, UK

Abstract

Touch is a potent communication tool. It has been suggested that a wide range of factors impact how we perceive social touch. No study has investigated which features we attend to when observing complex, naturalistic social touch expressions. To address this question, we curated 125 video clips showing a wide range of social interactions from the American TV series, Modern Family. Eighty participants watched 45 pseudo-randomly selected videos and performed a multiple arrangement task. This procedure produced the group-averaged pairwise similarity judgments for social touch expressions. Visual, social, and affective features were extracted from each video clip using artificial neural networks (ANN), behavioural experiments, and human annotations. The combination of multiple regression and variance partitioning analyses revealed that selected affective, social, high-level visual, and ANN features collectively explained 52% of the variance in the perception of social touch expressions. Among these, affective features uniquely accounted for 33% of the variance. The current findings suggest that affective features, specifically whether the touch is used to convey positive or negative emotions, drive human perceptions of social touch during naturalistic viewing. Conversely, ANN features explained the least variance, suggesting that the models, trained on action perception and facial expression recognition, may not be sufficient to decode social touch expressions.

Keywords: social touch; naturalistic stimuli; multiple arrangement, artificial neural network

Introduction

Social touch is fundamental to humans. It serves as a powerful communicative tool for expressing a wide range of emotions, such as affection, support, frustration, and anger (Hertenstein, Verkamp, Kerestes, & Holmes, 2006). The use of social touch also varies depending on the type of relationship one has with another person (Suvilehto, Glerean, Dunbar, Hari, & Nummenmaa, 2015). Despite its importance in social bonding and relationships, no previous studies have explored which features drive the representation of social touch during complex, naturalistic interactions that resemble those experienced in daily life. Previous work suggests that social contexts influence how touch is perceived, including comfort, valence, and appropriateness (Suvilehto et al., 2015; Mello, Fusaro, Aglioti, & Minio-Paluello, 2024). However, testing this with simple stimuli has been challenging, as touch videos often contain minimal information about the social relationship, typically only showing a simple stroke or a touch devoid of social context. The movie viewing paradigm is an effective approach to addressing this problem, as the touch depicted in movies is closer to what we experience in daily life. By adopting this paradigm, we aimed to reveal the features that drive human perception of social touch expressions in a more ecologically valid manner.

Methods

Naturalistic stimuli

We first gathered 661 video clips showing a wide range of social touch interactions from the American TV series Modern Family. Among them, only about 60 videos show negative touch. To balance the number of positive and negative touch, we curated 125 one-second video clips as the final set. According to valence ratings from 181 independent samples, 51 videos depict touch interactions conveying negative emotions, 59 videos depict touch interactions conveying positive emotions. 15 videos serve as control videos, showing social interactions without touch but with a mix of positive and negative interactions. The video set includes 48 interactions from couples, 38 from parent-child pairs, 21 from acquaintances, and additional interactions from other relationship types.

Feature extraction

Affective, social, high-level vision, and ANN features are extracted from each video. Affective features include perceived valence, arousal, and the dominance (Mehrabian & Russell, 1974) of touch based on human ratings. Social features include perceived closeness, relationship types (e.g., couple, parent-child pairs), the sex of the interacting people, and text descriptions of the video capturing the interaction within the social context (e.g., “Phil met a woman for the first time and shook hands when introducing themselves to each other.”). High-level visual features include actions (e.g., hug, slap, kiss), the number of people in the scene (52% dyads, 25% triads, and 23% groups of four or more), and whether the scene happens indoors or outdoors. Lastly, ANNs features were extracted. Currently, there are no ANN models specifically trained to recognise social touch. Social touch typically

involves specific actions, and facial expressions often provide emotional cues. Thus, we selected models pre-trained on human action and facial expression recognition. Specifically, activations from layers 1 and the final grand average pooling layer were extracted from each video using the 3D ResNet (Hara, Kataoka, & Satoh, 2017), pre-trained on the Moments in Time dataset (Monfort et al., 2019). Activations from the last fully connected layer were extracted from the middle frame of each video using EMONET (Toisoul, Kossaifi, Bulat, Tzimiropoulos, & Pantic, 2021), pre-trained on Affectnet (Mollahosseini, Hasani, & Mahoor, 2017). Euclidean distance was used to create the representational dissimilarity matrix (RDM) of each feature, except for the text descriptions. The pre-trained Sentence-BERT model ('all-MiniLM-L6-v2') (Reimers & Gurevych, 2019) was used to generate dense sentence embeddings, and Cosine similarity was used for this RDM. ANN features mostly show only mild correlations with social-affective features (Figure 1A).

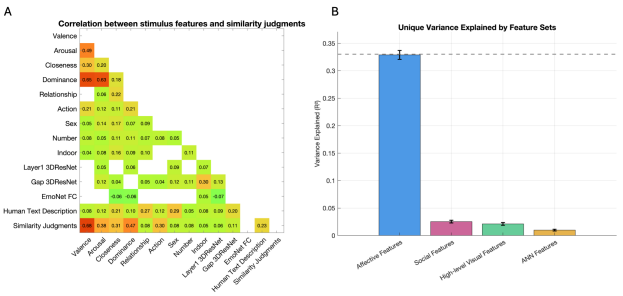

Figure 1: A. Correlation between features and similarity judgments, expressed as r-values (Red - high, green - low, white - no significant correlation). B. Unique variance explained by each feature group. The grey horizontal bar represents the noise ceiling. Each bar includes a confidence interval calculated with bootstrapping.

Procedure

80 participants watched a subset of video stimuli (N=45) and performed a multiple arrangement task (Kriegeskorte & Mur, 2012) on the Meadows Research platform, in which they were asked to place the videos that they perceived as similar close together, and those they perceived as dissimilar farther apart. This experiment yielded a behavioural RDM for each participant. Video pairs were compared by an average of 10 participants. The group-averaged behavioural RDM was used as a dependent variable in a multiple regression model, with all the selected features as independent predictors. We computed the unique variance explained by each predictor group (affective, social, high-level vision, and ANN features) by comparing the full model to reduced models with each feature group removed. The permutation and bootstrapping were used to measure the statistical significance and confidence intervals, respectively. A leave-one-out subject correlation was calculated to examine the reliability of the similarity judgments.

Results

Leave-one-out subject Spearman correlations (mean r = 0.33) and split-half correlations (mean r = 0.46) revealed that participants moderately agree on each other's similarity judgments. Multiple regression analysis revealed that selected affective, social, high-level visual, and ANN features collectively explained 52% of the variance in social touch perception. In particular, valence ($\beta = 0.59$) contributed the most, followed by action ($\beta = 0.14$), the text descriptions of the video ($\beta = 0.13$), and closeness between interacting people ($\beta = 0.08$). Based on the variance partitioning (Figure 1B), affective features, consisting of valence, arousal, and dominance of touch, uniquely accounted for 33% of the variance, which was identical to the lower bound of the reliability measure. Social features (2.6%), high-level visual features (2.1%), and ANN (0.9%) explained a small amount of variance, although all were statistically significant based on the permutation test.

Discussion

To our knowledge, this is the first study to implement a multiple arrangement task capturing human similarity judgments of a wide range of social touch expressions. While social touch expression datasets do exist (Lee Masson & Op de Beeck, 2018), the stimuli are devoid of social context, making it difficult to examine multiple high-level social features. By curating a novel naturalistic stimulus set, the current study demonstrates that affective features, particularly valence, drive human representations of social touch during naturalistic viewing of complex touch interactions within various social contexts. These findings extend previous work showing that touch can convey a wide range of affective states (Hertenstein et al., 2006; Hertenstein, Holmes, McCullough, & Keltner, 2009). Although social contexts have been suggested to be one of the driving factors in social touch perception (Suvilehto et al., 2015), its effect was relatively small compared to valence. ANN features explained the least variance, suggesting that the current models, which are trained on action perception and facial expression recognition, may not be sufficient to decode social touch. Social-affective features either weakly correlate with or do not correlate with ANN features. This suggests that, even though the stimuli are highly naturalistic, the stimulus set is curated in a way that minimizes visual confounding. Concerning affect computing, most studies have focused on facial expressions, with AI models demonstrating the ability to decode human facial expressions (Toisoul et al., 2021). In contrast, this capability does not extend to social touch, primarily due to the limited availability of training datasets and the insufficient understanding of how humans process social touch. Future work will include neuroimaging methods to investigate how these features are processed in the brain with the expectation that affective features of touch will be represented in the somatosensory cortex, as a vicarious response, and along the lateral visual pathway, including the superior temporal sulcus (Lee Masson, Van De Plas, Daniels, & Op de Beeck, 2018).

References

Hara, K., Kataoka, H., & Satoh, Y. (2017). Learning spatio-temporal features with 3d residual networks for action recognition. In *Proceedings of the ieee international conference on computer vision workshops* (pp. 3154–3160).

Hertenstein, M. J., Holmes, R., McCullough, M., & Keltner, D. (2009). The communication of emotion via touch. *Emotion*, *9*(4), 566.

Hertenstein, M. J., Verkamp, J. M., Kerestes, A. M., & Holmes, R. M. (2006, feb). The communicative functions of touch in humans, nonhuman primates, and rats: a review and synthesis of the empirical research. *Genetic, social, and general psychology monographs*, *132*(1), 5–94. doi: 10.3200/MONO.132.1.5-94

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*(JUL). doi: 10.3389/fpsyg.2012.00245

Lee Masson, H., & Op de Beeck, H. (2018, jan). Socio-affective touch expression database. *PLOS ONE*, *13*(1), e0190921. doi: 10.1371/journal.pone.0190921

Lee Masson, H., Van De Plas, S., Daniels, N., & Op de Beeck, H. (2018). The multidimensional representational space of observed socio-affective touch experiences. *Neuroimage*, *175*, 297–314.

Mehrabian, A., & Russell, J. A. (1974). The basic emotional impact of environments. *Perceptual and motor skills*, *38*(1), 283–301.

Mello, M., Fusaro, M., Aglioti, S. M., & Minio-Paluello, I. (2024). Exploring social touch in autistic and non-autistic adults via a self-report body-painting task: The role of sex, social context and body area. *Autism*, 13623613231218314.

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, *10*(1), 18–31.

Monfort, M., Andonian, A., Zhou, B., Ramakrishnan, K., Bargal, S. A., Yan, T., ... Vondrick, C. (2019). Moments in time dataset: one million videos for event understanding. *IEEE transactions on pattern analysis and machine intelligence*, *42*(2), 502–508.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Suvilehto, J. T., Glerean, E., Dunbar, R. I., Hari, R., & Nummenmaa, L. (2015). Topography of social touching depends on emotional bonds between humans. *Proceedings of the National Academy of Sciences*, *112*(45), 13811–13816.

Toisoul, A., Kossaifi, J., Bulat, A., Tzimiropoulos, G., & Pantic, M. (2021). Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nature Machine Intelligence*, *3*(1), 42–50. doi: 10.1038/s42256-020-00280-0