Human Self-Touch vs Other-Touch Resolved by Machine Learning⋆

Aruna Ramasamy[1,2][0000−0003−2343−9756], Damien Faux[1][0000−0003−4633−3372], Vincent Hayward[1,3][0000−0002−2102−1965], Malika Auvray[3][0000−0002−4173−8146], Xavier Job[4][0000−0002−6017−5983], and Louise Kirsch[5][0000−0002−8418−776X]

[1] Actronika SAS, 68 boulevard de Courcelles, 75017 Paris, France.
`aruna.ramasamy@actronika.com`
[2] École Normale Supérieure, CNRS, Laboratoire des Systèmes Perceptifs, 24 rue Lhomond, 75005 Paris, France
[3] Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, 4 Place Jussieu, 75005 Paris, France
[4] Karolinska Institutet, Department of Neuroscience, C4 Neurovetenskap, 171 77 Stockholm, Sweden
[5] University of Paris, CNRS, Integrative Neuroscience and Cognition Center, 45 Rue des Saint-Pères, 75006 Paris, France

Abstract. Using a database of vibratory signals captured from the index finger of participants performing self-touch or touching another person, we wondered whether these signals contained information that enabled the automatic classification into categories of self-touch and other-touch. The database included signals where the tactile pressure was varied systematically, where the sliding speed was varied systematically, and also where the touching posture were varied systematically. We found that using standard sound feature-extraction, a random forest classifier was able to predict with an accuracy greater than 90% that a signal came from self-touch or from other-touch regardless of the variation of the other factors. This result demonstrates that tactile signals produced during active touch contain latent cues that could play a role in the distinction between touching and being touched and which could have important applications in the creation of artificial worlds, in the study of social interactions, of sensory deficits, or cognitive conditions.

Keywords: Self-Touch · Touchant-touché · Social tactile interactions · Machine learning.

1 Introduction

Skin-to-skin touch is an important tactile interaction. This type of touch has attracted the attention of many authors (e.g. [31,24,9]) and motivated research across many fields; from philosophy [23,16], cognitive neuroscience [5,2,10,20,28],

---

to human development and well-being [4,1,8,25]. All these works are based on introspection or on behavioural observations since, by necessity, they cannot rely on the objectification of the mechanical consequences, hence of the sensory consequences of skin touching skin. It was however recently been realised that the objectification of tactile interactions is possible when hands actively interact with inanimate objects [30,12,29]. Motivated by this observation, some of us collected a database of vibration signals collected from a index finger interacting with another finger, or a forearm, with a view to provide objective data produced during skin-to-skin interactions [17]. This latter study demonstrated that the "tactile waves" measured on a touching finger bore features related to the interaction, to wit, the pressure applied (a tonic characteristic) and the sliding speed (a kinematic characteristic). The signals were shown to be relatively independent from the posture with which the interaction was effected, making this technique potentially useful for analyses about tactile behaviour [17].

1.1   Present Study

In the present study we advanced the hypothesis that information contained in single-channel vibration signals recorded from a finger in sliding contact with another finger, or with a forearm, contained information that would enable the discrimination between self-touch and touching another person. In the foregoing, we show that certain supervised machine classifiers can achieve a very high level of success in deciding whether tactile vibrations came from *self-touch* or from *other-touch* (touching another person). Supervised machine classifiers trained models through ground-truth labels which are indicated here by SELF and OTHER. Our findings could contribute to the study of behaviour in many domains, chief among them is the study of the role of touch in social interactions and investigations related to cognitive conditions such as autism or schizophrenia where self/other touch discrimination might be impaired [31,8]. It also bears the intriguing conclusion that the determining factors differentiating self from ordinary touch are not limited to a unique convergence of sensory and motor signals [23,3,14,15] but that the tactile inputs *per se* contain cues that are special to self-touch.

1.2   Signal Database

A key attribute of the database is that it included signals recorded in similar conditions of pressure and speed during self-touch but also when touching another person. The vibrations recorded from a finger sliding on skin clearly depended on the pressure applied and on the sliding speed [17]. On this account, it would be surprising if machine learning classifiers were not able to discriminate between categories of intensity or speed. In fact, the multichannel whole-hand recordings described in [29] contained sufficient information to enable a support vector machine classifier to categorise twelve different tactile gestures, three types of materials, as well as the shape of the objects being touched.

1.3   Feature Extraction

Since the data at hand were available in form of time-dependent signals arising from mechanical interactions between objects in contact, it stands to reason that techniques developed to classify sounds would also be appropriate to classify signals arising from sliding fingers. A first option was to train and then to test classifiers using raw data. Another possibility was to extract domain knowledge features from the signals.

Features frequently used in the processing of sound include: Maximum Mel Frequency Cepstral Coefficients (MFCC), a quantification technique for vibratory signals [11], minimum MFCC, mean MFCC, Zero-Crossing Rate which capture the rhythmic features of a signal [13], Chromograms which are commonly used for the analysis of musical sounds [27], Spectral Roll-Off which measures the right skewness of a spectrum [18], Spectral Flux which describes rate of change a time-varying spectrum arising from a non-stationary process [21], and Pitch which is a well-known instantaneous attribute of sounds [26]. Features do not contribute equally significant to a given classification problem. Their significance can be assessed through the decrease of accuracy in classification when a feature is dropped. To this end, GINI importance, or mean decrease in impurity (MIDI), may be used to evaluate the importance of each feature [6].

1.4   Performance Measures

The performance of a classifier is relative to a `test` dataset used to examine the model trained with a `train` dataset. Here we use standard performance metrics. Accuracy is measured by

$$\text{Accuracy} = (N_{\text{tp}} + N_{\text{tn}})/(N_{\text{tp}} + N_{\text{fp}} + N_{\text{fn}} + N_{\text{tn}}),$$

which accounts for the number, N, of predictions labeled as true positives (tp), true negatives (tn), false positives (fp), and false negatives (fn), commonly expressed in percent, while precision is defined by the proportion of true positive predictions to the total number of positive predictions, and recall which reaches one when there are no false negatives. We also used a statistical measure termed, F_1-Score, which is the harmonic mean of precision and recall. These metrics are recalled below.

$$\text{Precision} = N_{\text{tp}}/(N_{\text{tp}} + N_{\text{fp}}), \quad \text{Recall} = N_{\text{tp}}/(N_{\text{tp}} + N_{\text{fn}}),$$
$$F_1\text{-Score} = 2(\text{Recall} \cdot \text{Precision})/(\text{Recall} + \text{Precision}).$$

We used a graphical representation borrowed from Signal Detection Theory [22]. Here, a Receiver Operating Characteristics (ROC) curve plots the false positives vs. true positives. It can be interpreted as a plot of 1-sensitivity vs. sensitivity. Even though these curves may cross, any curve clearly above another is better. A single-number separability measure, the area under the ROC curve (AUC), follows from this representation. An AUC of 0.7 is said to be acceptable, excellent if it around 0.8, and outstanding above 0.9.

1.5   Ambiguity and Abstention

In the present study we employed a recently introduced technique which proposes that in case of ambiguity it is better to abstain rather than to make a prediction [7]. This technique can be implemented, for example, in a three-way random forest algorithm which can extract probabilities for each class. Given two probability thresholds, α and β, having value 1.0 for the ground-truth class and 0 for the incorrect class, predictions are declared positive when the score of the positive class is greater than α and the score the positive class is greater than the score of the negative class. Conversely, predictions are declared negative when the score of the negative class is greater than β and the score of the negative class is greater than the score of the positive class. When neither of these conditions are met, then there is an abstention. Typical values for α and β are 0.75.

2   Results

The skin-to-skin touch datasets described in [17] contained signal recorded with eighteen participants of balanced gender and hence captured a reasonable diversity of individual behaviours. The signals were recorded at audio-rate and down-sampled 10-fold. The initial one second interval of each recording was edited out to eliminate the energy burst due to the stick-to-slip transition. This deletion potentially eliminating useful information for the purpose of this study. The `pressure` dataset comprised ten 10-second recordings where participant touched ten times for each condition their own or the other participant's index finger with a gentle or firm touch, resulting in 720 trials. The `speed` dataset comprised similar recordings but the participants touched their own or the other participant's forearm at three different speeds giving rise to 1080 trials. The `posture` dataset comprised similar recordings but the participants touched their own or the other participant's index finger in two different orientation to vary the relationship between the sensor and the regions of skin contact, resulting in 720 trials.

2.1   Relative Performance of Classification Techniques

Table 1 shows the performance of various classification techniques using the `pressure` dataset suggesting that that the random forest classifier performed best compared to other classifiers. It produced negligible Mean Squared Error (MSE) for the `test` dataset with a classification accuracy of 81% greatly surpassing logistic regression, decision tree, Gaussian, and support vector classification.
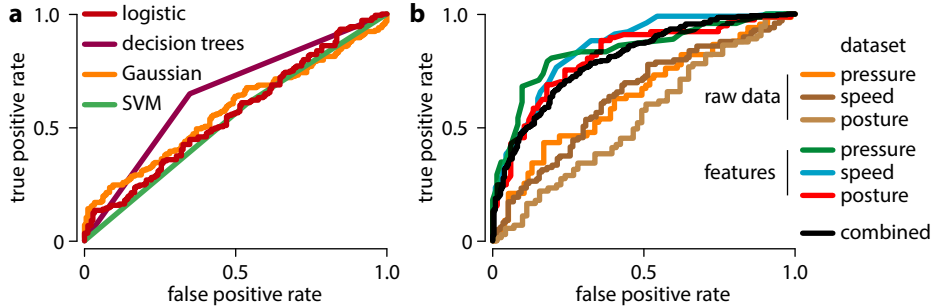
2.2   Importance of Feature Extraction

Tests conducted with the combined datasets to evaluate the contribution of extracted features vs raw data unequivocally confirmed the importance of providing the algorithms with extracted domain knowledge features. Most metrics,

Table 1. Classifier performance & Importance of feature extraction. Accuracy in %.

classifier w `pressure`	Accuracy	Precision	Recall	F_1-Score	`train`-MSE	`test`-MSE	AUC
logistic regression	52.5	0.52	0.52	0.52	0.41	0.47	0.56
decision tree	65.0	0.65	0.65	0.65	0.00	0.35	0.65
Gaussian	55.0	0.56	0.46	0.50	0.42	0.45	0.57
support vector	53.4	0.53	0.61	0.56	0.40	0.47	0.53
random forest	81.0	0.81	0.79	0.80	0.00	0.19	0.84
raw/features w dataset							
raw w `pressure`	61.0	0.58	0.63	0.60	0.02	0.39	0.64
raw w `speed`	62.0	0.57	0.67	0.62	0.17	0.38	0.64
raw w `posture`	53.0	0.54	0.56	0.55	0.10	0.47	0.54
features w `pressure`	81.0	0.81	0.79	0.80	0.00	0.19	0.83
features w `speed`	78.0	0.76	0.76	0.76	0.00	0.22	0.85
features w `posture`	76.0	0.78	0.75	0.76	0.01	0.24	0.81

Table 1, show low to unacceptable values when raw data was used. Figures 1a,b further indicate that classification models become skilful with the introduction of domain knowledge features since the figure shows across-the-board reduction in false positives rate.


Fig. 1. ROC curves. a, Weak performance of classifiers on the raw data of combined datasets. **b,** Overall effect of domain knowledge feature extraction on different datasets.

The results shown in Table 2 indicate that a three-way classification algorithm with abstention greatly improved discrimination between the labels SELF and OTHER. Accuracy was increased from 81% to 92% with the `pressure` dataset, from 78% to 97% with the `speed` dataset, and from 76% to 84% with the `posture` dataset. For the case of the combined dataset accuracy was increased from 73% to 90%, which is very significant.

Figure 2 summarizes the GINI importance of the different domain knowledge features used relatively to the datasets.

Table 2. Three-way classification of SELF and OTHER using abstention.

	Accuracy (%)	Precision	Recall	F1-Score	Train-MSE	Test-MSE	AUC
pressure	92.2	0.96	0.89	0.92	0.00	0.08	0.92
speed	97.6	1.00	0.95	0.87	0.00	0.02	0.98
posture	84.8	0.88	0.85	0.86	0.00	0.15	0.84
combined	90.2	0.87	0.97	0.91	0.00	0.01	0.90
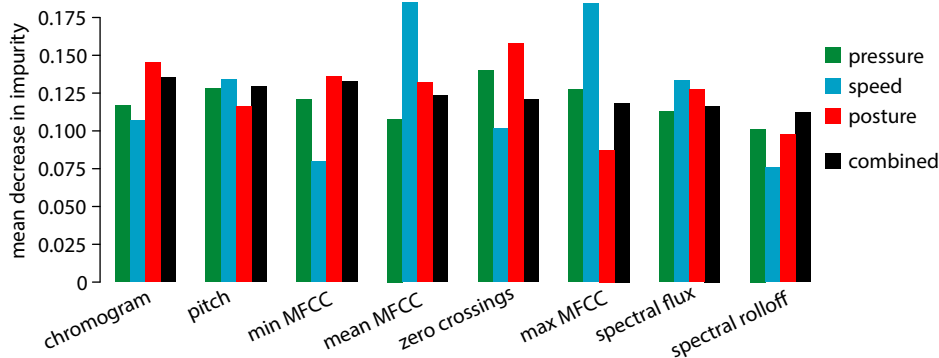
Fig. 2. GINI importance of features. This measure is plotted for each features during three-way random forest classification of different datasets. Features are ranked by order of decreasing importance with combined data.

2.3   Discussion and Conclusion

Overall, the best performance was achieved by the three-way random forest classification algorithm which makes use of an ensemble learning method though a multitude of decision trees. The mutually exclusive branches represent sub-categories of the input features. The random forest classifier overcomes over-fitting though voting to predict an output. The technique known as bootstrap aggregating de-correlates the decision trees corresponding to different training sets. Noise in single tree affects the performance of the model but not the average of many trees. This strategy was very successful in the classification of self-touch and other-touch (labels SELF and OTHER) from single-channel recordings of tactile waves in the index finger.

As a whole, the eight domain knowledge features showed little relative advantages over the others in the task of discriminating self-touch from other-touch. The mean and max MFCC features which made important contributions when the speed of sliding contact varied could be considered as exceptions. Also, the zero-crossing feature was important when the posture was changed. Chromogram and pitch had the highest importance in terms of classification with combined data. These findings indicates that none of the commonly used audio features preferentially revealed the latent characteristics of skin-to-skin friction-induced vibrations that can be used to distinguish self-touch from touching other people. It is possible that the removal of the stick-to-slip frictional transitions was responsible for this general lack of sensitivity. These findings therefore suggest

that further research is needed to discover better domain knowledge features for this type of data.

The AUC measure was found to be a preferred performance indicator over accuracy, an observation that was commented in [19] although it is recommended to consider different evaluation metrics to discuss performance of a classifier on a particular problem. This observation is validated by the ROC curve obtained under change in posture compared to that obtained under change in speed.

It is astonishing that some of the machine learning algorithms could reach such very high level of performance in discriminating self- from other-touch. Common sense would suggest that the applied pressure would be a factor but the results suggest otherwise. The same can be said of the speed of sliding. If it was an important factor, then the dataset where speed was purposefully varied would have led to poor performance, which was not the case. Thus, surprisingly, the latent characteristics that enabled discrimination between self- and other-touch were not related to neither the tonic nor the kinematic attributes of the gestures employed. Since it is not at all obvious which signal characteristics these algorithms exploited to achieve discrimination, future research will seek to identify which invariant properties hidden in the recorded tactile signals were used by the classifier to discriminate self- from other-touch.

References

1. Ardiel, E.L., Rankin, C.H.: The importance of touch in development. Paediatrics & child health **15**(3), 153–156 (2010)
2. Bays, P.M., Wolpert, D.M.: Predictive attenuation in the perception of touch. In: Haggard, P., Rosetti, Y., Kawato, M. (eds.) Sensorimotor foundations of higher cognition, vol. 22, pp. 339–358. Oxford University Press (2008)
3. Bermúdez, J.L.: The paradox of self-consciousness. MIT Press (2000)
4. Blackwell, P.L.: The influence of touch on child development: Implications for intervention. Infants & Young Children **13**(1), 25–39 (2000)
5. Blakemore, S.J., Wolpert, D.M., Frith, C.D.: Why can't you tickle yourself? Neuroreport **11**(11), R11–R16 (2000)
6. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
7. Campagner, A., Cabitza, F., Ciucci, D.: Three–way classification: Ambiguity and abstention in machine learning. In: International Joint Conference on Rough Sets. pp. 280–294. Springer (2019)
8. Cascio, C.J., Moore, D., McGlone, F.: Social touch and human development. Developmental Cognitive Neuroscience **35**, 5–11 (2019)
9. Classen, C.: The book of touch. Routledge (2020)
10. Crucianelli, L., Metcalf, N.K., Fotopoulou, A.K., Jenkinson, P.M.: Bodily pleasure matters: velocity of touch modulates body ownership during the rubber hand illusion. Frontiers in psychology **4**, 703 (2013)
11. Davis, S., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE transactions on acoustics, speech, and signal processing **28**(4), 357–366 (1980)
12. Delhaye, B., Hayward, V., Lefèvre, P., Thonnard, J.L.: Texture-induced vibrations in the forearm during tactile exploration. Frontiers in Behavioral Neuroscience **6**(37), 1–10 (2012)

13. Gouyon, F., Pachet, F., Delerue, O.: On the use of zero-crossing rate for an application of classification of percussive sounds. In: Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00), Verona, Italy. vol. 5 (2000)
14. Haggard, P., Clark, S., Kalogeras, J.: Voluntary action and conscious awareness. Nature neuroscience **5**(4), 382–385 (2002)
15. Hara, M., Pozeg, P., Rognini, G., Higuchi, T., Fukuhara, K., Yamamoto, A., Higuchi, T., Blanke, O., Salomon, R.: Voluntary self-touch increases body ownership. Frontiers in psychology **6**, 1509 (2015)
16. Husserl, E.: The constitution of psychic reality through the body. In: Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, pp. 151–169. Springer (1989)
17. Kirsch, L.P., Job, X.E., Auvray, M., Hayward, V.: Harnessing tactile waves to measure skin-to-skin interactions. Behavior Research Methods **53**(4), 1469–1477 (2021)
18. Kos, M., Kačič, Z., Vlaj, D.: Acoustic classification and segmentation using modified spectral roll-off and variance-based features. Digital Signal Processing **23**(2), 659–674 (2013)
19. Ling, C., Huang, J., Zhang, H.: Auc: a statistically consistent and more discriminating measure than accuracy. Proc. 18th Int'l Joint Conf. Artificial Intelligence (IJCAI) (2003)
20. Löken, L.S., Olausson, H.: The skin as a social organ. Experimental brain research **204**(3), 305–314 (2010)
21. Lu, L., Jiang, H., Zhang, H.J.: A robust audio classification and segmentation method. In: Proceedings of the ninth ACM international conference on Multimedia. pp. 203–211 (2001)
22. Marcum, J.I.: A statistical theory of target detection by pulsed radar. Tech. rep., Rand Corp Santa Monica, CA (1947)
23. Merleau-Ponty, M.: Phenomenology of Perception. Routledge (1962)
24. Morrison, I.: Keep calm and cuddle on: social touch as a stress buffer. Adaptive Human Behavior and Physiology **2**, 344–362 (2016)
25. Moscatelli, A., Nimbi, F.M., Ciotti, S., Jannini, E.A.: Haptic and somesthetic communication in sexual medicine. Sexual Medicine Reviews **9**(2), 267–279 (2021)
26. Nielsen, A.B., Hansen, L.K., Kjems, U.: Pitch based sound classification. In: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. vol. 3, pp. III–III. IEEE (2006)
27. Pelkowitz, L.: A generalization of the spectrogram for colored displays. IEEE Transactions on Acoustics, Speech, and Signal Processing **31**(1), 222–225 (1983)
28. Schütz-Bosbach, S., Musil, J.J., Haggard, P.: Touchant-touché: The role of self-touch in the representation of body structure. Consciousness and cognition **18**(1), 2–11 (2009)
29. Shao, Y., Hayward, V., Visell, Y.: Spatial patterns of cutaneous vibration during whole-hand haptic interactions. Proceedings of the National Academy of Sciences **113**(15), 4188–4193 (2016)
30. Tanaka, Y., Horita, Y., Sano, A.: Finger-mounted skin vibration sensor for active touch. In: Isokoski, P., Springare, J. (eds.) Haptics: Perception, Devices, Mobility, and Communication. pp. 169–174. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
31. Van Erp, J.B.F., Toet, A.: Social touch in human–computer interaction. Frontiers in Digital Humanities **2**, 2 (2015)